The role of DNA methylation and oligodendrocytes in neurodegeneration

Katherine Elizabeth Fodder

Department of Neurodegenerative Disease, Queen Square Institute of Neurology

University College London

Supervisors
Professor Conceição Bettencourt
Professor Rohan de Silva
Professor Tom Warner

This thesis is submitted for the degree of Doctor of Philosophy

January 2025

Declaration

I, Katherine Elizabeth Fodder, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Publications arising from this thesis, include:

- **Fodder, K**., Murthy, M., Rizzu, P. et al. Brain DNA methylomic analysis of frontotemporal lobar degeneration reveals OTUD4 in shared dysregulated signatures across pathological subtypes. Acta Neuropathol 146, 77–95 (2023). https://doi.org/10.1007/s00401-023-02583-z
- **Fodder K**, de Silva R, Warner TT, Bettencourt C. The contribution of DNA methylation to the (dys)function of oligodendroglia in neurodegeneration. Acta Neuropathol Commun. 2023 Jun 29;11(1):106. doi: 10.1186/s40478-023-01607-9. PMID: 37386505; PMCID: PMC10311741.

Publications not directly related to this thesis:

- Murthy M, **Fodder K**, Miki Y, Rambarack N, De Pablo Fernandez E, Pihlstrøm L, Mill J, Warner TT, Lashley T, Bettencourt C. DNA methylation patterns in the frontal lobe white matter of multiple system atrophy, Parkinson's disease, and progressive supranuclear palsy: a cross-comparative investigation. Acta Neuropathol. 2024 Jul 12;148(1):4. doi: 10.1007/s00401-024-02764-4. PMID: 38995454; PMCID: PMC11245434.
- Rambarack N, **Fodder K**, Murthy M , Toomey C , de Silva R , Humphrey J, Raj T , Lashley T, Bettencourt1 C. DNA methylation as a contributor to dysregulation of STX6

and other frontotemporal lobar degeneration genetic risk-associated loci. bioRxiv preprint doi: https://doi.org/10.1101/2025.01.21.634065

.

Abstract

Neurodegenerative diseases form a heterogeneous group of conditions characterised by the progressive degeneration of the structure and function of the central or peripheral nervous systems. The pathogenic mechanisms underlying these diseases are not fully understood. Various pathogenic mechanisms are thought to contribute to disease, and an increasing number of studies implicate dysfunction of oligodendrocytes (the myelin producing cells of the central nervous system) and myelin loss. Aberrant DNA methylation, the most widely studied epigenetic modification, has been associated with many neurodegenerative diseases, including Alzheimer's disease (AD), Parkinson's disease (PD), Progressive supranuclear palsy (PSP) and multiple system atrophy (MSA). Recent research highlights aberrant DNA methylation in oligodendrocyte (OLG)/myelin-related genes. In this thesis, we describe the results of investigations into the role of dysregulated DNA methylation within oligodendrocyte lineage genes across neurodegenerative diseases. Through complementary computational approaches, we identify key genes which have not previously been linked to altered DNA methylation as important in neurodegeneration. We also utilise gene expression datasets to follow up on findings, in order to uncover functional consequences of dysregulated DNA methylation. We find that several crucial myelin/OLG lineage genes such as *MBP* are differentially methylated and expressed across dementias. We also investigate DNA methylation changes during oligodendrocyte differentiation using DNA methylation profiles from iPSC derived cells across differentiation stages, and identify genes that we had found to be differentially methylated in disease as also being top differentially methylated genes during differentiation. This implicates aberrant DNA methylation as a mechanism that could be contributing to dysfunction of OLG lineage

progression. Finally, we investigate the oligodendrocyte-specific gene *MOBP*, which has been associated with multiple neurodegenerative diseases at the genetic level. *MOBP* has also been previously identified as aberrantly differentially methylated in multiple MSA. In this work, we highlight a shared genetic loci between ALS and PSP within *MOBP* that appears to be associated with changes in DNA methylation and gene expression in disease. Through this work, we have demonstrated dysregulation of DNA methylation affecting OLG lineage cells in neurodegeneration.

Impact Statement

There is currently no cure for neurodegenerative disease. The work in this thesis aims to demonstrate the importance of dysregulated DNA methylation within oligodendrocyte lineage cells genes across neurodegenerative diseases, which we carried out using computational approaches across multiple publicly available DNA methylation and gene expression datasets. The role of OLGs in neurodegeneration has thus far been overlooked. Here, we have provided evidence that genes related to myelin/OLG lineage genes are differentially methylated in disease and show functional consequences at the gene expression level, underpinning the importance of this area of research. We have also identified genes that have not been previously associated with neurodegeneration, justifying further investigation into their role in pathology. As DNA methylation is reversible, identification of disease relevant genes that are modified through DNA methylation in disease, the hope is that this work will contribute to the development of disease modifying therapies for neurodegenerative disease.

We have also investigated the role of DNA methylation in OLG lineage progression in human iPSC derived cells, which is so far not well understood. We identified genes that are important in the differentiation process that are also differentially methylated in disease, highlighting this as an area of research warranting further study. Work from Chapter 6 investigating *MOBP* has demonstrated the need to investigate this gene further across neurodegeneration.

We have leveraged multiple multiomic datasets to integrate publicly available data, maximizing the utility of previously generated datasets and exemplifying the power of data repurposing in neurodegenerative disease research. We have developed pipelines to investigate DNA methylation changes, code for which will be made publicly available.

Results from Chapters 3 and 4 have been partly published in *Acta Neuropathologica*, and manuscripts for other work are in progress.

UCL Research Paper Declaration Form (1)
referencing the doctoral candidate's own published work(s)

1. **For a research manuscript that has already been published** (if not yet published, please skip to section 2)

a) **What is the title of the manuscript?**

Brain DNA methylomic analysis of frontotemporal lobar degeneration reveals OTUD4 in shared dysregulated signatures across pathological subtypes

b) **Please include a link to or doi for the work**

https://doi.org/10.1007/s00401-023-02583-z

c) **Where was the work published?**

Acta Neuropathologica

d) **Who published the work?** (e.g. OUP)

SpringerNature

e) **When was the work published?**

07/05/2023

f) **List the manuscript's authors in the order they appear on the publication**

Katherine Fodder, Megha Murthy, Patrizia Rizzu, Christina E. Toomey, Rahat Hasan, Jack Humphrey, Towfique Raj, Katie Lunnon, Jonathan Mill, Peter Heutink, Tammaryn Lashley & Conceição Bettencourt

g) **Was the work peer reviewed?**

Yes

h) **Have you retained the copyright?**

Yes

i) **Was an earlier form of the manuscript uploaded to a preprint server?** (e.g. medRxiv). If 'Yes', please give a link or doi)

Yes - https://doi.org/10.1101/2022.10.21.513088

If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

*I acknowledge permission of the publisher named under **1d** to include in this thesis portions of the publication named as included in **1c**.*

2. **For a research manuscript prepared for publication but that has not yet been published** (if already published, please skip to section 3)

3. **For multi-authored work, please give a statement of contribution covering all authors** (if single-author, please skip to section 4)

K. Fodder undertook data analysis and drafted the manuscript. C. Bettencourt conceptualised and supervised the work. Other authors contributed with data included in this manuscript, including DNA methylation, transcriptomics and protein staining/proteomics data. All authors critically revised the manuscript.

4. **In which chapter(s) of your thesis can this material be found?**

Chapter 3 and 4

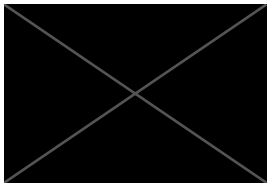
5. **e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

Candidate

Katherine Fodder

Date: 29/01/2024

Supervisor/ Senior Author (where appropriate)

Conceicao Bettencourt 

Date

30/01/2025

UCL Research Paper Declaration Form (2)
referencing the doctoral candidate's own published work(s)

1. **For a research manuscript that has already been published** (if not yet published, please skip to section 2)

   a) **What is the title of the manuscript?**

The contribution of DNA methylation to the (dys)function of oligodendroglia in neurodegeneration

   b) **Please include a link to or doi for the work**

https://doi.org/10.1186/s40478-023-01607-9

   c) **Where was the work published?**

Acta Neuropathologica Communications

   d) **Who published the work?** (e.g. OUP)

BMC

   e) **When was the work published?**

29/06/2023

   f) **List the manuscript's authors in the order they appear on the publication**

Katherine Fodder, Rohan de Silva, Thomas T.Warner, Conceição Bettencourt

   g) **Was the work peer reviewed?**

Yes

   h) **Have you retained the copyright?**

Yes

   i) **Was an earlier form of the manuscript uploaded to a preprint server?** (e.g. medRxiv). If 'Yes', please give a link or doi)

No

If 'No', please seek permission from the relevant publisher and check the box next to the below statement:

☒ The article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format

2. **For a research manuscript prepared for publication but that has not yet been published** (if already published, please skip to section 3)

3. **For multi-authored work, please give a statement of contribution covering all authors** (if single-author, please skip to section 4)

Katherine Fodder drafted the manuscript. Rohan de Silva, Thomas T. Warner and Conceição Bettencourt critically reviewed the manuscript. All authors read and approved the final manuscript

4. **In which chapter(s) of your thesis can this material be found?**

Chapter 1

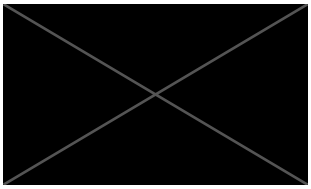
5. **e-Signatures confirming that the information above is accurate** (this form should be co-signed by the supervisor/ senior author unless this is not appropriate, e.g. if the paper was a single-author work)

Candidate

Katherine Fodder

Date: 29/01/2024

Supervisor/ Senior Author (where appropriate)

Conceicao Bettencourt 

Date

30/01/2025

Acknowledgements

I would first like to acknowledge all the individuals who have donated brain tissue to research, and to their families. I hope that in some way, this work contributes to research that leads to a treatment for neurodegenerative disease.

Firstly, a massive thank you to my supervisor **Sao Bettencourt**, who took on the challenge of supervising someone with very little knowledge of computational work or neuroscience. I hope you don't completely regret your decision! Your guidance and support over the last four years has turned the daunting prospect of undertaking a PhD into a rewarding and enjoyable experience. Thank you also for encouraging me to pursue teaching opportunities, conference presentations and travel opportunities, which have no doubt increased my confidence and abilities as a scientist. I am so excited to see what you have in store for the growing **Bettencourt Lab**. Next, to **Rohan de Silva** for his guidance and patience in the cell work, and for many words of wisdom over the years, both over wine at christmas parties and in the lab. Also to **Tom Warner** for giving me the opportunity to join the wider group, and direction throughout. I would also like to thank my committee thesis members **Dervis Salih** and **Regina Reynolds,** the meetings we have had have been incredibly useful in keeping me on track and preventing me from getting too overexcited by the many datasets I could have continued to analyse.

To all the **Lashley lab members**, I have loved being part of the group over the years, thank you for all of your support. I'm so pleased I have managed to drag several of you along to spin classes, whether you were willing or not.

I would also like to thank **Jack Humphrey** and **Kurt Farrell** for their support in getting me to New York where I had an amazing time at Mount Sinai learning all things genetics and karaoke. I also owe thanks to those whose work has informed mine. Thank you to **James Evans** and **Sonia Gandhi** for the provision of the iPSC derived cells, and to **Ignazio Piras** for passing through the gene lists fundamental to this thesis. Thanks also to all those who have published data that we use throughout this thesis.

On a more personal note, I would also like to thank my lab members and friends **Naiomi Rambarack, Yaz Buhidma, Ariana Gatt** and **Toby Curless** for their support during my PhD, not to mention the endless opportunities for coffee breaks. Outside the lab I owe a special mention to my friend **Amelia** for putting up with living with me for the majority of my PhD and valiantly attempting to understand concepts of computational epigenetics. Thank you to all of my other friends for the pints and chats, I promise I will no longer bore you with complaints about how difficult formatting a thesis is.

A final thank you to my wonderful family, to **Mum** for the provision of rosé and words of support, to **Dad** for endless offers of proofreading and the constant queries of ‘is it done yet’, and to my sister **Hattie**, I’m sorry I didn’t manage to slip your paper into my citations.

To everyone whose lives have been affected by neurodegenerative disease.

Abbreviations

AD Alzheimer's Disease

ALS Amyotrophic lateral sclerosis

bp Base Pair

bvFTD Behavioural variant frontotemporal dementia

ChAMP Chip Analysis Methylation Pipeline

CBD Corticobasal degeneration

CBS Corticobasal syndrome

CBR Cerebellum

CETYGO CEll TYpe deconvolution GOodness

CJD Creutzfeldt-Jakob disease

CNS Central nervous system

CpG Cytosine-phosphate-guanine

CRB Cerebellum

CTRL Control

DEG Differentially expressed gene

DLPFC Dorsolateral prefrontal cortex

DNA Deoxyribonucleic acid

eQTL Expression quantitative trait locus

EWAS Epigenome-wide association analysis

EWCE Enrichment-weighted cell-type expression

ERC Entorhinal cortex

FANS Fluorescence-activated nuclei sorting

FC Frontal cortex

FDR False discovery rate

FTD Frontotemporal dementia

FTLD Frontotemporal lobar degeneration

GCI Glial cytoplasmic inclusion

GRCh37 Genome Reference Consortium Human build 37

GRCh38 Genome Reference Consortium Human build 38

GWAS Genome-wide association study

HIPPO Hippocampus

iPSC Induced pluripotent stem cells

LBD Lewy body dementia

LD Linkage disequilibrium

MAF Minor allele frequency

MB Megabase

MM Module membership

mRNA messenger RNA

mQTL Methylation quantitative trait loci

MS Multiple sclerosis

MSA Multiple system atrophy

OLG Oligodendrocyte

OPC Oligodendrocyte precursor cell

PC Principal component

PCA Principal component analysis

PCR Polymerase chain reaction

PD Parkinson's disease

PMI Post-mortem interval

PP Posterior probability

PSP Progressive supranuclear palsy

qPCR Quantitative polymerase chain reaction

QTL Quantitative trait loci

RIN RNA integrity number

RNA Ribonucleic acid

ROS Reactive oxygen species

ROSMAP The Religious Order Study and the Memory and Ageing Project

SNP Single nucleotide Polymorphism

snRNA Single nuclei RNA-sequencing

SVD Singular value decomposition

UMAP Uniform manifold approximation and projection

TSS1500 Transcription Start Site 1500 bp

TSS200 Transcription Start Site 200 bp

WGCNA Weighted gene correlation network analysis

WHMs White matter hyperintensities

3'UTR 3 prime untranslated region

5hmC DNA hydroxymethylation

5mC DNA methylation

5'UTR 5 prime untranslated region

Chapter 1 - Introduction

1.1 Neurodegeneration

Neurodegenerative diseases form a diverse group of neurological disorders characterized by progressive degeneration of the structure and function of the central nervous system or peripheral nervous system ultimately leading to loss of neurons. Most neurodegenerative diseases are age-associated [1], which, in a rapidly aging society, poses a significant public health challenge.

Neurodegenerative diseases present with diverse clinical manifestations, brain region vulnerability and manner and length of progression. However, a central feature of many neurodegenerative diseases is aberrant aggregation of proteins [2]. Protein aggregations can manifest in several ways (**Figure 1.1 A**). Inclusions can occur within neurons or glial cells (non-neuronal cells of the central nervous system (CNS), either as intracytoplasmic aggregation or intranuclear aggregates. Proteins can also aggregate outside of cells in extracellular spaces, forming extracellular plaques. Mislocalization of proteins within cells can also occur, e.g. mislocalization of a nuclear protein to the cytoplasm of cells. Accumulation of proteins disrupts cellular functions and leads to death of neurons. So far, research into the pathology of neurodegenerative disease has been neurocentric, however there is growing evidence to suggest that understanding the involvement of glial cells is crucial [3]. Although neurodegenerative diseases share the central feature of protein aggregation, underlying pathological mechanisms driving and/or associated with such changes can be more disease-specific and remain incompletely understood. Features such as cell-type vulnerability [4], genetic associations [5] and regional susceptibility within the brain [6] are examples of such differences

across neurodegeneration. Clinical symptoms of neurodegenerative diseases are also diverse. Whilst many such diseases are associated with dementia (cognitive decline associated with memory loss), others are associated with motor impairments such as parkinsonism or weakness of limbs. Provided below is a brief overview of each of the neurodegenerative diseases that have been focused on throughout this thesis.

1.1.2.i Alzheimer's Disease

The most common form of neurodegenerative disease is Alzheimer's disease (AD) [7], which is also the leading cause of dementia worldwide [8]. AD pathology is characterised by the occurrence of extracellular plaques formed of amyloid-beta and intracellular neurofibrillary tangles composed of hyperphosphorylated microtubule-associated protein tau [7]. Such accumulation is associated with a range of pathological processes, including synaptic dysfunction, neuroinflammation, oxidative stress, vascular dysfunction and altered lipid metabolism [9]. AD presents clinically with impaired cognitive function, specifically severe memory loss during disease progression. The vast majority (95%) of AD cases are sporadic [10]. Such sporadic cases, also known as late-onset AD, have been associated with multiple genetic risk factors [11]. Numerous pathways such as lysosomal dysfunction, neuroinflammation, cholesterol metabolism, mitochondrial dysfunction and blood brain barrier malfunction have been implicated, but none have been defined as a primary mechanism driving disease pathology [9]. A minority of cases are familial and are caused by underlying genetic mutations, which has provided evidence for pathways involved, such as the finding of mutation of the *APP* and presenilin (*PSEN1* and *PSEN2*) genes in familial AD providing evidence for the amyloid cascade hypothesis of AD [12].

1.1.2.ii Frontotemporal lobar degeneration

Frontotemporal lobar degeneration (FTLD), is an umbrella term for a clinically, genetically and pathologically heterogeneous group of diseases. Frontotemporal dementia (FTD) encompasses a spectrum of clinically defined disorders that are found to have underlying FTLD pathology [13]. Disorders within the FTLD spectrum are characterised by atrophy and degeneration of the frontal and temporal lobes of the brain, but are diverse in their pathology and genetics. Most FTLD cases are characterised by 43 kDa transactive response DNA-binding protein (TDP-43) positive inclusions (FTLD-TDP), or with tau-positive inclusions (FTLD-tau), representing around 50% and 40% of FTLD cases, respectively [14,15]. Several diseases fall within the FTLD-tau spectrum, including frontotemporal dementia (FTD), corticobasal degeneration (CBD) and progressive supranuclear palsy (PSP). PSP is characterised by intracellular aggregation of the microtubule-associated protein tau in neurofibrillary tangles, tufted astrocytes and tau deposits in oligodendrocytes presenting as coiled bodies [16,17]. FTLD-TDP cases are defined by the loss of TDP-43 from its nuclear location, and aggregation of the protein within the cytoplasm [13]. Different subtypes of FTLD-TDP (including FTLD-TDP Types A, B and C, which are studied later in this thesis) are classified based on the distribution and type of pathological inclusions and lesions in cortical layers [18]. TDP-43 inclusion types include neuronal cytoplasmic inclusions, neuronal intranuclear inclusions, oligodendroglial inclusions and dystrophic neurites [18]. Clinically, the FTLD spectrum manifests as a variety of syndromes depending on the predominant brain regions affected and the underlying pathology. Broadly, FTLD encompasses behavioural variant frontotemporal dementia (bvFTD), primary progressive aphasia (PPA), and movement disorders, including corticobasal syndrome (CBS), PSP, and frontotemporal

dementia with motor neuron disease (FTD-MND) [14,18]. Each of these have distinct clinical manifestations (**Figure 1.1.B**).

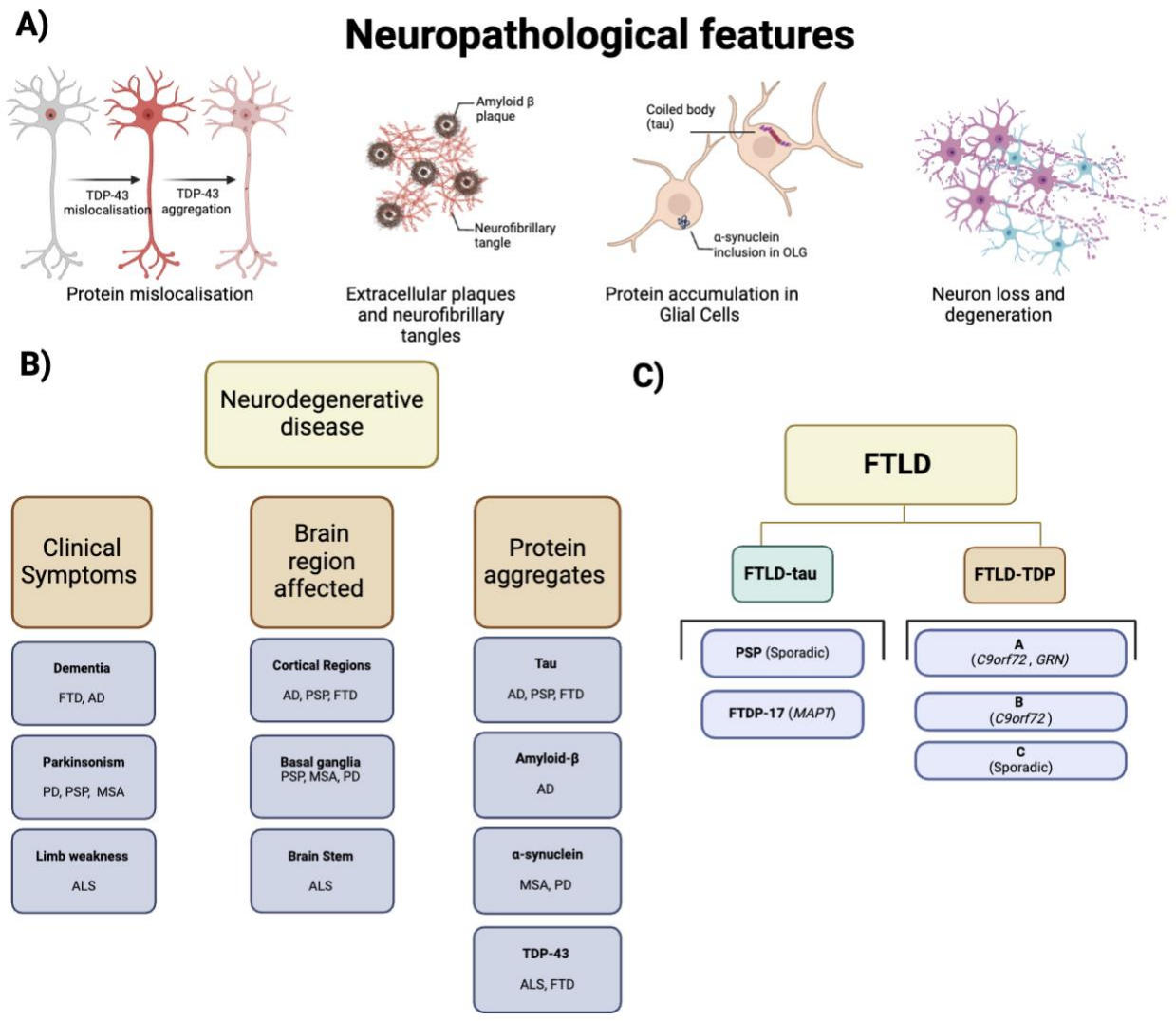
1.1.2.iii Other diseases highlighted

Multiple system atrophy (MSA) is a rare neurodegenerative movement disorder, which is characterized by accumulation of α-synuclein mostly in oligodendrocytes (OLGs) in the form of glial cytoplasmic inclusions [19], although how these are formed and their effect on neurons is not understood. MSA is characterised by very rapid decline, with death occurring ~10 years of onset of symptoms [20]. Although the aetiology of MSA remains elusive, efforts have been made to uncover genetic associations [21]. An interesting phenomenon in the activity of α-synuclein in MSA is that strains of α-synuclein (which are thought to be specific to MSA [22]) have been shown to have seeding abilities, which is the ability of misfolded proteins to induce aggregation in other functioning proteins, thus allowing rapid spread [23].

Parkinson's disease (PD) is also characterised by pathological inclusions of α-synuclein, which leads to neuronal loss in the substantia nigra in the basal ganglia, leading to loss of dopamine production [24]. Several genes have been implicated in PD, including *SNCA,* which codes the protein α-synuclein [25]. As in MSA, it has been suggested that seeding of α-synuclein may occur in PD [26]. Mechanistically, much research has implicated mitochondrial dysfunction in the pathology of the disease [27], as well as neuroinflammation [28]. The disease is classified as a movement disorder, with loss of neurons leading to motor impairment, however there are many non-motor symptoms, including dysregulation of sleep and mood disorders [29], and often dementia later on in disease progression[30].

Amyotrophic lateral sclerosis (ALS) is a neurodegenerative disease characterised by degeneration of motor neurons, leading to weakness of limbs and muscle wastage [31]. In the large majority of cases, ALS is a TDP-43 proteinopathy [15], where TDP-43 becomes mislocalized from the nucleus to the cytoplasm and forms insoluble aggregates [32]. ALS and FTLD-TDP are thought to be two ends of a disease spectrum, as there are overlapping clinical, genetic and pathological features. In terms of genetics, several genes are associated with both FTLD and ALS, including *C9orf72, TARDBP* (coding for TDP-43), *FUS* and *GRN* [33]. Mechanistic associations are also present with both ALS and FTLD, for example, associated with dysregulation of RNA pathways [34,35]. As with other neurodegenerative diseases, pathophysiology is complex and not fully understood, although mechanisms implicated include, along with dysregulated RNA metabolism, impaired protein homeostasis and oxidative stress responses, among others [36].

Figure 1.1. Features of neurodegenerative disorders



A) Key pathological hallmarks of neurodegenerative disease examined in this thesis, including protein mislocalization, protein aggregation, and neuronal loss and degeneration. **B)** Heterogeneity of neurodegenerative disease, non-exhaustive list illustrating hallmark clinical symptoms, brain regions affected and pathological protein aggregates. **C)** Overview of subclassification of frontotemporal lobar degeneration (FTLD) into FTLD-tau and FTLD-TDP based on underlying protein pathology relevant for this thesis. FTLD: frontotemporal lobar degeneration, AD: Alzheimer's disease, FTD: frontotemporal dementia, MSA: multiple system atrophy, PD: Parkinson's disease, PSP: Progressive supranuclear palsy, ALS: amyotrophic lateral sclerosis.

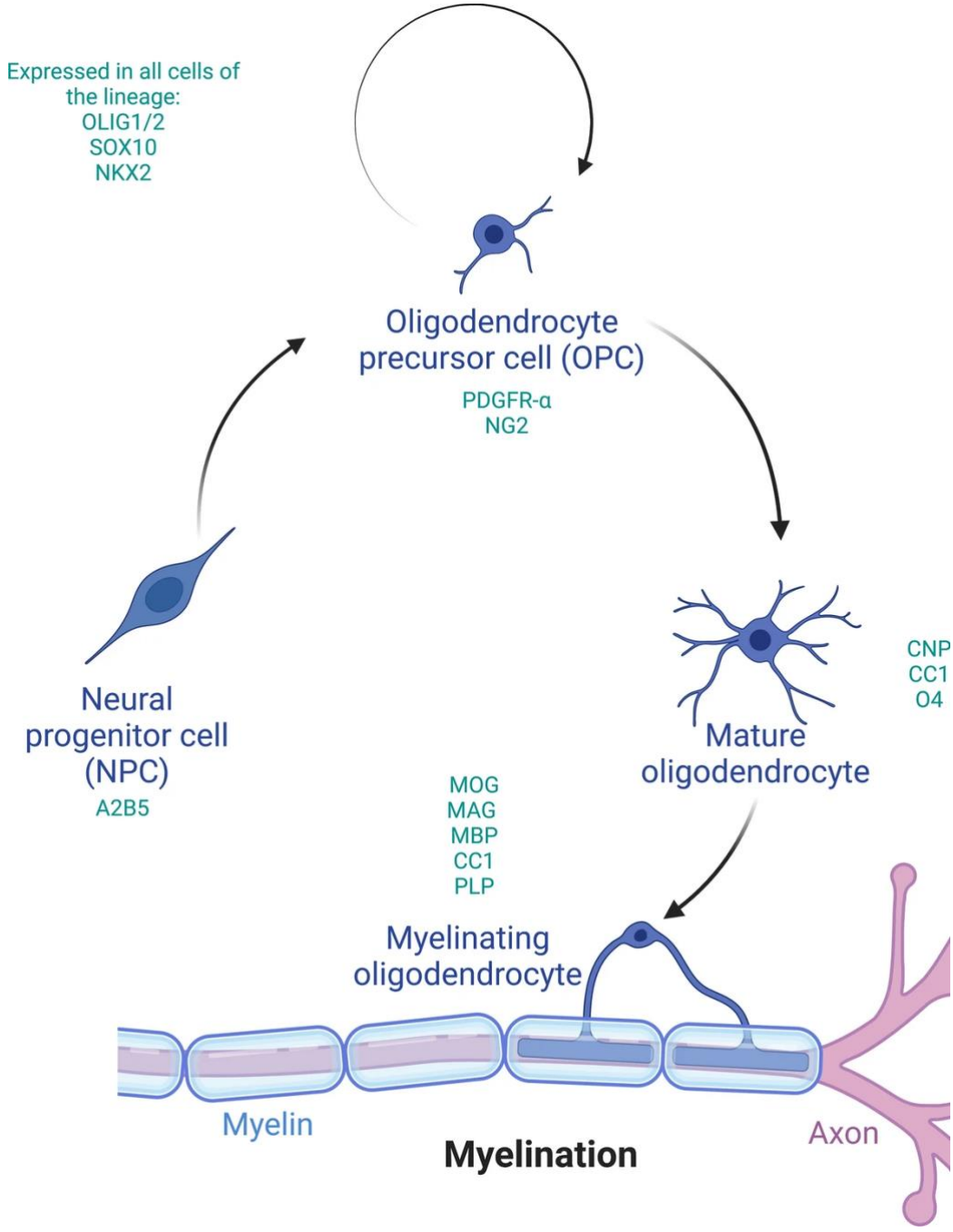
1.2 Oligodendrocyte lineage cells and myelin

OLGs are a major glial cell type in the CNS, which constitute around 75% of the CNS glial cell population [37]. OLGs are responsible for the production, stability, and maintenance of myelin [38], the lipid-rich, multilamellar membrane which wraps around axons and enables fast transmission of electrical signals. Structurally, the myelin sheath is an extension of the OLG plasma membrane that wraps around nerve axons in a concentric fashion [39]. The myelin sheath is not continuous along the neuron. Sections of myelinated axon are separated by nodes of Ranvier. These enable saltatory conduction, the 'hopping' of electrical impulses along axons, which allows for fast transmission of electrical signals. The importance of the myelin sheath is demonstrated by the consequences of its loss, notably in demyelinating diseases such as multiple sclerosis (MS), where it results in a range of neurological symptoms including visual, motor and sensory problems, with associated disability and reduced life expectancy [40]. OLGs are also involved in homeostasis, trophic support to neurons, provision of lactate to neurons, and the secretion of various growth factors [41].

OLGs arise from oligodendrocyte precursor cells (OPCs), which are characterised by the expression of PDGFR-α (platelet derived growth factor receptor α) and NG2 (neuron-glial antigen 2) [42]. It is known that OPCs, which arise in the ventricular zone during early development [43–45], proliferate and migrate, and differentiate in stages into myelinating OLGs [46] (**Figure 1.2**). Although most OPCs differentiate to form myelinating OLGs, some OPCs are retained in their proliferative stage. This results in OPCs accounting for 5–10% of all adult brain cells [47]. The main role of adult OPCs is thought to be the provision of a source of new mature, myelinating OLGs. However, recent studies show they also have other important roles, including their involvement in cell signalling, metabolic regulation and as immune modulators [48–50]. The maturation of OPCs into OLGs, though relatively well characterised in mice, is not well

described in humans (in health and/or disease) in part due to technical challenges of studying post-mortem brain tissue and/or limitations of current human OLG-like cell lines.

Figure 1.2 Schematic representation of the stages of OLG lineage differentiation



Schematic representation of the stages of OLG lineage differentiation. OPCs (PDGFRα high/NG2+) arise from NPCs (A2B5+), before forming mature OLGs (O4+/CNP+/CC1+) and then myelinating OLGs (MOG+/MAG+/MBP+/CC1+/PLP+). NPC: Neural progenitor cell, OLG: Oligodendrocyte, OPC: Oligodendrocyte precursor cell. Figure created with BioRender and reproduced from Fodder et al. [51]

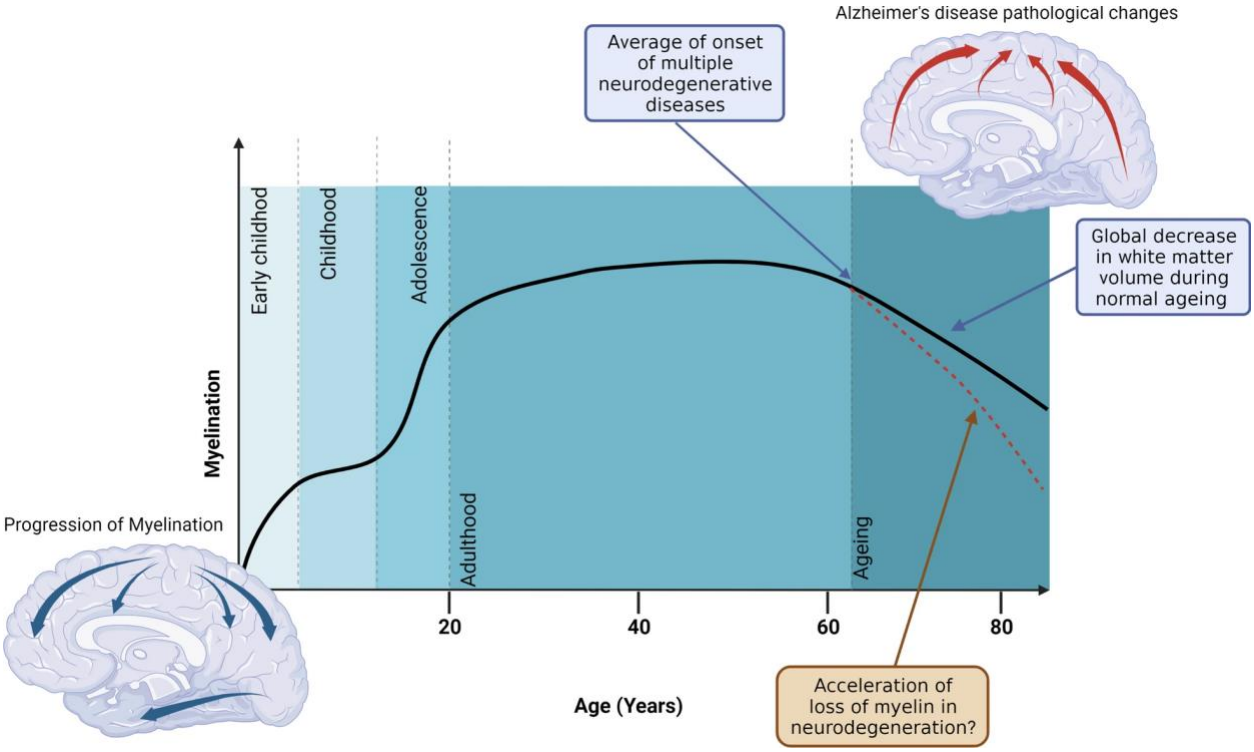
1.3 The involvement of myelin changes in neurodegeneration

Myelination is a dynamic process that continues throughout life. Most myelination takes place from early childhood through to adolescence, with the major part taking place in the first two years of life. However, myelination does continue into adulthood, followed by an age-related decline in myelination occurring around the sixth decade of life (**Figure 1.3**) [52]. Decreases in myelin with ageing are not uniform, with regions of the brain that are myelinated earlier in development (such as the primary motor and sensory regions) undergoing white matter decline later [37].

Although neurodegenerative diseases such AD are mainly associated with grey matter and neuronal damage, there is evidence for decline and involvement of white matter during disease progression. Disruption of myelin in AD was described at the beginning of the twentieth century by Alois Alzheimer [53]. It has also been noted that the typical age-of-onset of neurodegenerative diseases coincides with the time when age-related decline in myelination is observed (**Figure 1.3**) [52]. Moreover, early evidence of the disruption of myelin in AD suggested those regions of the cortex, such as the temporal and frontal lobes, that are myelinated later in development are more likely to present with AD pathology earlier [53,54]. This suggests that those regions that myelinate later are more vulnerable to pathogenic mechanisms which result in neurodegeneration. Further evidence of this involvement of myelin comes from the observations of white matter changes in brain imaging studies [55–57]. For example, white matter hyperintensities (WMHs), which are associated with loss of myelin integrity, have been shown to predict incident AD [55–57]. Brain imaging data has indicated that β-amyloid deposition may change white matter microstructure in early disease stages [58]. A study investigating causation versus causality in the link between β-amyloid deposition and myelin changes used a mouse

model to show that changes in myelination induced through knockout of crucial myelin genes enhanced β-amyloid deposition in a 5xFAD model of AD (a transgenic mouse model used to recapitulate β-amyloid pathological features of AD) [59]. This again adds strength to the hypothesis that white matter changes are not consequences of neurodegenerative processes, but are involved in driving pathogenesis. White matter abnormalities and myelin degradation are also described in other neurodegenerative diseases, including MSA [60], ALS [61] and PSP [62,63].

Figure 1.3 Schematic of the myelin changes throughout life



Shaded areas indicating myelination waves (as defined by de Faria et al. [52]). Also depicted are visualizations of the progression of cortical myelination and the progression of Alzheimer's disease related destruction. The average age of onset of multiple neurodegenerative diseases is also indicated and coincides with the start of normal ageing-related decline in myelination, which is hypothesized to be accelerated in neurodegeneration. Figure created with BioRender and reproduced from Fodder et al. [51]

1.4 The involvement of the oligodendrocyte lineage in neurodegeneration

1.3.1 Evidence from pathology

A direct role for OLGs in neurodegenerative disease is exemplified by the pathology of MSA, where, as mentioned earlier, glial cytoplasmic inclusions (GCIs) in OLGs are the pathological hallmark of the disease [64]. In MSA, these inclusions consist of aggregates of the synaptic protein α-synuclein. Whether α-synuclein is produced by the OLGs or propagated from neurons is not clear. In MSA, an increased number of OPCs is also reported in post-mortem brain tissue [65,66]. PSP and corticobasal degeneration (CBD) also display clear OLG pathology with disease hallmarks including tau deposits in OLGs, presenting as coiled bodies [7,17].

Although the precise role of OLGs in AD pathology is less clear, there is evidence from human post-mortem studies that there are alterations in the numbers and morphology of OLG lineage cells in this disease [57]. In post-mortem AD brain tissue, decreases in Olig2 + cells have been reported [67], as well as increased numbers of OPCs in white matter lesions [68]. Morphological changes in OLGs derived from AD post-mortem brains have also been seen, specifically a decrease in nuclear diameter in parahippocampal white matter [69]. A recent study, in apolipoprotein E-ε4 allele (*APOE*-ε4) carriers, also demonstrated aberrant deposition of cholesterol in OLGs and dysregulated myelination in AD [70].

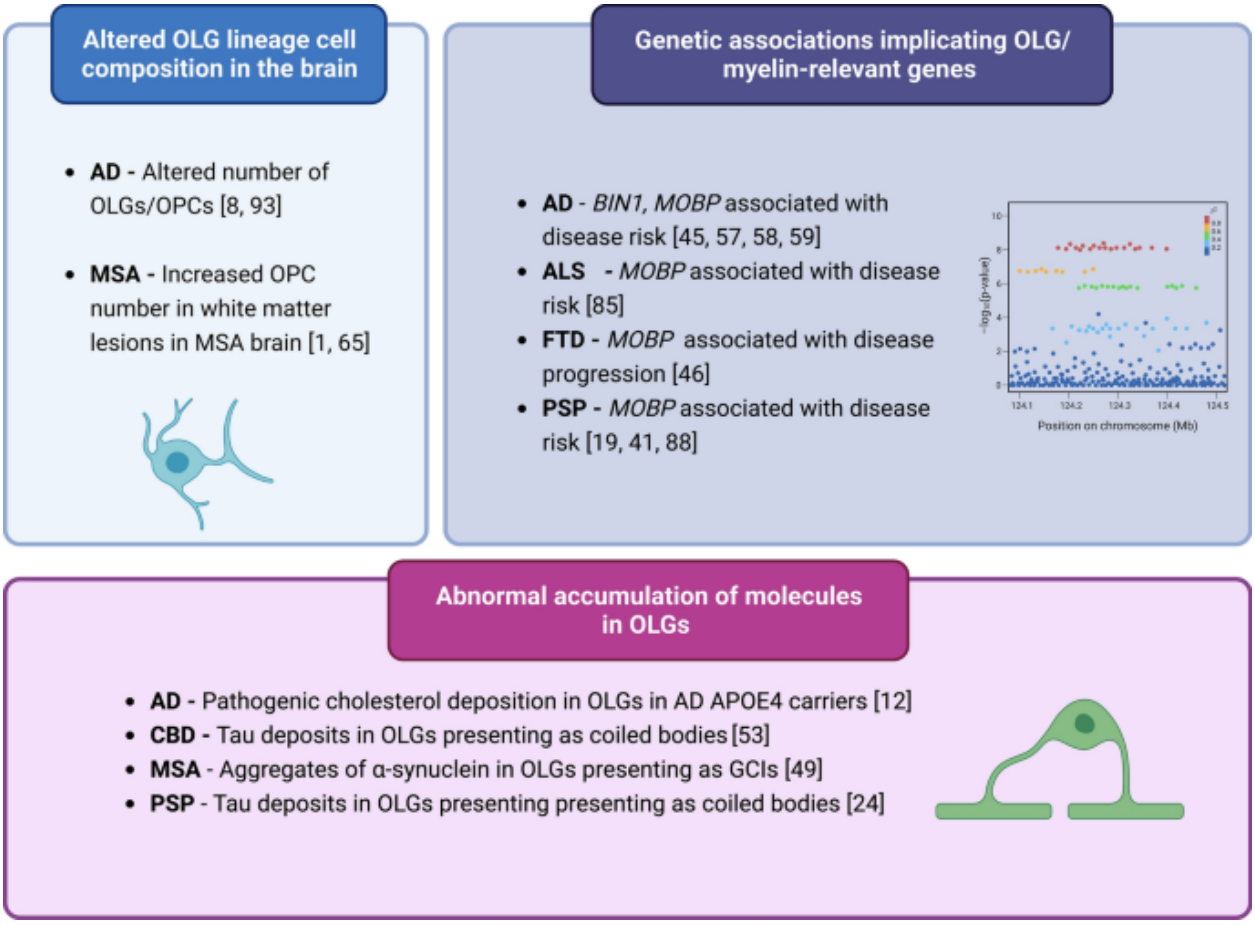
1.3.2 Evidence from genetics

GWAS have also implicated specific myelin/OLG-related genes in neurodegeneration, including the bridging integrator 1 (*BIN1*) gene, which is the second strongest genetic risk factor for late

onset AD (LOAD) [71–73] and known to be largely expressed by mature OLGs and localised to white matter tracts [74]. Increased expression of *BIN1* is reported in AD [75], although mechanisms behind the association of *BIN1* and AD are unclear. The myelin associated oligodendrocyte protein (*MOBP*) gene has been associated with disease risk in several neurodegenerative diseases, including PSP [76–79], CBD [80], AD *APOE*-ε4 carriers [81], ALS [82] and PD [83], and has also been reported to be associated with white matter degradation and increased rates of decline in executive function in behavioural variant frontotemporal dementia [84]. The functional repercussions of such associations remain unclear. However, in human brain tissue, the risk allele T, of the disease-associated single nucleotide polymorphism rs1768208, is also associated with increased expression of the *MOBP* gene in PSP [85].

Aside from such examples of myelin/OLG relevant genes identified through GWAS, transcriptomic analyses reveal gene expression changes in additional myelin-related genes in a broad range of neurodegenerative diseases, including AD [86], PSP [86], MSA [87], and frontotemporal lobar degeneration (FTLD) [88], further supporting the idea of myelination changes as a common pathway across these diseases. Examples of evidence supporting the importance of OLG/OPC involvement across several neurodegenerative diseases are given in **Figure 1.4**.

Figure 1.4 Non-exhaustive summary of evidence implicating the oligodendrocyte lineage across neurodegenerative diseases



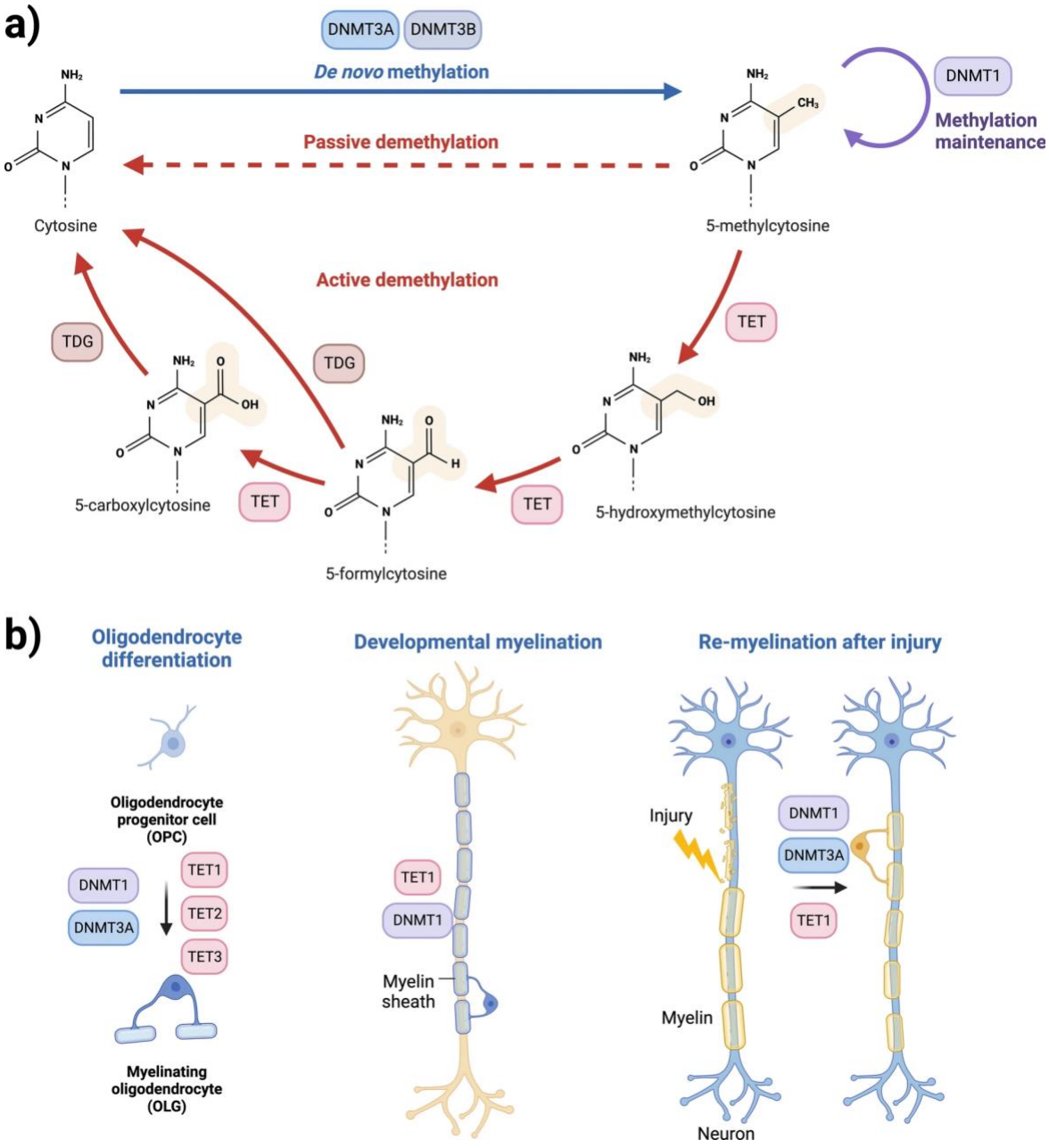
AD: Alzheimer's disease, ALS: Amyotrophic lateral sclerosis, *BIN1*: Bridging Integrator 1, CBD: Corticobasal degeneration, FTD: Frontotemporal dementia, GCI : Glial cytoplasmic inclusion, MSA: Multiple System Atrophy, *MOBP*: Myelin-associated oligodendrocyte protein, OLG: Oligodendrocyte, OPC: Oligodendrocyte precursor cell, PSP: Progressive supranuclear palsy.Figure created with BioRender and reproduced from Fodder et al. [51]

1.5 DNA Methylation

Epigenetic modifications are those which, without altering the underlying DNA sequence, bring around changes in gene expression. DNA methylation, the most widely studied epigenetic modification, involves the transfer of a methyl group to a cytosine nucleotide to form 5-methylcytosine (5mC) (**Figure 1.5**)[89]. This transfer is catalysed by a family of enzymes called DNA methyltransferases. The effect of DNA methylation on gene expression regulation is largely dependent upon genomic location [89,90]. For example, methylation within the gene body, i.e. protein coding exons and introns, more often results in increased gene expression, whereas methylation in the promoter region frequently leads to decreased gene expression [91–93]. DNA methylation, along with other epigenetic modifications, allow the intricate spatiotemporal control of gene expression and is crucial both during development and adult life.

DNA methylation has been implicated in many processes relevant for the brain, including in brain development, learning, memory, and brain cell-type specification [94]. DNA hydroxymethylation (5hmC), an oxidative derivative of DNA methylation (**Figure 1.5**), is also important. Having originally been presumed to be an intermediate mark before demethylation [95], evidence now supports a functional role for 5hmC [96]. Interestingly, the distribution of different methylation states varies in a tissue dependent manner, with 5hmC known to be enriched tenfold in the human brain compared to peripheral tissues [97]. Distribution of 5mC and 5hmC between brain cell-types has also been reported to be variable, with studies indicating that 5hmC may be enriched in neuronal cells compared to OLGs [98,99].

Figure 1.5 Schematic representation of the DNA modifications cycle, including factors responsible for the transitions between states



A) Overall DNA modifications cycle; B) Diagram summarising the known involvement of DNMTs and TET enzymes in OLG differentiation, developmental myelination, and in remyelination in response to injury. Evidence shows an age-dependent role for DNMT1 and DNMT3A, with the former suggested to be more important in developmental myelination, and the latter in the remyelination involving differentiation of adult OPCs [100,101]. Whilst it has been

suggested that all three TET enzymes are involved in oligodendrocyte differentiation [39], TET1 has been reported to be more important for myelination and remyelination after injury [95,102]. *DNMT1/3A/3B*:DNA methyltransferase 1/3A/3B, *TET*:ten-eleven translocation enzymes, *TDG*:thymine DNA glycosylase. Figure created with BioRender and reproduced from Fodder et al. [51]

1.5.1 DNA methylation in neurodegenerative diseases

Although epigenetic modifications such as DNA methylation allow for the control of gene expression that is fundamental to many cellular processes, changes in DNA methylation have also been associated with several diseases. Through hypermethylation or hypomethylation, alterations in DNA methylation may lead to changes in normal expression of genes, either through silencing or over-activation. Such effects may then lead to pathological outcomes.

Studies using immunodetection of 5mC or 5hmC have often failed to lead to consensus regarding the occurrence of global DNA methylation/hydroxymethylation changes in neurodegenerative diseases, possibly reflecting limitations of such techniques [103–105]. However, technological advances that allowed querying throughout the genome at specific sites, have empowered investigations of relevant candidate genes and epigenome-wide association studies (EWAS) to identify DNA methylation alterations in neurodegenerative diseases at single nucleotide resolution. In AD, EWAS studies utilising bulk brain tissue have identified multiple genes with DNA methylation changes associated with the disease and its pathological burden [106–109], and meta-analyses have identified significant changes across multiple brain regions [110–113]. Differentially methylated genes have also been identified in bulk brain tissue EWAS in other neurodegenerative diseases such as PD [114–116], PSP [117], MSA [118], FTLD [119] and Huntington's disease [120,121]. DNA methylation changes in AD and movement disorders (including PD, HD, PSP and MSA) have been reviewed by Smith et al. [111]and Murthy et al. [104], respectively.

Whilst most DNA methylation studies employ ‘bulk’ tissue analysis, more recent studies exploring DNA methylation changes in neurodegeneration have investigated cell-specific alterations. Gasparoni et al. carried out DNA methylation profiling on glial and neuronal fractions of nuclei in AD and control, and identified glial specific changes in DNA methylation[122]. Shireby et al also investigated AD-related DNA methylation signatures in purified brain nuclei and found that many AD-related DNA methylation changes that had previously been detected in ‘bulk’ tissue were indeed driven by changes in non-neuronal cells, including in OLGs [110]. This finding highlights the need for a deeper understanding of DNA methylation changes in OLGs which may also occur in AD and other neurodegenerative diseases. Below, we discuss several lines of evidence that support DNA methylation having a role in the dysfunction of OLGs and myelin in neurodegenerative diseases. These are summarised in **Figure 1.6**.

1.6 DNA methylation in oligodendrocyte dysfunction in neurodegenerative diseases

1.6.1 DNA methylation plays a crucial role in determining the fate of OPCs in health and neurodegenerative diseases

Gene expression changes determined by DNA methylation play an important role in the process of lineage specification from OPCs to mature OLGs [100,123]. Although most studies investigating the OLG life cycle have been conducted in animal models, there is significant evidence to suggest that DNA methyltransferases and DNA methylation are dynamic in the processes of OPC specification, survival, proliferation, differentiation, and myelination [100,123–126]. Proliferation of OPCs occurs in response to exogenous signals such as growth factors, and epigenetic modifications are important players in this regulation. In mice, ablation of DNA methyltransferases has been shown to result in a hypomyelinating phenotype through reduction in the OPC progenitor pool due to impaired OPC proliferation [124]. During differentiation of OPCs to OLGs in mice, decreased DNA methylation levels in myelin genes and increased methylation levels in cell cycle and neuronal genes were reported [124]. Given that the majority of the DNA methylation sites (CpGs) investigated in these studies were in promoter regions, and given the association of promoter region DNA hypermethylation with decreased gene expression, this supports an important role for DNA methylation in silencing cell cycle and proliferation genes and in activating myelin genes, thus enabling OPCs to leave their proliferative state and differentiate into myelinating OLGs. DNA methylation of specific genes involved in OPC differentiation has also been described. The DNA-binding protein inhibitors *Id2* and *Id4* showed

decreased expression during OPC differentiation in mice, which was correlated with hypermethylation of their promoter regions, suggesting a role for DNA methylation in the silencing of these genes to allow the expression of myelin genes during differentiation [126]. As described in MSA and AD brain tissue, the increased number of OPCs observed [41,65,68] could be reflective of an inability of the OPCs to mature and differentiate into myelinating OLGs, possibly in part due to defective DNA methylation. However, further investigations are required to shed light on such effects. It is also worth noting that DNA methyltransferases DNMT1, DNMT3a and DNMT3b have been shown to have distinct roles in various aspect of the OLG lineage cell cycle, myelination, and in remyelination after injury (**Figure 1.5 B**) [100,101,123–125].

DNA hydroxymethylation is also dynamic during the OLG life cycle, and TET1 is one of the enzymes involved in this process (**Figure. 1.5**). Slower cell cycle progression of OPCs was found in *Tet1* knock-out mice, an effect that appeared to be largely specific to the OLG lineage compared to neurons and astrocytes [96]. TET1 was also found to be implicated in processes of myelin repair through the regulation of genes important for the axon-myelin interface [102]. Interestingly, there is increased hydroxymethylation in adult OPCs compared to neonatal OPCs in mice, and evidence suggests that TET1 is essential for myelin repair after damage [102]. As with DNMTs, there is evidence for distinct and complex roles of the TET enzyme family in different aspects of the OLG lineage cell life cycle (**Figure 1.5 B**) [96,102,127]. Overall, these studies indicate that DNA modifications undergo dynamic changes between neonatal and adult OPCs and are relevant for the decrease in myelinating capacity that is observed in ageing OPCs [128].

More research is needed to elucidate further the importance of 5mC and 5hmC in OLGs, and to understand the complex roles of DNMTs and the TET family of enzymes. This should include investigation of changes in their catalytic activities, during OLG differentiation, in myelination, and in remyelination after injury.

1.6.2 Oligodendrocyte-related genes are differentially methylated in neurodegenerative diseases

As discussed above, OLG-related genes (e.g. *MOBP* and *BIN1*) have been associated in GWAS with the risk of developing neurodegenerative diseases. In addition, *MOBP* was shown to present aberrant DNA methylation in an EWAS of post-mortem MSA white matter [118]. Specifically, hypermethylation (i.e. increased methylation levels) of the promoter region of the gene was detected in MSA compared to controls. DNA methylation changes in *MOBP* were found even in brain regions very mildly affected by MSA pathology (e.g. occipital lobe), indicating that these may reflect early changes and contribute to disease pathogenesis. The methylation status of *MOBP* in MSA is linked to changes in *MOBP* expression levels [129], and the observed downregulation of this gene is likely driven by the hypermethylation of its promoter region. As MOBP protein is involved in the morphological differentiation of OLGs [130], changes in its expression levels due to aberrant methylation during OLG differentiation likely lead to functional impairment of these cells. As another example, *BIN1* is the second strongest genetic risk factor for late onset AD [71–73] and associations between AD neuropathology and the level of methylation at the *BIN1* locus have also been reported [106,131], with *BIN1* transcript levels being associated with β-amyloid load [131]. Given this, and that *BIN1* has been shown to be predominantly localised to white matter in the brain and expressed primarily in mature OLGs [74], it is reasonable to hypothesise that the involvement of *BIN1* gene in disease processes may be mediated through DNA methylation changes affecting OLGs.

1.6.3 Oligodendrocyte cell types and epigenetic age acceleration

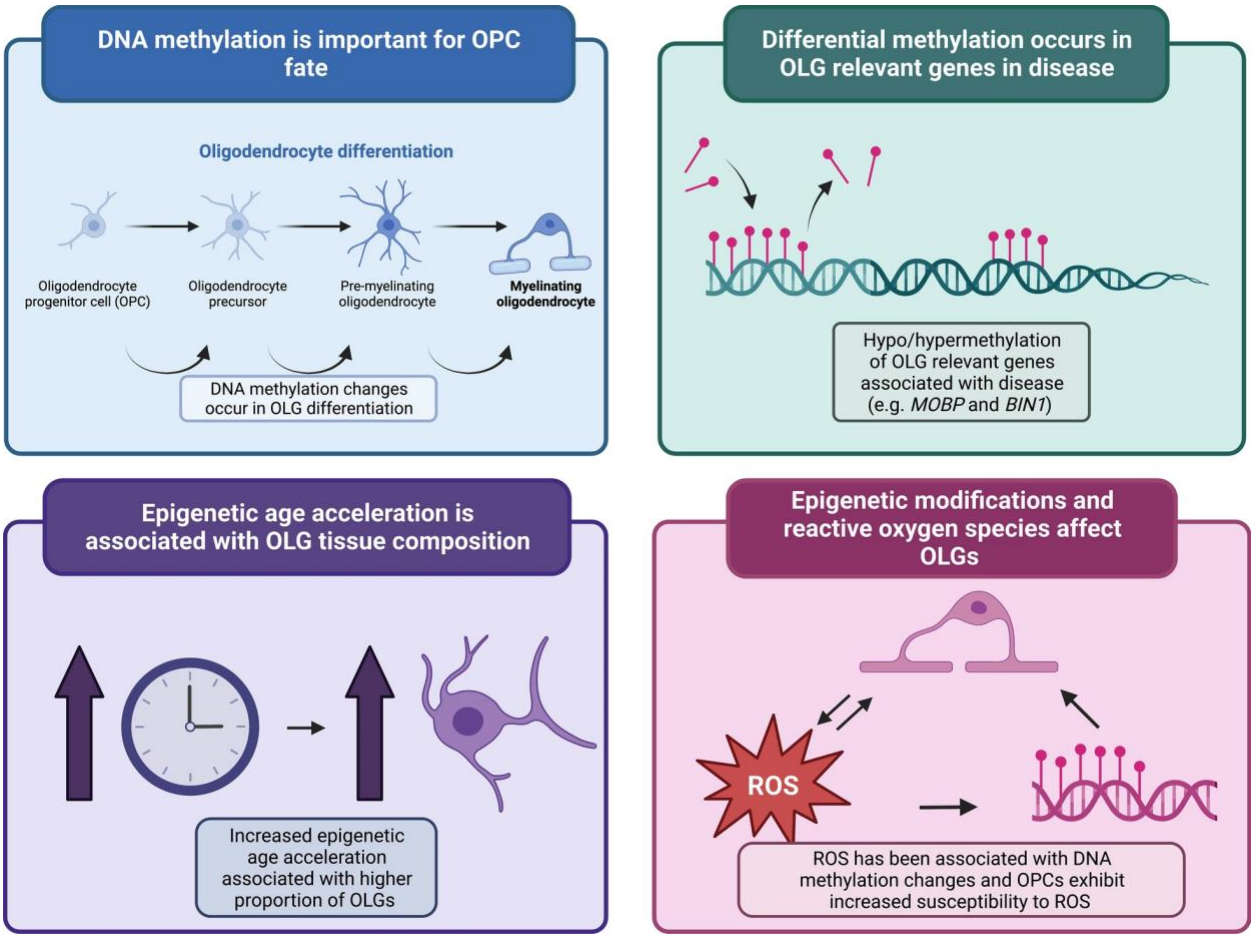
DNA methylation changes are known to occur during ageing, which is the major risk factor for neurodegeneration, with accelerated epigenetic ageing, as measured using epigenetic clocks, being reported in neurodegenerative diseases [120,132–134]. As epigenetic clocks allow estimations of biological ages based on the DNA methylome, it is then possible to infer the difference between the biological age, and the actual chronological age, i.e. epigenetic age acceleration. Age-related changes have been described in OLGs, notably the decrease in myelinating capacity with increased age [128], but it has also been suggested that there is a loss of epigenetic memory in these cells [135]. It has been proposed that intrinsic changes observed in ageing OLGs could be a result of changes on gene expression brought around by an altered epigenetic profile [135]. Indeed, a recent study investigating DNA methylation-based measures of accelerated ageing in post-mortem tissue from different brain regions in MSA and controls found that the relative frequency of OLGs in brain tissue is positively correlated with epigenetic age acceleration, which is opposite to that found in other brain cell types [136]. This relationship between OLG proportions and epigenetic age acceleration has also been found in some forms of FTLD [136]. These findings support a role for OLGs in pushing towards increased epigenetic/biological age, suggesting that this cell lineage ages faster than other brain cell types.

1.6.4 Vulnerability of oligodendrocyte lineage cells to reactive oxygen species via epigenetic modifications

A further role of OLGs (aside from myelination) is their involvement in iron equilibrium in the CNS. Iron is key for normal CNS function [137], and OLGs are important in maintaining brain iron homeostasis [138]. Dysregulation, and, specifically, increased iron levels in the brain are associated with neurodegenerative diseases such as AD, PD, and MSA [139]. Proposed mechanisms for the role of iron in neurodegeneration include increased oxidative stress, possibly due to enhanced generation of reactive oxygen species (ROS) associated with increased protein aggregation [41,139]. Given that OLGs are the principal iron-containing cells of the brain [140], it is reasonable to hypothesise that aberrant OLG function could contribute to neurodegeneration via dysregulation of brain iron levels. Indeed, investigations into brain gene expression in the context of neurodegeneration with brain iron accumulation (NBIA), and in mouse models of increased brain iron loading, implicate OLGs and myelin-related genes [141,142]. This may well be relevant for other neurodegenerative diseases. OPCs and OLGs are thought to be more vulnerable to oxidative stress than other brain cell types due to factors which include lower levels of antioxidant enzymes and free radical scavengers [143,144], as well as their high metabolic requirements [128,145]. Excessive oxidative stress can lead to OLG malfunction through the impairment of effective OLG differentiation [146], with such effects having been reported in neurodegenerative diseases [147]. Interestingly, a link between DNA methylation changes and presence of ROS has been suggested with the finding that increased ROS leads to oxidation of 5mC into 5hmC [148], likely leading to changes in gene expression regulation. Although speculative, it could be hypothesised that this proposed increase in susceptibility of OPCs to ROS-induced damage compared to other brain cell types could, at least in part, be driven by ROS induced alterations in DNA methylation in these cells [143,144,146]. However, causal

relationships between neurodegenerative processes, OPC dysfunction, DNA methylation and ROS are still unclear and require further investigations.

Figure 1.6 Potential roles for DNA methylation in the dysfunction of oligodendrocytes and myelin in neurodegenerative diseases



Panels illustrate different lines of evidence that implicate DNA methylation changes affecting OLGs/OLG-related genes and their relevance to neurodegeneration. *BIN1*:Bridging interactor 1, *MOBP*:Myelin associated oligodendrocyte protein, *OPC*:Oligodendrocyte precursor cell, *OLG*:Oligodendrocyte, *ROS*:Reactive oxygen species. figure created with BioRender and reproduced from Fodder et al. [51]

1.7 Thesis aims

From what is described above, there is mounting evidence showing that the efficient development, proliferation, differentiation, and maintenance of the OLG lineage cells may be disrupted in neurodegenerative diseases, and that aberrant DNA methylation may be implicated (**Figure 1.6**). However, the role of OLG lineage cells in the context of these diseases has not been given the attention it deserves. The overarching aim of the work I have carried out during my PhD is to investigate, using complementary computational approaches and a large number of DNA methylation datasets, the role of DNA methylation in dysregulation of OLG lineage/myelin relevant genes across neurodegenerative diseases.

The specific aims of this thesis are therefore:

1. To investigate differentially methylated genes relating to OLG lineage across neurodegenerative diseases, using network analysis as a complementary approach to uncover signatures of dysregulated DNA methylation associated with OLGs/OPCs across neurodegeneration (explored in Chapters 3 and 4).
2. To investigate DNA methylation during OLG differentiation and in healthy tissue to explore consequences of disease associated DNA methylation changes identified in Aim 1 (explored in Chapter 5).

3. To carry out a detailed investigation of the *MOBP* gene as an important OLG gene that has been implicated in multiple neurodegenerative diseases and found to be differentially methylated in MSA (explored in Chapter 6)

Chapter 2: General methods

In this chapter, we describe methods which are applicable across multiple chapters, including DNA methylation and gene expression datasets used throughout this thesis, detailed demographics of samples across these datasets, and general quality control and processing of these data. We also described the OLG and OPC gene lists which were used across this work to identify OLG/OPC relevant DNA methylation changes across neurodegeneration.

2.1 DNA methylation datasets and processing

2.1.1 Datasets used

Throughout this thesis, we describe the use of multiple DNA methylation datasets that were previously generated, derived from both bulk brain tissue, and from the glial fraction of sorted brain-nuclei samples. The datasets we analysed with disease/subtypes, brain regions, and a brief characterization of the samples is included in **Table 2.1**.

For FTLD1 (*N* = *23*), sorted-FTLD (*N* = *25*), and AD2-sorted (*N* = 11) all post-mortem tissue originated from brains donated to the Queen Square Brain Bank archives, where tissue is stored under a licence from the Human Tissue authority (No. 12198). Both the brain donation programme and protocols have received ethical approval for donation and research by the NRES Committee London—Central. All cases were characterised by age, gender, disease history (including disease onset and duration) as well as neuropathological findings. For FTLD2 (*N* = 48), all post-mortem tissues were obtained under a Material Transfer Agreement from the Netherlands Brain Bank, and MRC King's College London, as described by Menden et al. [149]. For FTLD3 (*N* = 163, after quality control), data made available by Weber et al. [117] was retrieved

from Gene Expression Omnibus (GEO) under accession code GSE75704. For AD1-sorted (*N* = *31*), data was made available by Gasparoni et al. [122], and retrieved from GEO under accession code GSE66351. For AD (*N* = *530*), data was made available by de Jager et al. [106] and was retrieved from Synapse (synapse ID syn7357283). For the AD brain region dataset (utilised in **Chapter 4**) data was made available by Semick et al.[150] and retrieved from GEO under accession code GSE125895. This dataset comprised samples across multiple brain regions, including cerebellum (CRB) (*N* = 67), hippocampus (HPPO) (*N* = 65), entorhinal cortex (ERC) (*N* = 69) and the dorsolateral prefrontal cortex (DLPFC) (*N* = 67). The brain regions from which the DNA methylation profiles were derived are detailed in **Table 2.1** and visualised in **Figure 2.1**.

Table 2.1 Pathological and demographic characteristics of the FTLD and AD cohorts used in this thesis for DNA methylation analyses.

Cohort	**Pathological FTLD subtypes and controls (N after quality control)**	**Mean age ± SD (years)**	**Sex**	**Regression models used for cohort-specific EWAS**	**Analyses this data was used for (corresponding Thesis Chapters)**	**Dataset Reference**	**Illumina Beadchip**
FTLD1 (Frontal Cortex)	FTLD (*N* = 15)	70.07 ± 5.59	7M/8F	*~ 0 + disease + age + sex + SOX10⁺ proportions + Double⁻ proportions + array* (0 surrogate variables detected)	EWAS and Network analysis (Chapters 3 and 4)	Fodder et al.[119]	Illumina Infinium MethylationEPIC (850K)
	FTLD-TDP type A (*C9orf72* mutation carriers, *N* = 7)	66.86 ± 4.85	3M/4F				
	FTLD-TDP type C (sporadic, *N* = 8)	72.88 ± 4.79	4M/4F				
	Controls (*N* = 8)	75.75 ± 5.63	3M/5F				
FTLD2 (Frontal Cortex)	FTLD (*N* = 34)	63.18 ± 7.92	14M/20F	*~ 0 + disease + age + sex + SOX10⁺ proportions + Double⁻ proportions + array + slide* (0 surrogate variables detected)	EWAS and Network analysis (Chapters 3 and 4)	Menden et al.[149]	Illumina Infinium HumanMethylation450 (450K)
	FTLD-TDP type A (*GRN* mutation carriers, *N* = 7)	65.57 ± 7.63	2M/5F				
	FTLD-TDP type B (*C9orf72* mutation carriers, *N* = 14)	64.57 ± 8.41	5M/9F				
	FTLD-tau (*MAPT* mutation carriers, *N* = 13)	60.92 ± 7.60	7M/6F				
	Controls (*N* = 14)	78.43 ± 11.76	5M/9F				

FTLD3 (Frontal Cortex)	FTLD (*N* = 93)	71.16 ± 5.32	54M/39F	*~ 0 + disease + age + sex + SOX10*$^{+}$ *proportions + Double*$^{-}$ *proportions + array + slide + surrogate variable* (1/1 surrogate variables detected)	EWAS and Network analysis (Chapters 3 and 4)	Weber et al.[117]	Illumina Infinium HumanMethylation450 (450K)
	FTLD-Tau (sporadic PSP)						
	Controls (*N* = 70)	76.17 ± 7.93	45M/25F				
FTLD-sorted (Frontal Cortex)	FTLD (N = 19)	67.3 ± 12.2	10M/9F	*~ 0 + disease + age + sex + NEUN*$^{+}$ *proportions + array + surrogate variables (4/4 detected surrogate variables)*	EWAS and Network analysis (Chapters 3 and 4)	Bettencourt lab, unpublished	Illumina Infinium HumanMethylation450 (450K)
	FTLD-TDP type A (*C9orf72* mutation carriers, *N* = 7)	64 ± 9.38	3M/4F				
	FTLD-TDP type C (sporadic, *N* = 5)	70.2 ± 3.96	3M/2F				
	FTLD-TDP type A (sporadic, *N* = 7)	68.5 ± 9.38	4M/3F				
	Controls (*N* = 6)	59.5 ± 12.2	3M/3F				
AD (Dorsolateral Prefrontal Cortex)	AD (*N* = 201)	85.4 ± 5.39	120M/81F	*~0 + disease + slide + bs_conversion + NEUN*$^{+}$ *proportion + Double*$^{-}$ *proportion + array + PC1+ PC2*	EWAS, Methylation/Gene expression correlations, Methreg (Chapters 3 and 5)	De Jager et al.[106]	Illumina Infinium HumanMethylation450 (450K)
	Controls (*N* = 329)	87.3 ± 3.91	212M/117F				
AD1-sorted (Occipital Cortex)	AD (*N* = 15)	80.27 ± 7.21	5M/10F	*~ 0 + disease + age + NeuN+ proportions + Double*$^{-}$ *proportions + slide + sex*	EWAS and Network analysis (Chapters 3 and 4)	Gasparoni et al. [122]	Illumina Infinium HumanMethylation450 (450K)
	Controls (*N* = 16	69.63 ± 21.3	9M/7F				

AD2-sorted (Frontal Cortex)	AD (*N* = 5)	69 ± 7.97	2M/3F	*~0 + disease + NeuN+ proportions + sex*	EWAS and Network analysis (Chapters 3 and 4)	Bettencourt lab, unpublished	Illumina Infinium HumanMethylation450 (450K)
	Controls (*N* = 6)	59.5 ± 12.2	3M/3F				
AD-HIPPO	AD (*N* = 17)	81.5 ± 9.16	8M/9	*~0+Sample_group+plate+DoubleN+sex+NeuNP+age+array*	Network analysis (Chapter 4)	Semick et al.[150]	Illumina Infinium HumanMethylation450 (450K)
	CTRL (*N* = 48)	61.7 ± 7.66	29M/19F				
AD-DLPFC	AD (*N* = 21)	79.9 ± 9.46	11M/10F	*~0+Sample_group+neuNP+age+sex+plate*			
	CTRL (*N* = 46)	61.3 ± 7.56	27M/19F				
AD-ERC	AD (*N* = 20)	79.7 ± 9.64	9M/11F	*~0+slide+DoubleN+Sox10P+age + 1/1 surrogate variables detected*			
	CTRL (*N* = 49)	61.6 ± 7.61	29M/20F				
AD-CRB	AD (*N* = 24)	79.5 ±10.0	11M/13F	*~0+Sample_group+slide+sex+age+neuNP+doubleN*			
	CTRL (*N* = 43)	60.6 ± 7.04	24M/19F				

FTLD, Frontotemporal lobar degeneration; FTLD-TDP, FTLD with 43 kDa transactive response DNA-binding protein (TDP-43) positive inclusions; FTLD-Tau, FTLD with tau-positive inclusions; PSP, progressive supranuclear palsy; AD, Alzheimer's Disease; SD, Standard deviation; F, Females; M, Males; Double$^-$ proportions, NeuN$^-$/SOX10$^-$ proportions; PC1, 1st principal component; PC2, 2nd principal component, HIPPO; hippocampus, DLPFC; dorsolateral prefrontal cortex, ERC; entorhinal cortex, CRB; cerebellum

Figure 2.1 Brain regions analysed across DNA methylation cohorts

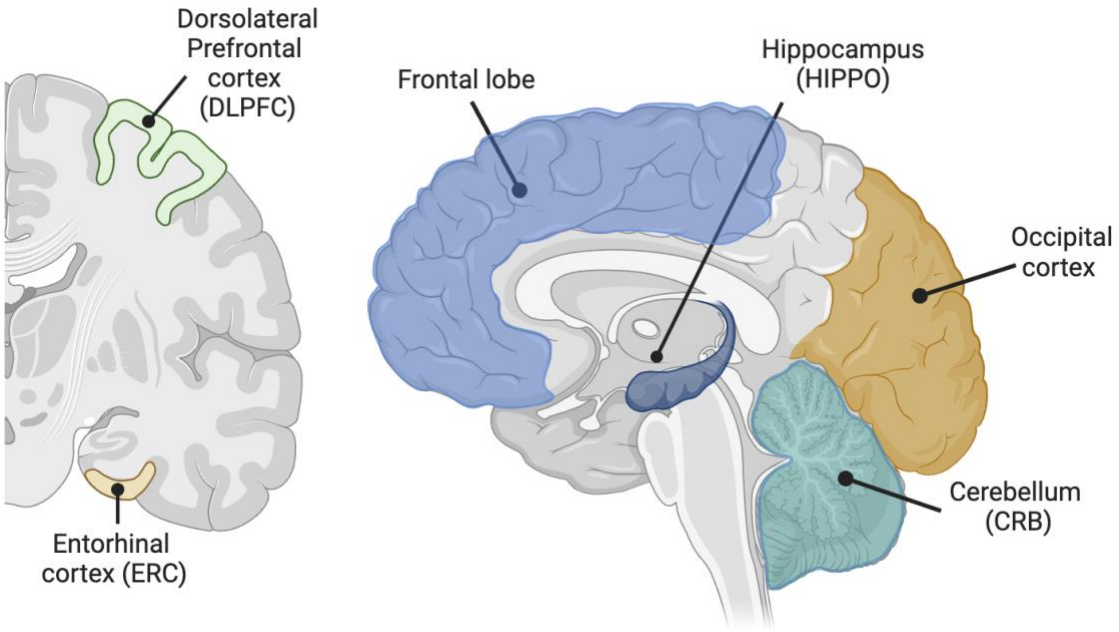


Illustration highlights the brain regions analysed in DNA methylation cohorts. Each labeled region represents a distinct brain area. **Dorsolateral Prefrontal Cortex (DLPFC)** - marked in green. **Hippocampus (HIPPO)** - marked in blue. **Entorhinal Cortex (ERC)** - marked in orange. **Cerebellum (CRB)** - marked in teal. **Occipital cortex** - marked in yellow. **Frontal lobe** - marked in mid blue.

2.1.2 Quality control

DNA methylation profiles that we used throughout this thesis were generated using the Illumina Infinium HumanMethylation450 (450K) or MethylationEPIC v1 (850K) arrays (aside from those detailed in Chapter 5). Briefly, after DNA extraction from samples, bisulfite conversion is carried out, through which unmethylated cytosines are converted to uracil, while methylated cytosines remain unchanged. Bisulfite-converted DNA is then hybridized to illumina beadchip arrays, where methylation level at CpG sites is measured through the use of fluorophore incorporation

based on methylation level. Raw intensity levels are passed as IDAT files, which contain fluorescence intensity values which can then be used to generate beta-values ranging from 0 to 1 (approximately 0% to 100% methylation, respectively). Beta-values are thus used to estimate the methylation levels of each CpG site using the ratio of intensities between methylated and unmethylated alleles. Coverage of CpG sites varies between Illumina Infinium HumanMethylation450 (450K) and MethylationEPIC v1 (850K) arrays. The Illumina Infinium HumanMethylation450 (450K) array covers ~ 450,000 CpG sites across the genome, largely focusing on promoter regions and CpG islands, whilst the Illumina Infinium MethylationEPIC (850K) array covers ~ 850,000 CpG sites including 90% of the CpG sites covered by the IlluminaMethylation450 (450K). The Illumina Infinium MethylationEPIC (850K) array, as well as increased coverage of regions analysed in the IlluminaMethylation450 (450K) array, also features increased coverage of enhancer regions [151]. For the two arrays, 99% of RefSeq genes are probed.

For each of the abovementioned cohorts, the following pipeline was carried out for quality control and data processing prior to its use in various downstream analyses.

All cohorts were subjected to harmonised quality control checks and pre-processing. Briefly, raw data (idat files) were imported and subjected to rigorous pre-processing and thorough quality control checks using minfi [152], wateRmelon [153], and ChAMP packages [154]. Examples of images used in quality control are given in **Figure 2.2.A-B.** The following criteria were used to exclude methylation sites that did not pass quality control checks from further analysis: (1) poor quality, (2) cross reactive, (3) included common genetic variants, and (4) mapped to *X* or *Y* chromosome. In addition, samples were dropped during quality control if: (1) they presented with a high failure rate (≥2% of methylation sites), (2) the predicted sex did not match the

phenotypic sex, and (3) they clustered inappropriately on multidimensional scaling analysis.

Beta-values were normalised with ChAMP using the Beta-Mixture Quantile (BMIQ) normalisation method (**Figure 2.2.C)**, to normalise for the technological factors during DNA methylation profiling. *M*-values, computed as the logit transformation of beta-values, were used for all statistical analysis, as recommended by Du et al. [155], owing to their reduced heteroscedasticity (as opposed to beta-values) and improved statistical validity for differential methylation analysis.

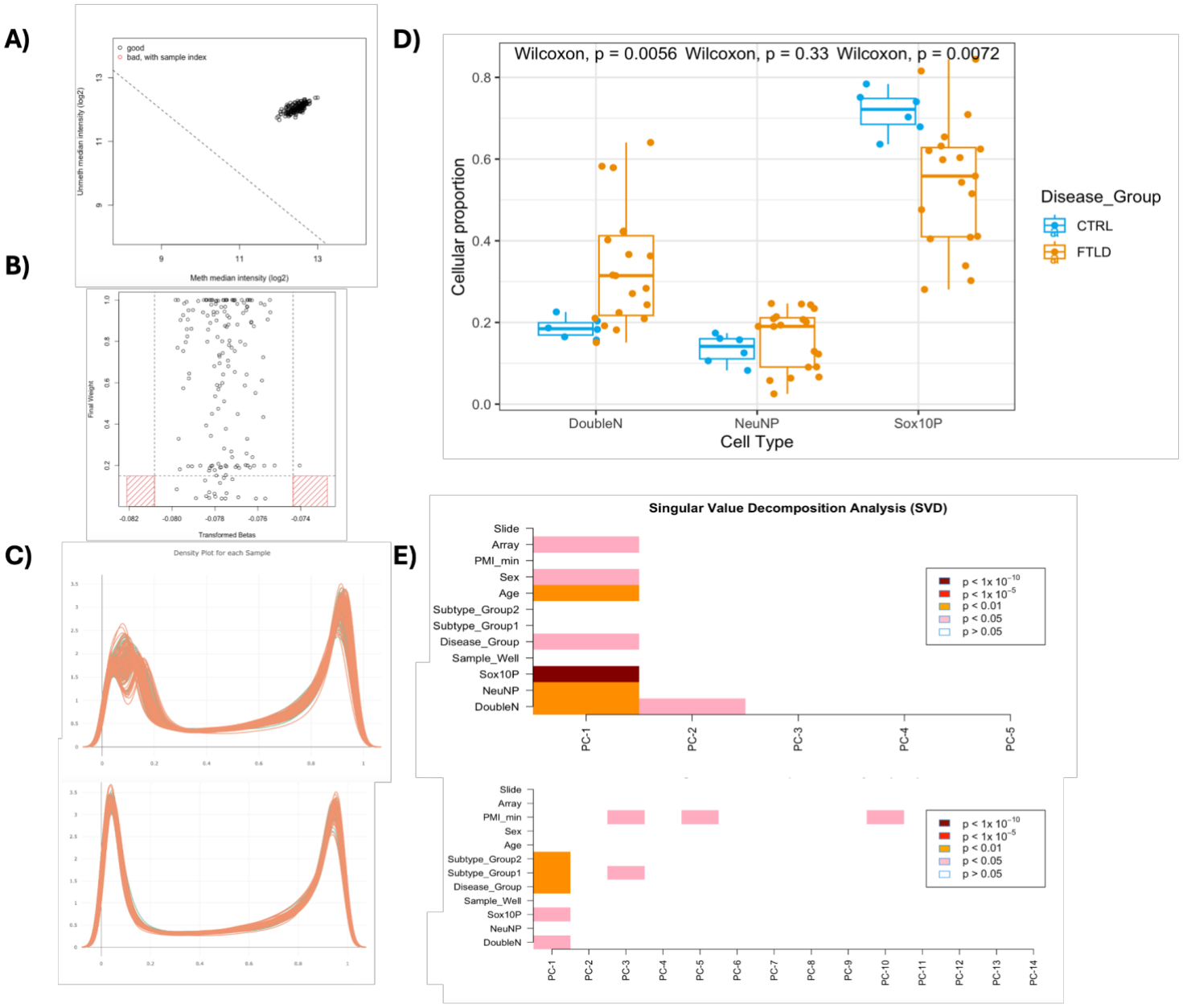
2.1.3 Cell type deconvolution estimation using DNA methylation data

For a significant amount of the analysis undertaken in this thesis, we have utilised DNA methylation data derived from 'bulk' brain tissue. As DNA methylation patterns are subject to high levels of variation between distinct cell types, the proportion of each cell type in a sample could thus lead to DNA methylation changes being detected that merely reflect, for example, an increased/decreased proportion of a particular cell type in the sample being processed. Therefore, we have used a novel cell-type deconvolution algorithm described by Shireby et al. [110] which has used reference panel data obtained from fluorescence activated sorted nuclei from cortical brain tissue to estimate the relevant proportions of the following cell groups: NeuN+ nuclei (representing neuronal cell-types), SOX10+ nuclei (representing oligodendrocytes), and NeuN-/SOX10- nuclei (representing other glial cell types). Using this cell-type sorted reference dataset and the CETYGO (CEll TYpe deconvolution GOodness) package (https://github.com/ds420/CETYGO), we estimated, for each sample in our datasets, the relative proportions of each three cell populations (example - **Figure 2.2.D**). We used these estimations as factors in our quality control steps and covariate correction.

2.1.4 Differential methylation analysis

We applied linear regression models (**Table 2.1**) using the *M*-values as the input to identify associations between DNA methylation variation at specific CpG sites and FTLD using the limma package [156]. For FTLD1, we have accounted for possible confounding factors, such as age and sex as well as factors detected in principal components 1 and 2 as seen in Singular Value Decomposition (SVD) plots (ChAMP package[154]) (**Figure 2.2.E**), which included cell proportions (SOX10+and Double−) and sample position in the array. Using this regression model, no surrogate variables were detected with the num.sv function of the SVA package [157], meaning there were no remaining unknown, unmodelled, or latent sources of variation [156]. The same process was applied to FTLD2 and FTLD3. The model for FTLD2 was further adjusted for slide, whereas for FTLD3, the model was further adjusted for slide and one surrogate variable (**Table 2.1**). For the FTLD-sorted cohort, we included in the model age, sex, cell proportions (NeuN+), array and 2 surrogate variables. For AD1 (sorted), we included age, cell proportions (NeuN+ and DoubleN), slide and sex. The regression model for AD2 only included cell proportions (NeuN+) and sex. For the bulk AD dataset, we included, as well as slide, array and cell proportions (NeuN+ and DoubleN), principle components 1 and 2, and bisulfite conversion efficiency (technical variability arising from processing of DNA prior to sequencing). For the multi-region AD datasets, the following covariates were included: DLPFC - cell proportions (NeuN+), age, sex, plate, ERC - slide, cell proportions (DoubleN + SOX10+), age one surrogate variable, HIPPO - plate, cell proportions (DoubleN and NeuN+) sex, age, array and CRB - slide, sex, age and cell proportions (NeuN+ and DoubleN) (**Table 2.1**).

Figure 2.2 Quality control steps example figures



A) Scatter plot showing the relationship between unmethylated (Unmeth) and methylated (Meth) median intensities for all samples generated using minfi [152]**.** Good-quality samples are represented as black circles, while low-quality samples are highlighted as red circles with sample indices. **B) Scatter plot of transformed beta values versus final weights generated with wateRmelon** [153]. Red-hatched regions indicate transformed beta values

outside the acceptable range. **C) Density plots displaying the distribution of beta values (ranging from 0 to 1) for each sample in the dataset.** Top panel shows density of samples pre-normalisation, whilst bottom panel shows density of beta values after normalisation using the BMIQ method in ChAMP [154]. **D) Results from cell-type deconvolution analysis.** Bar plot illustrates proportions of each cell type: NeuN+ nuclei (representing neuronal cell-types), SOX10+ nuclei (representing oligodendrocytes), and NeuN-/SOX10- nuclei (representing other glial cell types) for samples across control and frontotemporal lobar degeneration (FTLD) groups. **E) SVD plots illustrating which covariates contribute to variation within the dataset.** The top panel illustrates the diagnostic SVD plot as seen by principal components (PC), showing that batch and biological effects are contributing to data variation. The bottom panel illustrates covariate contribution after adjustment.

2.2 Gene expression datasets

2.2.1 RNA-sequencing datasets

Throughout this thesis, we utilise three RNA-sequencing datasets (**Table 2.2**). For all of these RNA sequencing datasets, quality control and processing was performed as follows. Normalization factors were calculated using the limma package [156] to account for differences in library sizes across samples, ensuring comparability of expression values. Genes with low expression levels were filtered out (i.e. genes whose maximum counts per million (CPM) value across all samples was less than 1). The voom function from the limma package was then applied to the filtered data to model the mean-variance relationship and transform the counts data into log2 counts per million (log-CPM) values, making them suitable for linear modeling. A linear model was fitted to the transformed data used to adjust for covariates (as determined necessary using SVD analysis). Regression models and corresponding covariates used in each case are described in **Table 2.2**.

Table 2.2 Pathological and demographic characteristics of the FTLD and AD cohorts used in this thesis for gene expression analyses.

Cohort	Samples	Mean age ± SD (years)	Sex	Regression models used for differential expression analysis
FTLD1-Expression	FTLD-TDP (*N* = 44)	66.8 ± 13.2	25M/19M	*~0 + disease + age + sex + pmi*
	CTRL (*N* = 24)	70 ± 9.46	11M/11M	
FTLD2-Expression	FTLD (*N* = 30)	62.6 ± 8.10	14M/16F	*~0 + disease + age + sex + pmi*
	FTLD-TDP (*N* = 19)	63.9 ± 7.98	7M/12F	
	FTLD-tau (*N* = 11)	60.5 ± 8.23	7M/4F	
	CTRL (*N* = 14)	78.4 ± 11.8	5M/9F	
AD-Expression	AD (N = 201)	83.1 5.26	81M/120F	*~0 + disease + age + sex + pmi + sequencing batch*
	CTRL (N = 329)	85.0 4.07	117M/212F	

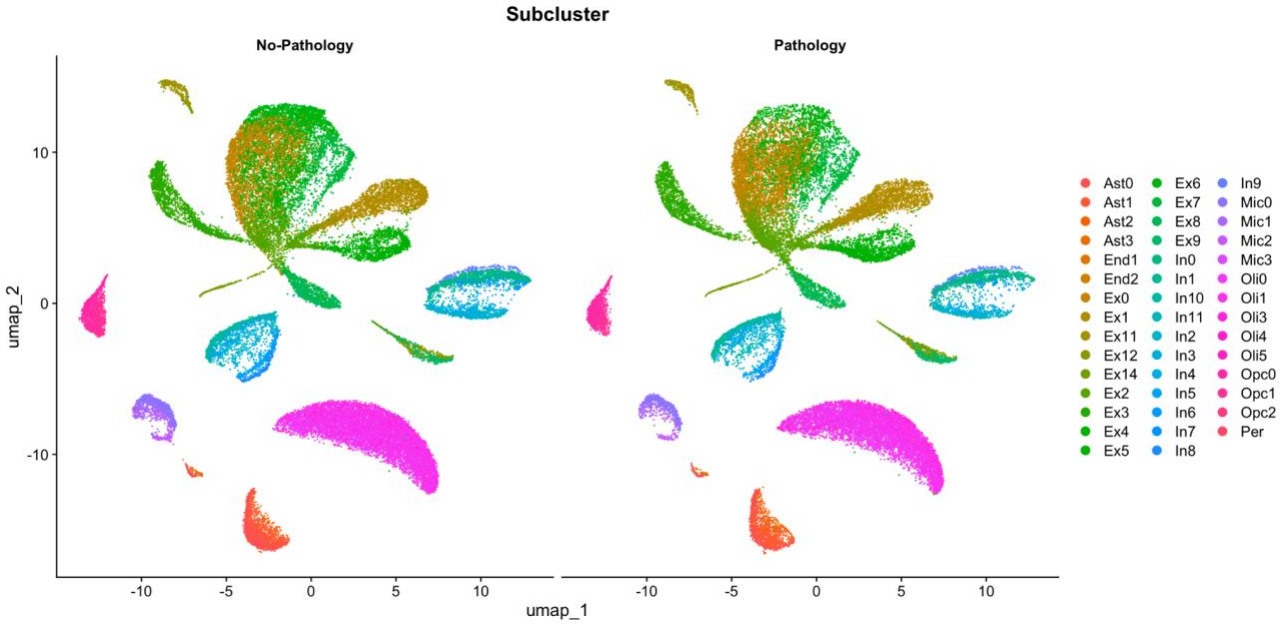
FTLD, Frontotemporal lobar degeneration; FTLD-TDP, FTLD with 43 kDa transactive response DNA-binding protein (TDP-43) positive inclusions; FTLD-Tau, FTLD with tau-positive inclusions; PSP, progressive supranuclear palsy; AD, Alzheimer's Disease; SD, Standard deviation; F, Females; M, Males

2.2.2i Single-nuclei RNA sequencing dataset

We also used single-nuclei RNA sequencing data from the previously published and publicly available ROSMAP study[158] (https://www.synapse.org/#!Synapse:syn18485175). We obtained filtered cell counts from 24 individuals showing little or no pathology and 24 showing mild to severe AD pathology (*N* = 48 total). Filtered data consisted of 70,634 droplet-based single-

nucleus covering a total of 17,926 genes. After downloading, we used Seurat to normalise the data with the function “NormalizeData” with the option “LogNormalize”, using a scale factor of 10,000. We then assigned cell cluster identities as identified by Mathys et al [158], which contained 8 cell-type clusters; Astrocytes (Ast), Endothelial cells (End), Excitatory neurons (Ext), Inhibitory neurons (In), Microglia (Mic), Oligodendrocytes (Oli), Oligodendrocyte precursor cells (Opc) and pericytes (Per). Subclusters of these cells were also defined, which for OLGs and OPCs were Oli_1,3,4 and 5, and Opc_0,1,2 (**Figure 2.3**).

Figure 2.3 Single-nuclei Expression data used



UMAP plot illustrating the distribution of various cell subclusters under two conditions: *No-Pathology* (left) and *Pathology* (right). Each point represents a single nuclei, and cells are grouped into distinct subclusters based on gene expression patterns. Ast: Astrocytes, End: Endothelial cells, Ext: Excitatory neurons, In: Inhibitory neurons, Mic: Microglia, Oli: Oligodendrocytes, Opc: Oligodendrocyte precursor cells, Per: pericytes, UMAP: Uniform Manifold Approximation and Projection

2.3 Oligodendrocyte and oligodendrocyte precursor cell gene lists

For investigation into which genes showing differential methylation across diseases were relevant to OLGs and OPCs, we utilised a list of genes kindly provided by collaborator Dr Piras, who curated the gene lists from human snRNA-sequencing data generated by Mathys et al [158]. As previously described by Dr Piras [159], data was imported through Seurat, and genes were assigned to a cell type class using a linear regression model where the dependent variable and predictor were gene expression level and cell type clusters respectively. Genes were assigned to a cell type if 1) adjusted p-value (false discovery rate (FDR)) was less than 0.05, and 2) the ratio of the regression coefficient of the most enriched cell type compared to the second most enriched cell type was ≥ 1.76 (as defined by Piras et al, for more details, see reference[159]).

For various analyses throughout this thesis, we utilized different thresholds to define oligodendrocyte (OLG) and oligodendrocyte precursor cell (OPC)-relevant gene lists, as provided by Dr. Piras. Specifically, in Chapter 3, which examines differential methylation of OLG/OPC-relevant genes, a more stringent cutoff was applied. Genes were selected if the ratio of the regression coefficient of the most enriched cell type to the second most enriched cell type was ≥ 3. Through this more stringent method, 251 and 123 genes were included as 'stringent OLG genes' and 'stringent OPC genes' respectively (**Appendix A**). This stringent filtering ensured a focus on highly specific cell-type markers. In Chapter 4, which assesses cell-type enrichment within co-methylation modules, the full list of all genes identified through the regression-based cell type assignment method (described earlier) was used, where 444 and 447 genes were included as 'OLG genes' and 'OPC genes' respectively (**Appendix A**). This less stringent gene list was justified to uncover patterns of cell-type enrichment within co-methylation networks modules.

Chapter 3: Oligodendrocyte lineage genes show differential methylation and expression across dementias

3.1 Introduction

Differential methylation is an analytical approach commonly used to identify differences in DNA methylation patterns between disease and control groups. There have been several epigenome-wide association studies (EWAS) which have identified genes in which aberrant DNA methylation is associated with neurodegenerative disease (described in Chapter 1, section 1.4.1). Numerous genes have been found to be consistently differentially methylated across datasets, for example *ANK1*, which is consistently found to be dysregulated in AD [106,107,160]. Several genes relating to oligodendrocytes have also been identified, for example *BIN1* and *MOBP* have been found to be differentially methylated inAD [106,131] and MSA [118,129] respectively. However, as far as we know, no studies comprehensively focusing on OLG and OPC relevant genes within EWAS of bulk brain tissue have been carried out. Meta-analyses have also been carried out across DNA methylation datasets, for example in AD[112]. Meta-analyses take results from multiple studies/datasets and combine them, allowing increased statistical power to detect novel differentially methylated sites. In the context of DNA methylation analysis, meta-analyses are a particularly valuable tool[71,73]; DNA methylation studies are often limited by small sample size and large batch effects due to technical variation.

In this chapter, we used both bulk brain tissue data and sorted brain-nuclei data. Bulk DNA methylation datasets refers to datasets derived from a piece of brain tissue with a variable mix of neuronal and glial cells, frequently used due ease and cost-effectiveness as there is no need for complex and expensive cell/nuclei-sorting techniques. There are advantages to the use of data from such studies - namely that sample sizes tend to be greater, which is important when analysing DNA methylation data as referred to above. However, techniques to sort brain nuclei into cell-type specific fractions, mainly through Fluorescence-Activated Nuclei Sorting (FANS), have enabled the isolation of cell-type specific nuclei fractions, from which cell-type specific DNA methylation profiles can be generated. Gasparoni et al. used brain-nuclei sorting to profile DNA methylation changes associated with AD in neurons and glia, and identified genes which showed differential methylation in only one cell-type population that had not been detected in previous bulk tissue studies, underlining the importance of such techniques [122]. A recent study went further and profiled three populations of cells - neurons, OLGs and astrocyte/microglia [110,122]. Interestingly it was found that most of the DNA methylation changes were identified in non-neuronal cell populations [110,122]. This indicates that DNA methylation changes identified in bulk tissue may be highly driven by changes in OLGs, as well as other glial cell populations, and highlights the need for approaches studying cell-specific DNA methylation patterns in neurodegeneration.

Another use of the generation of sorted brain-nuclei datasets is that they have enabled the development of reference based cell-type deconvolution algorithms, which allow populations of distinct cell types in bulk DNA methylation datasets to be estimated. The cell-type algorithm we use in this work was described by Shireby et al. [110]. Briefly, FANS was used to obtained three purified nuclei populations: neurons (NeuN+), oligodendrocytes (SOX10+), and microglia/astrocytes (NeuN–/SOX10–), which were used as reference data. This cell-type

specific data was used, with an algorithm created using a method described by Houseman et al. [161], to obtain DNA methylation markers of specific cell types. The methylation levels of the marker methylation sites were used to construct reference profiles for each of these cell-types. Quadratic programming was then applied to bulk data to estimate cell types based on the difference between observed bulk methylation levels and the predicted levels based on reference profiles. Such algorithms allow for the estimation of different cell type proportions which can then be added as variables to regression models. In differential DNA methylation analysis, this step is crucial. It is important that differences in DNA methylation between control and disease are not simply a reflection of distinct cell composition between sample groups. In studying neurodegeneration, where specific populations of cells are affected[4], this is even more important.

Here, we use differential methylation analysis to uncover which genes relating to OLGs and OPCs show dysregulated patterns of DNA methylation between controls and AD and/or FTLD. As described, brain-nuclei sorted DNA methylation datasets are invaluable as they enable the reduction of 'noise' signals coming from multiple cell types and DNA methylation changes that affect just one cell type are less diluted. In this work, where we are interested in DNA methylation changes affecting OLG lineage genes, the use of this type of data is particularly valuable. Due to complexity in generating data however, there are far fewer datasets available (both in house and publicly available), and such datasets tend to have smaller sample sizes. Therefore, we utilised both bulk and brain-nuclei sorted DNA methylation data as complementary approaches.

3.2 Methods

3.2.1 DNA methylation datasets used in this chapter

In this chapter, we utilised 7 DNA methylation datasets across FTLD subtypes and AD, both bulk-tissue and sorted brain-nuclei derived. We have described the demographic characterisation of these datasets in detail in Chapter 2 section 2.1. A brief overview of the datasets is provided in **Table 3.1.** To investigate DNA changes in FTLD relating to OLGs and OPCs, we carry out cohort-specific EWAS on three bulk FTLD datasets (FTLD1-3), followed by a meta-analysis of these three cohorts. We then utilise a brain-nuclei sorted dataset to follow up on findings from the bulk datasets (FTLD-sorted). In our investigation of DNA methylation changes associated with AD in OLGs and OPCs, we utilised one bulk AD dataset (AD) and two brain-nuclei sorted datasets (AD1-sorted, AD2-sorted) (**Table 3.1**).

Table 3.1 Brief overview of the DNA methylation datasets used in this chapter

Cohort	**Samples included after quality control**	**Reference**
FTLD1	FTLD (*N* = 15) and CTRL (*N* =8)	Fodder et al.[119]
FTLD2	FTLD (*N* = 34) and CTRL (*N* = 14)	Menden et al.[149]
]\		
FTLD3	FTLD (*N* = 93) and CTRL (*N* = 70)	Weber et al.[117]
AD	AD (*N* = 201) and CTRL (*N* = 329)	De Jager et al.[106]
FTLD-sorted	FTLD (*N* = 19) and CTRL (*N* = 6)	Bettencourt lab, unpublished
AD1-sorted	AD (*N* = 15) and CTRL (*N* = 16)	Gasparoni et al.[122]
AD2-sorted	AD (*N* = 5) and CTRL (*N* = 6)	Bettencourt lab, unpublished

FTLD, Frontotemporal lobar degeneration;AD, Alzheimer's Disease.

3.2.2 Data processing and differential methylation

Data processing was conducted as described in **Section 2.1.2**. Cell type deconvolution was carried out as described in **Section 2.1.3**. As described in **Section 2.1**, we carried out differential methylation analysis using regression modeling, where we include estimates of cell-type proportions as covariates. The models used for differential methylation analysis are described in **Section 2.1.4** and **Table 2.1**.

3.2.3 Meta-analysis

In the case of the three FTLD cohorts, and the two sorted AD datasets, we also carried out a meta-analyses. We used the estimated coefficients and SEs obtained from the regression models, described above for the three FTLD cohorts and the sorted-AD cohorts, to undertake an inverse variance meta-analysis using the metagen function from the meta R package [162]. Only methylation methylation sites present in all datasets ($N = 363{,}781$ for the FTLD meta-analysis, and N = 376,028 for the sorted AD meta-analysis) were considered for this analysis. When reporting differentially methylated sites, a conservative Bonferroni significance was defined as $p < 1.374 \times 10^{-7}$ ($p < 0.05/363{,}781$) and p < $p < 1.330 \times 10^{-7}$ (p < 0.05/376,028) for the FTLD and AD meta-analyses, respectively, to account for multiple testing. We report random-effects meta-analysis results as the three cohorts included different FTLD subgroups/subtypes according to their neuropathological classification possibly leading to high heterogeneity in the meta-analysis.

3.2.4 Oligodendrocyte and oligodendrocyte precursor cell gene lists

For investigation into which genes showing differential methylation across diseases were relevant to OLGs and OPCs, we utilised a list of genes kindly provided by collaborator Dr Piras. In order to only investigate genes strongly associated with OLGs and OPCs, in this analysis we filter for only those genes with a more stringent cut off for OLG/OPC association, as described in **Chapter 2 Section 2.3**, in order to limit findings to those genes highly relevant to our cell types of interest.

3.2.4 Comparisons of DNA methylation hits with gene expression data

To examine the gene expression patterns of genes we found to be differentially methylated, we used several gene expression datasets (**Table 3.2**), which have been described in detail in **Chapter 2 section 2.2**. In this chapter, we utilise two bulk FTLD RNA-sequencing datasets, which overlap with a subset of samples from the FTLD1 and FTLD2 DNA methylation datasets, and one bulk AD RNA-sequencing dataset, also overlapping with samples from the AD DNA methylation dataset. We also utilise a single-nuclei RNA sequencing dataset derived from AD and control samples. The quality control and processing of these datasets is described in **Chapter 2 section 2.2**.

3.2.5 Further Investigation of DNA methylation hits

To investigate any hits we have found in our EWAS, we have utilised online databases such as Protein Atlas (https://www.proteinatlas.org/). We have also carried out functional enrichment of groups of differentially methylated genes using Gene Ontology (GO) enrichment analysis [163,164]. GO enrichment analysis was conducted through ClusterProfiler [165].

A schematic for the work carried out in this Chapter is provided in **Figure 3.1.**

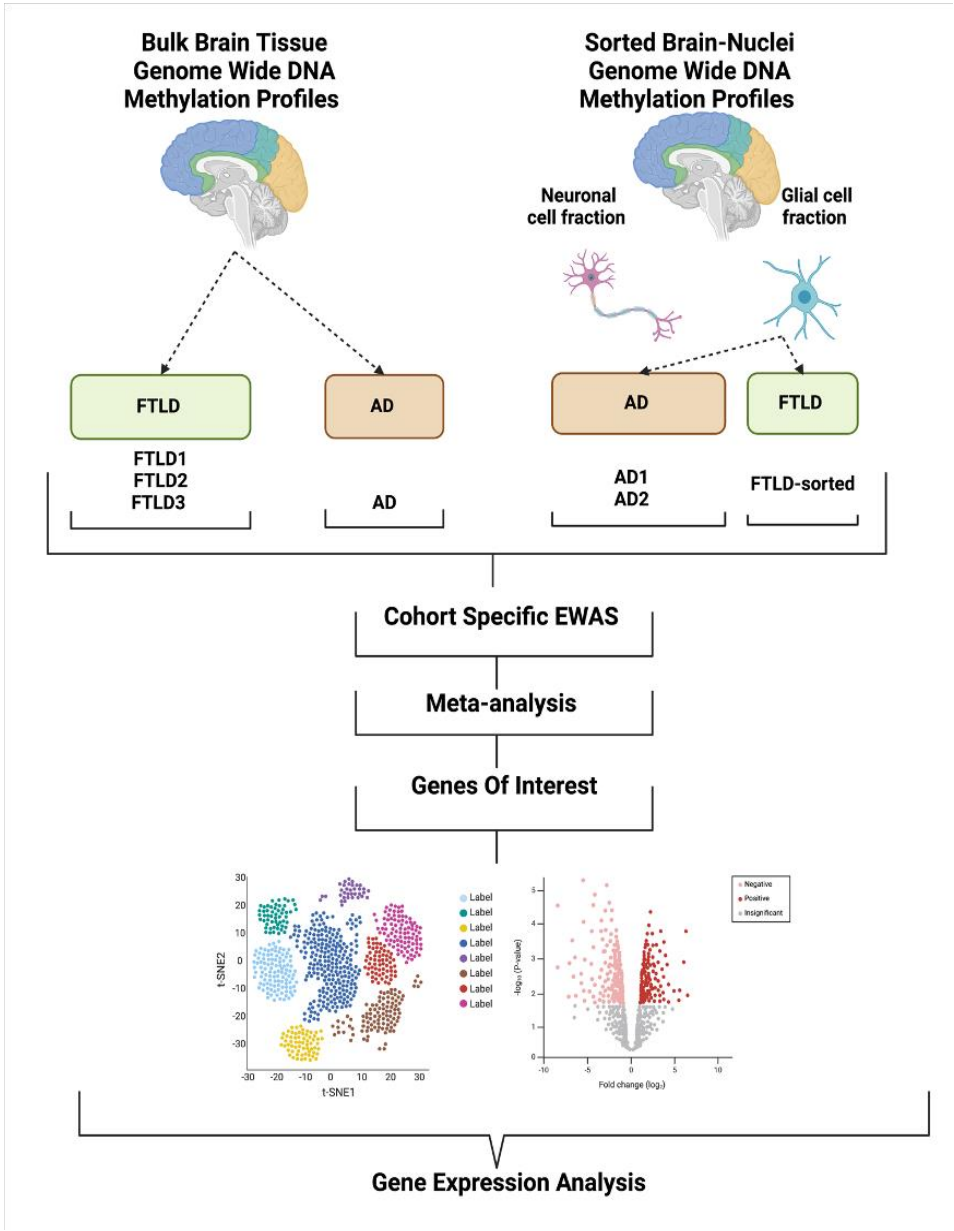


Figure 3.1 Schematic illustration work carried out in Chapter 3

Cohort specific EWAS analysis were carried out on four bulk brain tissue DNA methylation cohorts; FTLD1, FTLD2, FTLD3 and AD. Cohort specific EWAS were also carried out on three sorted brain-nuclei DNA methylation cohorts; AD1, AD2 and FTLD-sorted. Meta-analyses were carried out on the FTLD1, FTLD2, and FTLD3 datasets, and AD1 and AD2 datasets. Genes of interest were identified, and changes in gene expression were investigated through multiple gene expression datasets. Functional enrichment on differentially methylated genes was also carried out. FTLD: frontotemporal lobar degeneration, AD: Alzheimer's disease, EWAS: epigenome-wide association study.

3.3 FTLD results and discussion

3.3.1 EWAS of three bulk FTLD datasets reveals that oligodendrocyte lineage genes are differentially methylated in FTLD and show downstream changes in gene expression

To investigate DNA methylation changes relating to oligodendrocyte lineage cells in our three FTLD cohorts, we used linear regression models to perform cohort-specific case-control EWAS, and then investigated effects within OLG/OPC relevant genes (as defined in **Section 2.3**). In these cohort-specific EWAS, likely due to limited sample size, no methylation sites mapping to genes within our OLG and OPC relevant gene lists reached genome-wide significance after multiple testing corrections. We thus discuss methylation sites that showed an arbitrary nominal significance of unadjusted P-value < 0.01 in each cohort (**Appendix B)**.

The top differentially methylated CpGs in FTLD1, FTLD2 and FTLD3 that mapped to OLG relevant genes were cg13010326 in *ANLN* ($p = 6.74e^{-05}$), cg08407007 in *HSPA2* ($p = 4.27e^{-05}$) , and cg11965880 in *PIP4K2A* ($p = 1.31e^{-05}$). *ANLN* codes for Anilin, which is a cytoskeletal gene known to be important in the process of cytokinesis (the division of cells post-mitosis)[166]. The gene was identified through snRNA-sequencing as being upregulated in OLGs in AD compared to controls [167]. Two methylation sites within this gene were found to be significantly differentially methylated in the FTLD1 dataset, and one was found within the FTLD2 cohort (**Appendix B**). *ANLN* did not show changes in either of the FTLD expression datasets analysed.

HSPA2 codes for Heat shock Protein Family A Member 2, which acts as a molecular chaperone [168]. *HSPA2* has been found to be upregulated in PD and MSA [169], and has also been linked to other neurodegenerative diseases. A study found that *HSPA2* expression was associated with both amyloid plaque and neurofibrillary tangle density, and expression of the gene was also significantly different between control and AD OLGs in single nuclei RNA sequencing data [170].

PIP4K2A codes for the enzyme phosphatidylinositol-5-phosphate 4-kinase type-2 α, which has a role in the trafficking of cholesterol to peroxisomes from lysosomes [171]. The gene was found to be part of a core set of OLG genes (from transcriptome co-expression networks)[172], and is also genetically associated with AD [172]. The gene showed upregulation at the gene expression level in the FTLD1-expression dataset (**Figure 3.2.D, Appendix C**).

We investigated whether any genes showed differences in DNA methylation across all three FTLD cohorts. Differential methylation in *CTNNA3, DNAH17* and *SCD* was found in all three cohorts (**Figure 3.2, Appendix C**). *CTNNA3* encodes the adhesion protein catenin-α 3, which has been (with some controversy) associated with AD [173–175], and is part of the Wnt-signalling pathway [176]. *DNAH17* codes for dynein axonemal heavy chain 17, a motor protein that has been associated with PD [177,178]. Interestingly, in a study incorporating multi-omics data to investigate key molecular players in myelin-related dysregulation, both *CTNNA3* and *DNAH17* were identified as hits [172]. *SCD* codes for the enzyme Stearoyl-CoA Desaturase, an enzyme involved in fatty acid biosynthesis. It was shown, in an AD mouse model, that inhibition of SCD (at the protein level) led to some restoration of hippocampal function [179].

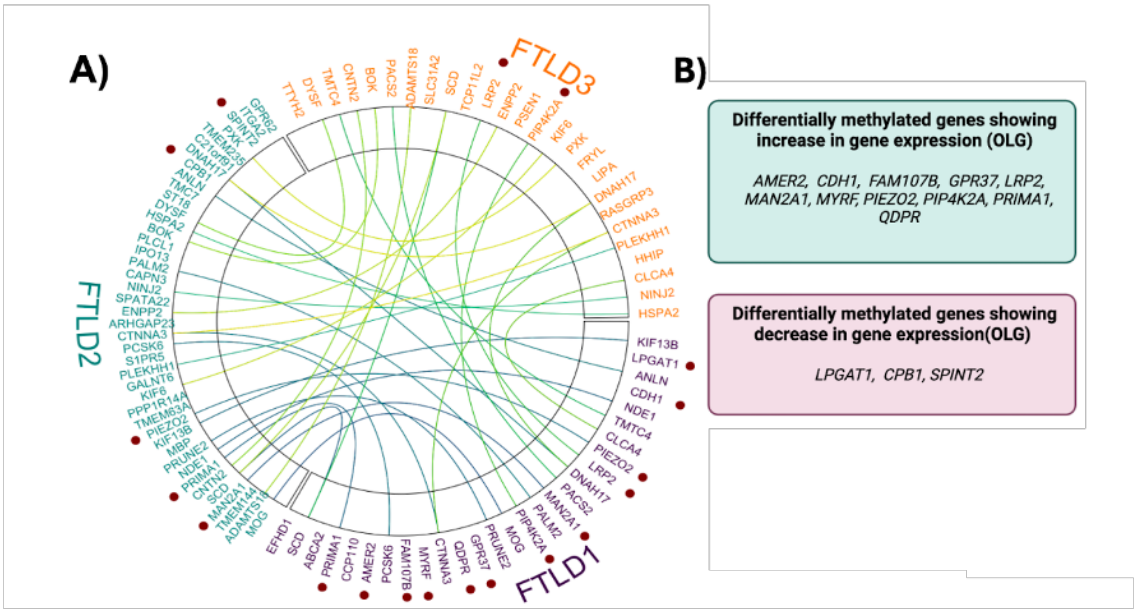
When examining regions of dysregulated DNA methylation patterns within these three genes, we saw that all 6 differentially methylated sites across the three FTLD cohorts mapping to *CTNNA3* were located within the gene body (**Figure 3.2 C, Appendix B**). Differentially

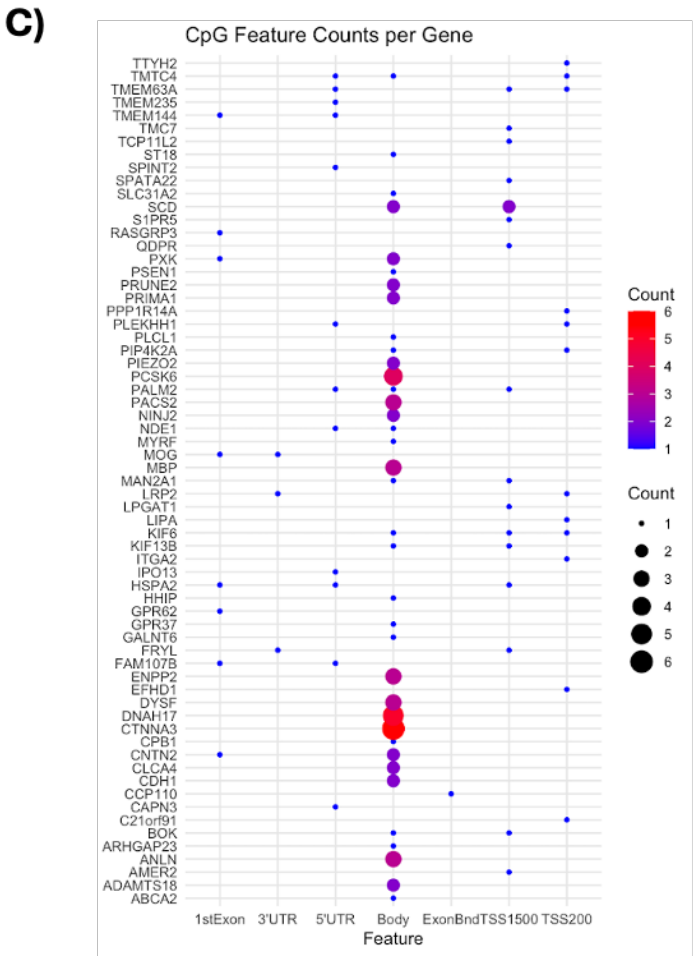
methylated sites mapping to *DNAH17* were also all located within the gene body (**Appendix B**). In contrast, the methylation sites mapping to *SCD* were positioned both within the promoter region (TSS1500) and within the gene body but close to the 5' end of the gene (**Appendix B**). In the case of both *CTNNA3* and *DNAH17*, there appears to be a general dysregulation of DNA methylation across a large section of the gene bodies, with both hypermethylation and hypomethylation. In the case of *SCD*, there was a tendency towards hypomethylation around the promoter region of the gene. Several other OLG genes showed dysregulated methylation across multiple datasets (**Figure 3.2 A, Appendix B**).

To explore any potential downstream consequences of the DNA methylation patterns observed, we investigated, where possible, FTLD transcriptomic data. We utilised two FTLD expression datasets, FTLD1-expression and FTLD2-expression, which overlapped with samples from the FTLD1 and FTLD2 DNA methylation datasets. Several genes that showed differentially methylated sites also showed changes in gene expression between FTLD and controls (**Figure 3.2.A,C, Appendix C**). The gene *MYRF* (myelin-regulatory factor), a gene of high importance to OLGs due to its role in maintenance of mature myelinating phenotypes, showed differential methylation at the body of the gene, and showed an upregulation at the gene expression level in the FTLD1-expression data (**Figure 3.2 B,C, Appendix C**).

For the genes we identified as being differentially methylated across all three datasets: *CTNNA3, DNAH17* and *SCD*, we saw no significant changes in expression in the available gene expression datasets. However, particularly in the case of the former two genes, differentially methylated sites were clustered within the gene bodies rather than the promoter regions, where the role of DNA methylation on gene expression is less well understood [89,180]. It is possible that DNA methylation in the gene body may be having an effect that is beyond the scope of this thesis, for example in differential splicing.

Figure 3.2 Differentially methylated oligodendrocyte genes from FTLD1, FTLD2 and `FTLD3 EWAS





(A) Circos plots representing genes that are differentially methylated across the three FTLD datasets (FTLD1, FTLD2, and FTLD3) that map to OLG genes. Each FTLD dataset is represented by a distinct colour; FTLD1: Purple, FTLD2: Teal, FTLD3: Orange. Coloured lines within the circle connect genes that are shared (overlapping) between two or more FTLD subtypes. DEGs Highlighted: Genes marked with red points outside the circular layout indicate genes that are differentially expressed (DEGs) in addition to being differentially methylated. **B) Differentially Methylated genes that show increased or decreased expression in corresponding gene expression data.** Boxes depict genes that are upregulated or downregulated in either of the FTLD expression datasets analysed. **C) Bubble plot displaying CpG feature counts for differentially methylated genes across genomic features.** The y-axis lists the genes, and the x-axis represents various genomic features (e.g., 1st Exon, 3' UTR, 5' UTR, Body, TSS1500, TSS200). The colour and size of each bubble indicates the count of CpG sites associated with a given feature for each gene.

For OPC relevant genes, the top differentially methylated CpGs were cg19594305 in *CHST8* (p = 0.00088), cg05602183 in *TRAF4* (p = 0.00013) and cg24825027 in *CREB3L1* (p = 0.00073) in FTLD1, FTLD2 and FTLD3, respectively (**Appendix B**). No genes relating to OPCs contained differentially methylated methylation sites (P < 0.01) across all three datasets, however there were several showing differential methylation in at least 2/3 datasets (**Figure 3.3.B**).

CHST8 codes for carbohydrate sulfotransferase 8, which is not known to be linked to neurodegeneration, however we did see that this gene exhibited downregulation at the gene expression level in our FTLD2-expression data (**Figure 3.3.D, Appendix C**).

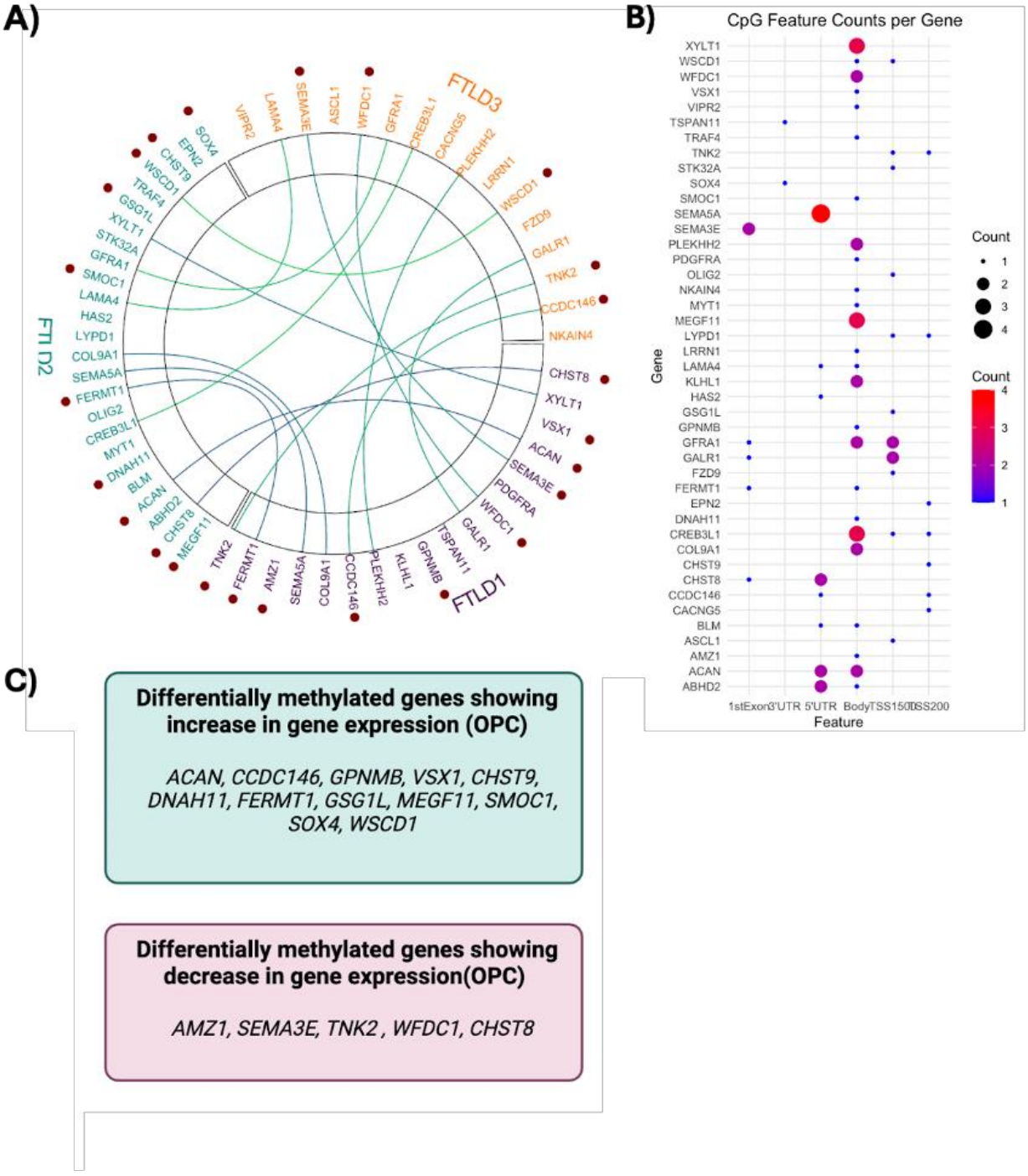
TRAF4 codes for Tumor Necrosis Factor Receptor-Associated Factor 4 which is a ubiquitin ligase involved in cell proliferation amongst other processes [181]. *TRAF4* has been shown to be present at all stages of OLG differentiation, but most highly expressed in early OPCs in mice [182], suggesting the gene may be important in the early stages of the OLG life cycle. The gene has also been found to be involved in the Wnt/β-catenin signalling pathway, which is known to be

important in OLG development [183]. It has also been found that TRAF4-KO mice exhibit significant loss in myelin integrity [182]. We did not see differential expression of this gene in our FTLD expression datasets.

CREBL codes for CAMP responsive element binding protein 3 like 1, a protein involved in the endoplasmic reticulum stress response, and has been found to show decreased gene expression in AD [184]. The gene has also been implicated as being involved in glioma pathogenesis [185].

We also found that several of the OPC genes we identified in our EWAS analysis showed differential expression in FTLD-expression datasets 1 and 2. To highlight a few findings; the gene *ACAN* was differentially methylated in both FTLD1 and FTLD2 EWAS and differentially expressed (showing upregulation)in FTLD1-expression and FTLD2-expression, respectively. *ACAN* codes for Aggrecan, a crucial component of the extracellular matrix [186]. We had found terms related to 'extracellular matrix' to be enriched across all differentially methylated OPC genes found in our three FTLD EWAS (**Figure 3.3.C**). It has been found that variable number tandem repeat polymorphisms in this gene are associated with AD [187]. Another gene of interest was *SOX4,* which was differentially methylated and downregulated in FTLD2, and is a transcription factor expressed in OPCs that is associated with inhibition of OPC differentiation [188,189], making this a potentially key gene in myelin dysregulation in FTLD.

Figure 3.3 Differentially methylated oligodendrocyte precursor genes from FTLD1, FTLD2 and FTLD3 EWAS



(A) Circos plots representing OPC genes that are differentially methylated across the three FTLD datasets (FTLD1, FTLD2, and FTLD3) that map to OPC genes. Each FTLD dataset is represented by a distinct colour; FTLD1: Purple, FTLD2: Teal, FTLD3: Orange. Coloured lines within the circle connect genes that are shared

(overlapping) between two or more FTLD subtypes. DEGs Highlighted: Genes marked with red points outside the circular layout indicate genes that are differentially expressed (DEGs) in addition to being differentially methylated. **B) Bubble plot displaying CpG feature counts for differentially methylated genes across genomic features.** The y-axis lists the genes, and the x-axis represents various genomic features (e.g., 1st Exon, 3' UTR, 5' UTR, Body, TSS1500, TSS200). The colour and size of each bubble indicates the count of CpG sites associated with a given feature for each gene. **C) Differentially Methylated genes that show increased or decreased expression in corresponding gene expression data.**

3.3.2 Meta-analysis of all three bulk tissue datasets reveals differentially methylated oligodendrocyte lineage relevant genes

We also carried out a meta-analysis of all three FTLD cohorts, to enable the identification of OLG/OPC relevant differential DNA methylation signatures across 234 individuals (142 FTLD cases and 92 controls). No methylation sites within OLG/OPC relevant genes passed significance after a conservative Bonferroni adjustment for multiple testing ($p < 1.37 \times 10^{-7}$).

However, we ranked methylation sites by a conservative random effect P-value, subset those below the selected threshold (p-value < 0.01), and 31 and 12 methylation sites within OLG and OPC relevant genes, respectively, were considered to be significantly differentially methylated. These methylation sites and their positions within the genes are shown in **Appendix D**.

Unsurprisingly, OLG genes methylation sites in *CTNNA3, DNAH17,* and *SCD,* identified in the cohort-specific EWAS were found to be in those top differentially methylated from the meta-analysis (**Figure 3.4.A, Appendix D)**, strengthening the association of dysregulation of DNA methylation in these regions across FTLD subtypes. Methylation sites in *HSPA2* and *PIP4K2A* (those genes containing the top differentially methylated methylation site in FTLD2 and FTLD3, respectively) were also found to be among the top hits from the meta-analysis (**Appendix D**). It

was notable that 3/31 nominally significantly differentially methylated OLG methylation sites mapped to *HSPA2* (**Appendix D**).

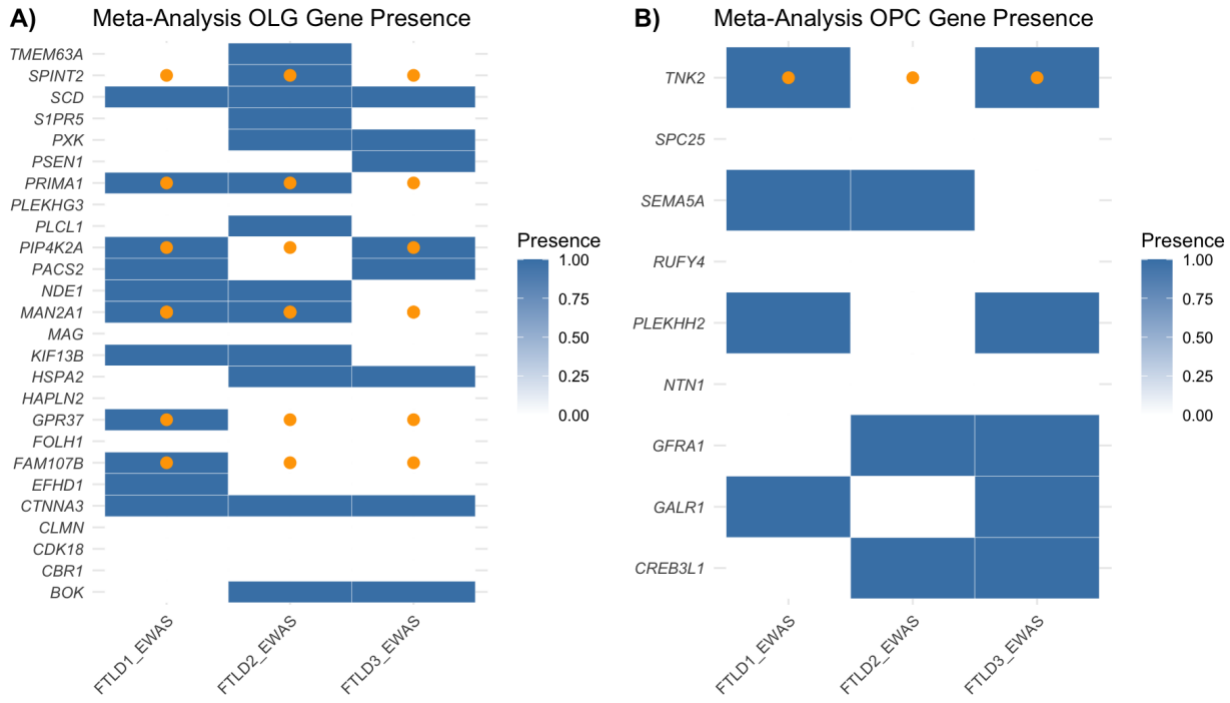
Twelve methylation sites were differentially methylated in our meta-analysis that mapped to genes within our OPC list that were nominally significantly differentially methylated within our chosen threshold ($p < 0.01$) (**Figure 3.4, Appendix D**). The most significantly differentially methylated methylation site cg18716096 mapped to the promoter region of *TNK2*. This gene was also highlighted in the FTLD1 and FTLD3 EWAS (**Figure 3.2.B, Appendix B**), and showed a decrease in expression in the FTLD1-expression data (**Figure 3.3.D**). *TNK2,* which codes for ACK1 (activated CDC42 kinase 1), a non-receptor tyrosine kinase, which is associated with several forms of cancer [190] and has been found to be mutated in familial PD [191].

There were several genes found within the OPC meta-analysis results that contained more than one differentially methylated site (**Appendix D**), possibly indicating robustness of association of aberrant DNA methylation within that gene. Three methylation sites mapping to *GALR1* (2 within TSS1500 and 1 at TS200) and 2 methylation sites mapping to *GFRA1* (both within TSS1500) were in those top identified through the meta-analysis. *GALR1* codes for Galanin receptor 1, which is a G-protein coupled receptor (cell surface receptor) for galanin, a neuropeptide involved in many signalling pathways in the CNS [192]. Galnin overexpression has been described in AD, although it is contested whether this contributes to disease or acts as a neuroprotective mechanism [192]. Galanin has been described to act as a growth and survival factor in OLGs [193]. Furthermore, in a mouse model to replicate demyelination in MS, gene expression changes in GalR1 during demyelination and remyelination was observed [192,194], indicating that this receptor is important in allowing correct myelination to occur. Additionally, increased galanin expression in a transgenic mouse model was associated with elevated expression of *MBP* (myelin basic protein), a marker of mature, myelinating oligodendrocytes [193]. Although no significant gene

expression changes in *GALR1* were observed in either of our FTLD-expression datasets, the differential DNA methylation detected at this gene suggests potential regulatory changes in OPC life/myelination processes. *GFRA1* codes for GDNF (glial cell line-derived neurotrophic factor) family receptor α-1. GDNFs and their receptors are known to have important roles in proliferation and differentiation of OLG lineage cells [186,195,196], and *GFRA1* has been implicated as a risk gene in AD [197].

Looking at the gene expression of genes mapping to methylation sites that were differentially methylated in the meta-analysis, we saw several changes in gene expression of several OLG genes (**Figure 3.4**).

Figure 3.4 EWAS Meta-analysis OLG/OPC gene presence across FTLD1, FTLD2 and FTLD3



Meta-Analysis Gene Presence for A) OLG genes and B) OPC genes. Bar chart displaying the presence of differentially methylated OLG/OPC genes detected in the FTLD EWAS meta-analysis including the FTLD1, FTLD2 and FTLD3 data. Orange dots indicate those genes that are differentially expressed in FTLD1 and/or FTLD2-expression datasets. OLG: oligodendrocyte, OPC: oligodendrocyte precursor gene, FTLD: frontotemporal lobar degeneration, EWAS: epigenome-wide association study.

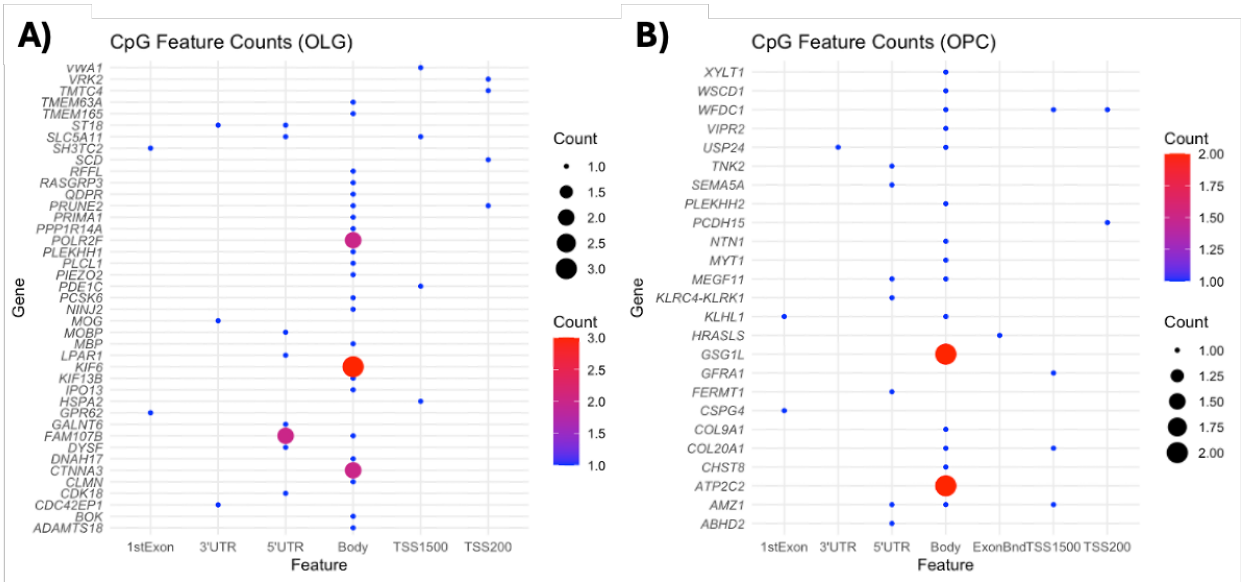
3.3.3 EWAS of glial fraction of FTLD sorted brain-nuclei data show differentially methylated OLG/OPC genes

As with each of the bulk FTLD datasets described above, we used a linear regression model to perform an EWAS with the glial fraction of a cell-type nuclei sorted FTLD dataset (which contained overlapping brain donors with the FTLD1 dataset). As described in **Section 3.1,** the use of sorted brain-nuclei data should significantly remove 'noise' from neuronal cell types, allowing increased power to detect DNA methylation changes relevant to OLGs at finer resolution. The top differentially methylated CpG relevant to OLG genes was cg10415442 which mapped to the 5'UTR region of the gene *ST18,* a gene which has been identified as a susceptibility loci of brain atrophy in AD [198] (**Appendix E**). *ST18*, which codes for suppression of tumorigenicity 18 has been found to be hypermethylated in cancer which results in a decrease in *ST18* expression, indicating that DNA methylation changes at this disease are associated with changes in gene expression.

Several genes that were found to contain differentially methylated methylation sites in the three bulk FTLD EWAS and the corresponding meta-analysis were also found to be differentially methylated in this glial fraction of the sorted data, strengthening the likelihood of their importance to glial cells/OLGs in this disease context. These genes include the three genes containing differentially methylated methylation sites across all three datasets – *CTNNA3*, *DNAH17* and *SCD.* As with the bulk data, the methylation sites mapping to the first of these two genes were found within the gene body, and for *SCD* within the promoter region. Functional enrichment analysis of OLG FTLD-sorted EWAS hits included very relevant terms such as "myelination", "glial cell differentiation" and "gliogenesis".

In regards to OPC genes, the top-most differentially methylated methylation site from this FTLD nuclei-sorted EWAS mapped to the promoter region of *PCDH15* (TSS200) (**Appendix D**). This gene was not found to show differential methylation in any of the bulk FTLD datasets. *PCDH15* codes for protocadherin-15, which has interestingly been described to have a role in OPC proliferation [199] [200]. Despite an increase in DNA methylation in FTLD vs controls (Delta-M = 0.7) (**Appendix B**), the gene did not show differential expression in either FTLD1-expression or FTLD2-expression (data not shown). As well as identifying new genes, we also saw methylation sites in genes such as *TNK2* and *GFRA1,* in which we had identified differentially methylated sites in the bulk FTLD EWAS. Functional enrichment terms relating to OPC genes from the FTLD-sorted EWAS included many terms relating to cell growth, developmental maturation and also those relating to the regulation of peptidyl-tyrosine phosphorylation (**Figure 3.5**). Interestingly, the regulation of tyrosine-phosphorylation has been reported as being key to OLG differentiation [201]

Figure 3.5 CpG genomic location of the differentially methylated OLG and OPC genes in the FTLD-sorted EWAS



Bubble plot displaying CpG feature counts for differentially methylated genes across genomic features for A) OLG and B) OPC genes. The y-axis lists the genes, and the x-axis represents various genomic features (e.g., 1st Exon, 3' UTR, 5' UTR, Body, TSS1500, TSS200). The colour and size of each bubble indicates the count of differentially methylated sites associated with a given feature for each gene. OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell, FTLD: frontotemporal lobar degeneration.

3.4 AD results and discussion

3.4.1 EWAS of one bulk AD dataset reveals oligodendrocyte lineage genes are differentially methylated in AD and show downstream changes in gene expression

To investigate DNA methylation changes relating to oligodendrocyte lineage cells in our AD cohort, we performed an EWAS and investigated effects within OLG/OPC relevant genes.

Three differentially methylated sites within this EWAS mapped to OLG genes which reached a more stringent measure of significance (FDR-adjusted p-value < 0.05) - methylation sites within *PACS2, PCSK6* and *RFFL* (**Table 3.5**). *PACS2*, which we had also found to contain differentially methylated methylation sites within in our FTLD EWAS meta-analysis (**Appendix D**), codes for phosphofurin acidic cluster sorting protein 2, and is part of the tissue expression cluster 'oligodendrocytes - myelination' (PACS2 expression), is involved in the functioning of the mitochondria-associated endoplasmic reticulum membranes, and in secretory pathway trafficking [202]. Mutations in the gene are associated with epilepsy, however its role in OLGs in neurodegeneration has not, as far as we are aware, been investigated. Of the 60 nominally differentially methylated methylation sites (P < 0.01), another 4 methylation sites also mapped to *PACS2,* all within the body of the gene, with 3 of the 4 methylation sites showing decreased DNA methylation in AD vs Control (**Appendix F**). The gene showed a decrease in gene expression in AD vs control corresponding gene expression data from the same samples (**Figure 3.6, Appendix F**).

PCSK6 codes for proprotein convertase subtilisin/kexin type 6, and was found to be a member of a cluster of upregulated AD-specific OLG genes [203]. *PCSK6* was also found to be differentially

methylated in multiple sclerosis compared to control in white matter tissue [204]. Another methylation site mapping to this gene showed nominally significant differential methylation in AD compared to controls (**Appendix F**), and both methylation sites mapping to this gene were located within the gene body and showed an increase in DNA methylation in AD vs controls. *PCSK6* also showed a decrease in expression in AD compared to controls (**Figure 3.6, Appendix G**).

The gene *RFFL*, which codes the E3 ubiquitin-protein ligase rififylin, showed an increase in DNA methylation in the promoter region (TSS1500) in AD compared to controls (**Table 3.2**). There was a second nominal significant methylation site within the same region (TSS1500) showing DNA methylation in the same direction (**Appendix F**). This gene displayed a corresponding significant decrease in gene expression in AD in matching samples (**Appendix G**), and has previously been reported to show downregulation in AD [205].

Table 3.2 Genome-wide significant differentially methylated OLG CpGs in bulk AD data

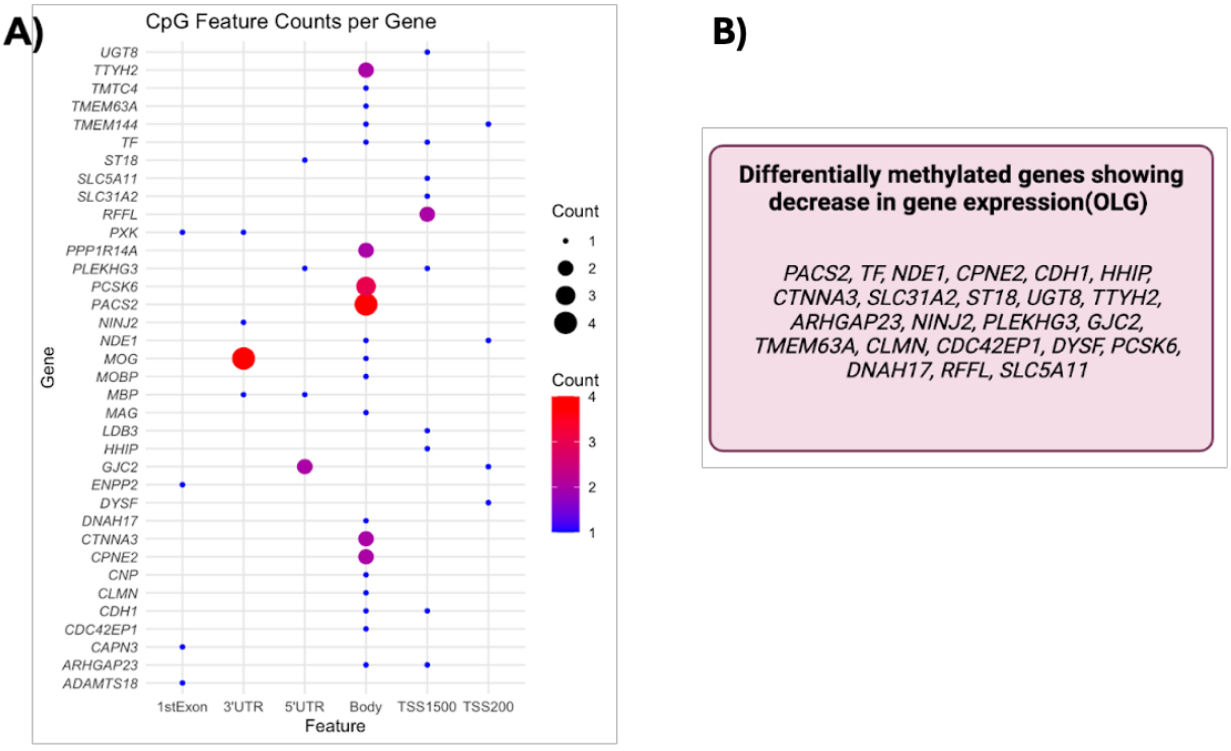
Methylation site	Gene	Delta M-value	P value	Adjusted P value (FDR < 0.05)	Feature
OLG Genes					
cg01941881	*PACS2*	-0.10263	1.00×10^{-5}	0.02123	Body
cg16179521	*PCSK6*	0.07483	3.32×10^{-05}	0.03306	Body
cg10464462	*RFFL*	0.08819	8.00×10^{-5}	0.04923	TSS1500

AD: Alzheimer's disease, OLG: oligodendrocyte

There were 60 methylation sites which passed the less stringent threshold for significance, which mapped to 36 distinct genes (**Appendix F**). We found that several genes previously

identified in our FTLD analysis were also differentially methylated in this bulk AD cohort, including *CTNNA3* and *DNAH17* (**Appendix F**). There were 23/36 differentially methylated OLG genes which also showed differential expression (P < 0.05) (**Figure 3.6, Appendix G**). Interestingly, all of these genes were downregulated in AD compared to controls.

Figure 3.6 Differentially methylated oligodendrocytes genes from AD EWAS



A) Bubble plot displaying CpG feature counts for differentially methylated genes across genomic features for OLG genes in the AD EWAS. B) Differentially expressed genes showing DNA methylation changes in the AD EWAS. AD: Alzheimer's disease, OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell, EWAS: epigenome-wide association study.

For OPC genes, 3 methylation sites passed a more stringent threshold for significance (FDR-adjusted p-value < 0.05) which mapped to *FZD9, PLEKHH2* and *TNK2* (**Table 3.3**). The CpG that was most significantly associated with changes in DNA methylation levels between AD and

control was located within the 1st exon of *FZD9* (**Table 3.3**), showing decreased DNA methylation in AD compared to controls. This gene showed a decrease in expression in our corresponding expression data (**Figure 3.7, Appendix G**). *FZD9* codes for Frizzled-9, a receptor involved in the Wnt signalling pathway [188].

PLEKHH2 was found to show decreased DNA methylation in the 3'UTR region in AD compared to controls (**Table 3.3**). We had also found a methylation site mapping to the body of this gene within our FTLD meta-analysis (**Appendix D**). *PLEKHH2* was found to be part of a cluster that was upregulated in the context of tau pathology, and is regulated by *BIN1* [206], which, as described in Chapter 1, is a strong genetic risk locus for late-onset AD[71]. However, this gene did not show changes within the available gene expression data.

The final genome-wide significant differentially methylated OPC methylation site mapped to the body of *TNK2* (**Table 3.3**). We had previously described methylation sites mapping to the promoter region of *TNK2* as being differentially methylated in FTLD (**Appendix B**), where the gene was also downregulated in FTLD compared to controls in the gene expression analysis. In the available AD expression data, *TNK2* showed no changes in gene expression between AD and controls.

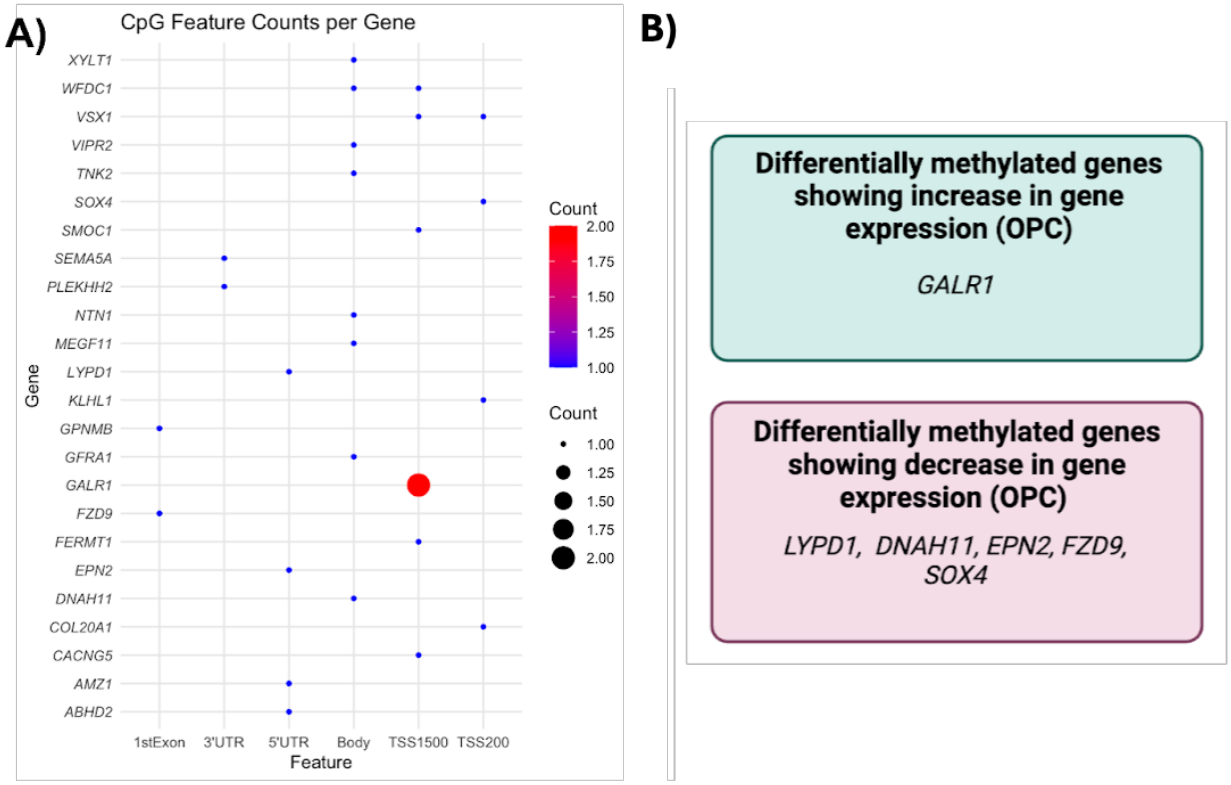
Table 3.3 Genome-wide significant differentially methylated OPC methylation sites in bulk AD data

Methylation site	**Gene**	**Delta M-value**	**P value**	**Adjusted P value (FDR < 0.05)**	**Feature**
OPC Genes					
cg20692569	*FZD9*	-0.116	4.18×10^{-6}	0.0150	1stExon
cg00916179	*PLEKHH2*	-0.084	1.00×10^{-4}	0.0429	3'UTR
cg17640485	*TNK2*	0.101	1.00×10^{-4}	0.0488	Body

AD: Alzheimer's disease, OPC: oligodendrocyte precursor cell

Using our less stringent measure of significance ($p < 0.01$), we found that in total 27 OPC methylation sites showed differential methylation between AD and control, which mapped to 24 distinct genes (**Figure 3.7, Appendix F**). We found that 7 of the methylation sites, mapping to 6 genes, within our significantly differentiated OPC methylation sites matched to genes which were differentially expressed; *LYPD1, GALR1, DNAH11, EPN2, FZD9* and *SOX4* (**Figure 3.7, Appendix G**). Aside from *GALR1,* all of these genes showed downregulation in AD compared to controls. *GALR1* (described above for the FTLD EWAS), which showed promoter hypermethylation, showed increased gene expression, which we would not necessarily expect given the described role of increased DNA methylation at promoter regions more often leading to a decrease in gene expression. *SOX4* had been previously identified as being differentially methylated and differentially expressed in FTLD, although in FTLD we had seen upregulation of *SOX4* expression, in contrast to the downregulation we see in this AD dataset.

Figure 3.7 Differentially methylated oligodendrocyte precursor genes from AD EWAS



A) Bubble plot displaying CpG feature counts for differentially methylated genes across genomic features for OPC genes in the AD EWAS. B) Differentially expressed OPC genes showing DNA methylation changes in the AD EWAS. AD: Alzheimer's disease, OPC: oligodendrocyte precursor cell, EWAS: epigenome-wide association study.

3.4.2 EWAS of two sorted brain nuclei AD datasets reveals differential methylation of oligodendrocyte lineage genes which show differential expression at single-nuclei resolution

We also carried out differential methylation analysis on two brain-nuclei sorted AD datasets; AD1 (N=31) and AD2 (N=11). The top-most differentially methylated OLG CpGs in AD1 and AD2 were located in genes *IPO13* (5'UTR) and *CRYAB* (TSS200), respectively (**Figure 3.8,**

Appendix H). There were five genes that contained differentially methylated sites across both datasets: *HHIP, DNAH17, SCD, ARHGAP23* and *MBP*, and one methylation site that passed our significance level in both datasets; cg24402667 in the transcription start site (TSS1500) of the gene *ARHGAP23* (**Figure 3.8, Appendix H**). This methylation site showed an increase in methylation in both datasets. *SCD* and *DNAH17* have already been described above as genes differentially methylated across FTLD datasets (**Figure 3.2**).

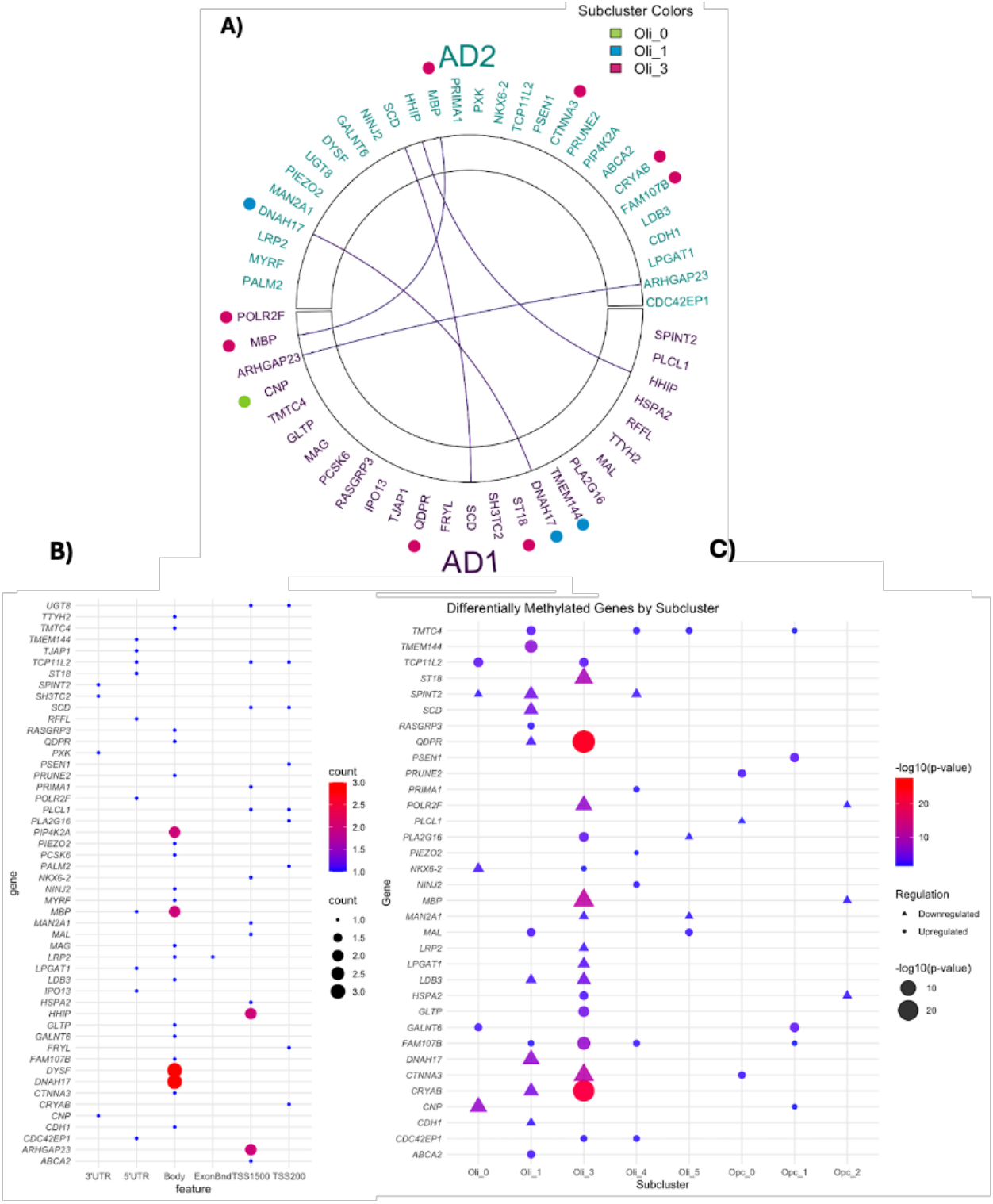
IPO13, which codes for Importin-13, is known to have important roles in the response to oxidative stress, which we have discussed as important in OPCs/OLGs in Chapter 1 [188,189,207]. *CRYAB* codes for αβ-crystallin, which is part of a heat shock family of proteins, and acts to counteract aggregation of abnormal proteins in cells [208] . *CRYAB* has been found to be upregulated in glial cells with cytoplasmic inclusions in PD [209]. In AD, it was found the concentration of the αβ-crystallin protein was increased in the temporal and frontal lobes in AD compared to controls and was found to be localised within OLGs[210].

To investigate whether changes in DNA methylation we were observing across our brain-nuclei sorted AD datasets were having effects at the gene expression at higher cell-type resolution, we analysed a publically available single-nuclei RNA-sequencing dataset comprised of 24 samples with “no-pathology”, i.e. tau Braak stage < 3, and 24 cases showing pathological hallmarks of AD; Braak stages 3-6 (**Chapter 2 Section 2.2.2**). Of the top OLG genes identified in the AD1 and AD2 EWAS mapping to OLG genes, only *CRYAB* showed differential expression, being significantly upregulated in one OLG cell cluster and downregulated in another. The most significantly differentially expressed gene of our differentially methylated genes was *QDPR,* which showed hypermethylation in the body of the gene in AD1 (**Figure 3.8, Appendices G**

and H). *QDPR,* which codes for quinoid dihydropteridine reductase, is a gene highly expressed in OLGs and, in publicly available data repositories (The Human Protein Atlas), is within a myelination cluster based on neighbouring RNA tissue expression (QDPR expression) and was upregulated (**Figure 3.8, Appendix I**) in one OLG subcluster, but downregulated in another (less significant). Although there is little data in the literature about the importance of this gene in AD pathology, it was found to be a marker of pathogenic OLG cells in a snRNAsequencing dataset, and that there are high levels of QDPR protein expression in the white matter of AD-pathology individuals [158].

MBP was also differentially expressed, and was downregulated in the same OLG cluster that *QDPR* was upregulated in - Oli-3 (**Figure 3.8**). *MBP* contained differentially methylated sites in the AD1 and AD2 across multiple gene regions; 3'UTR and Body (**Figure 3.8, Appendix H**). *MBP* is a key marker of mature OLGs, and the expression of key myelin markers such as *MBP* has been reported to be decreased in AD [211]. This could reflect a decrease in the ability of cells to express genes needed for myelination, and our finding that *MBP* is consistently differentially methylated across datasets could implicate aberrant DNA methylation as a contributing factor to such decreases in expression. The gene *CTNNA3* was also downregulated in the same gene cluster - Oli-3. *CTNNA3* showed changes in DNA methylation in the gene body in the brain-nuclei sorted data and in the bulk data, and it was also one of the three genes which showed differential methylation across all three FTLD datasets as described above (**Figure 3.8, Appendices G and H**). *DNAH17*, another gene consistently differentially methylated across sorted and bulk FTLD and AD datasets, was also differentially expressed - being downregulated in Oli-1 (**Figure 3.8**).

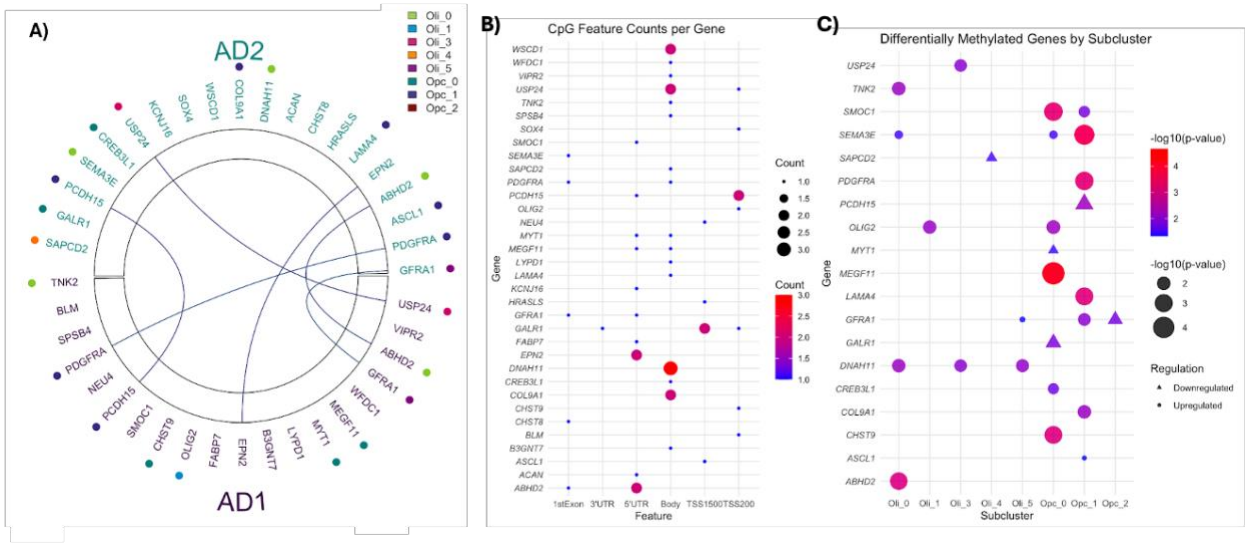
Figure 3.8 Differentially methylated oligodendrocyte genes from the glial-fraction of two AD brain-nuclei sorted EWAS



(A) Circos plots representing genes that are differentially methylated across the two brain-nuclei AD datasets (AD1 and AD2) that map to OLG genes. Each AD dataset is represented by a distinct colour; AD1: Purple, AD2: Teal. Coloured lines within the circle connect genes that are shared (overlapping) between the two AD datasets. Genes marked with coloured points outside the circular layout indicate genes that are differentially expressed (adjusted p-value < 0.05) in addition to being differentially methylated. The colour of the point indicates which single-nuclei RNA-seq cluster the gene was most significantly differentially methylated in. **C) Bubble plot displaying CpG feature counts for differentially methylated genes across genomic features.** The y-axis lists the genes, and the x-axis represents various genomic features (e.g., 1st Exon, 3' UTR, 5' UTR, Body, TSS1500, TSS200). The colour and size of each bubble indicates the count of CpG sites associated with a given feature for each gene. **D) Differentially Methylated genes that show increased or decreased genome-wide differential expression in corresponding gene expression data across single-cell subclusters.** The y-axis lists the differentially expressed and differentially methylated, and the x-axis represents the subclusters in which these genes are differentially methylated. The size of the bubbles corresponds to the magnitude of statistical significance, with larger bubbles indicating higher significance. Triangles indicate genes which show decreased expression, circles indicate genes which are upregulated. AD: Alzheimer's disease, OLG: oligodendrocyte.

Of the methylation sites mapping to OPC relevant genes, the top-most differentially methylated CpGs in AD1 and AD2 mapped to genes *VIPR2* (Body) and *GALR1* (TSS1500), respectively. There was one methylation site which passed significance in both datasets – cg2515694, located in the 5'UTR region of the gene *ABHD2* and hypermethylated in both datasets. This gene was upregulated (nominal significance, $p < 0.05$) in the Oli-1 subcluster in the single-nuclei RNA-sequencing dataset, but interestingly showed no significant differences in any of the OPC clusters. Other genes we have previously identified were also present in this analysis; *GAL1R* contained a differentially methylated methylation site in the 3'UTR in AD1 and was nominally differentially expressed (downregulated) in OPC cluster Opc_0 (**Figure 3.9, Appendices G and H**) and *GFRA1* which showed differential methylation in both datasets.

Figure 3.9 Differentially methylated oligodendrocyte precursor genes from the glial-fraction of two AD brain-nuclei sorted EWAS



(A) Circos plots representing genes that are differentially methylated across the two brain-nuclei AD datasets (AD1 and AD2) that map to OPC genes. Each AD dataset is represented by a distinct colour; AD1: Purple, AD2: Teal. Coloured lines within the circle connect genes that are shared (overlapping) between the two AD datasets. Genes marked with coloured points outside the circular layout indicate genes that are differentially expressed (adjusted p-value < 0.05) in addition to being differentially methylated. The colour of the point indicates which single-nuclei RNA-seq cluster the gene was most significantly differentially methylated in. **B) Bubble plot displaying CpG feature counts for differentially methylated genes across genomic features.** The y-axis lists the genes, and the x-axis represents various genomic features (e.g., 1st Exon, 3' UTR, 5' UTR, Body, TSS1500, TSS200). The colour and size of each bubble indicates the count of CpG sites associated with a given feature for each gene. **C) Differentially Methylated genes that show increased or decreased nominal differential expression in corresponding gene expression data across single-cell subclusters.** The y-axis lists the differentially expressed and methylated genes, and the x-axis represents the subclusters in which these genes are differentially methylated. The size of the bubbles corresponds to the magnitude of statistical significance, with larger bubbles indicating higher significance. Triangles indicate genes which show decreased expression, circles indicate genes which are upregulated. AD: Alzheimer's disease, OPC: oligodendrocyte precursor cell.

3.4.3 Meta-analysis of glial fraction of AD datasets reveals differentially methylated oligodendrocyte lineage genes

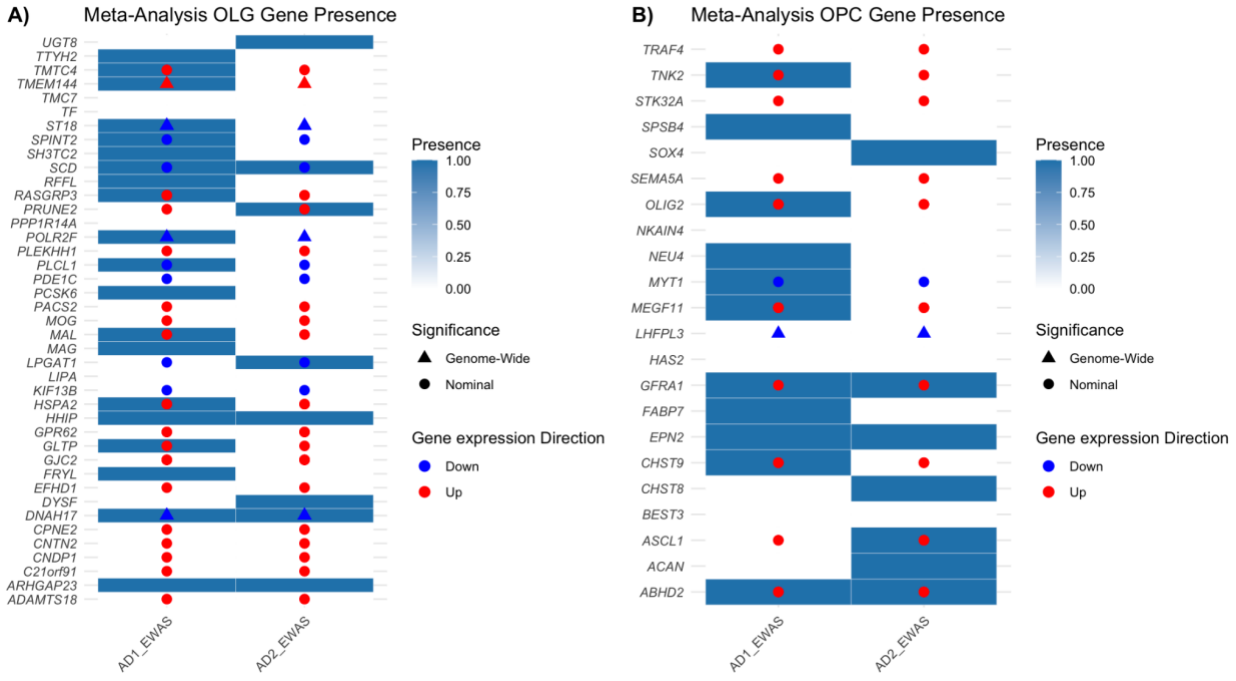
As well as carrying out the cohort specific EWAS of AD1 and AD2, we also ran a random-effect meta-analysis of the glial fraction of both brain-nuclei sorted AD datasets to identify DNA methylation signatures relevant to OPC/OLG across 42 individuals (20 AD cases and 22 controls). We found 58 methylation sites mapping to OLG genes and 30 methylation sites mapping to OPC genes to investigate further that had p-value < 0.01. These mapped to 41 and 22 distinct genes for OLGs and OPCs, respectively **(Figure 3.10, Appendix J**).

The top-most differentially methylated OLG CpG from the meta-analysis was cg24402667 in the promoter region (TSS1500) of the gene *ARHGAP23,* the same methylation site that was the top differentially methylated methylation site in the AD1 dataset. We also identified several genes which were not identified from either the AD1 or AD2 OLG individual EWAS analyses, including but not limited to the gene *PPP1R14A* (which we had previously identified as being dysregulated in FTLD), and *MOG* (a key marker of OLGs).

The top-most differentially methylated OPC CpG was cg25592910 in promoter region (TSS200) of the gene *PCDH15*, which, within the same region, contained the the top-most differentially methylated methylation site from the FTLD-sorted EWAS. Again, we were interested to see that OPC genes not found within our significance threshold in either the AD1 or AD2 EWAS were found within this meta-analysis, notably the gene *TRAF4,* which was the differentially methylated OPC gene in the FTLD1 EWAS.

We also examined, as with the AD1 and AD2 EWAS, expression of these genes in the AD snRNA-sequencing data. In terms of genes which we have not discussed above in the context of the EWAS, we saw the *MOG* was nominally upregulated in AD compared to controls, as was *PACS2* and *PLEKHH1* (which were also differentially methylated in FTLD). For OPC genes, only one gene that was present in the meta-analysis was genome-wide significantly differentially expressed - *LHFPL3,* which was downregulated in AD compared to controls. This gene has been identified as a marker of different OLG lineage states, specifically of OPCs [200].

Figure 3.10 Meta-analysis of differentially methylated genes in OLGs and OPCs and occurrence in AD1-EWAS and AD2-EWAS datasets.



Meta-Analysis OLG Gene Presence for A) OLG genes and B) OPC presence. Bar chart displaying the presence of differentially methylated OLG/OPC genes detected in the AD meta-analysis in the AD1-EWAS and AD2-EWAS datasets. Blue and red dots indicate whether the gene is alo differentially expressed in snRNA-sequencing data, with red points denoting upregulation, and blue points denoting downregulation. Triangles indicate those genes which are

differentially expressed at genome-wide significance in the snRNA-sequencing data, and circles denote those which are differentially expressed at nominal significance. OLG: oligodendrocyte, OPG: oligodendrocyte precursor cell, AD: Alzheimer's disease, EWAS: epigenome-wide association study, snRNA-sequencing: single-nuclei RNA sequencing.

3.5 Commonalities across AD and FTLD

Given that we had seen multiple genes occurring across these analyses, we were interested to explore this in more detail. Forty-two OLG genes contained differentially methylated sites in at least 3 of the 7 total datasets that we analysed (FTLD1, FTLD2, FTLD3, FTLD-sorted, AD, AD1-sorted and AD2-sorted) (**Figure 3.11**). Of the bulk datasets across FTLD (FTLD1-3) and AD, there were two genes which were always differentially methylated; *DNAH17* and *CTNNA3.* In the brain-nuclei sorted datasets; AD1, AD2 and sorted-FTLD, there were 3 genes which always showed differential methylation; *DNAH17, SCD* and *MBP.* The gene *DNAH17* showed differential DNA methylation across all 7 datasets.

It was interesting to see that key OLG genes including *MOG* and *MBP* - well described markers of OLGs, were differentially methylated in 4 and 5 datasets respectively. Both of these genes code for proteins crucial for OLGs [212213]. *MBP* was also found to be differentially expressed in the snRNA-sequencing data, suggesting functional consequences of aberrant DNA methylation. This finding highlights the importance of DNA dysregulation affecting highly relevant OLG genes in neurodegeneration.

In terms of OPC genes that showed consistent dysregulated DNA methylation across multiple datasets, we found 29 genes differentially methylated in at least 3/7 of the datasets (**Figure 3.11**). *GFRA1* showed dysregulation in 6/7 datasets analysed at our significance thresholds.

Genes *ABHD2, TNK2* and *WFDC1* all showed dysregulated DNA methylation in 5 of the 7 DNA methylation datasets. We also note that *PDGFRA* and *SOX4,* two genes known to be crucial in OPC differentiation [189,214], were differentially methylated in multiple datasets. This could indicate that OPC differentiation is a process affected by pathology associated with neurodegenerative disease, possibly contributing to altered proportions of OPCs/OLGs observed. It is also possible that these findings reflect compensatory mechanisms with attempts at upregulation/downregulation of genes necessary for OPC differentiation in response to processes associated with diseases that lead to malfunctioning of OLGs.

Figure 3.11 Gene occurrence across bulk and sorted FTLD and AD dataset for OPC and OLG genes.

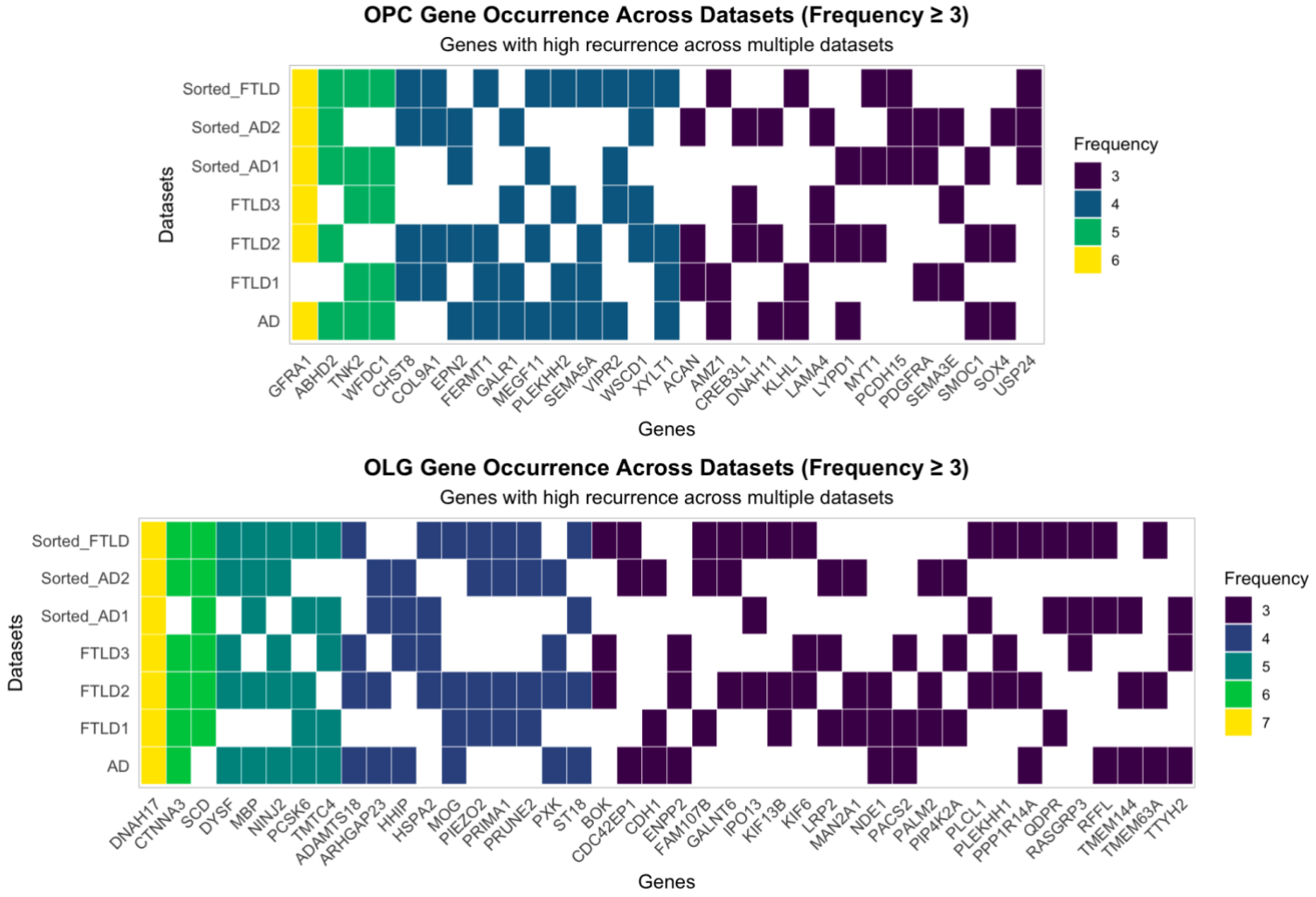


Figure illustrates the distribution of gene occurrences across FTLD1, FTLD2, FTLD3, sorted FTLD, sorted AD1, sorted AD2 and bulk AD datasets for OPC (top) and OLG (bottom) cell types. Each row corresponds to a specific dataset, and each column represents a gene that appears in three or more datasets. FTLD: frontotemporal lobar degeneration, AD: Alzheimer's disease, OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell.

3.6 General discussion

In this work, we have used differential methylation analysis, both through EWAS (epigenome-wide association studies) and meta-analyses, to investigate genes showing differential methylation in dementias, including FTLD and AD. An interesting observation was the substantial number of shared genes between AD and FTLD. Although roughly half of FTLD cases are characterised by tau pathology (which is also associated with AD pathology), a roughly equal number are also characterised by pathological inclusions of TDP-43. In our datasets, FTLD1 was composed solely of FTLD-TDP cases, FTLD2 was a mixed cohort of FTLD-TDP and FTLD-tau cases, and FTLD3 was composed of sporadic PSP cases (FTLD-tau). Whilst AD is traditionally defined as a tauopathy, it also has amyloid-beta pathology. We might have expected to see distinct differentially methylated genes across these diverse pathologies, however many genes were found to be dysregulated across multiple datasets and pathologies. Others have compared differential methylation signatures across cases with co-pathologies. Shireby et al[110] analysed whether AD associated differentially methylated sites were also associated with TDP-43 status, and found consistent effects. Such observations could indicate that there may be shared underlying mechanisms and/or vulnerabilities of OLGs across neurodegenerative diseases, and this could be shared through altered epigenetics.

Three genes; *CTNNA3, SCD* and *DNAH17*, all contained methylation sites that were differentially methylated across all 5 modes of analysis of FTLD DNA methylation data; the three bulk FTLD EWAS, the meta-analysis of these three cohorts, and in the EWAS of the glial fraction of the sorted FTLD data. *CTNNA3* (also known as *VR22*) codes for Catenin α-3, an actin-filament binding protein. The gene is of interest as it has been genetically associated with AD [173–175] and with AD particularly in females [215]. It is worth noting that this finding has been controversial, although it has also been suggested that the effect of this gene on susceptibility to

AD is dependent upon the *APOE* status [175]. *CTNNA3* was identified as a significant hit in a myelin dysregulation in multi-modal investigation of myelin dysfunction in AD[175], and was also identified through exome sequencing of MS families as a candidate gene[177]. The gene is part of the Wnt signalling pathway, dysfunction of which has been linked to aberrant (re)myelination [216]. In addition to DNA methylation changes, we found that *CTNNA3* showed downregulation in the AD single-nuclei RNA sequencing data (**Figure 3.8**). Throughout this chapter, we have identified several differentially methylated genes that have relevance to Wnt-signalling pathways, including *TRAF* as the top differentially methylated gene in FTLD2 EWAS, present in the AD-sorted meta analysis and nominally upregulated in AD snRNA-sequencing data. Another gene with functions related to Wnt signalling was *FZD9*[213,217], which was also downregulated at the gene expression level in the bulk AD RNA-sequencing dataset we analysed. Wnt signalling is an important pathway in the regulation of cell proliferation, and has been implicated specifically in the role of OPC maturation and differentiation[183]. Wnt signalling has been reported to be downregulated with age[218,219], and it has been reported that in AD, AD associated mutations such as in *APP* and *PSEN1* are associated with altered Wnt/beta-catenin signalling[218]. Wnt signalling has also previously been found to be an enriched process in an investigation of DNA methylation changes in the white matter of MSA, PSP and PD [116].

SCD, another gene that was consistently dysregulated, codes for the enzyme Stearoyl-CoA Desaturase, which is involved in fatty acid biosynthesis, particularly of the conversion of saturated fatty acids to oleic acid (and other monounsaturated fatty acids)[220]. In the brains of AD patients, elevated expression of *SCD*, and increased levels of monounsaturated fatty acids has been reported [179,221,222], although we did not see significant changes in gene expression of *SCD* in our data. It was shown that inhibition of this enzyme led to restoration of microglia activation, synaptic loss and learning and memory deficits in the 3xTf mouse (used to emulate AD pathological features)[179]. Expression of *SCD* was also found to be reduced in OLGs from white

matter in MSA cases compared to controls [223]. Although we did not see altered expression of *SCD* in our gene expression datasets, it is possible that aberrant DNA methylation of this gene in neurodegeneration is associated with other consequences of DNA methylation, for example alternative transcription. The same is true of the gene *DNAH17,* which was consistently dysregulated at the DNA methylation level without showing (in the gene expression datasets we analysed) altered gene expression. It could also be the case that limitations relating to cell compositions of RNA sequencing datasets could be disguising differential expression of these genes.

DNAH17, which codes for Dynein Axonemal Heavy Chain 17. Recently *DNAH17* was identified in a study investigating OLG genes associated with AD through analysing gene co-expression networks enriched for OLGs, and subsequently identification of which genes within these modules was nominally significantly associated with AD from GWAS statistics[172]. *DNAH17* has also been associated with PD [224], although functional relevance to neurodegeneration of this gene is unclear. Interestingly, in this analysis, other genes that we have identified as showing differential methylation in this chapter were also present in the list of 43 genes identified as part of the gene co-expression networks and associated with AD in the abovementioned study; *PIP4K2A* and *MBP*[172].

We were also interested to see which genes showed most significant and/or substantial differential expression in corresponding datasets we analysed. *QDPR* was differentially methylated in AD and FTLD, and showed significant increases in both diseases. It was notable that this was one the top most differentially expressed genes in a disease associated OLG subcluster of cells in the single-nuclei RNA sequencing data. Another gene showing differential expression patterns was *CRYAB*, to which the top-most differentially methylated site mapped to in the AD2 EWAS. This gene was the second most differentially expressed in the same cluster

of OLG genes as *QDPR*. In the original study presenting this snRNA-sequencing data, the finding of high levels of *CRYAB* and *QDPR* in AD-pathology associated OLG lineage cells was investigated at the protein level, and it was found that in the white matter of AD individuals, there was higher levels of these two proteins [158,225]. It was interesting to see that two key genes showing significant differential expression are found in this investigation of aberrant DNA methylation in OLGs. *QDPR* codes for the enzyme quinoid dihydropteridine reductase, and has a role in the regeneration of tetrahydrobiopterin (BH4) which is needed for the production of some neurotransmitters [226]. Interestingly, BH4 has been linked to reactive oxygen species (ROS), with low BH levels associated with damage through ROS [227]. The gene *IPO13,* the top differentially methylated gene in the AD1 EWAS, is also known to be important in the response to oxidative stress [207]. As we have discussed in Chapter 1 Section 1.5.3, OLGs are thought to be particularly vulnerable to oxidative stress [143,144], and oxidative stress has been linked to dysregulation of OPC differentiation [147,207].

CRYAB codes for αβ-crystallin, which is a molecular chaperone acting to protect cells from protein aggregation. Heat shock proteins, including αB-crystallin, are implicated in neurodegenerative disease pathology[228], and αβ-crystallin specifically has been found within OLG tau inclusions[229]. Interestingly, Dabir et al. [229] also report that increased expression of this protein was more marked in those diseases with greater glial pathology, indicating that this response/interaction with tau and this protein could be related to OLG lineage cells. Furthermore, it was found that in mice expressing human tau protein, infusion of αβ-crystallin was protective, leading to reduced neuronal loss[208]. Another heat-shock protein, *HSPA2,* was found to be differentially methylated in FTLD and AD, and was the top differentially methylated OLG loci in the FTLD2 EWAS.

MBP was also one of the top genes to be differentially expressed in the same cell cluster as *CRYAB* and *QDPR* (the disease associated cluster characterised by Mathys et al [158]). *MBP* was differentially methylated across all of the brain-nuclei sorted datasets. *MBP,* coding for myelin basic protein, has previously been associated with AD, with the expression of MBP increased in AD compared to controls, and evidence to suggest that MBP localises with amyloid plaques in AD [230].

We also saw multiple genes relating to OPC proliferation occurring throughout these EWAS studies, including *SOX4* and *PDGFRA* which were differentially methylated across multiple datasets. We explore the concept of differential methylation in relation to OLG lineage cell differentiation further in Chapter 5.

There are several limitations to the methods used in this chapter. An important point to note is the fact that we have not been able to use genome-wide significant *p*-values. Genome-wide significant *p*-values are the gold standard in studies to mitigate the interpretation of false-positives as real results. Although we have had to use a more relaxed threshold in this analysis in order to explore effects of DNA methylation changes across OLG/OPC specific genes, we have attempted to mitigate the occurrence of false-positives by repeating our analysis across independent datasets, and carrying out meta-analyses, and paying close attention to those results which are consistent across datasets to increase robustness.

Although we have taken steps to adjust for important confounding factors such as cell-type heterogeneity, it is still possible that we have not been able to account completely for different samples containing different cellular composition. As described in Chapter 3.1, cell-type composition of samples strongly influences DNA methylation, as DNA methylation has such a crucial role in cell differentiation and specificity. As well as using deconvolution to account for

such factors, we have also included in our analysis the use of glial fraction portions of brain-nuclei sorted DNA methylation data, which should be less affected by noise from neuronal populations.

Although we have, as far as possible, integrated gene expression data in order to understand the effects that DNA methylation changes may be having on OLG/OPC genes during disease processes, there are many limitations, firstly lack of data available. It is also true that even when we do have access to matching DNA methylation and gene expression data from the same brain donors, sampling methods may mean that different tissue is analysed and correlations between DNA methylation changes and gene expression changes may be missed/obscured.

Another important note would be that we have chosen to focus to a moderate degree on which genes contain differentially methylated sites rather than individually differentially methylated CpG sites. There are limitations to this approach - importantly that gene size will influence this. Larger genes will be more likely to contain methylation sites that are differentially methylated, a point confounded by the fact that we are having to use non-genome wide significant *p*-values. However, we note that several of the genes that we report to show differential methylation consistently show effects across one region of the gene, for example *SCD* was consistently differentially methylated around the promoter region of the gene rather than across the whole gene, strengthening the likelihood that this represents a real biological effect rather than rather than a random artefact driven by gene size or the inclusion of multiple CpG sites within larger genes.

Future work set by this work described in this chapter would include further validation of genes identified that show aberrant DNA methylation. Whilst we have attempted to investigate gene expression changes, we have not investigated changes such as alternative splicing, which has

been associated with altered DNA methylation [231,232]. Examining effects of DNA methylation on transcript expression could elucidate less well characterised functions of DNA methylation in the context of OLG dysfunction in neurodegeneration. Therefore, if warranted, follow up studies aimed at investigating functional consequences of DNA methylation changes of specific genes on gene expression would be necessary.

In this chapter, we have described OLG/OPC genes which show differential methylation across dementias, and have demonstrated that many such genes show downstream changes in gene expression. While several of these genes have been previously implicated in dementia pathogenesis, many have not been associated with mechanisms involving DNA methylation changes. This finding highlights the potential role of epigenetic regulation as a key driver in OLG/OPC dysfunction and implicates DNA methylation as a critical pathway contributing to disease pathology in these cell types, which has not been investigated to a high degree. This chapter thus reinforces the contribution of DNA methylation in the pathology of neurodegenerative diseases and highlights several genes as candidates for further investigation in the role of DNA methylation alterations in OLG lineage cells in neurodegeneration.

Chapter 4: Using network analysis to uncover disease-associated DNA methylation signatures relevant to oligodendrocyte lineage cell types across neurodegenerative diseases

4.1 Introduction

Network analysis is an agnostic systems biology approach to identify higher order relationships between biological entities. Weighted-gene correlation network analysis (WGCNA) is a widely used form of network analysis which takes into account strength of connections between such biological entities [233]. In the context of this thesis, a network is a representation of the relationship between levels of DNA methylation in certain genomic sites and/or expression of genes, and how such relationships may be involved in neurodegenerative disease. Such analysis is a powerful tool to identify clusters of genes and DNA methylation sites that show similar patterns of expression and/or DNA methylation across samples, and therefore thought to be involved in similar functions or pathways. These clusters, also called modules, can then be probed to identify relevance to traits such as disease status, pathology status or features, and/or genetic signals. Network analysis is also a powerful tool in investigating an enrichment of cell-type specific signatures where single-cell/single-nuclei data is unavailable or to complement such data. Presence of such enrichment would support that such a signature would be affecting that particular cell-type more than others.

Network analysis has been widely used to study gene and protein expression in disease to identify disease relevant expression changes that may not be detected using more basic analyses considering gene by gene such as differential expression. A few examples will be described next. In AD, co-expression network analysis was used on post-mortem derived transcriptomic data from LOAD by Zhang et al., who highlighted an immune/microglia module which was strongly associated with AD pathology, and identified the gene *TYROBP* as a regulator of genes within this module, thought to be involved in amyloid-beta turnover [234]. Similarly, Chen et al. applied co-expression network analysis to identify hippocampus-related AD markers, which were subsequently validated in an independent cohort, demonstrating the utility of network analysis in developing diagnostic biomarkers and therapeutic targets [235]. Beyond transcriptomics, network analysis has also been applied to protein co-expression. For instance, Zhang et al. examined AD-associated relationships between co-expressed proteins, facilitating the discovery of previously unrecognized disease-associated proteins and pathways and thus expanding our understanding of AD pathology. Indeed, a key strength of network analysis is the ability to investigate groups of co-regulated biological entities, enabling identification of functional modules of groups that may be missed by more reductionist approaches. The studies described above provide examples which underscore the value of network analysis in elucidating disease mechanisms, identifying biomarkers, and uncovering therapeutic targets.

Network analysis is often used as a complementary approach to analysis such as differential expression. In the context of DNA methylation studies, this is applicable as well. The use of analyses such as differential methylation (as described in Chapter 3) to investigate disease associated DNA methylation changes, whilst no-doubt invaluable, may lead to important genes being overlooked, possibly due to small numbers of methylation sites reaching significance (due to DNA methylation changes being small), and/or large numbers of genes being identified, and

key genes being lost. In contrast to diseases such as cancer, where changes in epigenetic signatures including DNA methylation tend to show larger effect sizes and are more easily detectable, changes in DNA methylation in neurodegenerative disease tend to be more subtle and gene specific. Studying DNA methylation changes in neurodegenerative diseases is indeed complex for many reasons, including the inherent heterogeneity of the brain, which is composed of distinct cell types, including neurons, astrocytes, microglia, oligodendrocytes, and others, each with its own methylation profile and contributing to the overall methylation landscape. Further complexity is added due to cell-type composition changes over time - loss of neurons, activation of astrocytes and other glial cells, etc. Given this complexity, the use of more holistic systems biology approaches, complementary to traditional differential methylation analyses, can bring new insights to the understanding of DNA methylation changes in neurodegenerative diseases.

Relatively few studies have taken the approach of creating DNA methylation networks to uncover DNA methylation signatures that are related to disease. One of the first studies to describe the application of network analysis to DNA methylation data was conducted by Hovarth et al. [236], who utilised the method to identify signatures of co-methylated genes that were associated with ageing in blood and brain [236]. Our group has also recently published studies investigating DNA methylation signatures in neurodegenerative diseases. Bettencourt et al. used network analysis to investigate DNA methylation changes in white matter in the α-synucleinopathy MSA [118], leading to identification of a co-methylation module including a methylation site in the gene *SNCA* (the gene encoding α-synuclein), lending the first evidence for DNA methylation changes contributing to the pathology of MSA. We have also co-methylation network analyses across FTLD subtypes, which identified ubiquitin and glutamatergic signaling as key processes in FTLD pathology. This analysis also highlighted *OTUD4* and related genes as genes of significant interest in FTLD [119].

Although network analysis is implemented on one dataset at a time, useful methods have been produced to allow users to assess 'replication' of one network in another dataset - using module preservation analysis. Module preservation analysis is useful not only for replicability analysis (i.e. is a signature present in one dataset also present in another dataset of the same disease), but also in cross-region or cross-disease analysis. Cross-region analysis is particularly useful when studying neurodegenerative diseases where region-specific pathology is present, for example in AD. Understanding distinct regulatory changes that are present in distinct brain regions could add insight into disease progression. Cross-disease preservation analysis has utility in identifying both shared dysregulated signatures across distinct diseases, and those which are unique to a particular disease or pathology [116].

The role of OLGs/OPCs and their molecular dysregulation in neurodegeneration, as described in Chapter 1, is not well described, and the use of tools such as network and preservation analyses can be useful. In this chapter, we use network analysis to uncover signatures present in DNA methylation datasets derived from post-mortem brain tissue that are associated with disease and are found to be relevant to OLG lineage cells. Through this approach, we identify modules of genes showing similar DNA methylation patterns. By identifying such signatures and carrying out hub gene (i.e. highly interconnected genes) and functional enrichment analysis, we then uncover OLG relevant pathways that may be altered by DNA methylation changes across and within neurodegenerative diseases. For this, we have leveraged all datasets previously used in our EWAS analyses described in Chapter 3. In this chapter, we also describe the use of an additional bulk AD dataset where DNA methylation profiles are available for multiple brain regions; the ERC (entorhinal cortex), DLPFC (dorsolateral prefrontal cortex), HIPPO (hippocampus) and CRB (cerebellum). As discussed in Chapter 3 Section 3.1, both bulk and

brain-nuclei sorted datasets have their uses, and as in Chapter 3, we utilise both to investigate OLG/OPC enriched signatures across neurodegeneration here.

4.2 Methods

4.2.1 Datasets Used

In this chapter, we describe our use of multiple DNA methylation datasets that were previously generated using the Illumina 450K or EPIC arrays, derived from both bulk brain tissue, and from the glial fraction of brain-nuclei sorted samples. The datasets we analysed are briefly described in **Table 4.1**, and a detailed characterisation of the cohorts is given in **Table 2.1** (**Methods Section 2.1**).

For the co-methylation networks, we used the datasets FTLD1, FTLD2, FTLD3, FTLD-sorted and AD-sorted datasets which we used also in Chapter 3. We have described the quality control and data processing in Chapter 3, as these datasets were also used for EWAS analysis. In place of the AD dataset (bulk tissue) that we utilised in Chapter 3, we instead utilise in this chapter a bulk AD DNA methylation dataset which contained samples from different brain regions; cerebellum (CRB), hippocampus (HPPO), entorhinal cortex (ERC) and the dorsolateral prefrontal cortex (DLPFC). This dataset was made publicly available by Semick et al. [150]. The quality control and processing of these DNA methylation datasets is described in detail in **Chapter 2, section 2.1**. Linear regression models were then applied for covariate adjustment of the data. Adjusted M-value matrices were then used as input in downstream network construction.

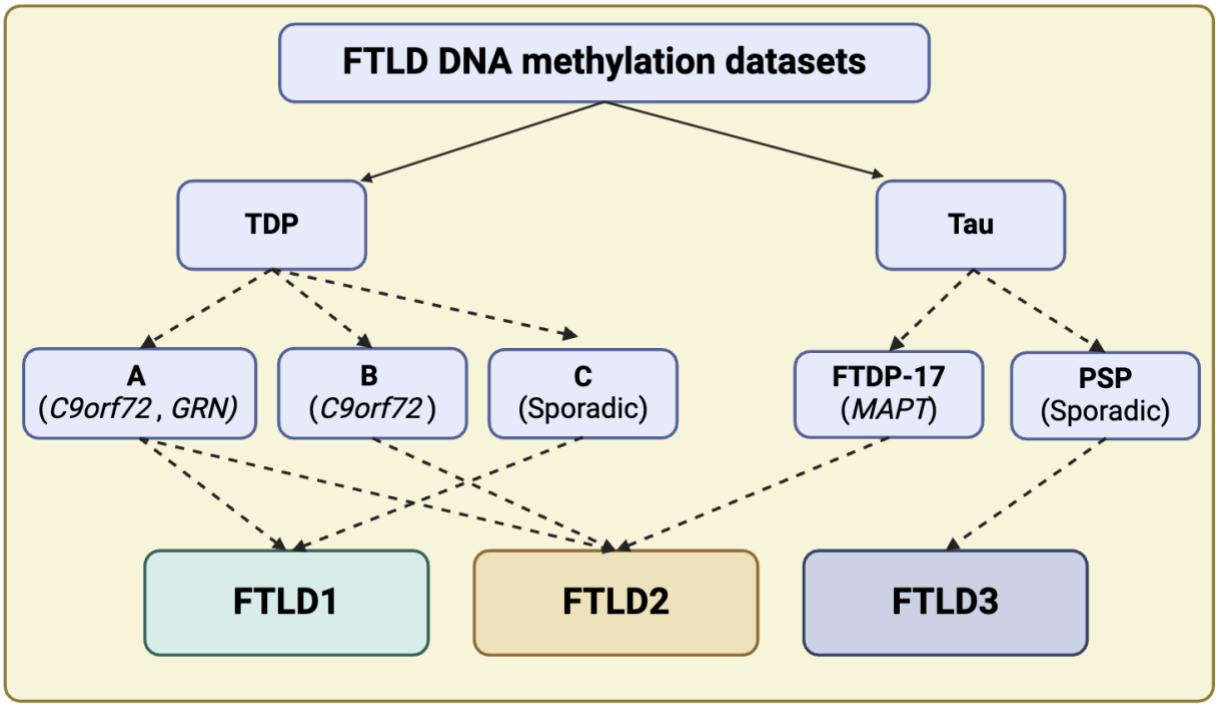
Table 4.1 Overview of datasets analysed using co-methylation network analysis in this chapter

Cohort	Pathological FTLD subtypes and controls included after quality control	Soft-thresholding power used	Number (%) of methylation sites included	Reference
FTLD1 (Frontal Cortex)	FTLD (*N* = 15) and Controls (*N* = 8)	16	56,001 (20%)	Fodder et al.[119]
FTLD2 (Frontal Cortex)	FTLD (*N* = 34) and Controls (*N* = 14)	10	56,001 (20%)	Menden et al.[149]
FTLD3 (Frontal Cortex)	FTLD (*N* = 93) and Controls (*N* = 70)	12	56,001 (20%)	Weber et al.[117]
FTLD-sorted (Frontal Cortex)	FTLD (N = 19) and Controls (N = 6)	12	53, 410 (10%)	Bettencourt lab, unpublished
AD-HIPPO	AD (*N* = 17) and Controls (*N* = 48)	12	44,609 (20%)	Semick et al.[150]
AD-DLPFC	AD (*N* = 21) and Controls (*N* = 46)	14	53,826 (20%)	
AD-ERC	AD (*N* = 20) and Controls (*N* = 49)	12	52,296 (20%)	

AD-CRB	AD (*N* = 24) and Controls (*N* = 43)	16	52, 712 (20%)	
AD1-sorted (Occipital Cortex)	AD (*N* = 15) and Controls (*N* = 16)	16	57,620 (20%)	Gasparoni et al.[122]
AD2-sorted (Frontal Cortex)	AD (*N* = 5) and Controls (*N* = 6)	12	52,637 (20%)	Bettencourt lab, unpublished

FTLD, Frontotemporal lobar degeneration; AD, Alzheimer's Disease; HIPPO; hippocampus, DLPFC; dorsolateral prefrontal cortex, ERC; entorhinal cortex, CRB; cerebellum

Figure 4.1 Outline of FTLD cohorts used in analysis



FTLD, Frontotemporal lobar degeneration; PSP, Progressive supranuclear palsy.

4.2.2 Co-methylation network analysis

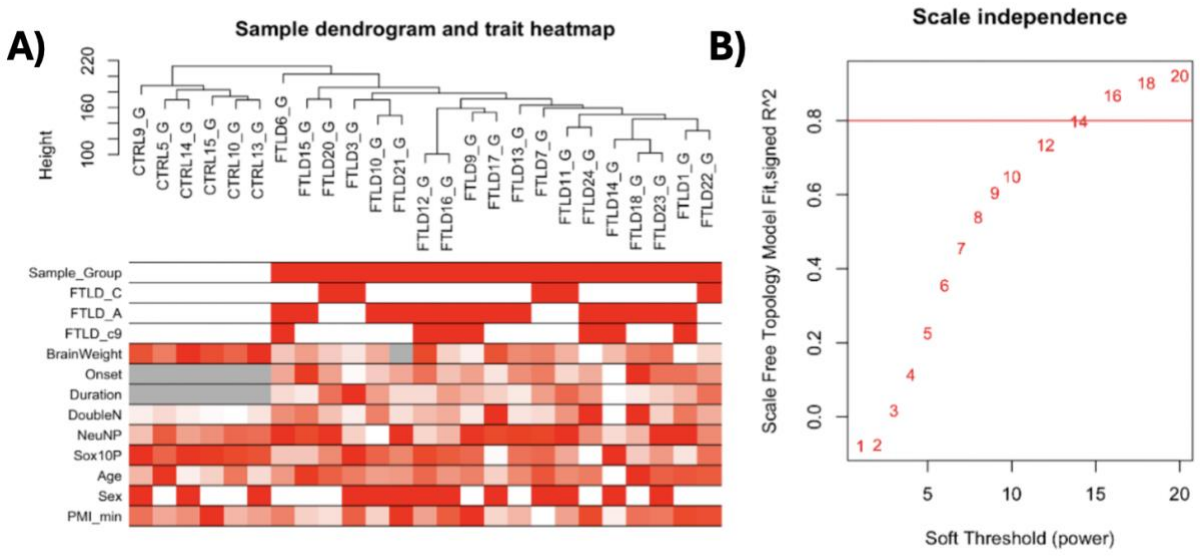
We used a systems biology approach to identify co-methylation modules in our datasets (i.e. clusters of highly correlated DNA methylation sites - CpGs). For this we have used weighted gene correlation network analysis (WGCNA [233]). For all datasets, we excluded those CpGs annotated to intergenic regions (as such methylation sites would be less biologically informative in this context), and selected only those showing the highest variance (top 10-20% depending on dataset) across individuals within each dataset (again to increase ability to find disease relevant effects). The percentage of total methylation sites and number of methylation sites included for each analysis is given in Table 3.1. We used adjusted M-values (described in Chapter 2 section 2.1) for such high variance CpGs as input for this analysis. Before network construction, we assessed clustering of samples based on traits (i.e. disease status, age, sampling, batch effects) in order to confirm that networks would represent biologically meaningful relationships rather than dataset artifacts (**Figure 4.2 A**). To construct the weighted co-methylation networks, pairwise methylation expression correlations were calculated to create a similarity matrix[233]. Soft-thresholding was applied by raising the correlations to a power which puts more emphasis on strong correlations whilst still maintaining weaker correlations. Optimal soft-thresholding power is selected, using the pickSoftThreshold function in WGCNA, for which scale-free topology is reached (**Figure 4.2.B**). Scale-free topology is a network structure wherein most of the nodes (i.e. CpGs) have few connections, but few CpGs have many connections. Once this is chosen, networks are constructed using the WGCNA blockwiseModules function, and following the WGCNA package workflow. In this context, key components of the biological network are nodes, edges and modules. Nodes are the CpG sites where DNA methylation is measured, edges are the interaction between nodes (i.e. correlation

between methylation at CpG sites), and modules (also known as clusters) represent groups of nodes within the network that are more densely connected to each other than to other nodes in the network.

Module membership (MM) was then reassigned for each network using the applyKMeans function of the CoExpNets package [237]. Module membership is a measurement of the relationship between a CpG and it's co-methylation module within the network - i.e. how strongly related a CpG's methylation is associated with the module's eigengene (a single representative methylation profile for the module, derived through summarising the methylation of all CpGs within a module). Reassignment through this method employs k-means clustering to reclassify CpGs in order to minimize within-cluster variance and maximise between-cluster separation. CpGs with high MM are considered hub genes, meaning they are highly connected, and likely key biological drivers within that module.

After co-methylation modules are identified, we carry out module-trait correlation analysis (as part of the WGCNA workflow[233]) to understand the relationship between co-methylation modules and disease traits. Each co-methylation module's eigengene is correlated with the disease trait (e.g presence or absence of disease), and each module is given a correlation coefficient and significance value (i.e. p-value). For each of the modules within our networks, we calculate the correlations and significance with disease status to identify disease relevant modules. We apply Bonferroni multiple testing correction, i.e. correct correlation p-value over the number of modules detected, to minimize the finding of false-positives.

Figure 4.2 Preprocessing of data pre-network analysis examples



A) Sample dendrogram and trait heatmap. Hierarchical clustering dendrogram showing the relationships among samples based on DNA methylation profiles. Branch heights indicate the similarity/dissimilarity between samples, with shorter branches representing closer relationships. The associated heatmap illustrates the distribution of sample traits, including sample group classifications (CTRL, FTLD_C, FTLD_A, FTLD_c9) and covariates such as cell type proportions, age and sex. The intensity of red shading in the heatmap corresponds to the magnitude or presence of the trait value, while gray shading represents missing data. **B) Diagnostic determination of soft-threshold power for network construction.** The plot shows the scale-free topology model fit (signed R) as a function of the soft-threshold power. Each point represents a tested power value, with the corresponding R^2 value indicated. The red horizontal line at R^2=0.8 represents the threshold for scale-free topology. A power value of 14 or higher achieves the desired R^2, indicating the optimal soft-threshold power for constructing a scale-free network in this dataset.

4.2.3 DNA methylation cross-network preservation analysis

As a method for differential network analysis, i.e. to identify which co-methylation modules in each of the generated networks are preserved (i.e. shared) or perturbed (i.e. unique) in the other relevant datasets, we employ module preservation analysis, as described by Langfelder et al. [238] . For each network (taken as the “reference dataset”), module preservation in the other

datasets (the “test data”) can be calculated using the modulePreservation function implemented in WGCNA. In all instances, the “test data” contains methylation values (adjusted M-values) for CpG sites used to construct the “reference dataset” network. As a measure of module preservation, we use Z-summary statistics (a composite measure to summarise multiple preservation statistics). A Z-summary greater than 10 indicates strong preservation of this module in the “test data”, a Z-summary of between 2 and 10 indicates moderate preservation, and a Z-summary less than 2 indicates no preservation. The Z-summary integrates two measures of module preservation; connectivity-based preservation statistics and density-based preservation statistics. Preservation of connectivity ($Z_{connectivity}$) is a measure of the inter-modular connectivity (i.e. the relationship between the modules in the test and reference dataset). This measure is based on the following:

$$Z_{connectivity} = median(Zcor.kIM, Z\,cor.kME, Zcor.cor)$$

Where $Z_{cor.kIM}$ quantifies preservation of intramodular connectivity - i.e. how strongly a particular CpG site is connected to other sites within the same module, $Z_{cor.kME}$ as a measurement of the preservation of the correlation between module membership in the reference and test datasets, and $Z_{cor.cor}$ as a measure of how much relationships between individual CpG sites is retained in the test dataset compared to the reference dataset.

Preservation of density measures how much the modules are retained between datasets, i.e. a comparison of the adjacency between CpG sites in the reference datasets with the same CpG sites in the test datasets:

$$Zdensity = median(Z\,meanCor, ZmeanAdj, ZpropVarExpl, ZmeanKME)$$

Where $Z_{meanCor}$ is a measure of the pairwise correlations between CpG sites within a module, i.e. whether the correlations between sites in the test dataset are similar to those in the reference dataset, $Z_{meanAdj}$ is a measure of the preservation of adjacency (derived using soft-thresholding as described above), $Z_{propVarexpl}$ is a measure of how much the module eigengene captures a similar proportion of variance in the test dataset as in the reference dataset, and $Z_{meanKME}$ as a measure of preservation of the module membership (i.e. how strongly each CpG site is associated with the module eigengene) - calculated as the correlation between the CpG site and module eigengene.

$Z_{summary}$ is then defined as:

$$Zsummary = \frac{Zdensity + Zconnectivity}{2}$$

Where equal weighting is given to each of the $Z_{density}$ and $Z_{connectivity}$ statistics.

For more details see the original publication by Langfelder et al. [238]

4.2.4 Functional and cell-type enrichment analyses

To investigate disease-associated modules relevant to OLG/OPCs, we have carried out cell-type enrichment analysis on disease associated modules using the package EWCE[120] and

associated data[121]. EWCE requires an input gene list (which here is the list of genes mapping to CpGs within a module) and a reference expression dataset (which in this case is a single-cell mouse RNA sequencing dataset). The package then enables a bootstrap-based cell-type enrichment test, which tests whether observed enrichment of a gene list in specific cell types deviates from the null distribution generated through random sampling. Two measures are then given - the fold change (i.e. how much more expressed the genes are in a cell type compared to random expectation), and a p-value, which indicates the statistical significance of such an enrichment. To validate any cell-type enrichment results, we also checked for enrichment of OLGs and OPCs within modules using curated gene lists (courtesy of Piras et al.[122]) and a Fisher's exact test (henceforth called gene list enrichment). The gene lists we have been provided with were described in Chapter 2 section 2.4. We also also checked modules within one cohort for enrichment of peroxisome related genes. We obtained a list of peroxisome relevant genes downloaded from GO, and tested for enrichment of these within modules using the same approach as the gene list cell-type enrichment analysis.

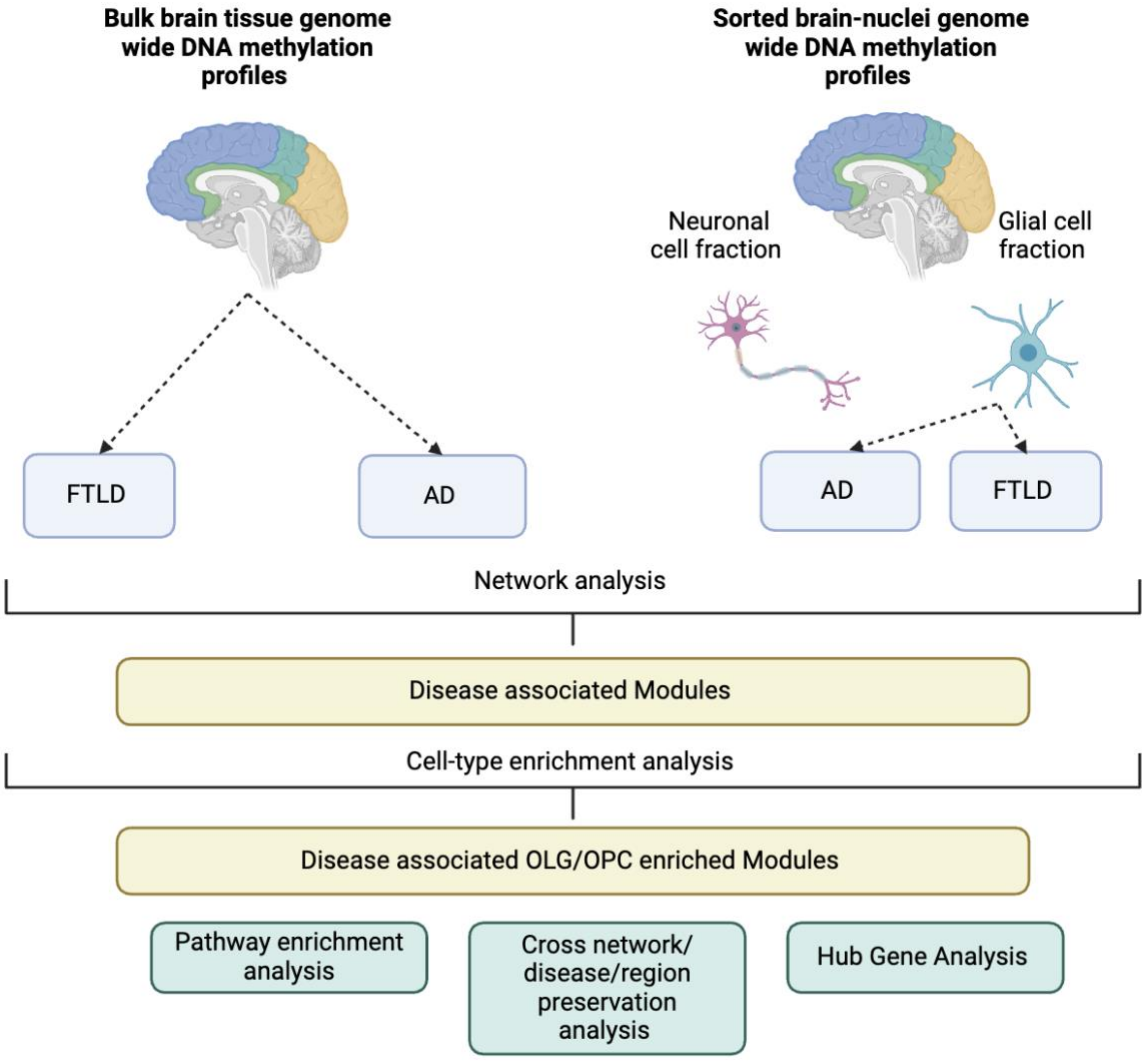
Functional enrichment analysis to investigate the function of genes found within disease associated OLG/OPC enriched co-methylation network modules was then performed using HumanBase (https://hb.flatironinstitute.org/), which allows for tissue specific enrichment analysis to be performed. In the case where the number of genes mapping to modules was less than 3,000 (the maximum input for this database), we have used all genes present within the module. In the case where the number of genes exceeds 3,000, we have chosen the top 3,000 genes which methylation sites show the highest module membership. An example of the output of module enrichment analysis using HumanBase is given in **Appendix K**. All enrichment terms reported had significance value $q < 0.05$, with q used as a statistical measure used to account for multiple hypothesis testing, given the large number of genes being tested.

4.2.4 Follow-up Datasets

We used several gene expression datasets to investigate gene expression changes in genes identified through co-methylation network analysis, including the FTLD1 and FTLD2 expression datasets, and the AD snRNA-sequencing dataset used in Chapter 3 and described in detail in Chapter 2 Section 2.2.

A schematic illustrating the approaches utilised in this chapter is given in **Figure 4.1**.

Figure 4.3 Schematic of the analysis workflow in Chapter 4



The workflow includes two primary sources of genome-wide DNA methylation profiles: (1) Bulk brain tissue DNA methylation profiles from FTLD and AD samples and (2) Sorted brain nuclei DNA methylation profiles separated into neuronal and glial cell fractions for AD and FTLD samples. Both datasets undergo network analysis to identify disease-associated modules. Subsequently, a cell-type enrichment analysis is conducted to extract disease-associated Oligodendrocyte (OLG)/Oligodendrocyte Progenitor Cell (OPC) enriched modules. Further downstream analyses include Pathway Enrichment Analysis, Cross-network/disease/region preservation analysis, and Hub Gene Analysis. AD; Alzheimer's disease, FTLD; frontotemporal-lobar degeneration, OLG; oligodendrocyte, OPC; oligodendrocyte precursor cell

4.3 Results

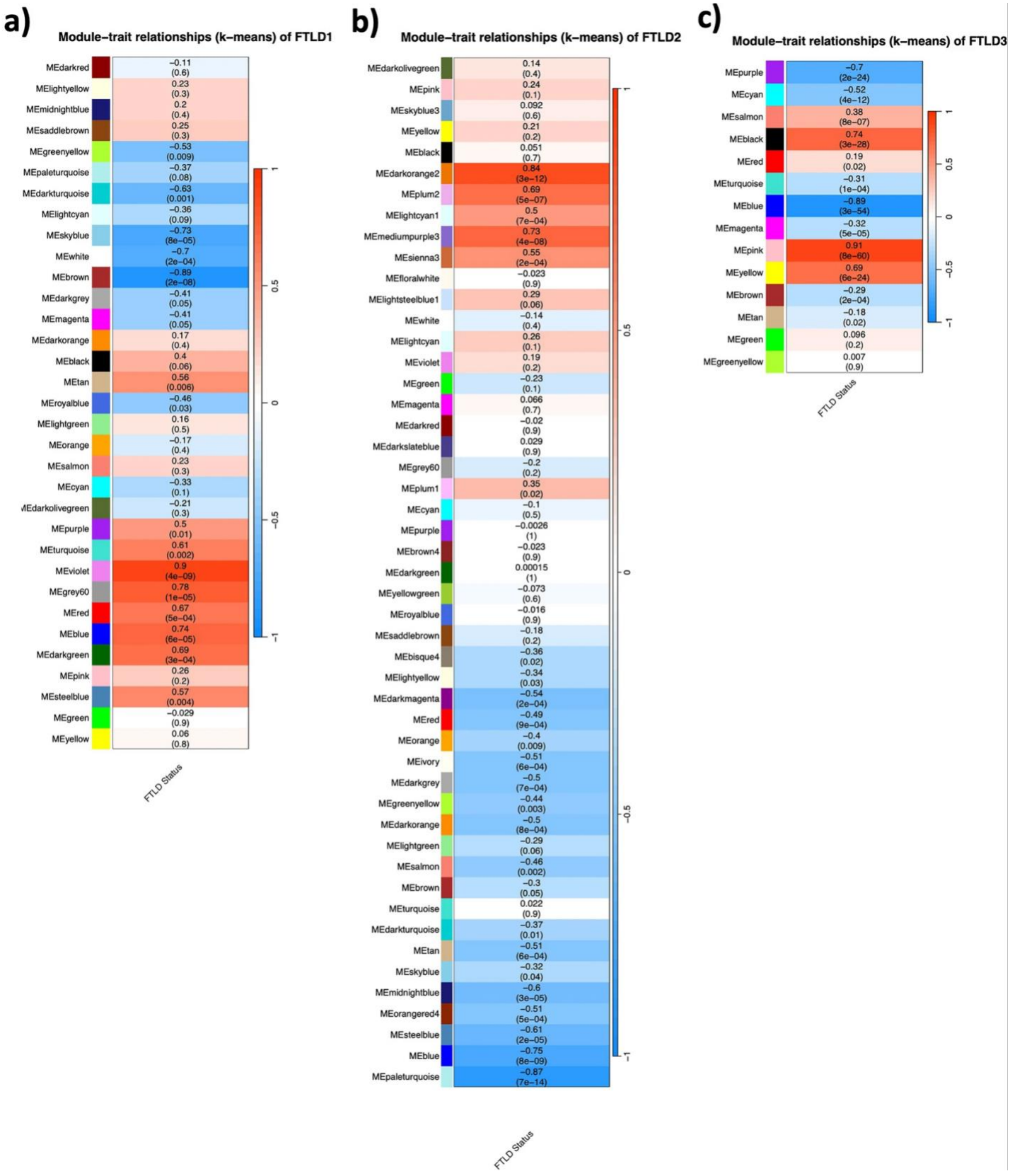
4.3.1 Bulk Tissue Networks

4.3.1.1 FTLD Networks

We used three FTLD cohorts to generate networks encompassing a range of FTLD pathological subtypes in order to investigate the role of DNA methylation and oligodendrocytes dysregulation/dysfunction across the pathological spectrum. For each of these three cohorts, we generated a distinct network.

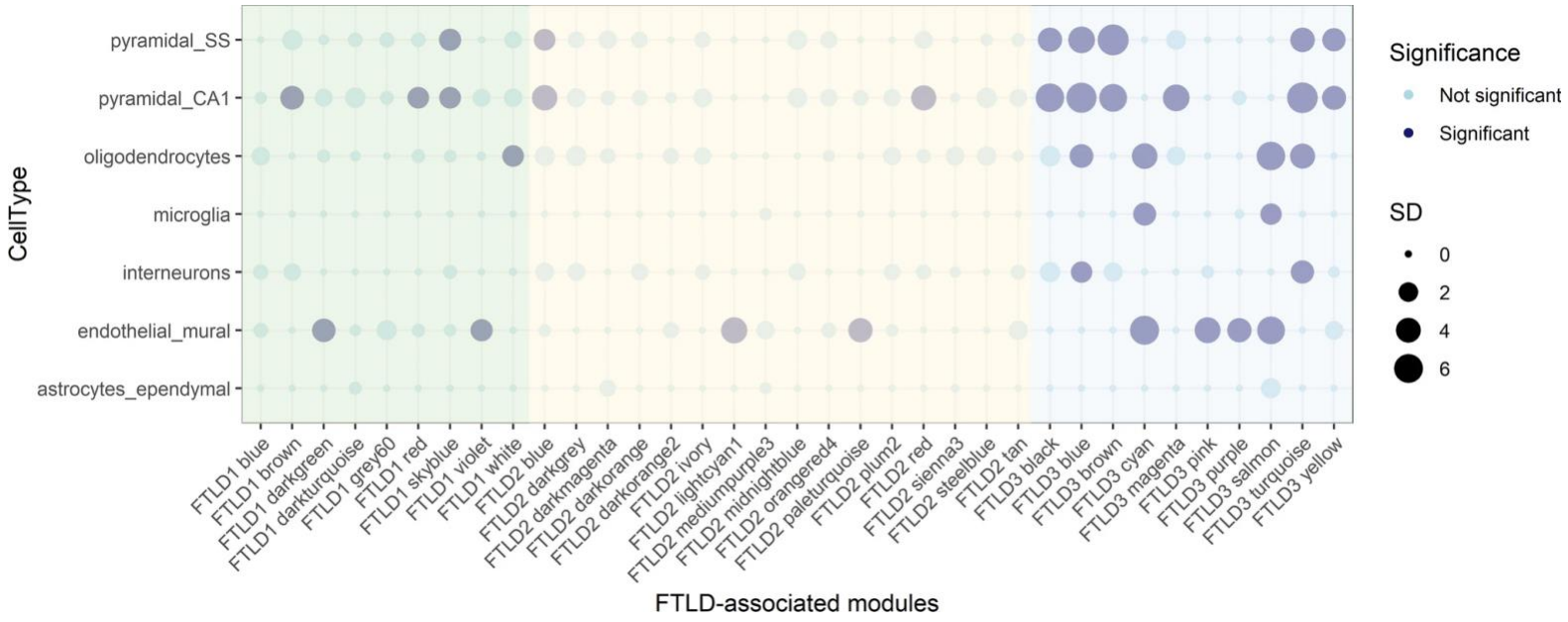
After multiple testing corrections, 9/33 ($p < 0.002$, 0.05/33 modules), 16/49 ($p < 0.001$, 0.05/49 modules) and 10/14 ($p < 0.004$, 0.05/14 modules) co-methylation modules were found to be associated with the disease status (i.e. FTLD or control), for the FTLD1, FTLD2 and FTLD3 networks, respectively (**Figure 4.4 A–C**). Several disease-associated modules were enriched for OPC/OLG relevant genes in the FTLD1 and FTLD3 networks, but not in the FTLD2 networks (as further detailed below) (**Figures 4.5 and 4.6**).

Figure 4.4 Module-trait correlations for the FTLD co-methylation networks



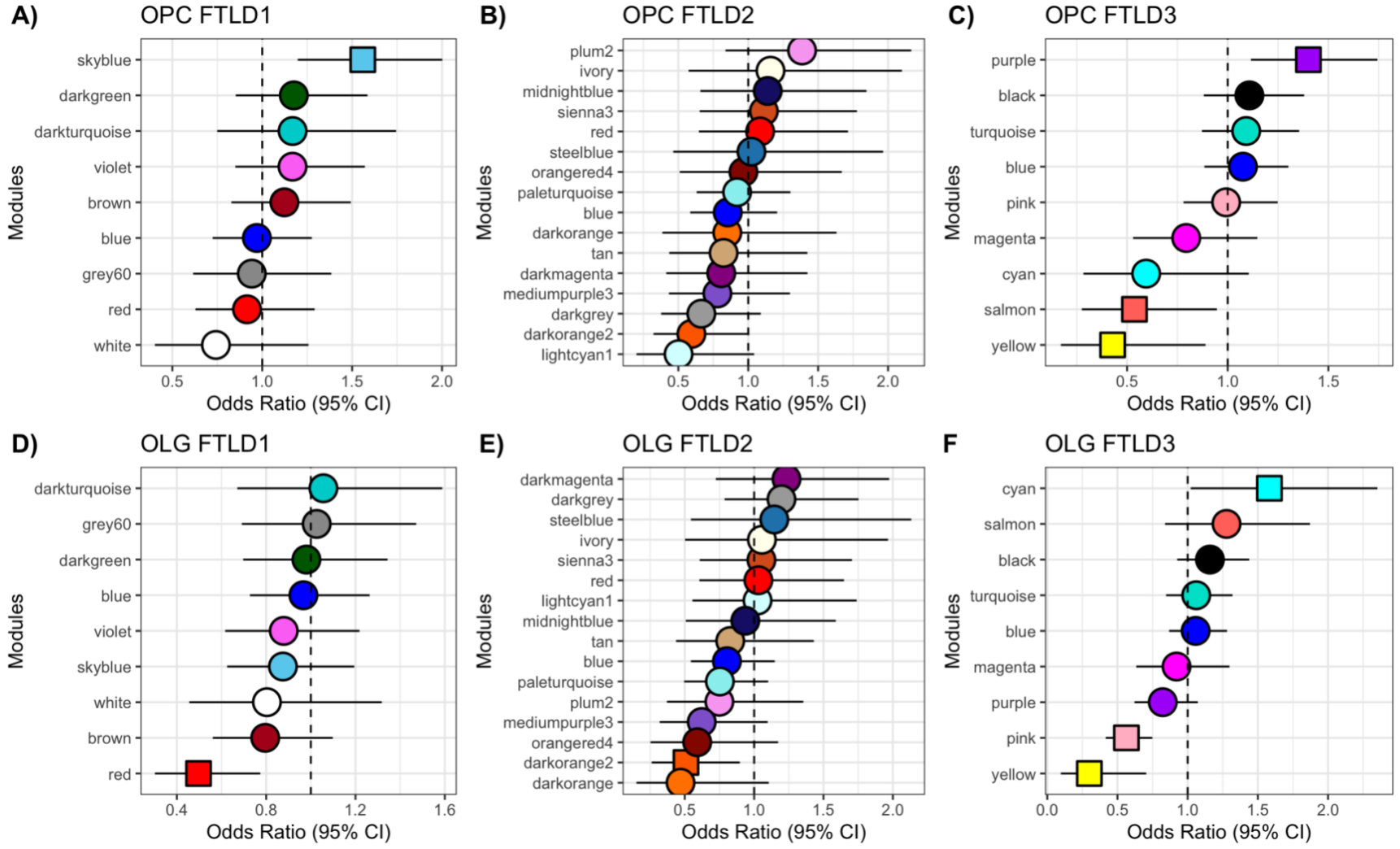
Module trait relationships for networks A) FTLD1 B) FTLD2; C) FTLD3. The rows represent the co-methylation module Eeigengenes (ME) and their colours, and the column represents the correlation of the methylation levels of CpGs in each module with the disease status. *p*-values are presented within each cell and the colour scale at the right indicates the strength of the correlation (darker cells depict stronger correlations, with blue representing negative and red representing positive correlations). FTLD: frontotemporal lobar degeneration.

Figure 4.5 Cell-type enrichment for all FTLD-associated co-methylation modules across the three co-methylation networks using EWCE.



Enrichment of modules using EWCE. Green denotes FTLD-associated modules in the FTLD1 network; Yellow denotes FTLD-associated modules in the FTLD2 network; Blue denotes FTLD-associated modules in the FTLD3 network. Dark filled circles highlight the cell types found to be significantly enriched with adjusted p<0.05 after Bonferroni correction over all cell types within each module; the size of the circles represents the number of standard deviations (SD) from the mean. Cell-type enrichment analysis on the FTLD-related modules was performed using the package EWCE [239] and associated single-cell transcriptomic data[240]. FTLD: frontotemporal lobar degeneration, OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell.

Figure 4.6 Gene list enrichment for all FTLD-associated co-methylation modules across the three co-methylation networks.



Enrichment analysis using curated lists of OLG and OPC relevant genes which CpGs within the FTLD associated modules of each network map to. A-C) Enrichment of methylation sites mapping to OPC relevant genes within FTLD1, FTLD2 and FTLD3 networks within disease associated modules. **D-F)** Enrichment of methylation sites mapping to OLG relevant genes within FTLD1, FTLD2 and FTLD3 networks within disease associated modules. Odds Ratios and 95% confidence intervals were calculated using Fisher's exact test. Asterisks indicate significant Enrichment ($p < 0.05$). FTLD: frontotemporal lobar degeneration, OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell.

FTLD1 network modules

For the FTLD1 network, one module was significantly enriched for OPC-relevant genes, the skyblue module, and one for OLG-relevant genes, the white module, from the gene list enrichment (**Figure 4.6**).

The top hub gene of the skyblue module was *TMEM168*, which codes for transmembrane protein 168. From publicly available gene clustering analysis, *TMEM168* is part of a gene cluster highly associated with white matter/myelination (TMEM expression), and the gene showed the highest expression in OLGs and OPCs compared to all other brain cell types (TMEM expression). *TMEM168* has been found to be involved in glioblastoma (a tumour of glial cell origin in the brain) pathogenesis through involvement in the Wnt/beta-catenin signalling pathway [241]. Functional enrichment terms within this module included "sphingolipid translocation" and "phosphatidylinositol-3-phosphate biosynthetic process" (**Table 4.2**). Sphingolipids include sphingomyelins, which are a crucial component of myelin. Sphingolipid pathways are known to be important in the pathology of MS [242], making this functional enrichment term notable.

Through the use of EWCE, the FTLD1 white module was additionally identified as being enriched for OLG genes (**Figure 4.5**). The hub gene of this module was *CTDSP2*, which codes for CTD (Carboxy-Terminal Domain, RNA Polymerase II, Polypeptide A) Small Phosphatase 2.

It is thought to be involved in the suppression of neural transcripts in non-neuronal cells [243]. Functional enrichment of genes within this module included “amyloid precursor protein biosynthetic process” and “regulation of amyloid precursor protein biosynthetic processes” (**Table 4.2**).

FTLD3 Modules

In the FTLD3 network, 5 modules were found to be significantly enriched for OLG/OPC relevant genes. Through gene list enrichment analysis, the cyan module was identified as being enriched for OLG genes and the purple module for OPCs. In the analysis using EWCE, the blue, turquoise, salmon and cyan modules (the latter of which was also identified in the gene list enrichment analysis) were all identified as being enriched for OLG cell type genes (**Figure 4.5, Table 4.2**).

The hub gene of the FTLD3 cyan module was *DDIT4L (*also known as *REDD2),* which codes for DNA-damage-inducible transcript 4 like, which is a stress response protein. REDD2 is a known inhibitor of mTOR signaling pathways [244], which play crucial roles in OLG differentiation and myelination [245]. In line with this, functional enrichment terms from this module included “negative regulation of cell differentiation” and “negative regulation of cell proliferation” (**Table 4.2**)

The hub gene of the FTLD3 salmon module was *PCK1*, which codes for phosphoenolpyruvate carboxykinase 1, and has been reported to be a risk gene for AD [246] and has also been found to influence brain atrophy in MS patients [247]. PCK1 is an enzyme involved in gluconeogenesis, a process whereby glucose is created from non-carbon sources such as lipids or proteins, and mice with reduced *Pck1* expression exhibit dysregulated lipid metabolism [248]. The finding that this gene is dysregulated in FTLD as well could indicate common processes involving

dysregulated lipids and myelin across these neurodegenerative diseases. Functional enrichment terms within this module included “regulation of superoxide anion generation” (**Table 4.2**), suggesting that genes within this module may be important in influencing how reactive oxygen species are produced and/or cleared.

Additionally identified from EWCE analysis was the FTLD3 blue module (**Figure 4.5, Table 4.2**). The hub gene of the FTLD3 blue module was *GDAP1,* which codes for ganglioside induced differentiation associated protein 1. Mutations of this gene are associated with a demyelinating subtype of Charcot Marie Tooth disease (an inherited peripheral neuropathy)[249]*.* Functional enrichment analysis of the blue module revealed terms including “ganglioside biosynthetic processes” and “ganglioside metabolic processes” (**Table 4.2**). The presence of genes and terms relating to gangliosides within this module is notable due to their importance in the stability in myelin structure [250]. Other enriched terms of interest to us were “glycolipid metabolic processes”, “glycolipid biosynthetic processes” and “glycosphingolipid metabolic processes”, due to the importance of both glycolipids and glycophospholipids in myelin [251,252] The blue FTLD3 module notably included the methylation site mapping to *PIP4K2A* that was found to be the most significantly differentiated methylation site mapping to OLG relevant genes in the FTLD3 dataset (**Chapter 3, Section 3.1, Appendix B**). As described in Chapter 3, this gene has a role in cholesterol trafficking between peroxisomes and lysosomes. Interestingly, knockdown of *GDAP1* (the FTLD3-blue hub gene) was found to alter peroxisome function and morphology in vitro [253]. Due to the presence of the *PIP4K2A* gene in this module and the link between *GDAP1* and peroxisomes, we were interested to see if any other genes relating to the function of peroxisomes were enriched within this module, and found that indeed this module was significantly enriched for these terms (**Appendix L**).

The purple module, enriched in the gene list analysis for OPC genes (**Figure 4.6**), had the hub gene *FMN1*, which codes for the protein Formin 1. It has been suggested that formins may have links to neurodegeneration through tau biology [254,255] We found nothing in the literature linking this gene to neurodegeneration,*FMN1* in particular has also been implicated in glioblastoma invasion [256]. Enrichment terms from this module included “positive regulation of reactive oxygen species metabolic process”. As discussed, OPCs are known to be particularly vulnerable to reactive oxygen species, making this enrichment term notable.

The final module enriched for OLG genes was the FTLD3 turquoise module. The hub gene of which is *FUT8*, coding for the enzyme Fucosyltransferase 8. From public databases, the gene is highly expressed in oligodendrocytes and is part of a tissue expression cluster enriched for OLGs and associated with myelination (FUT8 Tissue expression). *FUT8* codes for an enzyme involved in glycosylation (a post-translational modification of proteins). This module contained many functional enrichment terms of interest, including “plasma membrane bounded cell projection morphogenesis” and “phosphatidylinositol metabolic processes” (**Table 4.2**). OLGs extend cell projections which wrap around axons in order to form the myelin sheath, and as components of cell membranes phosphatidylinositols are likely relevant [257] **.**

Table 4.2 Hub genes and functional enrichment of disease associated OLG lineage gene enriched modules across FTLD-networks

Module	**Disease trait(s) associated with**	**Hub Gene**	**Key enrichment terms relevant to OLG/OPCs and/or neurodegeneration**	**Type of enrichment analysis**
FTLD1 skyblue	FTLD	*TMEM168*	"sphingolipid translocation" "phosphatidylinositol-3-phosphate biosynthetic process"	EWCE
FTLD1 white	FTLD	*CTDSP2*	"amyloid precursor protein biosynthetic process" "regulation of amyloid precursor protein biosynthetic process"	EWCE
FTLD3 cyan	FTLD	*DDIT4L*	"negative regulation of cell differentiation" "negative regulation of cell proliferation" "cholesterol homeostasis" "protein targeting to lysosome" "protein localization to lysosome"	Gene list OLG and EWCE
FTLD3 purple	FTLD	*FMN1*	"positive regulation of reactive oxygen species metabolic process"	Gene list OPC
FTLD3 salmon	FTLD	*PCK1*	"regulation of ion transmembrane transport" "regulation of ion transport" "regulation of superoxide anion generation"	EWCE
FTLD3 blue	FTLD	*GDAP1*	"ganglioside biosynthetic processes" "ganglioside metabolic processes"	EWCE
FTLD3 turquoise	FTLD	*FUT8*	"neuron death in response to hydrogen peroxide" "regulation of mitochondrial membrane potential" "plasma membrane bounded cell projection morphogenesis" "regulation of axogenesis" "phosphatidylinositol metabolic process" "regulation of plasma lipoprotein particle levels"	EWCE

FTLD: frontotemporal lobar degeneration, OLG: Oligodendrocyte, OPC: oligodendrocyte precursor cell, EWCE: expression weighted cell type-enrichment analysis.

We were next interested to see if any of the identified hub genes were differentially expressed in either the FTLD1-expression or FTLD2-expression datasets (described in Chapter 2, Section 2.2). Four of the eight hub genes of the FTLD associated OLG/OPC enriched co-methylation network modules were significantly (genome-wide) differentially expressed in at least one of these two expression datasets. The gene *GDAP1,* the hub gene of the FTLD3-blue module , was significantly downregulated (**Table 4.3**). *DDIT4L,* hub gene of the FTLD3-cyan module was significantly upregulated in both FTLD datasets (**Table 4.3**). *CTDSP2* showed genome-wide significant upregulation in the FTLD2-expression dataset, and nominally significant upregulation in the FTLD1-expression dataset (**Table 4.3**). *PCK1* showed significant upregulation in the FTLD1-expression dataset and nominally significant upregulation in the FTLD2-expression dataset (**Table 4.3**). The identification of gene expression changes in hub genes uncovered through DNA co-methylation network analysis adds support to our findings implicating their dysregulation in disease pathogenesis.

Table 4.3 Expression of hub genes in FTLD RNA-sequencing datasets

	FTLD1-Expression		**FTLD2-Expression**	
Gene	**logFC**	**Adjusted P-value**	**logFC**	**Adjusted P-value**
GDAP1	-0.369	0.022	-0.891	0.007
DDIT4L	1.144	0.001	1.386	0.007
CTDSP2	0.204	0.086	0.499	0.018
PCK1	1.190	0.034	1.855	0.097

FTLD: Frontotemporal lobar degeneration, logFC: log Fold change, FDR Adjusted P-value: p-value adjusted for multiple testing correction, *GDAP1: DDIT4L: CTDSP2: PCK1.*

Oligodendrocyte-enriched, FTLD-associated co-methylation modules are preserved between sub-types

To assess replication of FTLD-associated co-methylation modules across datasets, we then ran preservation analysis for each dataset against each of the networks. As previously outlined, the three FTLD datasets we have analysed here are composed of different subtypes of FTLD. In brief, FTLD1 is composed of only FTLD-TDP cases (subtypes A and C), FTLD2 is composed of both FTLD-TDP (subtypes A and B) and FTLD-tau cases, and FTLD3 is composed only of sporadic FTLD-tau cases (PSP cases). Interestingly, all four disease-associated OLG/OPC enriched FTLD3 modules were highly preserved in the FTLD2 data (**Table 4.4**). The salmon module was also highly preserved in the FTLD1 data. This was interesting given that its hub gene was *PCK1,* a gene implicated in lipid dysregulation that has also been associated with the neurodegenerative diseases MS and AD [247]. Combined with the high preservation found across FTLD subtypes, this could indicate that this signature represents common dysregulation of OLG/OPCs and lipids across a spectrum of neurodegenerative diseases and pathological subtypes. It was interesting to see that the FTLD3 blue module, which we had identified as a module implicated in peroxisome function, whilst not being preserved in the FTLD1 data, was highly preserved in the FTLD2 data. As described, whilst FTLD2 contains mixed cases of FTLD-TDP and FTLD-tau pathology, whilst FTLD1 only contains FTLD-TDP. It is possible that the FTLD3-blue module signature is associated with tau pathology, hence it being preserved in a cohort containing this pathology and not preserved in the FTLD1 data. The same is true of the FTLD3-cyan module.

Table 4.4 Module preservation between OLG enriched FTLD co-methylation modules in FTLD1, FTLD2 and FTLD3 networks

	Preservation in FTLD1	**Preservation in FTLD2**	**Preservation in FTLD3**
FTLD1			
skyblue		Moderate	Not Preserved
white		Not Preserved	Moderate
FTLD3			
blue	Not preserved	High	
turquoise	Not preserved	High	
salmon	High	High	
cyan	Moderate	High	

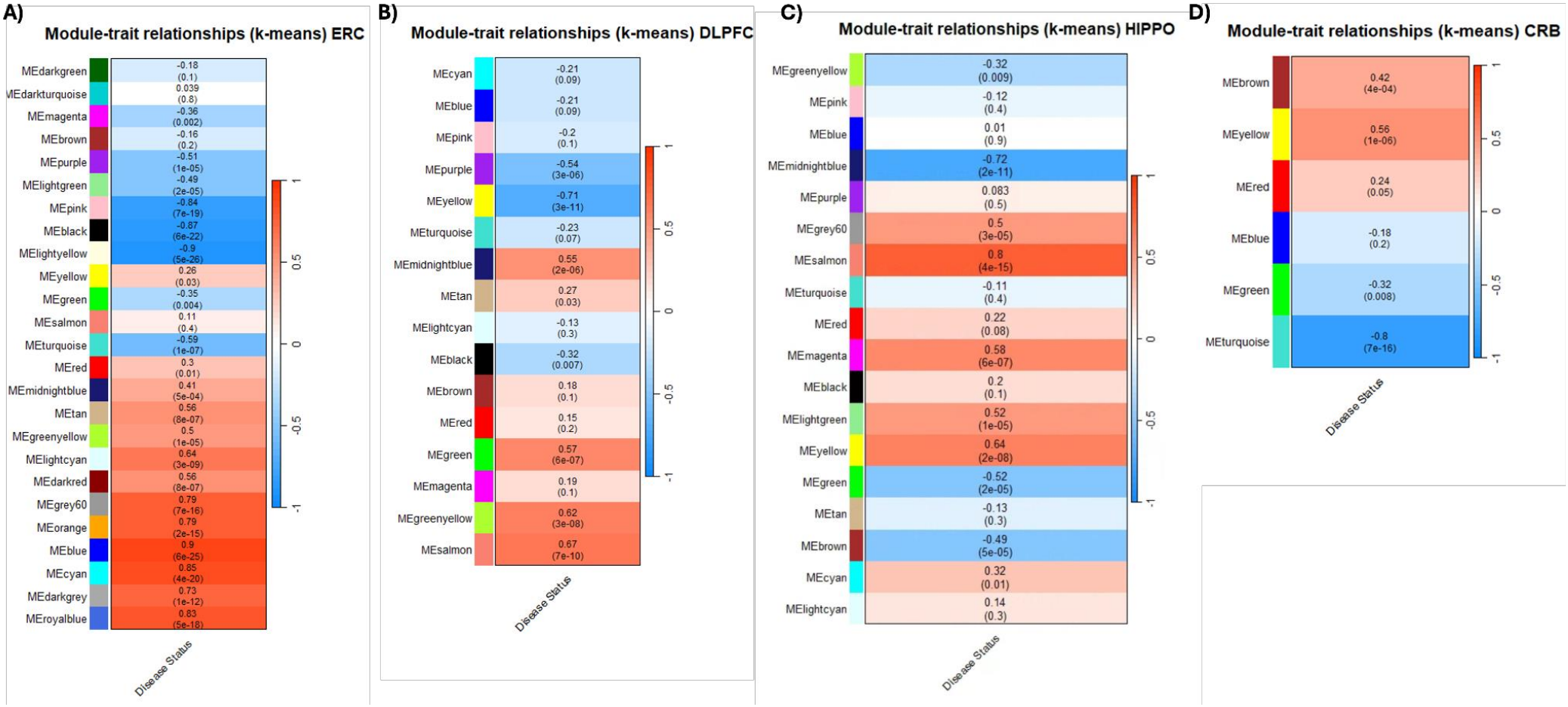
FTLD: Frontotemporal lobar degeneration

4.3.1.2 Alzheimer's Disease Networks

For the analysis into oligodendrocyte-related co-methylation signatures in AD, we utilised a dataset composed of DNA methylation profiles from multiple brain regions; entorhinal cortex (ERC) , hippocampus (HIPPO), cerebellum (CRB) and dorsolateral prefrontal cortex (DLPFC) (**Figure 4.1**).

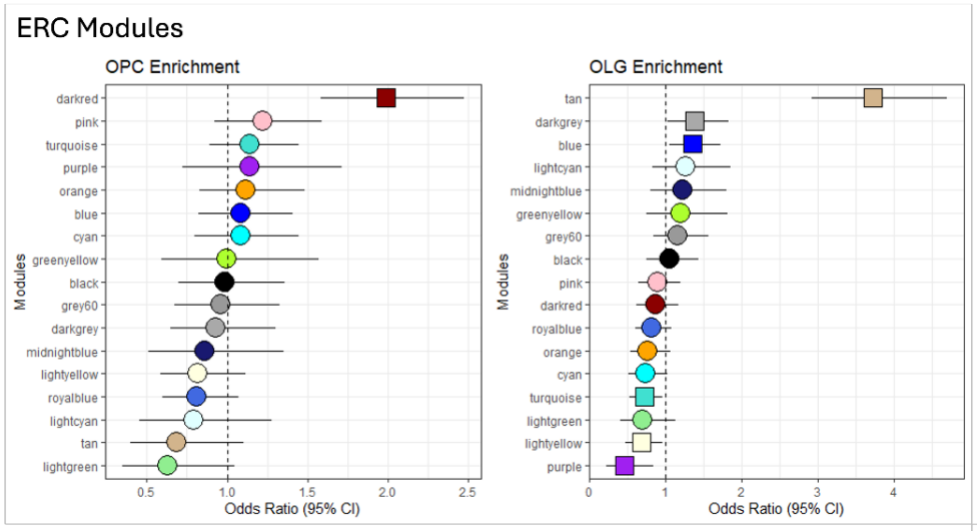
For each of these brain regions, we generated co-methylation networks (**Figure 4.7**). For the ERC, DLPFC, HIPPO and CRB, networks, 17/25 ($p < 0.002$, 0.05/25 modules), 6/16 ($p < 0.003$, 0.05/16 modules), 7/18 ($p < 0.003$, 0.05/18 modules) and 4/6 ($p < 0.008$, 0.05/6 modules), co-methylation modules were found to be associated with the disease status (i.e. AD or control), respectively (**Figure 4.7 A-D**). We next analysed which of these modules were enriched for OLG/OPC relevant genes, again using both gene list enrichment methods and the package EWCE as described for FTLD (**Figures 4.8 and 4.9**)

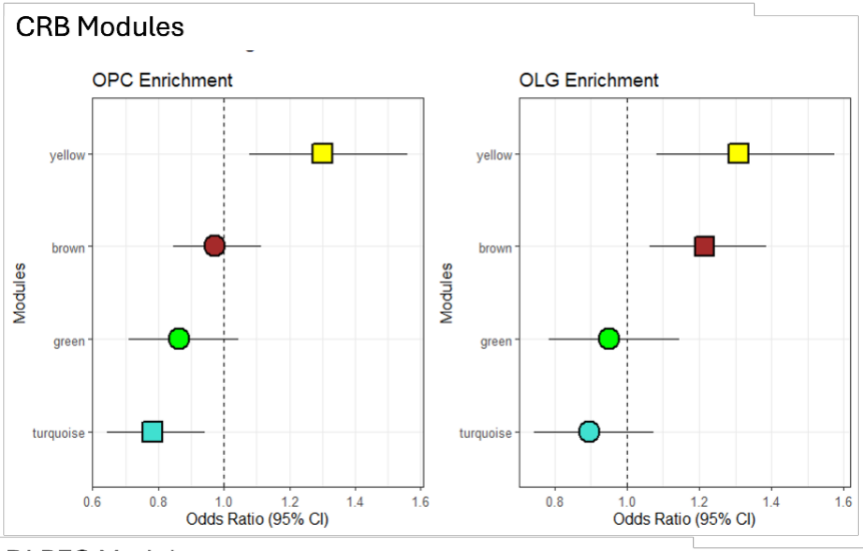
Figure 4.7 Module-trait correlations for the AD brain region co-methylation network

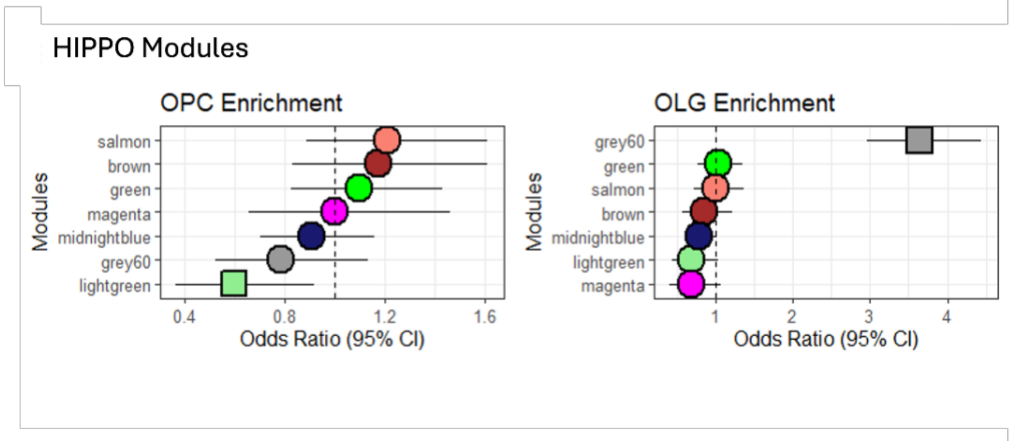


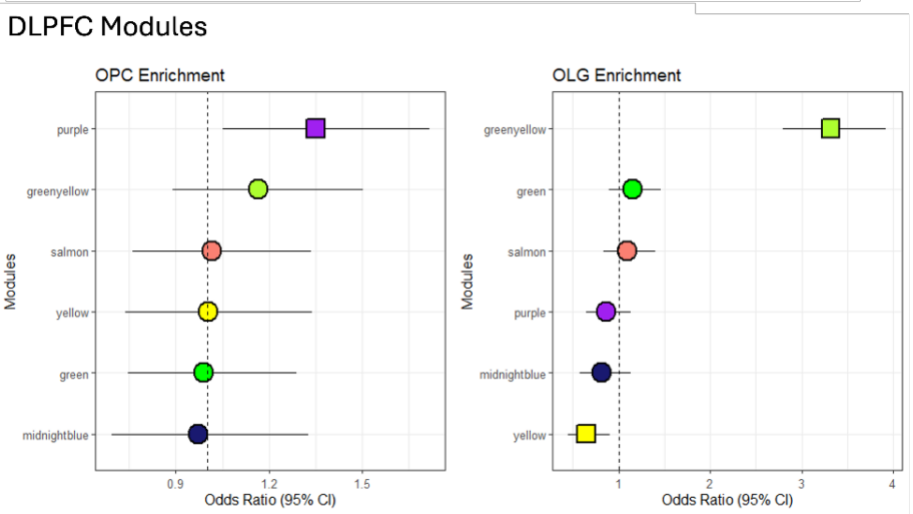
Module trait relationships for A) ERC, B) DLPFC, C) HIPPO, and D) CRB networks. The rows represent the co-methylation module eigengenes (ME) and their colours, and the column represents the correlation of the methylation levels of CpGs in each module with the disease status. *p*-values are presented within each cell and the colour scale at the right indicates the strength of the correlation (darker cells depict stronger correlations, with blue representing negative and red representing positive correlations). ERC: entorhinal cortex, DLPFC: dorsolateral prefrontal cortex, HIPPO: hippocampus, CRB: cerebellum

Figure 4.8. Cell-type enrichment for all AD-associated co-methylation modules across the four brain-region co-methylation networks



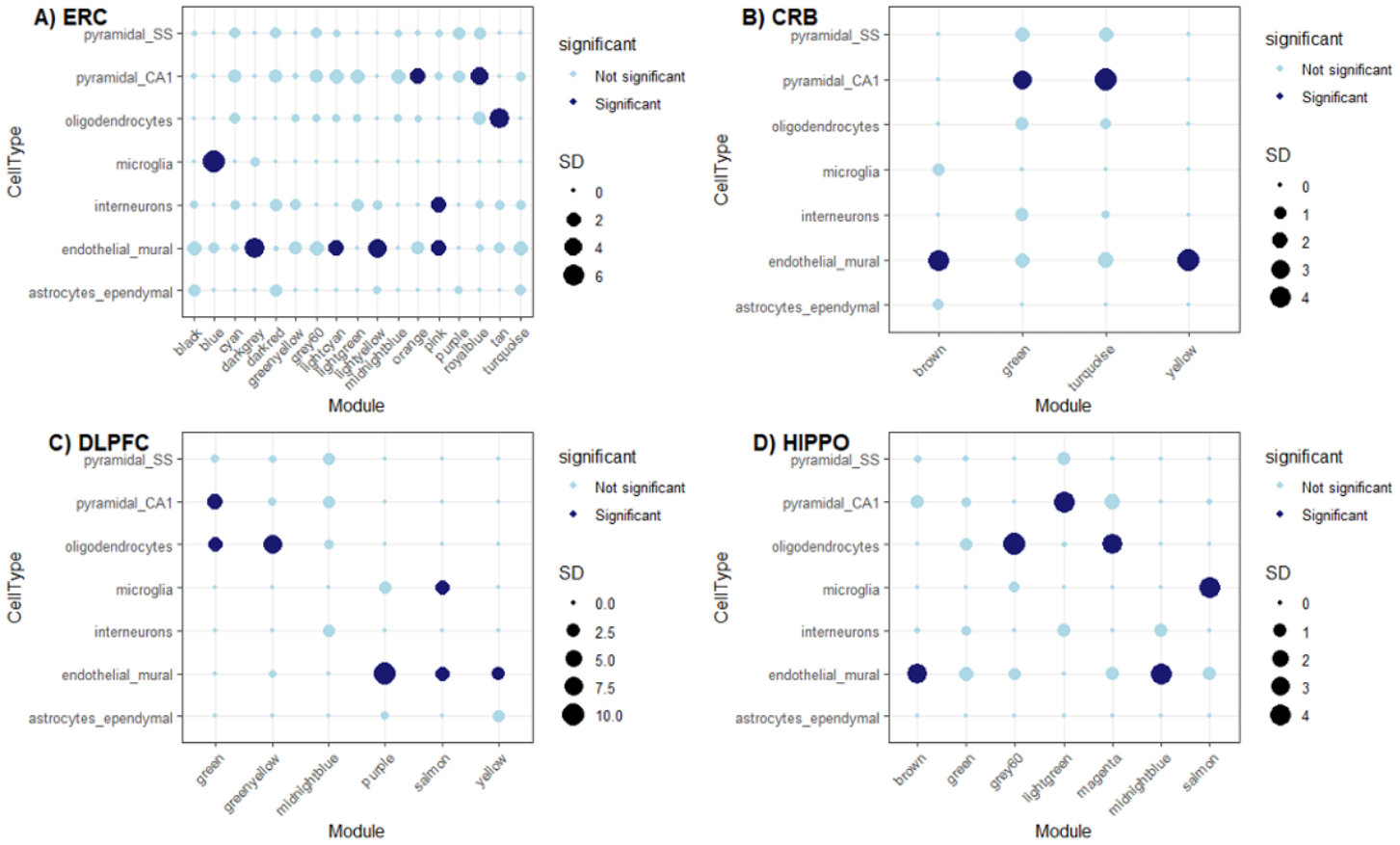






Enrichment analysis using curated lists of OLG and OPC relevant genes which CpGs within the AD associated modules of brain-region network map to. For all plots, the left panel indicates the enrichment of methylation sites mapping to OPC relevant genes the ERC, CRB, DLPFC and HIPPO modules, the right panel indicates enrichment of methylation sites mapping to OLG relevant genes for the same modules. For each module within each network, Odds Ratios and 95% confidence intervals were calculated using Fisher's exact test. Squares indicate significant enrichment (Odds Ratio $p < 0.05$), circles indicate non-significant results (Odds Ratio $p > 0.05$)

Figure 4.9 Cell-type enrichment for all AD-associated co-methylation modules across the four brain-region co-methylation networks using EWCE



Enrichment of modules using EWCE for A) ERC, B) CRB, C) DLPFC and D) HIPPO networks. Dark filled circles highlight the cell types found to be significantly enriched with adjusted $p<0.05$ after Bonferroni correction over all cell types within each module; the size of the circles represents the number of standard deviations (SD) from the mean. Cell-type enrichment analysis on the FTLD-related modules was performed using the package EWCE[239] and associated single-cell transcriptomic data [240]. EWCE,

expression-weighted cell-type enrichment, ERC: entorhinal cortex, DLPFC: dorsolateral prefrontal cortex, HIPPO: hippocampus, CRB: cerebellum, AD: Alzheimer's disease.

DLPFC modules

The greenyellow module was significantly enriched for OLGs, both from the gene list enrichment analysis and from analysis using EWCE (**Figure 4.8, Table 4.6**). The hub gene of this module was *MOG* (Myelin oligodendrocyte glycoprotein), a well described marker of mature OLGs [258] which has been found to present with differential methylation in chronic demyelinated lesions in multiple sclerosis [259]. We also found that this gene was significantly differentially methylated in our meta-analysis of brain nuclei-sorted AD datasets, as well as being nominally differentially expressed in AD snRNA-sequencing data (Chapter 3, Section 3.3.2.iii). The finding of such a well described marker of OLGs as the hub gene of this OLG enriched module adds strength to the hypothesis that this module does indeed represent a group of genes relevant to OLGs that are differentially methylated in disease. The CpG with the second highest module membership (MM) (after *MOG*) in this module mapped to the gene *ATP11A* (ATPase phospholipid transporting 11A). This was also of particular interest due to this gene having been reported to be mutated in hypomyelinating leukodystrophy [260]. We noted that several methylation sites with high MM in this greenyellow module mapped to the gene *ATP11A.* Another methylation site with high MM (top 10) in this module mapped to the gene *PIP4K2A,* which we have previously described as showing dysregulation in neurodegeneration, and present within disease associated co-methylation modules in FTLD. Enrichment of terms relevant to OLGs/neurodegenerative diseases was limited, but included terms related to ageing of cells (**Table 4.6**). It has previously been described by Murthy et al. [136] associating OLGs with increased epigenetic ageing (as discussed in Chapter 1) .

Another module from the DLPFC network that was enriched for OLG lineage cells was the purple module, which was significantly enriched for OPC-relevant genes in our gene list enrichment analysis only (**Figure 4.8, Table 4.6**). The hub gene of this module was *PRKAG2,* which codes for PRKG2 (cGMP-dependent protein kinase 2). The gene has previously been linked to AD, with elevated protein expression being associated with increased levels of amyloid-beta [261]. Functional enrichment terms within this module included those relevant to lipid function, including “sphingolipid translocation”.

The final module enriched for OLG related genes was identified with EWCE, and was the DLPFC green module (**Figure 4.9, Table 4.6**). *FAM69A* (also known as *DPIK1A*) was the hub gene of this module, genetic mutations of which have been associated with MS [262]. The gene is known to be highly expressed within OLGs, and has been also reported to show altered gene expression in MS [263]. Enriched terms within this module included those related to lysosomes (**Table 4.6**).

ERC Modules

The disease associated modules in the ERC network that were significantly enriched for OLG lineage genes were the darkred, tan, darkgrey and blue modules (**Figures 4.8 and 4.9, Table 4.6**). The only ERC module that was significantly enriched for OLG relevant genes within both forms of enrichment analysis was the tan module. Interestingly, the hub gene of this module was *C11ORF9* (also known as *MYRF*), which codes for myelin regulatory factor, an important driver of post-OLG differentiation myelination and myelin maintenance [264]. Again, the finding of such a highly relevant OLG gene as the hub gene of a disease associated module is of high relevance within the context of this work. The methylation site with the second highest MM for the tan module also mapped to the same gene. In a previous study using gene expression and

proteomic networks, *MYRF* was found to be a key driver of AD-associated disrupted OLG signatures [172]. We also identified *MYRF* as being differentially methylated and differentially expressed (upregulated) in our analysis of differentially methylated genes in FTLD (Chapter 3). Another gene to which many methylation sites mapped to was the gene *ATP11A*, which was also present within the DLPFC-greenyellow module, with 4 methylation sites within the top 50 CpGs with highest MM in this tan module (data not shown). We also note the presence of a methylation site mapping to *MBP* (a well described marker of mature OLGs) with high module membership in the module. Functional enrichment terms within this module included those related to acetyl-coA metabolic processes, and those involved in reactive oxygen species (**Table 4.6**).

Another module with very high enrichment of OLG lineage genes in the ERC module was the darkred module, which was enriched for genes relevant to OPCs in the gene list analysis (**Figure 4.8, Table 4.6**). The hub gene of the darkred module was *KIAA0556.* A gene associated with microtubule function [265] that is mutated in Joubert syndrome - a disease associated with abnormal brain development that has been described as a congenital cerebellar ataxia [266]. We also note that within this module several of the methylation sites with highest MM mapped to the gene *TNK2,* which we had identified as an OPC gene frequently showing differential methylation in Chapter 3. Enrichment terms within this module included, again, those relating to lysosomes (**Table 4.6**).

The darkgrey module was also enriched for OLG genes from our gene list analysis (**Figure 4.8, Table 4.6**). The hub gene of this module was *NUAK1.* This gene codes for an AMPK-related kinase which is known to stabilise the tau protein, and it has been found that reduction of this protein in a mouse model resulted in alleviation of tau-related deficits [267]. Enrichment terms within this module included “lysosomal transport” (**Table 4.6**).

The final ERC module enriched for OLG genes was the blue module, identified in the gene list enrichment analysis, of which the hub gene was *WDR81* (**Figure 4.8, Table 4.6**). This gene has previously been associated with cerebellar ataxia [268] and hypermethylation of the promoter region of this gene has been reported following demyelination in the hippocampus of MS patients [269].

HIPPO modules

Within the HIPPO network, the grey60 module was significantly enriched for OLG genes (in both the gene list enrichment and EWCE analyses) (**Figures 4.8 and 4.9, Table 4.6**). The hub gene of this module was *MGC14436,* which maps to chromosome 12 and is not well characterised. Several methylation sites with high MM in this module mapped to this gene. Interestingly, the same gene was present within the ERC tan module and the DLPFC greenyellow module. This disease-associated OLG module also contained several methylation sites mapping to the gene *ATP11A*, as we had seen with the ERC-tan and the DLPFC-greenyellow module. Enrichment genes within this module included “glycosphingolipid metabolic processes” (**Table 4.6**).

From the EWCE analysis, we also saw that the magenta module was enriched for OLG genes (**Figure 4.9, Table 4.6**). The hub gene of this module was *RTTN.* From The Human Protein Atlas, we saw that *RTTN* is highly expressed in OLGs, and is part of cluster ‘Oligodendrocytes - Unknown function’ (RTTN Protein Atlas). Mutations in *RTTN* have been reported to result in dysregulated cortical development and microcephaly [270].

CRB modules

In our enrichment analysis using EWCE of the modules in the CRB co-methylation networks, none were significantly enriched for OLG cell types. In the gene listl enrichment analysis however, we found that the yellow and brown modules were enriched for OPC (yellow) and OLG (yellow and brown) relevant genes (**Figure 4.8, Table 4.6**).

The hub gene of the CRB-yellow module was *C1orf92*, also known as *LRRC71.* There was relatively little in the literature as to the function or relevance of this gene, although it has been found to be upregulated in the white matter of AD samples [271]. Functional enrichment terms from this module “medium-chain fatty-acyl-CoA metabolic process” (**Table 4.6**). The hub gene of the CRB-brown module was *FLJ10213*, for which there is also limited information in the literature.

Table 4.5 Hub genes and functional enrichment of disease associated OLG lineage gene enriched modules across AD-networks

Module		**Highlighted enrichment terms relevant to OLG/OPCS and/or neurodegeneration**	**Type of enrichment analysis**
DLPFC-greenyellow	*MOG*	"cellular senescence" "cell ageing"	Gene list enrichment OLG, EWCE
DLPFC-purple	*PRKAG2*	"sphingolipid translocation" "phosphatidylinositol-3-phosphate biosynthetic process"	Gene list enrichment OPC
DLPFC-green		"lysosomal transport" "cytosolic transport"	EWCE
ERC-darkred	*KIAA0556*	"lysosomal transport" "dermatan sulfate proteoglycan" "biosynthetic process" "golgi to lysosome transport"	Gene list enrichment OPC
ERC- tan		"negative regulation of reactive oxygen species biosynthetic process" "acetyl-CoA biosynthetic process" "acetyl-CoA metabolic process"	Gene list enrichment OLG, EWCE
ERC-darkgrey		"lysosomal transport " "signal transduction involved in cell cycle checkpoint"	Gene list enrichment OLG
ERC- blue		"maintenance of protein location in cell"	Gene list enrichment OLG
HIPPO-grey60		"maintenance of unfolded protein" "glycosphingolipid metabolic process"	Gene list enrichment OLG, EWCE
HIPPO-magenta	*RTTN*	"inositol lipid-mediated signaling"	EWCE
CRB-yellow	*C1orf92*	"medium-chain fatty-acyl-CoA metabolic process" "establishment of protein localization to plasma membrane" "Golgi to plasma membrane protein transport"	Gene list enrichment OLG and OPC
CRB-brown	*FLJ10213*	"regulation of DNA biosynthetic process" "positive regulation of RNA splicing"	Gene list enrichment OLG

AD: Alzheimer's disease, DLPFC: dorsolateral prefrontal cortex, ERC: entorhinal cortex, HIPPO: hippocampus, CRB: cerebellum, OPC: oligodendrocyte precursor cell, OLG: oligodendrocyte, EWCE: expression weighted cell-type enrichment

As with the FTLD network hub genes, we were also interested to see if any of our AD network hub genes showed differential expression. We utilised the AD snRNA-sequencing dataset that we have described in **Chapter 3** to investigate this. We found that only three of the hub genes were differentially expressed in at least one subcluster of OLG/OPC cells in our snRNA-sequencing dataset. The gene *WDR81,* which was the hub gene of the ERC-blue module, was significantly ($p < 0.05$) downregulated one OLG cluster (**Figure 4.10.A**). The gene *MOG,* which was the hub gene of the DLPFC-greenyellow module, was differentially expressed across 4 clusters, showing increased expression in 3 and a decrease in expression in one OLG cluster (**Figure 4.10.A**). Finally, the gene *FAM69A*, hub gene of the DLPFC-green module showed an increased expression in one OLG cluster (**Figure 4.10.A**). None of the genes were significantly differentially expressed in any OPC modules.

To assess whether these identified OLG DNA methylation signatures identified in one brain region were present across other brain regions, we carried out preservation analysis of all network modules against all other brain regions (**Table 4.7**). In general, modules showed preservation across brain regions - only one disease associated OLG enriched module showed no preservation in another dataset; the ERC-darkrey module in the DLPFC. Given that the data is derived from the same post-mortem brain samples but in different brain regions, this is perhaps unsurprising. It was interesting to see that the DLPFC-greenyellow and ERC-tan modules, which we had identified as representing a potentially shared cross-region OLG disease signature (discussed above), were highly preserved across all datasets.

Table 4.6 Module preservation between OLG enriched AD co-methylation modules across brain region DLPFC, ERC, HIPPO and CRB networks

Module	Preservation in HIPPO	Preservation in CRB	Preservation in DLPFC	Preservation in ERC
DLPFC				
purple	High	Moderate		High
greenyellow	High	High		High
ERC				
tan	High	High	High	
darkgrey	Moderate	High	Not preserved	
blue	High	High	High	
darkred	High	High	Moderate	
HIPPO				
grey60		High	Moderate	Moderate
CRB				
yellow	Moderate		Moderate	Moderate
brown	High		High	High

AD: Alzheimer's disease, DLPFC: dorsolateral prefrontal cortex, ERC: entorhinal cortex, HIPPO: hippocampus, CRB: cerebellum

Comparison of modules across brain region networks

Given that we had identified recurring hub genes that methylation sites mapped to across the ERC-tan, DLPFC-greenyellow and HIPPO-grey60 modules, and high preservation in particular of the ERC-tan and DLPFC-greenyellow modules across datasets, we decided to investigate

similarities across these modules further. We found that methylation sites mapping to *MYRF* (hub gene of ERC-tan) and *MGC14436* (hub gene of HIPPO-grey60) were present across all three modules. Seventy four methylation sites mapping to 54 unique genes were present across in all three of these modules. Notably, 8/74 methylation sites mapped to the gene *ATP11A.* Other genes overrepresented amongst this list of common methylation sites were *INPP5A* (5/69 methylation sites) and *FAM107B* (3/74 methylation sites). *INPP5A* (inositol polyphosphate-5-phosphatase A), which has been implicated in the pathology of spinocerebellar ataxias [272] shows relatively low brain-cell type specificity, but of the non-neuronal brain cell types, is most highly expressed in OLGs (INPP5A Protein Atlas). *FAM107B* is highly expressed in OLGs in comparison to all other brain cell types (FAM107B Protein Atlas), and has been shown to present with upregulation at the gene expression level in AD in the temporal cortex [257,273]. Another gene of interest that contained a methylation site across all three of these modules was *BIN1,* which, as discussed, is a risk factor for AD[71], of high relevance to OLGs [74], and previously found to be differentially methylated in disease [106,131].

We then decided to investigate if any of the genes identified across this 3 module signature showed differential expression using the AD snRNA-sequencing dataset previously described (**Chapters 2 and 3**). Seven genes identified from the OLG AD DNA methylation signatures showed differential expression in at least one OLG subcluster; *ADARB2, FAM107B, QDPR, RASGRF2, UNC5C, CAV1* and *SLC5A11* (**Figure 4.10.B**). Five of these seven genes showed differential expression in the same subcluster; Oli3. Although only these seven genes showed significance after adjustment for multiple testing, 27/54 genes from our signature also showed nominally significant differential expression patterns across at least one OLG/OPC subcluster/.

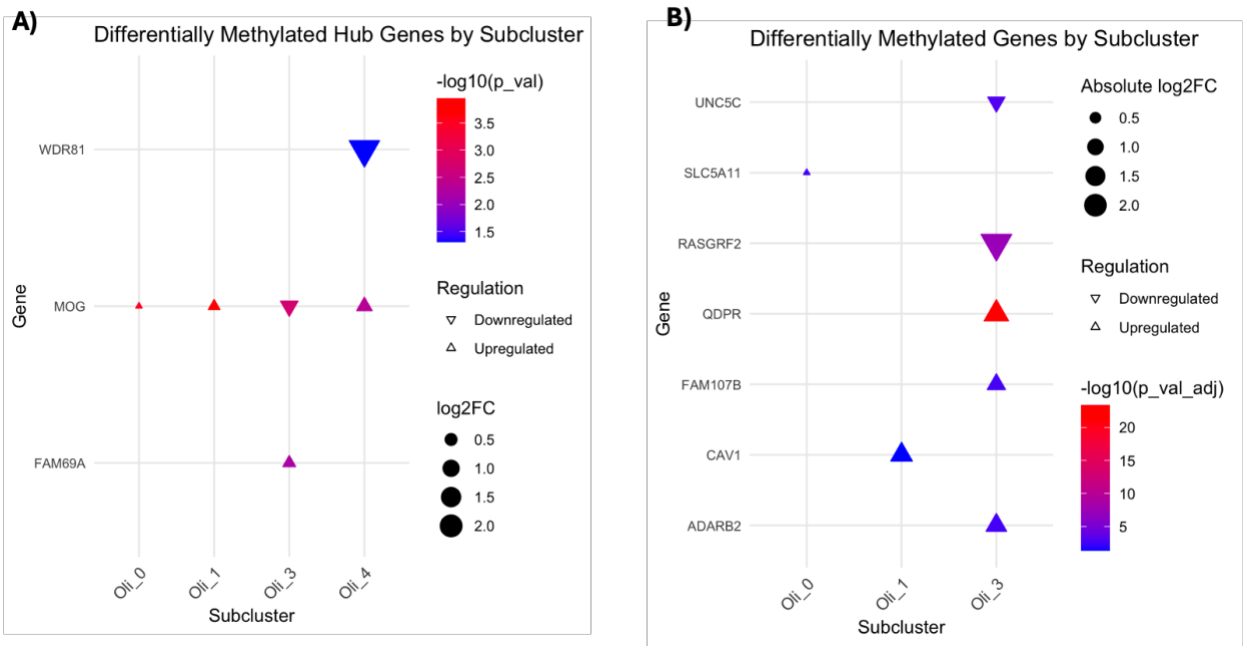
One of the genes showing significantly different patterns of expression in AD was *QDPR,* to which one methylation site had mapped to in our signature. This gene had been found to be

differentially methylated in FTLD and AD, and differentially expressed (upregulated) in our FTLD-expression data. Interestingly, Mathys et al. found that this gene was a marker of AD pathology in OLG subclusters and showed with immunohistochemistry that *QDPR* exhibited higher expression in white matter of AD individuals [158].

Another gene showing significant differential expression patterns was *ADARB2,* which has previously been reported to show differential methylation in the hippocampus of AD individuals compared to controls [274].

SLC5A11, which was present in our signature and upregulated in the snRNA-sequencing data, had also been found to contain differentially methylated sites across our bulk FTLD meta-analysis, sorted brain-nuclei FTLD EWAS, and bulk AD methylation EWAS, where it was also found to show a decrease in expression in the corresponding bulk RNA-sequencing data.

Figure 4.10. Expression of genes of interest in AD brain region co-methylation network modules



A) Expression levels of AD co-methylation OLG/OPC enriched module hub genes that were significantly differentially expressed in at least one snRNAseq subcluster (adjusted p-value < 0.05). B) Expression levels of genes found across our three module signatures that were significantly differentially expressed in at least one snRNAseq subcluster (adjusted p-value < 0.05). Each triangle represents a gene, with the shape indicating the direction of regulation: upregulated (▲) or downregulated (▽). The size of the triangle corresponds to the absolute log2 fold change (log2FC), indicating the magnitude of differential methylation. The colour gradient reflects the significance level of the adjusted p-values (-log10(p_val_adj)), with red indicating highly significant genes.

4.3.1.3 Comparison of disease signatures across Neurodegenerative Diseases in bulk co-methylation networks

In order to assess whether the disease signatures that were enriched for OLG/OPC genes were disease specific, or represented a general dysregulation of neurodegeneration, we carried out cross-disease preservation analysis. First, we compared each of the AD brain region networks against each of the FTLD1, FTLD2 and FTLD3 datasets, focusing firstly on the modules of interest (i.e. disease-associated and OLG lineage relevant) as described above. For the HIPPO network, the grey60 module (OLG enriched) was moderately preserved in the FTLD1 and FTLD2 datasets, but highly preserved in the FTLD3 network (**Table 4.7**). For the DLPFC network, the greenyellow module (OLG enriched) was moderately preserved in the FTLD1 and FTLD2 data and highly preserved (the most highly preserved of all modules) in the FTLD3 data (**Table 4.7**). Following a similar pattern, the tan module (OLG enriched) in the ERC network data, whilst being moderately preserved across the FTLD1 and FTLD2 networks, was again highly preserved and the most preserved module in the FTLD3 data (**Table 4.7**).

An explanation for this higher preservation of AD associated OLG enriched modules with the FTLD3 data compared to the FTLD1/2 datasets could well be linked to the underlying pathology of the FTLD subtypes present across the datasets. As described, the FTLD3 data is composed of FTLD-tau cases solely (sporadic PSP). As AD is also a tauopathy, it could be that the molecular changes that are represented by these AD co-methylation modules highlight potential shared OLG dysfunction that is related in some way to tau pathology. The distinct pathological mechanisms in FTLD-TDP subtypes may result in less overlap with AD-associated molecular pathways, explaining less preservation of OLG-enriched modules in these datasets. However, it is of note that there is some degree of overlap, highlighting potential shared mechanisms of DNA methylation driven dysregulation across OLG cells in neurodegeneration.

Other modules of interest from the AD brain region networks that showed high preservation between datasets was the purple DLPFC module (significantly enriched for OPC cell types), which showed high preservation in all three FTLD datasets (**Table 4.7**). However, for the FTLD1 and FTLD2 data, this DLPFC-purple module was by far the most highly preserved (**Table 4.7**), whilst it was 7th most highly preserved module in the FTLD3 data (**Table 4.7**). The DLPFC-purple module contained enrichment terms associated with “sphingolipid translocation” (**Table 4.7**), which we had also seen as enriched within several FTLD1 and FTLD2 OLG/OPC enriched disease-associated modules (**Chapter 4 Section 4.3.1**). The ERC darkred module (also enriched for OPC cell types), was also highly preserved across all three FTLD datasets (**Table 4.7**). We had highlighted that this module contained many methylation sites mapping to the gene *TNK2* (a gene from our OPC list) which showed high module membership. We had identified this gene as being frequently differentially methylated and expressed in FTLD and AD in our EWAS analysis in Chapter 3. This could point towards dysregulation represented by this module being shared across distinct pathologies and representing general dysfunction of OPCs in neurodegeneration.

Table 4.7 Module Preservation Analysis of FTLD Brain Region Networks against AD Data

Module	Preservation in FTLD1	Preservation in FTLD2	Preservation in FTLD3
DLPFC-greenyellow	High	Moderate	High (highest of all modules, regardless of disease status)
DLPFC-purple	High (highest of all, regardless of disease status)	High (highest of all, regardless of disease status)	High
DLPFC-green	Not preserved	Moderate	High
ERC-darkred	High	High	High
ERC- tan	Moderate	Moderate	High (highest of all modules, regardless of disease status)
ERC-darkgrey	Moderate	Moderate	High
ERC- blue	High	High	High
HIPPO-grey60	Moderate	Moderate	High
HIPPO-magenta	Moderate	Moderate	Moderate
CRB-yellow	Moderate	Moderate	High
CRB-brown	High	High	High

FTLD: frontotemporal lobar degeneration, AD: Alzheimer's disease, DLPFC: dorsolateral prefrontal cortex, ERC: entorhinal cortex, HIPPO: hippocampus, CRB: cerebellum

We also carried out the analysis vice-versa, and analysed the preservation of FTLD1, FTLD2 and FTLD3 network modules against the AD brain region networks.

In the FTLD1 networks, there was one OPC enriched disease-associated module - the skyblue module. The skyblue module was moderately preserved in the HIPPO, DLPFC and CRB, but not preserved in the ERC data (**Table 4.8**). In the FTLD2 networks, the singular OPC enriched module (orangered4), was moderately preserved across the HIPPO, DLPFC and CRB data but not preserved in the ERC data (**Table 4.8**), following a similar pattern to the FTLD1-skyblue module.

There were five modules of interest from the FTLD3 networks; cyan, salmon, turquoise and blue. The blue module was highly preserved across HIPPO, DLPFC, and CRB but not preserved in the ERC data (**Table 4.8**).

The salmon module was highly preserved across all brain regions (which was interesting given we had also seen preservation of this module across all three FTLD datasets), and the turquoise and cyan modules were either moderately or highly preserved across all brain regions (**Table 4.8**).

In general, we see higher preservation of disease-associated OLG/OPC modules across the FTLD3 datasets in comparison to the FTLD1 and FTLD2 networks against the AD brain region datasets (**Table 4.8)**. As discussed above, such increase in preservation between FTLD3 and AD networks could be attributed to similar underlying pathologies - i.e. the presence of tau, which is present to a lesser extent in the FTLD2 dataset (a mixed TDP and tau cohort) and not in the FTLD1 dataset (composed solely of FTLD-TDP cases).

The FTLD3 salmon module, which we saw to be highly preserved across all AD brain regions, and was also highly preserved across all FTLD subtypes (**Table 4.5**). The hub gene of this module was *PCK1,* which as described above has been previously associated with

neurodegenerative diseases such as MS and AD [247]. It is interesting therefore to see this FTLD module well preserved across all brain region networks of this AD dataset, strengthening the hypothesis that this module could represent common dysregulation of processes relating to OLGs across neurodegeneration.

Table 4.8 Module Preservation Analysis of FTLD Brain Region Networks against AD Data

Module	Preservation in DLPFC data	Preservation in ERC data	Preservation in HIPPO data	Preservation in CRB data
FTLD1-skyblue	Moderate	Not preserved	Moderate	Not preserved
FTLD1-white	Moderate	Not preserved	Moderate	Not preserved
FTLD3-blue	Moderate	Not preserved	Not preserved	Not preserved
FTLD3-turquoise	High	Moderate	High	High
FTLD3-cyan	High	Moderate	High	Moderate
FTLD3-salmon	High	High	High	High
FTLD3-purple	High (highest of all modules, regardless of disease status)	High	High (highest of all modules, regardless of disease status)	High

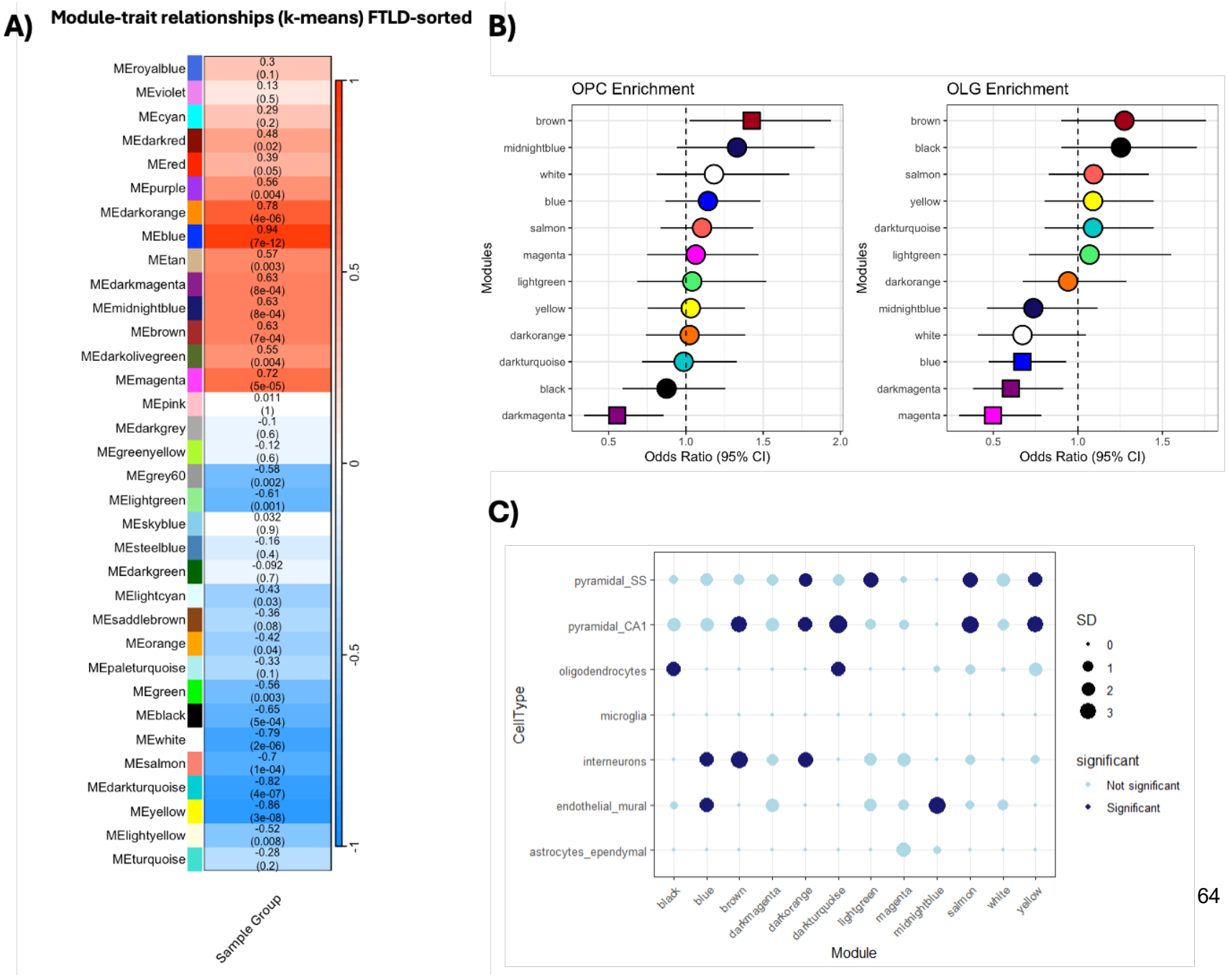
FTLD: frontotemporal lobar degeneration, AD: Alzheimer's disease, DLPFC: dorsolateral prefrontal cortex, ERC: entorhinal cortex, HIPPO: hippocampus, CRB: cerebellum

4.3.2 Brain-nuclei Sorted Networks

4.3.2.1 FTLD-sorted Network

We next carried out co-methylation network analysis on the glial fraction of a brain nuclei-sorted FTLD dataset. We found 12/34 ($p < 0.002$, 0.05/34 modules) co-methylation modules that were associated with the disease status (i.e. FTLD or control) (**Figure 4.11.A**). Cell-type enrichment analysis using gene list enrichment revealed one module that was significantly enriched for OPC-relevant genes (the brown module), but none for OLG relevant genes ((**Figure 4.11.B**). Using EWCE, we identified the black and darkturquoise modules as being enriched for OLGs (**Figure 4.11.C**)

Figure 4.11 Module-trait correlations and celltype enrichment within brain-nuclei sorted FTLD network



A) Module-trait correlations for the sorted FTLD co-methylation networks. The rows represent the co-methylation module eigengenes (ME) and their colours, and the column represents the correlation of the methylation levels of CpGs in each module with the disease status. p-values are presented within each cell and the colour scale at the right indicates the strength of the correlation (darker cells depict stronger correlations, with blue representing negative and red representing positive correlations). **B) Gene list enrichment analysis using curated lists of OPC (left panel) and OLG (right panel) relevant genes with CpGs within the sorted-FTLD associated modules.**. For each module within each network, Odds Ratios and 95% confidence intervals were calculated using Fisher's exact test. Square shape indicates significant enrichment (Odds Ratio ($p < 0.05$)). **C) Cell-type enrichment for all FTLD-associated co-methylation modules.** Dark filled circles highlight the cell types found to be significantly enriched with adjusted $p<0.05$ after Bonferroni correction over all cell types within each module; the size of the circles represents the number of standard deviations (SD) from the mean. Cell-type enrichment analysis on the FTLD-related modules was performed using the package EWCE[239] and associated single-cell transcriptomic data [240].

The hub gene of the brown module (enriched for OPCs in the gene list enrichment analysis (**Figure 4.11.B**)) was *PARM1*, which codes for Prostate androgen-regulated mucin-like protein 1, has been reported to be dysregulated in AD and ALS [184,275]. It has been suggested that the gene plays a role in the regulation of telomerase [275]. One of the terms enriched in genes from this module was "cell ageing", which is interesting given the role of telomerase in cell senescence. Other terms enriched in the brown module were "cell cycle phase transition", "regulation of mitotic cell cycle" and "nervous system development" (**Table 4.9**).

The hub gene of the black module was *LOC100288798,* which codes for a long non-coding RNA, which is uncharacterised. It was also notable that this module contained several terms relating to peroxisomes, which we had also seen in the FTLD3-blue module in the bulk FTLD networks (**Table 4.9**).

The hub gene of the darkturquoise module was *RABEP2, Rab-GTPase binding effector protein 2.* Rab proteins are involved in membrane trafficking There is little in the literature relating this

gene to OLGs or neurodegeneration, however it was identified as a substrate for GSK3 [276], Glycogen synthase kinase 3, which has been linked to AD and PD [277]. Functional enrichment terms of this module included those relating to fatty acid metabolism, notable due to the role of fatty acids in myelin function and structure [278] (**Table 4.9**).

Table 4.9 Hub genes and functional enrichment of disease associated OLG lineage gene enriched modules across brain-nuclei sorted FTLD network

Module	**Hub Gene**	**Highlighted enrichment terms relevant to OLG/OPCS and/or neurodegeneration**	**Type of enrichment analysis**
FTLD-sorted brown	*PARM1*	"glycosphingolipid metabolic process"	Gene list enrichment OPC
FTLD-sorted black	*LOC100288798*	"fatty acid derivative biosynthetic process" "cholesterol transport" "intracellular cholesterol transport" "peroxisomal membrane transport " "protein targeting to peroxisome" "peroxisomal transport" "positive regulation of amyloid-beta formation" "positive regulation of amyloid precursor protein catabolic process" "positive regulation of stress-activated protein kinase signaling cascade"	EWCE
FTLD-sorted dark turquoise	*RABEP2*	"fatty acid elongation" "fatty acid biosynthetic process" "positive regulation of neuron apoptotic process" "very long-chain fatty acid metabolic process"	EWCE

FTLD: frontotemporal lobar degeneration, OPC: oligodendrocyte precursor cell, OLG: oligodendrocyte, EWCE: expression weighted cell-type enrichment

Finally, we checked the expression of these hub genes in our FTLD1-expression and FTLD2-expression datasets (**Table 4.10**). Only *PARM1,* the hub gene of the brown module, showed significant differential expression between FTLD and controls. In the FTLD1-expression dataset, the gene was genome-wide significantly downregulated in FTLD compared to controls, whilst in the FTLD2-expression data, the genome was nominally significantly downregulated.

Table 4.10 Expression of hub genes in FTLD RNA-sequencing datasets

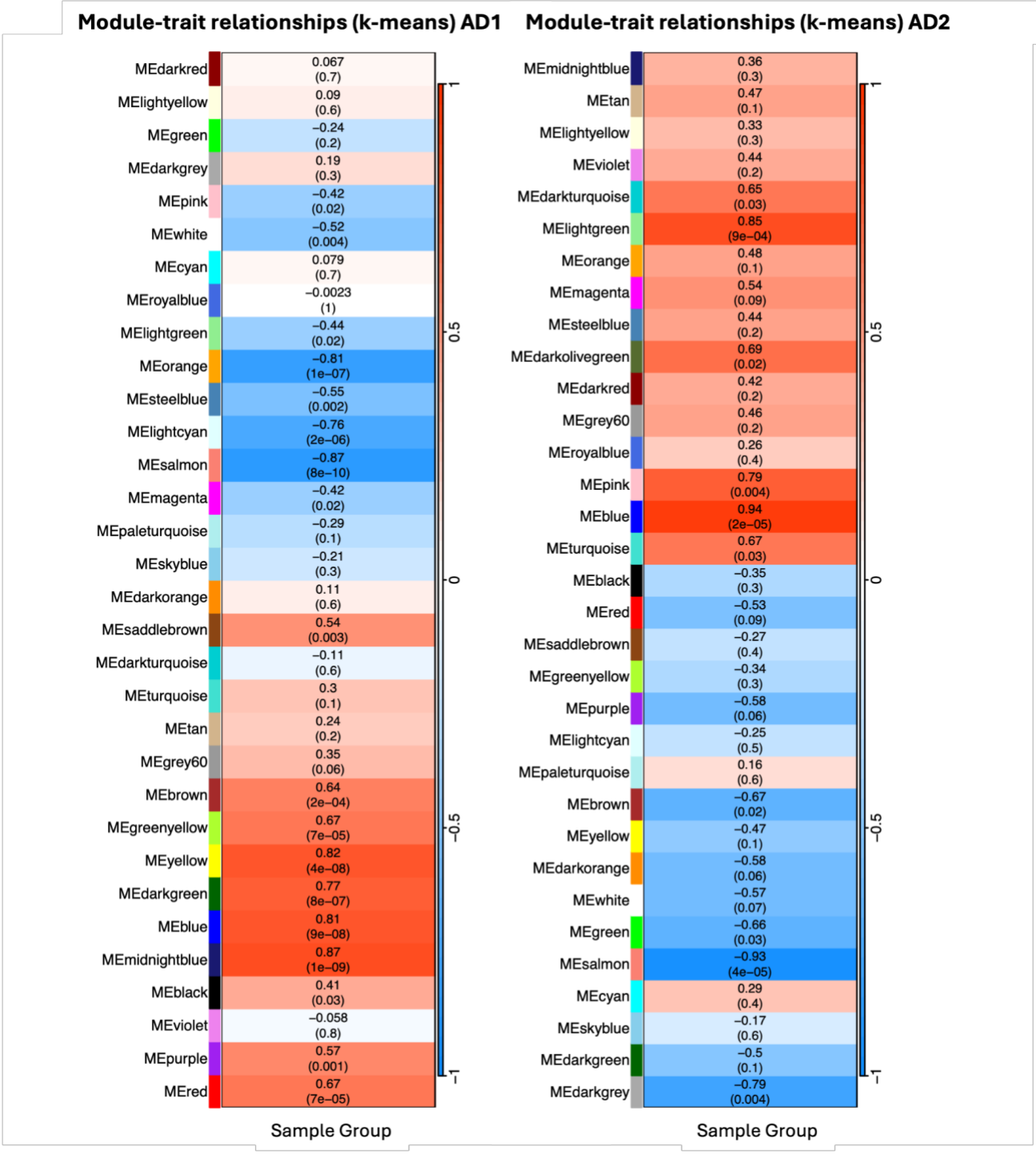
	FTLD1-Expression		**FTLD2-Expression**	
Gene	logFC	Adjusted P-value	logFC	Adjusted P-value
PARM1	- 0.429	0.041	-0.825	0.093

FTLD: Frontotemporal lobar degeneration, logFC: log Fold change, FDR Adjusted P-value: p-value adjusted for multiple testing correction

4.3.2.2 AD-sorted Networks

We also carried out co-methylation network analysis on the glial fraction of each of the two brain nuclei-sorted AD datasets separately. For the AD1 and AD2 networks respectively, 11 and 3 modules significantly associated with disease status ($p < 0.0016$, 0.05/32 modules, and $p < 0.0015$, 0.05/33 modules for AD1 and AD2, respectively) (**Figure 4.12**). The package EWCE identified 2 oligodendrocyte enriched modules in the AD1 dataset that were significantly associated with disease status; the blue and brown modules (**Figure 4.13.A, Table 4.11)**. In the AD2 data, EWCE identified the blue module as being enriched for OLGs (**Figure 4.13.B, Table 4.11**). The second method, using gene list enrichment, showed results consistent with the first; the AD1 brown module and the AD2 blue module were found to be significantly enriched for OLG relevant genes (**Figure 4.13.C,D**, **Table 4.11**). Using this method no other modules significantly associated with disease status were revealed to be enriched for OLG or OPC cell types.

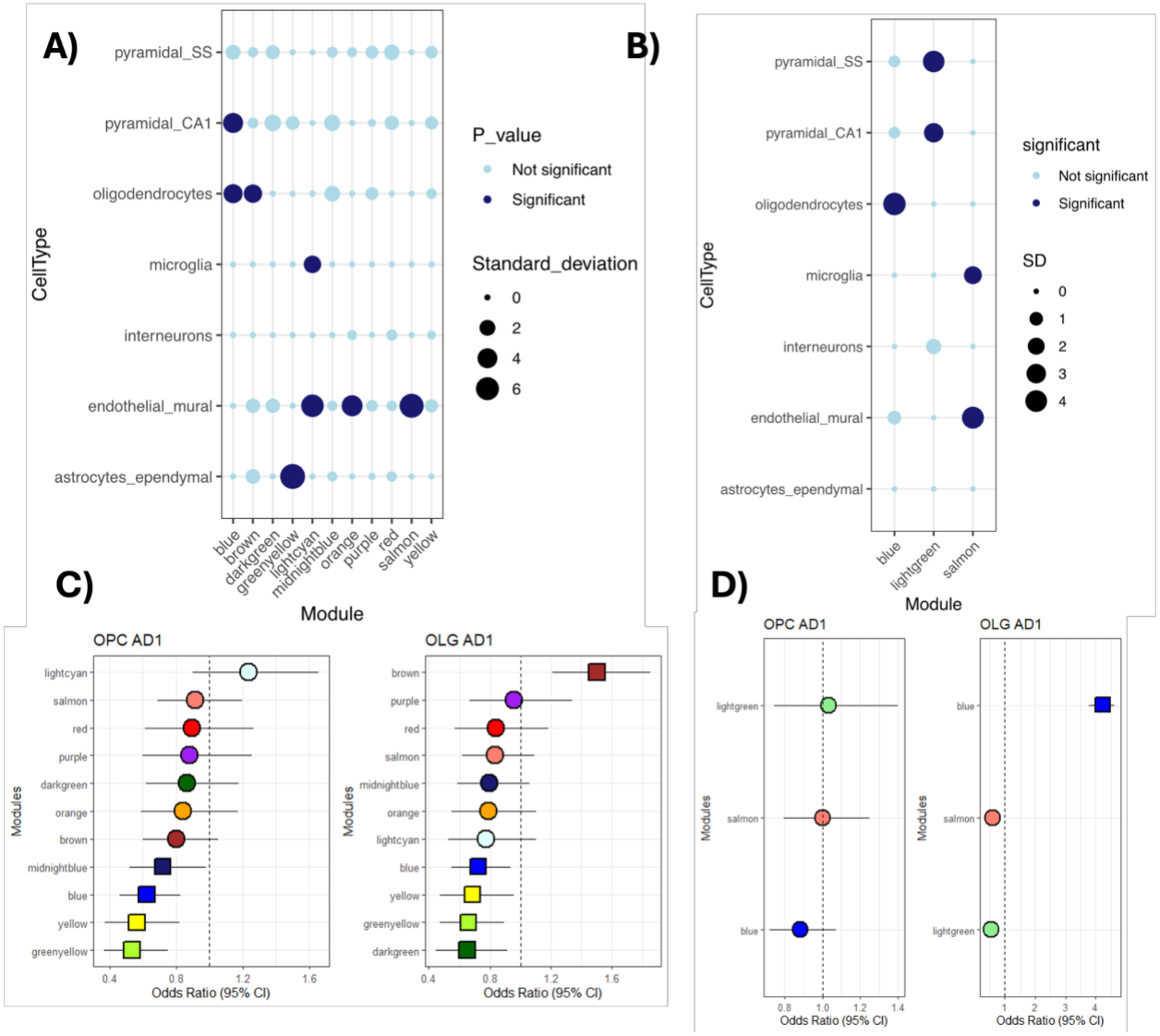
Figure 4.12 Module-trait correlations AD-sorted Networks



Module-trait correlations for the AD co-methylation networks for AD1 and AD2. The rows represent the co-methylation module eigengenes (ME) and their colours, and the column represents the correlation of the methylation levels of CpGs in each module with disease status. P values are presented within each cell and the

colour scale at the right indicates the strength of the correlation (darker cells depict stronger correlations, with blue representing negative and red representing positive correlations).

Figure 4.13. Cell-type enrichment across AD-sorted co-methylation network modules



Cell type enrichment in those co-methylation modules showing significant association with disease status. Results using EWCE for A) AD1 and B) AD2. Dark filled circles highlight the cell types found to be significantly enriched with adjusted p<0.05 after Bonferroni correction over all cell types within each disease associated module; the size of the circles represents the number of standard deviations (SD) from the mean. Cell-type enrichment analysis on the FTLD-related modules was performed using the package EWCE[239] and associated single-cell transcriptomic data [240]..

Enrichment analysis using curated lists of OLG and OPC relevant genes within AD1 (C) and AD2 (B) co-methylation disease associated modules. Enrichment of methylation sites mapping to OPC relevant genes within FTLD2 associated modules For each module within each network, Odds Ratios and 95% confidence intervals were calculated using Fisher's exact test. Square shape indicates significant enrichment (Odds Ratio ($p < 0.05$)).

Table 4.11 Hub genes and functional enrichment of disease associated OLG lineage gene enriched modules across brain-nuclei sorted AD networks

Module	**Hub Gene**	**Highlighted enrichment terms relevant to OLG/OPCS and/or neurodegeneration**	**Type of enrichment analysis**
AD1-brown	*CTNNA1*	"establishment of Golgi localisation" "regulation of cell shape" "Positive regulation of oxidative stress-induced cell death" "cell death in response to oxidative stress"	EWCE and gene list enrichment OLG
AD1-blue	*ZNF143*	"regulation of transcription from RNA polymerase II promoter in response to stress" "positive regulation of stress-activated protein kinase signaling cascade" "lysosomal transport"	EWCE
AD2-blue	*FAAH*	"regulation of neuron differentiation"	EWCE and gene list enrichment OLG

AD1 Blue module

The hub gene of the AD1 blue module was *ZNF143,* coding for a transcription factor for which abnormal expression is associated with cell proliferation and differentiation in cancer [279]. Enrichment terms within this module included "regulation of transcription from RNA polymerase II promoter in response to stress", indicating that other genes within this module could be linked to transcriptional activity as well (**Table 4.11**).

AD1 Brown module

The AD1 Brown module was found to be enriched for OLG lineage cells with both methods. The hub gene of this module *CTNNA1* ((**Table 4.11**), which codes for catenin (cadherin-associated protein) α-1, which has a role in cell adhesion and has been identified as a gene amongst those elevated in OLGs derived from post-mortem AD patients [280]. Notably, *CTNNA1* was also identified as being within a differentially hydroxymethylated locus in AD patients [281]. The occurrence of this gene as a hub gene in this OLG associated AD module is therefore interesting and warrants further investigation. Within this module we also noted that several methylation sites with high module membership (MM) mapped to the gene *ATP11A* (data not shown), which we had described as being present across a three module signature in the bulk AD brain region networks.

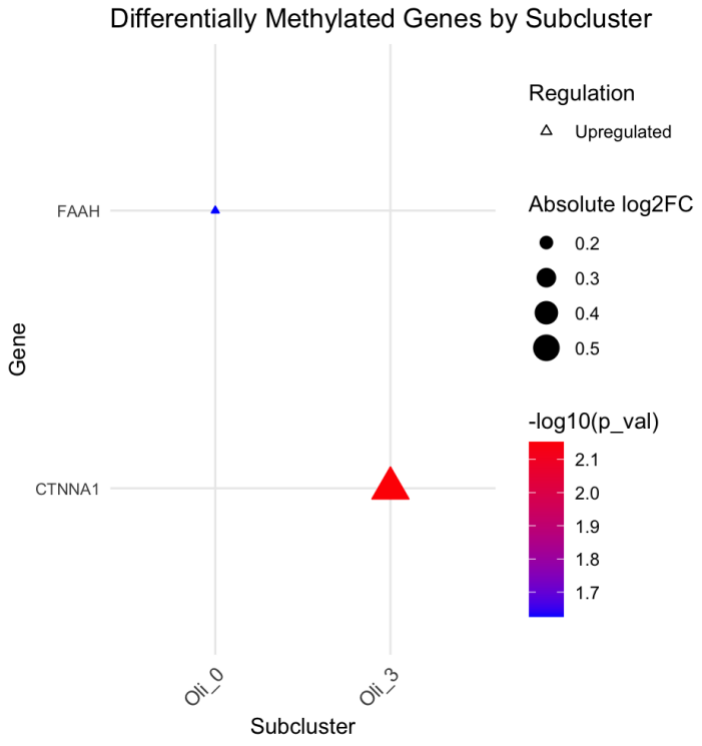
AD2 Blue module

The AD2 Blue module was the only module to be enriched with OLG cell types and significantly associated with AD. The hub gene of AD2 blue was the gene *FAAH*, which codes for fatty-acid amine hydrolase 1, which is involved in lipid metabolism and in the brain is most highly expressed in oligodendrocytes (FAAH expression). *FAAH* has been shown to be overexpressed at the protein level in glial cells that are associated with increased inflammatory processes in AD [282]. In this module, as with the AD1-brown module, we also found several methylation sites mapping to the gene *ATP11A.*

Again, we also used the above described AD single-nuclei dataset to investigate expression patterns of these genes across OLG/OPC subclusters and the single-nuclei level (**Figure 4.14**). *CTNNA1* and *FAAH* both showed significant differential expression in one subcluster. *CTNNA1*

which has previously been reported as being upregulated in AD brains vs controls [280], was also upregulated in an OLG cluster in this dataset in AD compared to controls.

Figure 4.14 Differential expression of brain-nuclei sorted AD co-methylation network hub genes in snRNA-seq data



Plot shows genes that are significantly differentially methylated in various Oli subclusters. The y-axis represents gene names, and the x-axis represents subclusters (e.g., Oli_0, Oli_3). The size of the points corresponds to the significance of the differential methylation (-log10 of the p-value). The colour gradient of the points represents the magnitude of the absolute log2 fold change (absolute log2FC), ranging from blue (lower absolute log2FC) to red (higher absolute log2FC). Triangular shapes indicate the direction of regulation, with "Upregulated" genes represented in this case.

As with previous datasets, we carried out module preservation analysis to analyse whether our identified OLG signatures were replicated across datasets. We found that the AD1 Brown module (hub gene *CTNNA1*) was the most highly preserved in the AD2 data, and that the AD2 blue module (hub gene *FAAH*) was the second most highly preserved AD2 module against the AD1 dataset (**Table 4.12**). This finding suggests that this strong OLG signature is not a dataset artefact but in fact represents a replicable signature related to dysregulation of OLG genes in AD.

We investigated overlap of the methylation sites and genes present in these two modules of interest further. There were 447 genes which had methylation sites present in both modules, representing 25% and 21% of the unique genes in the AD1-brown and AD2-blue modules respectively.

Table 4.12: Module Preservation between brain-nuclei sorted AD datasets AD1 and AD2

Module	**Preservation in AD1**	**Preservation in AD2**
AD1-brown		**High (highest of all modules, regardless of disease association)**
AD1-blue		**Moderate**
AD2-blue	**High**	

4.4 Discussion

In this chapter, we have employed network analysis across a number of DNA methylation datasets and across multiple neurodegenerative diseases, in order to investigate the effect of DNA methylation on OLG relevant genes in disease. We have identified several modules across FTLD and AD that are associated with disease and enriched for OLG/OPC genes. To investigate these signatures, we have carried out hub gene and functional enrichment analysis, and also investigated gene expression profiles of the genes of interest. Below, we highlight the most notable findings and discuss their significance in the context of the broader scientific literature.

To gain insight into the role of OLGs and myelin-related genes across different subgroups/subtypes of FTLD, we investigated three different FTLD DNA methylation datasets comprising different FTLD pathological subgroups and subtypes. In the FTLD1 dataset, composed of FTLD-TDP type A (*C9orf72* mutation carriers) and FTLD-TDP type C (sporadic) cases, we saw one disease-associated OPC enriched module, the FTLD1-skyblue module. The hub gene of this module was *TMEM168*, which is part of the brain expression cluster related to - 'White matter - Myelination' (*TMEM168* expression), and is known to be a gene involved in cell cycle progression via regulation of the Wnt/β-catenin pathway in glioblastoma cells [241]. Enriched terms within the FTLD1 skyblue module included "sphingolipid translocation". Several other modules across all networks were also enriched for terms relating to sphingolipids; enrichment terms in the FTLDsorted-brown module (enriched for OPCs), the DLPFC-greenyellow module and the HIPPO-grey60 module included "glycosphingolipid metabolic processes", the term "sphingolipid translocation" was also found in enrichment analysis for the DLPFC-purple module (enriched for OPC genes). Sphingolipids are an important component of plasma membranes,

particularly in the nervous system, and include the group of lipids sphingomyelins. Changes in sphingolipid composition/structure are known to be important in diseases in which altered myelination occurs [283]. In an investigation into myelin lipid dysregulation in FTLD, Marian et al. found that sphingolipid metabolism was significantly altered across FTLD-*GRN* and FTLD-*C9orf72* cases compared to controls [284].

Given the well characterised and distinct pathology relating to OLGs in PSP (FTLD-Tau/sporadic), it was unsurprising that we found the greatest number of modules enriched for OLG genes within the FTLD3 network of all three bulk FTLD co-methylation networks. The FTLD3 cyan module had the hub gene of *DDITL4/REDD2*, a known regulator of the mTOR pathway [285], and we saw that this gene showed an increase in expression in two FTLD gene expression datasets. The mTOR pathway is known to be vital in allowing the correct differentiation of OLGs [286]. Supporting the evidence that this module could represent a dysregulated signal of OLG differentiation; enrichment terms within this module included "negative regulation of cell differentiation" and "negative regulation of cell proliferation". Furthermore, *DDITL4* also has a role in DNA damage response [285,286], which is essential for cells under stress conditions such as oxidative stress, to which OLGs are particularly vulnerable (as discussed in Chapter 1). The FTLD3 cyan module was seen to be moderately or highly preserved across all FTLD and AD brain region datasets.

The FTLD3 blue module shows enrichment for genes related to OLGs with the hub gene *GDAP1*, which is a mitochondrial fission factor that is mutated in a demyelinating subtype of Charcot Marie Tooth disease [287]. We saw that this gene showed downregulation at the gene expression level in both FTLD RNA-sequencing datasets. *Gdap1* knockout mice exhibit changes in mitochondrial morphology and oxidative stress, as well as hypomyelination [288]. A study investigating GDAP1 protein activity found that, as well as altering mitochondrial fission,

loss of GDAP1 also resulted in the altered morphology of the organelles peroxisomes [253]. Peroxisomes are organelles that are essential for myelination, and play important roles in myelination, and also in the response to oxidative stress [289]. This finding was notable to us given that, when investigating top OLG/OPC genes differentially methylated across datasets (Chapter 3), we identified the CpG mapping to the gene PIP4K2A as the topmost differentially methylated OLG/OPC relevant gene. This gene, which also showed increased expression in the FTLD1 RNA-sequencing dataset, codes for an enzyme crucial in the functioning of peroxisomes. Specifically, *PIP4K2A* codes for phosphatidylinositol 5-phosphate 4-kinase type-2 α which is involved in the regulation of intracellular cholesterol transport. In the central nervous system, ~70-80% of cholesterol is found within myelin membranes [290,291], therefore rendering findings related to cholesterol trafficking of particular relevance to oligodendrocytes. Indeed, the demyelinating disease MS has been linked to disruption of CNS cholesterol metabolism [292], and cholesterol transport has also recently been associated with altered myelination and AD through a study investigating effects of *APOE4* (coded for by APOE, the strongest genetic risk factor for AD) on the human brain [70]. In this study, it was found that cholesterol was aberrantly deposited in oligodendrocytes, and that this altered deposition was associated with reduction in myelin. Given that we noted several genes within this module that have known relevance to peroxisome biology, and also the fact that the CpG mapping to *PIP4K2A* was the most differentially methylated from the OLG PSP EWAS, we checked to see whether the blue module was enriched for terms relating to peroxisome. Indeed we found that, when using a list of peroxisome relevant genes (from publicly available databases), the FTLD3 blue module was significantly enriched for peroxisome function. Peroxisomes are also involved in the production of reactive oxygen species (ROS) [293], which, as we have previously discussed (Chapter 1), could be potentially highly damaging to OPCs/OLGs due to their high metabolic oxygen demand [51]. The blue module was moderately preserved in the FTLD1 dataset, and highly preserved in the FTLD2 dataset, indicating that although such pathogenic mechanisms discussed may have

greater relevance to tau biology. The FTLD3-blue module was also highly preserved across the HIPPO, DLPFC and CRB AD brain region datasets. In terms of AD, PIP4K2A was found to be a pTau interacting protein [294], further strengthening the possibility that the mechanism through which peroxisomes dysfunction is related to tau biology. It is also important to note that whilst the FTLD1 dataset is derived from post-mortem grey matter, the FTLD2 dataset contains both grey and white matter, possibly contributing to the greater preservation of such myelin signatures in the latter dataset. Whether peroxisome dysfunction could be resulting in changes to ROS (or other) pathways which in turn leads to alterations in DNA methylation in OLG lineage cells, or whether DNA methylation changes could be causative in potential malfunction of these organelles is warrants further investigation. Interestingly, we also saw that the FTLD-sorted black module was functionally enriched for terms relating to peroxisomes, again adding evidence for the involvement and importance of these organelles in the context of OLGs in FTLD.

Next, we analysed AD associated DNA methylation signatures relating to OLGs/OPCs across four brain regions; HIPPO, DLPFC, ERC and CRB. We generated networks for each of the four brain regions separately in order to assess how such signatures may differ between regions with distinct susceptibility to AD pathology. Whilst the ERC and HIPPO are known to exhibit changes early on in AD progression, DLPFC is thought to become affected later on, and the CRB is not thought to be affected by "conventional" AD pathology to a great degree [295].

When investigating the network modules for these brain region datasets, we found several modules of interest. The DLPFC-greenyellow and the ERC-tan module were both highly enriched for OLG lineage markers, and strongly positively associated with AD disease status, as well as being preserved across brain regions. Both of these modules had hub genes which are known important regulators of myelination; *MOG* for DLPFC-greenyellow and *MYRF* for ERC-

tan. *MYRF* is a transcription factor that is essential for myelin maintenance [296] and a target of *SOX10* - another crucial activator of OLG differentiation-related genes [297]. *MOG* is a gene expressed by mature OLGs. It is interesting that these two cross-region modules exhibit hub genes crucial to myelination and mature OLG function. When investigating these two modules, as well as the HIPPO-grey60 module (which also showed strong enrichment for OLG genes and was strongly positively associated with AD disease status), we found that several genes occurred recurrently through these modules. We noticed that across all three of these modules, several methylation sites were present with high module membership (MM) that mapped to the gene *ATP11A*, a gene implicated in a hypomyelinating leukodystrophy [260] Mechanisms behind such effects of mutations in this gene were suggested to involve disruption of phospholipids in the cell membrane which led to, amongst other effects, disrupted cell cholesterol homeostasis. Given the known importance of cholesterol functioning in OLGs mediated through the *APOE-ε*4 mutation in AD (Chapter 1)[70], this was of interest. When investigating this gene, we noticed that this gene is part of a cluster of genes involved in white matter signal transduction, and that its nearest neighbour based on RNA tissue expression correlation is *MYRF* - the hub gene of the ERC-tan module (*ATP11A* expression), which is notable to the recurrent finding of these genes together in disease-associated co-methylation modules. In transcriptomic analysis of AD, *ATP11A* has been found to be upregulated in AD compared to controls [298]. When investigating other genes that were present across all three modules in this signature, we also found genes that had known relevance to AD pathology and white matter perturbations. The gene *QDPR,* which codes for quinoid dihydropteridine (an enzyme important in the regeneration of tetrahydrobiopterin (BH4), which is important in nitric oxide production) was found to present across all three modules, which was interesting as we have previously identified this gene as being differentially methylated in FTLD and AD, and showing increased expression in AD compared to controls (Chapter 3A. 3.ii), therefore the finding of this gene within this signature is notable.

We also carried out network analysis on the glial fraction of three sorted brain-nuclei datasets, 2 AD and one FTLD. As described in Chapter 3 Section 3.1, there are many advantages in the use of brain-nuclei sorted data over bulk DNA methylation datasets. Through our analysis of the glial fraction of these DNA methylation datasets, we expect that there will be less dilution of disease signals through the presence of neuronal cell types. As mentioned above, we identified a module in this network, the black module, that was enriched for terms relating to peroxisomes, which was a key finding also in the bulk tissue analysis, confirming the relevance of bulk tissue findings.

In our analysis of the two brain-nuclei sorted AD datasets, we found a strong signature enriched for OLG genes that was highly preserved between the two datasets. The hub genes of the two modules, AD1-brown and AD2-blue, were *CTNNA1* and *FAAH* respectively. *CTNNA1* is a cell-adhesion protein and part of the brain expression cluster 'White matter - Signal transduction' (CTNNA1 expression). Interestingly, its second nearest neighbour in terms of brain RNA expression correlation is the gene *PIP4K2A* (0.9228 correlation based on brain RNA expression), found within our FTLD3-blue module and discussed above. *CTNNA1* has been found to be significantly upregulated in OLGs in AD [280] and also showed increased expression in the AD snRNA-sequencing data we have utilised throughout this Chapter and Chapter 3. The gene has also been identified as being differentially hydroxymethylated in AD [281]. It is also worth highlighting here that the gene *CTNNA3* is frequently presented with aberrant DNA methylation across AD and FTLD EWAS analysis in Chapter 3. *CTNNA1* and *CTNNA3* are two of three α-catenin genes expressed in humans. It was notable therefore that we find that the hub gene of an OLG enriched AD associated module is also a gene important in α-catenin biology.

FAAH is part of the brain expression cluster - 'White matter - Myelination', and interestingly has

the highest RNA expression correlation with *APLP1* (Amyloid beta precursor like protein 1) (0.928 correlation based on brain RNA expression) (*FAAH* Expression). This correlation could represent a functional link between pathways involving/regulated by *FAAH* and amyloid precursor pathways, which are well known to be implicated in AD pathology [299]. In the AD1-brown and AD2-blue modules, we also found methylation sites mapping to the gene *ATP11A,* which we previously discussed, strengthening the likelihood that this gene is an important component of dysregulation of DNA methylation affecting OLGs in AD.

There are, of course, several limitations to methods we have used in this chapter, specifically in the use of cell-type and functional enrichment analysis. Whilst both can provide important information, there are drawbacks specifically related to limitations of annotations available. Cell type markers are continuously evolving, and whilst we have attempted to limit the overlooking of OLG/OPC specific modules through our use of two different methods of cell-type enrichment analysis, available cell type markers might not cover all possible cell types, especially rare or newly characterised types, limiting the depth of cell-type enrichment analyses. There is also the possibility that markers are expressed across multiple cell types or subtypes, leading to ambiguity when assigning genes or pathways to specific cell populations. Some modules are enriched for multiple cell types, which could either represent a signature that is not cell-specific, or ambiguity in annotations. A further problem that is specific to DNA methylation analysis is that in contrast to gene expression, multiple DNA methylation sites (CpGs) may map to one singular gene. This means that one gene can be represented in more than one co-methylation module. Added complexity is given by the fact that the effect of DNA methylation is highly dependent upon genomic context. Whilst a module could contain both genes relevant to, for example, astrocytes and OLGs, it could be that the signature leads to a repression of astrocyte relevant genes and an increase in expression of OLG genes that has become dysregulated.

In conclusion, throughout the work in this chapter, using co-methylation analysis complemented with additional datasets, we have identified signatures representing disruption of DNA methylation affecting OLG/OPCs across neurodegenerative diseases. Across our signatures, we saw functional enrichment terms of high relevance to OLG/OPCs and neurodegeneration, highlighting the utility of the use of network analysis to identify disease relevant effects. This method has allowed us to identify genes that recurrently appear in signatures across neurodegenerative diseases and datasets, e.g. *ATP11A, QDPR,* and *MYRF*. Overall, it was interesting to see that many of our disease-associated modules that had hub genes with known relevance to AD and/or FTLD were preserved across datasets, and contained genes we had identified in our EWAS analyses. This could indicate that dysregulation of OLG genes across these dementias is more similar than previously thought. We have also been able to use functional enrichment analysis to multiple pathways as being effected, for example peroxisome pathways, altered lipid metabolism/homeostasis and responses to oxidative stress. Such pathways have been previously linked to AD and/or OLG dysfunction, but we have provided evidence as to the contribution of DNA methylation. The work in this chapter demonstrates the utility of the use of network analysis to identify disease and cell type relevant genes that show dysregulated DNA methylation in disease, and highlights several key genes and pathways and generates hypotheses that warrant further investigation.

Chapter 5 - Investigating the role of DNA Methylation in gene regulation in oligodendrocytes: insights from human cell models and tissue towards a better understanding of neurodegenerative disease-associated changes

5.1 Introduction

DNA methylation is crucial in the control of gene expression both in health and disease, acting in a spatiotemporal manner to allow intricate control of when and where genes are expressed. It is known that DNA methylation is crucial in cell fate determination. By selectively repressing or activating genes, DNA methylation patterns establish and maintain cell-type specificity, ensuring that specific genes required for a particular cell's function are expressed, while others are silenced [89,90]. As we have described in Chapter 1, although there have been several studies investigating the role of DNA methylation in OLG lineage progression from immature OPCs to mature myelinating OLGs in mice [124], as far as we are aware studies investigating this process in human cells are limited. As there is evidence for aberrant proportions of OPCs and OLGs in several neurodegenerative diseases (Chapter 1), understanding the role of DNA methylation changes during OPC differentiation might be key in understanding how aberrant DNA methylation may be contributing to disease processes affecting OLGs in neurodegeneration.

In previous chapters, we have identified OLG lineage relevant genes showing differential methylation in neurodegenerative diseases and, where possible, investigating downstream

effects using corresponding gene expression data. However, due to limited numbers of datasets available including overlapping samples, and confounding factors such as tissue variability, elucidating the causal effect of DNA methylation is difficult. The role of DNA methylation on gene expression, as previously described (Chapter 1), is best characterised to occur in the promoter region of genes, where increased promoter DNA methylation is more often associated with a decrease in gene expression[89]. The mechanism behind this is not well understood. Methyl-CpG-binding domain (MBD) proteins are a family of proteins known to 'read' the epigenome, and are thought to lead to gene repression through recruitment of transcription factors and chromatin remodelers [300]. However, we have, in addition to finding DNA methylation changes within promoter regions, identified many differentially methylated sites within the gene body and other non-promoter regions. We aimed here to investigate effects on gene expression of DNA methylation across genes which we had found to be differentially methylated, in order to add to understanding of the role of DNA methylation on OLG/OPC genes. We therefore investigated which DNA methylation sites showed high correlation with gene expression in healthy tissue using matched DNA methylation and gene expression datasets from a large cohort of control brains.

Building upon previous findings of differential DNA methylation in neurodegenerative diseases, this chapter explores the role of DNA methylation during OLG differentiation and in regulating gene expression in post-mortem control tissue, leveraging insights from cell models and healthy tissue to assess the potential functional relevance of DNA methylation changes observed in neurodegeneration. We utilized DNA methylation profiles from Induced pluripotent stem cells (iPSCs) derived cells throughout OLG differentiation stages (kindly provided by Dr Evans from the Gandhi lab). iPSCs are stem cells created through the reprogramming of human adult cells to a pluripotent state, allowing them to develop into nearly any cell type in the body [301]. Such cells are highly useful in studying differentiation, as they can be programmed to differentiate

from neural stem cells to cells expressing mature OLG markers [301]. We have investigated DNA methylation profiles at three stages of OLG differentiation, allowing us to study changes in DNA methylation through OLG lineage fate progression. The overarching goal of this Chapter was to investigate genes that undergo changes in DNA methylation during OLG lineage progression in order to understand if changes in DNA methylation in neurodegeneration, identified in Chapters 3-4, could reflect dysfunction of this process. We also undertook a bioinformatic approach to investigate DNA methylation/expression of disease associated OLG/OPC genes in healthy tissue, in order to explore the functional consequences of DNA methylation changes in disease.

5.2 Methods

5.2.i iPSC model of OLG differentiation

Dr James Evans (Gandhi lab, Francis Crick Institute, London), kindly provided us with cell pellets from hiPSC-derived OLGs at three differentiation stages. OLIG2-positive neural cells (NSCs) were generated through dual SMAD inhibition (inhibition of TGF-beta (SMAD2/3) and of BMP (SMAD1/5/8) pathways). SOX10-induced cells (USIs) were produced through introduction of lentiviral mediated overexpression of OLG transcription factor *SOX10*. O4-positive cells (O4s) were purified from the SOX10-induced culture using magnetic-activated cell sorting (MACS). We therefore analysed cells at three states of OLG maturation; NSC, USIs (OPC like) and O4s (OLG like), O4+ being a marker of more mature OLGs [302]. The hiPSC cell lines used for DNA methylation analysis in this thesis (as provided by Dr James Evans), included both control cell lines and lines from donors carrying the *SNCA* A53T mutation, which is associated with Parkinson's disease (**Table 5.1**). Characterization of these cell lines is further detailed in the PhD thesis of Dr James Evans (Thesis publication).

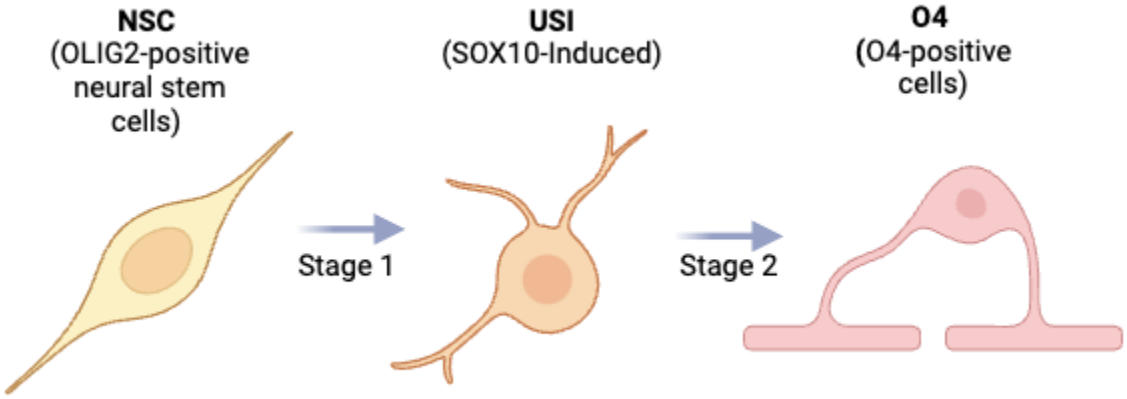
Table 5.1 Summary of hiPSC donor lines for which DNA methylation data was derived for this Chapter

Sample	**Mutation**	**PD Diagnosis**	**Stages of OLG analysed**
C1	-	Negative	NSC, USI and O4
C2	-	Negative	NSC and USI
C3	-	Negative	NSC, USI and O4

C4	*SNCA* A53T corrected	-	NSC, USI and O4
D1	*SNCA* A53T	Positive	NSC, USI and O4
D2	*SNCA* A53T	Prodromal	NSC, USI and O4
D3	*SNCA* A53T	Positive	NSC, USI and O4

PD: Parkinson's disease, OLG: Oligodendrocyte, NSC: OLIG2-positive neural stem cells, USI: SOX10-induced cells, O4: O4-positive cells, *SNCA: alpha-synuclein gene.*

Figure 5.1 Differentiation stages of OLGs as generated from hiPSCs



Schematic illustration demonstrating the OLG model of differentiation through which DNA methylation changes were measured in this chapter. Stage 1 represents the transition from NSC (representing the most immature stage of OLG) to USI (an intermediate stage of OLG differentiation), whilst Stage 2 represents the transition from USI to O4-positive cells (representing the most mature stage of OLG differentiation). NSC: OLIG2-positive neural stem cell, USI: *SOX10*-induced cells, O4: O4-positive cells.

5.2.ii DNA methylation profiling of iPSC-derived cells

Genomic DNA was extracted from iPSC derived cell pellets using the Quick-DNA/RNA Miniprep kit, following the manufacturer's instructions (Zymo Research, product code D7001). Five hundred ng of DNA per sample were then sent to UCL Genomics for bisulfite conversion and DNA methylation profiling. Genome-wide methylation profiles were generated using the Infinium HumanMethylationEPIC Version 2 BeadChip (Illumina).

Beta-values ranging from 0 to 1 (approximately 0% to 100% methylation, respectively), were used to estimate the methylation levels of each CpG site using the ratio of intensities between methylated and unmethylated alleles. Data analysis was conducted using several R Bioconductor packages. Due to differences in array versions and inability to load data from EPIC v2 with the current version of the ChAMP package[154], we modified the method of loading and quality control from that described in Chapter 2 Section 2.1, relying on the minfi and wateRmelon R packages only, and manually filtering out methylation sites that showed: (1) poor quality, (2) cross reactivity, (3) included common genetic variants, and (4) mapped to *X* or *Y* chromosome. As with other DNA methylation datasets described in Chapter 2 Section 2,1, samples were dropped during quality control if: (1) they presented with a high failure rate (≥2% of methylation sites), or (2) they clustered inappropriately on multidimensional scaling analysis.

Beta-values were normalised with wateRmelon using the dasen normalisation method. *M*-values, computed as the logit transformation of beta-values, were again used for all statistical analysis, as recommended by [155], and further described in Chapter 2.

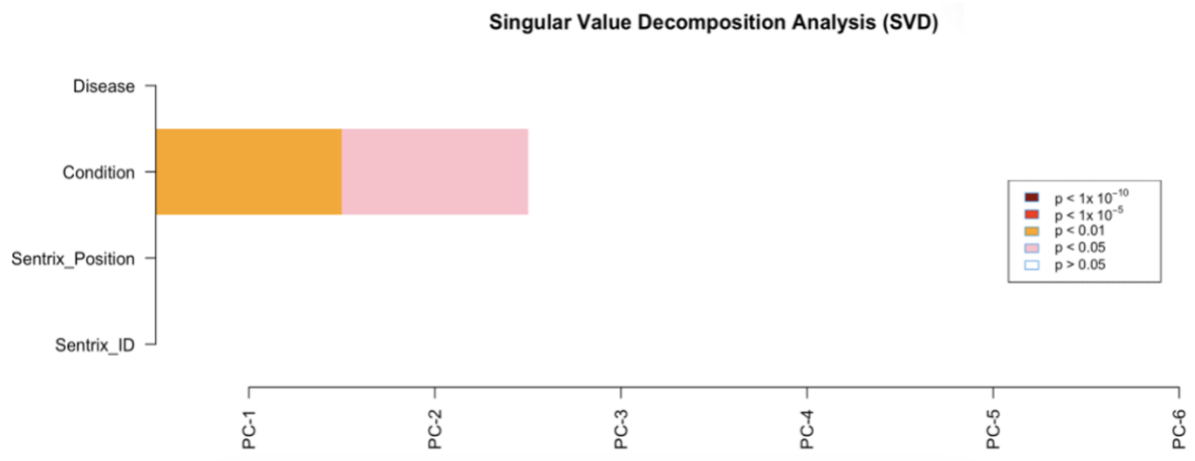
5.2.iii Differential methylation analysis

We applied linear regression models using the *M*-values as the input to identify associations between DNA methylation variation at specific CpG sites and differentiation stage (condition) using the limma package [156] using the following model:

$$\sim 0 + condition + slide$$

The following comparisons were performed USI versus NSC and O4 versus USI. False discovery rate (FDR) adjusted *p*-values < 0.05 were considered genome-wide significant. Given that we did not see on SVD plots any separation between samples with or without the *SNCA* A53T mutation (**Figure 5.2)**, all samples were analysed together in a single model to increase statistical power with a larger sample size, without stratifying for the presence of the mutation. We adjusted the data for slide, as we had seen that this technical batch effect contributed to data variance (data not shown).

Figure 5.2 Singular Value Decomposition (SVD) analysis of DNA Methylation Data in iPSC derived OLG Cells



SVD analysis visualising both technical and biological variances explaining data variance after quality control, normalisation and covariate adjustment. PC: Principal component, Disease - presence/absence of A53T mutation,

Condition - OLG differentiation stage (i.e. NSC, USI or O4+), NSC: neural stem cell, USI: unstimulated SOX10-induced, O4+: O4 antigen expressing cells, SVD: singular value decomposition, Sentrix_Position: slide position (as detailed above), Sentrix_ID (array).

5.2.iv Post-mortem brain tissue DNA methylation - gene expression correlations

To analyse the role of DNA methylation on gene expression within OLG/OPC genes, we chose the largest brain derived control dataset for which we had matching DNA methylation and gene expression profiles, which was control samples from our bulk AD dataset (N = 518) described in Chapter 3 (and described in methods section 2.1). Processing of this data is described in Chapter 2.2, where we also provide a description of the quality control and processing of the RNA-sequencing data matching with this DNA methylation dataset.

5.2.v Curation of gene lists of interest

In order to investigate the disease-associated differentially methylated genes from previous Chapters 3 and 4, we curated a list of 'genes of interest' to examine. As well as looking solely at differentially methylated methylation sites from the EWAS analyses described in Chapter 3, we included in this list hub genes of disease associated OLG/OPC modules from co-methylation networks described in Chapter 4. We also included genes which, although not themselves hub genes, showed consistent presence in disease-associated co-methylation modules and showed differential expression in disease. Gene list and reasons for inclusion are outlined in **Appendix M.**

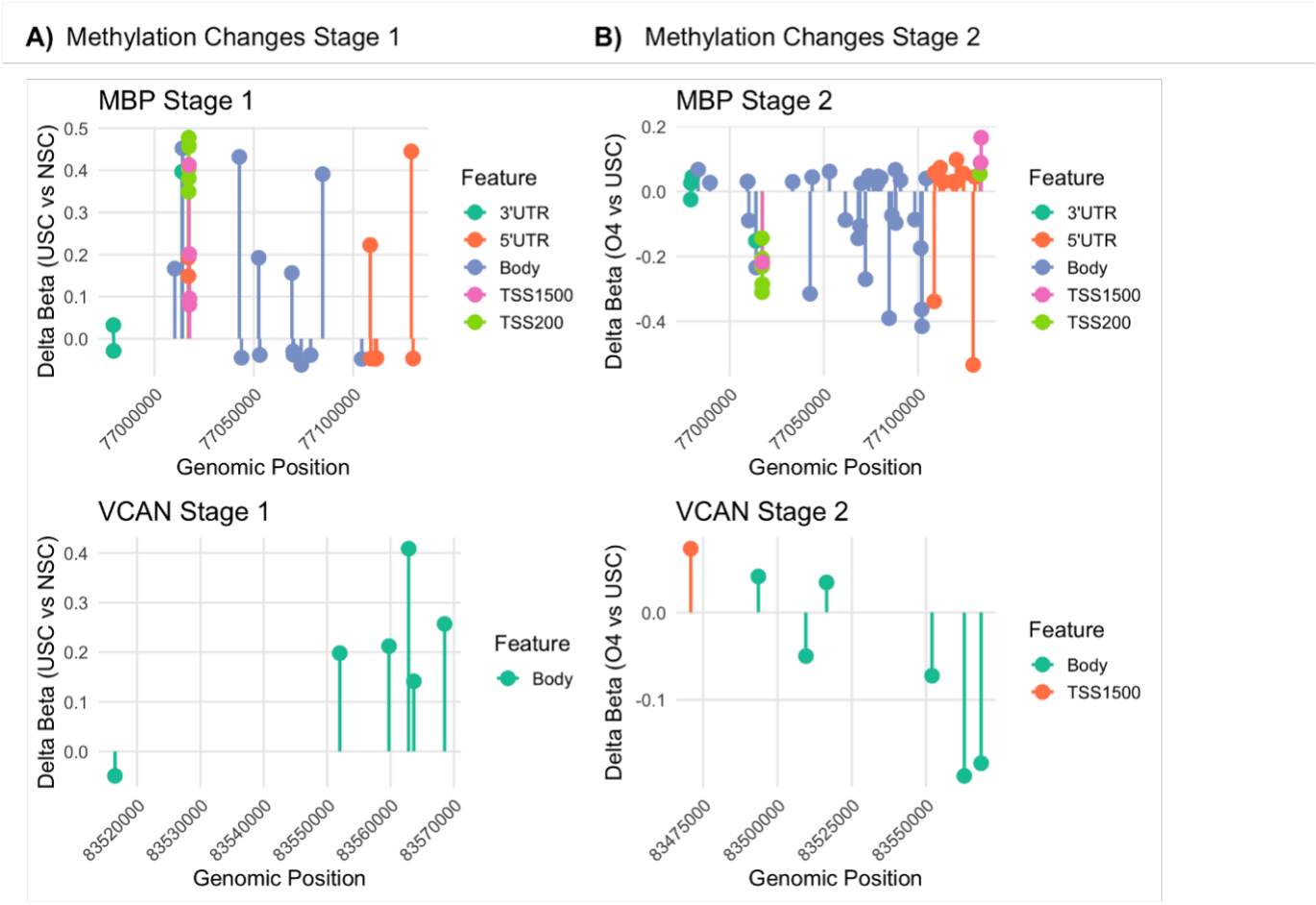
5.3 Results

5.3.1 DNA methylation changes during Oligodendrocyte Differentiation using iPSC derived Models

5.3.1.i Characterisation of DNA methylation profiles across OLG differentiation stages

Firstly, we undertook differential methylation analysis to investigate which genes contained differentially methylated sites at the two stages of differentiation - NSC to USI (OPC-like) (Stage 1), and USI to O4 (OLG-like) (Stage 2). We first checked DNA methylation changes in two known markers of OLG differentiation states, to add confidence to this model of differentiation derived from iPSCs, namely at *VCAN* and *MBP. VCAN* is a well described marker of OPCs and has previously been used to ascertain differentiation in a human cell model of OLG differentiation [303]. *MBP* codes for myelin basic protein which is exclusively expressed by myelinating OLGs [304]. We saw methylation changes at key sites in these genes across the two differentiation stages (**Figure 5.3)**. For *VCAN,* we saw hypermethylation of the promoter region during Stage 2 of differentiation - from OPC-like to OLG-like. Given that increased DNA methylation at the promoter region of genes is associated with decreases in gene expression - this aligns with repression of this OPC gene in more mature cells. For *MBP,* we again saw significant changes in promoter region DNA methylation in both Stages 1 and 2. In the change from NSC to USI/OPC-like cells, the promoter region is hypermethylated, whilst the region becomes hypomethylated in the transition from USI (OPC-like) to O4+ (OLG-like). These changes in *MBP* DNA methylation support silencing of gene expression during the first stages of differentiation, and then activation during the change from OPC-like cells to OLG-like cells. Since *MBP* is expressed by myelinating mature OLGs, this finding suggests that this model of OLG differentiation in iPSC recapitulates key features of OLG differentiation processes.

Figure 5.3 Differential methylation changes across OLG maturation markers at two stages of OLG differentiation



Differential methylation changes in key oligodendrocyte genes (e.g., *MBP* and *VCAN*) during Stage 1 (A) and Stage 2 (B) of differentiation. Each point represents a methylation methylation site mapped to a specific genomic position. The y-axis shows the magnitude of methylation change (values above zero represent hypermethylation and below zero show hypomethylation) measured as Delta Beta values, while the x-axis represents the genomic coordinates. Methylation sites are coloured by their genomic features.

Looking more generally at which genes contained differentially methylated sites across the two stages of OLG differentiation, we found a striking number of differentially methylated sites at both stages. Of the 704,510 methylation sites analysed, 89,370 and 198,043 for stages 1 and 2, respectively, showed significantly different methylation (adjusted p-value < 0.05). The top most differentially methylated methylation site in stage 1 of OLG differentiation mapped to the gene

body of *RERE* (**Table 5.2)**. *RERE* has been implicated as a mediator of neural progenitor cell proliferation [305]. The top-most differentially methylated methylation site in stage 2 of OLG differentiation mapped to the gene *EPHB2* (**Table 5.2)**. *EPHB2* codes for ephrin type-B receptor 2 which is a transmembrane protein, ephrins and ephrin receptors have been shown to be important during OPC differentiation [306]. Interestingly, in a mouse model of OPC differentiation, *Ephb2* was found to be differentially methylated between OPCs and OLGs, as well as being differentially expressed [124]. These findings of top differentially methylated genes, including *RERE* and *EPHB2*, having described roles in OPC differentiation adds confidence to the biological relevance of this iPSC model of differentiation.

Interestingly, many of the top differentially methylated genes in both stages showed opposite directions of effect in the other stage (**Figure 5.4, Table 5.2**). For example, several genes mapping to the gene *CTNNA1* showed significant increase in DNA methylation during Stage 1 of differentiation - in the transition from NSC to USI, but then showed significant decrease in DNA methylation during Stage 2 - from USI to O4 (**Figure 5.4, Table 5.2**). This was true for multiple genes, as is visualised in **Figure 5.4C**. When investigating the top significantly differentially methylated sites across both stages, all show opposing directions of DNA methylation change. This highlights the importance of DNA methylation changes during OLG maturation. The dynamic nature of these changes in DNA methylation status may regulate processes that are activated at one stage (e.g., during the NSC to USI transition) and deactivated at the other (e.g., during the USI to O4 transition), or vice versa. The 10 top-most differentially methylated sites in Stages 1 and 2 of the model of OLG differentiation are provided in **Table 5.2**. Interestingly, 7/10 of these sites are the same, but show opposite directions of effect. This could be explained through a potential reversal in dynamics between the early and late stages of OLG differentiation - potentially a shifting mechanism whereby DNA methylation allows expression of genes essential for one stage of OLG differentiation and then re-

establishes repression once that stage of OLG differentiation is complete. This aligns with the concept of gene switching[307], where genes are activated in one stage and deactivated at a later stage, or vice versa. It has been found that the activation of the *Wnt* signalling pathway (which we have discussed in Chapters 3 and 4 in relevance to OPC proliferation and discussed below) inhibits specification of OPCs from NSCs, but is required for differentiation of OPCs into OLGs [183].

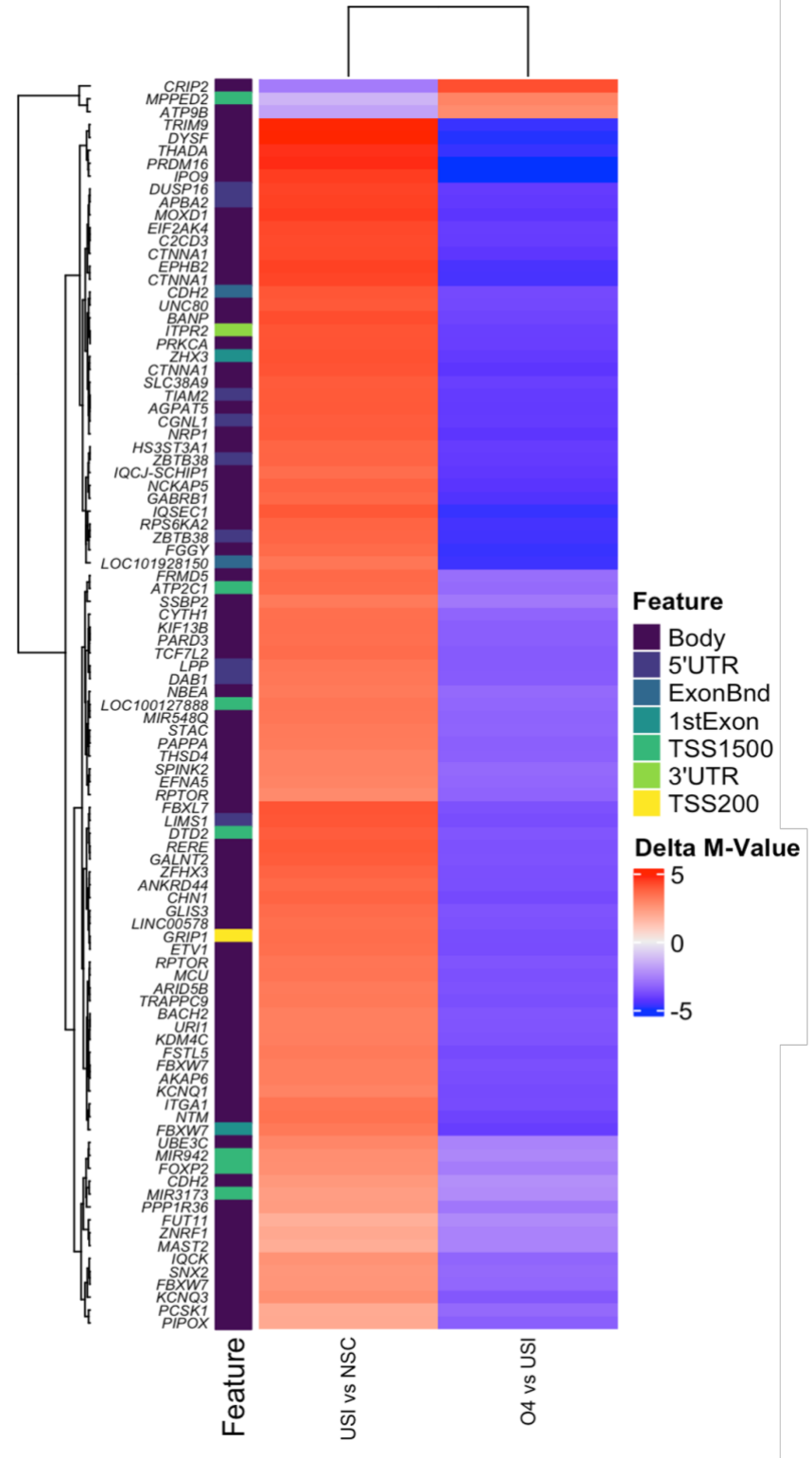


Figure 5.4 Top Differentially methylated sites (DMPs) between two stages of oligodendrocyte differentiation.

Heatmap of differentially methylated sites (DMPs) across oligodendrocyte differentiation comparisons. Methylation differences (delta M-values) for the top 100 significant methylation sites across two comparisons: USI vs NSC (SOX10-induced stem cells vs neural stem cells) and O4 vs USI (mature oligodendrocytes vs SOX10-induced stem cells). Rows represent individual genes associated with the DMPs. Columns represent the two comparisons (USI vs NSC and O4 vs USI). Red indicates increased methylation (positive delta M-value). Blue indicates decreased methylation (negative delta M-value). White represents no change (delta M-value near zero). Coloured bars on the left hand side represent the genomic feature of each methylation site (e.g., Body, TSS200, 5'UTR, etc.), as indicated in the legend. NSC: neural stem cell, USI: SOX10 induced cell, O4: O4+ cell.

Table 5.2 Top 10 differentially methylated sites from differentiation stages in iPSC model of Oligodendrocyte differentiation

CpG	CHR	Gene	Feature	Delta M-Value	P-value (adjusted)
Top differentially methylated sites Stage 1 Differentiation					
cg07145988	1	*RERE*	Body	4.0077202	5.86E-09
cg15447913	1	*EPHB2*	Body	4.45407123	5.86E-09
cg06606949	5	*CTNNA1*	Body	4.3140927	2.21E-08
cg22333841	2	*THADA*	Body	4.68778293	2.59E-08
cg14196395	2	*DYSF*	Body	5.03555759	2.86E-08
cg00285941	2	*UNC80*	Body	4.0069706	3.49E-08
cg06211743	15	*CGNL1*	5'UTR	3.93739665	3.80E-08
cg11285029	3	*ATP2C1*	TSS1500	3.61385993	4.34E-08
cg00066640	15	*FRMD5*	Body	3.64116213	4.44E-08
cg03173797	1	*CAPN9*	Body	2.78572712	4.44E-08
Top differentially methylated sites Stage 2 Differentiation					
cg15447913	1	*EPHB2*	Body	-4.5801337	2.48E-11
cg07145988	1	*RERE*	Body	-3.7089158	5.99E-11
cg06606949	5	*CTNNA1*	Body	-4.2827302	9.43E-11
cg22333841	2	*THADA*	Body	-4.7802765	9.43E-11
cg06211743	15	*CGNL1*	5'UTR	-4.1864351	9.86E-11
cg03795776	6	*BACH2*	Body	-3.6034639	1.26E-10
cg14196395	2	*DYSF*	Body	-4.9093266	1.45E-10

cg21902252	1	*MAST2*	Body	-2.5181932	1.55E-10
cg00285941	2	*UNC80*	Body	-3.914429	1.83E-10
cg23184518	5	*CTNNA1*	Body	-4.3013814	1.83E-10

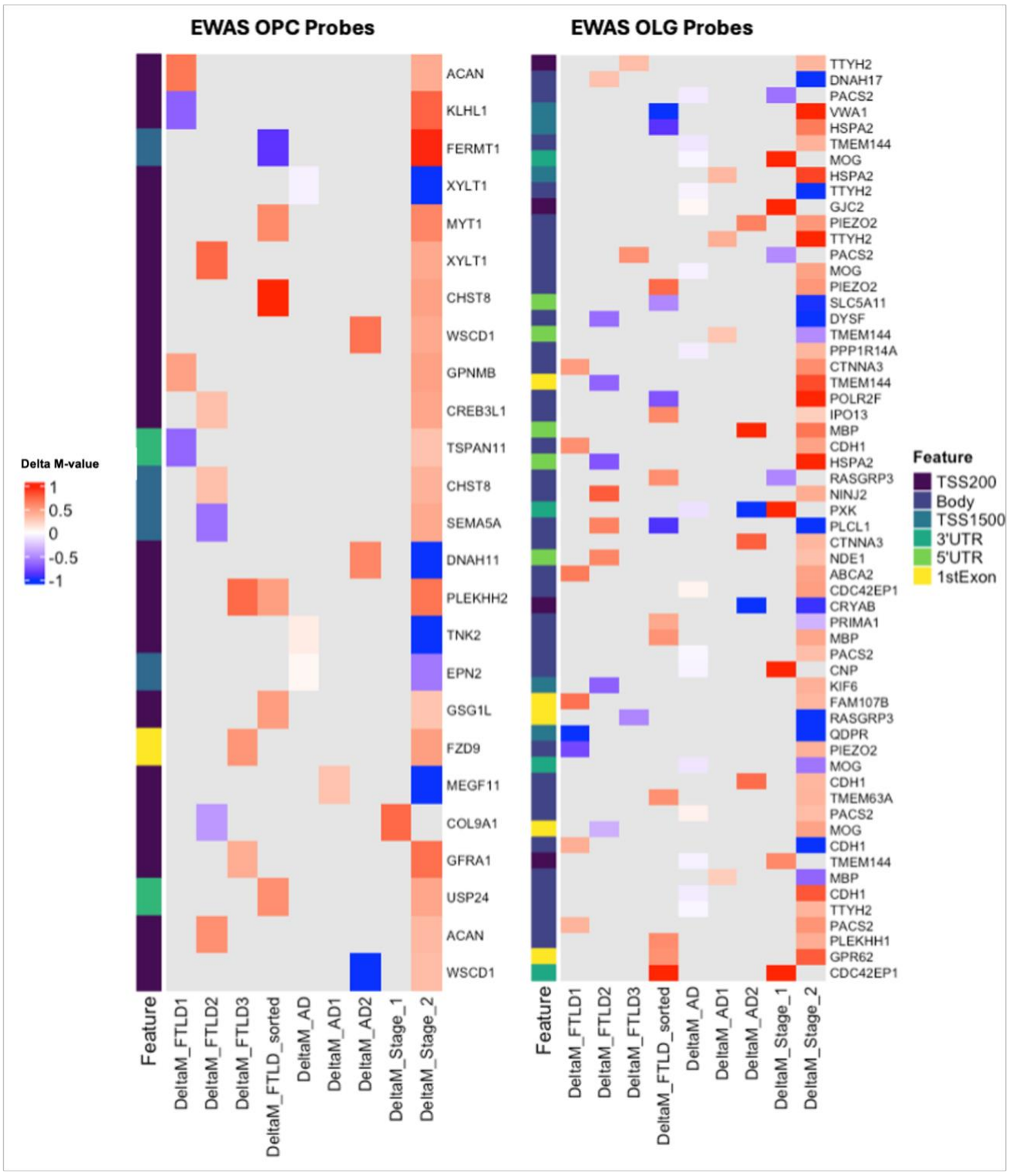
Next, we investigated whether any of the specific DNA methylation sites that were significantly differentially methylated during these stages of differentiation were also differentially methylated in any of the AD and FTLD EWAS analyses (**Chapter 3**). We found that there were several methylation sites which showed differential methylation at both OLG differentiation and in disease. To highlight some observations, we saw that a DNA methylation site in *FZD9* that was the top-most differentially methylated OPC site in the bulk AD EWAS was differentially methylated in Stage 2 of differentiation (**Figure 5.5**). *FZD9* codes for Frizzled-9, a receptor involved in the Wnt signalling pathway, which is known to undergo stage specific changes during OPC differentiation [183]. We also saw that the gene *FERMT1* was differentially methylated during differentiation (**Figure 5.5**). *FERMT1* was found to be differentially methylated in the FTLD1, FTLD2, FTLD-sorted and AD datasets, and upregulated at the gene expression level in the FTLD1-expression data. It has been found that *FERMT1* activates the Wnt signalling pathway through interaction with β-catenin [308]. Additional genes found to be differentially methylated in disease and differentiation that have been proposed to be involved in Wnt signalling include *GPNMB* [309], *SEMA5A* [310] and *EPN2* [311], highlighting the relevance of this signalling pathway.

Looking at OLG EWAS disease-associated methylation sites and sites differentially methylated during differentiation, we again saw overlap. Only one specific methylation site was present

across multiple EWAS and differentially methylated during differentiation - a methylation site within the 3'UTR of the gene *PXK.* Although this gene showed an increase in DNA methylation during the Stage 1 of OLG differentiation, this methylation site showed a decrease in DNA methylation in the bulk AD and sorted AD2 analysis.

For both OPC and OLG genes, in some cases, the direction of methylation change in disease (i.e. disease vs control) was the same as in differentiation stages, i.e. the methylation site showed increased DNA methylation through the differentiation stage (either Stage 1 or Stage 2), and also showed increase in DNA methylation in disease vs control (**Figure 5.5**). However, there were also instances in which the direction was different (**Figure 5.5**). Where directions are opposite, this could indicate that processes by which DNA methylation acts to allow differentiation of OLGs are being disrupted by disease-associated DNA methylation changes. It could also indicate altered proportions of OLGs/OPCs in disease tissue. If there is a higher proportion of OPCs in disease samples, that would tend towards DNA methylation profiles showing more similar patterns of DNA methylation to earlier stages of the OLG lineage.

Figure 5.5 Heatmaps of significantly differentially methylated sites in EWAS analyses of oligodendrocyte precursor cells and oligodendrocytes and in stages of oligodendrocyte differentiation

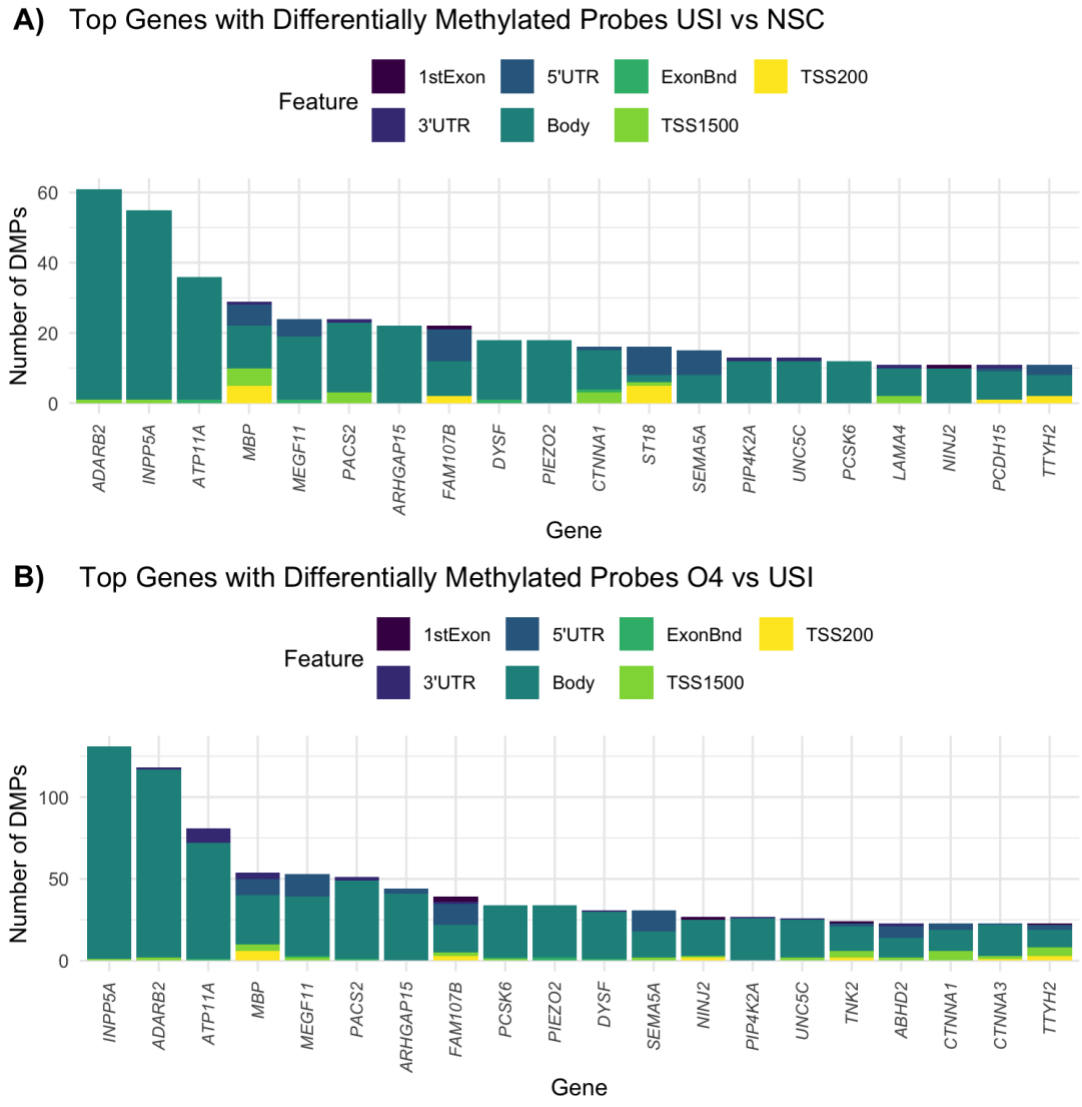


Heatmaps of differentially methylated sites (DMPs) in EWAS analyses of oligodendrocyte precursor cells (OPCs) and oligodendrocytes (OLGs). Left Panel (EWAS OPC methylation sites): Heatmap displaying the Delta M-values of DMPs associated with OPC genes across multiple EWAS datasets and Delta M-values of those DMPs in stages 1 and 2 of OLG differentiation. Right Panel (EWAS OLG methylation sites): Heatmap displaying the Delta M-values of DMPs associated with OLG genes across multiple EWAS datasets and Delta M-values of those DMPs in stages 1 and 2 of OLG differentiation. Columns represent specific EWAS datasets, including FTLD1, FTLD2, FTLD3, FTLD-Sorted, AD, AD1, AD2, and differentiation stages (Stage_1 and Stage_2). Rows indicate genes linked to differentially methylated sites found in OPCs/OLGs. Colour Scale - red indicates positive mean difference of M-value (hypermethylation), blue indicates negative mean difference of M-value (hypomethylation), white indicates no change, grey indicates that the CpG was not significantly associated with disease or differentiation stage. AD: Alzheimer's disease, FTLD: frontotemporal lobar degeneration, OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell, EWAS: epigenome-wide association study.

Aside from looking at specific DNA methylation sites that were found in both disease (EWAS from Chapter 3) and differentiation associated, we were also interested more generally in the behaviour of key genes identified across our network analysis (Chapter 4) (**Figure 5.7**). For this, we used a curated list of genes showing high relevance to disease (**Appendix M, Methods Section 5.2**). In the differentiation Stage 1, the gene of interest which contained the highest number of differentially methylated sites was *ADARB2* (**Figure 5.7**). This gene was of interest due to its consistent occurrence in our DNA methylation signatures associated with OLGs across multiple brain regions in AD (Chapter 4). *INPP5A* was another gene of interest that was included due to its presence across this disease-associated cross-brain region signature, as was *ATP11A, UNC5D* and *FAM107B.* All of these genes showed multiple methylation sites that were significantly differentially methylated across both stages of differentiation. As was described in Chapter 4, hub genes of the co-methylation modules containing the above mentioned genes, included *MYRF,* and *MOG*, which are important drivers of OLG maturation

and myelination [312]. The identification of multiple sites mapping to these genes highlights and supports potential functional consequences of DNA methylation changes in differentiation and disease.

Figure 5.7 Distribution of differentially methylated sites across top genes and associated genomic features in oligodendrocyte differentiation.



A) Bar plot visualising the top genes with DMPs in the comparison between SOX10-induced stem cells (USI) and neural stem cells (NSC). B) Bar plot visualising the top genes with DMPs in the comparison between mature

oligodendrocytes (O4-positive cells) and SOX10-induced stem cells (USI). For both, the x-axis represents genes, and the y-axis shows the number of DMPs mapped to each gene. Colours correspond to specific genomic features (e.g., TSS200, Body, 3'UTR), as shown in the legend. NSC: neural stem cell, USI: SOX10 induced cell, O4: O4+ cell.

As we had described above, a gene which contained some of the most significantly differentially methylated sites in both stages of OLG differentiation (although in opposite directions in each of these stages) was *CTNNA1,* which codes for α-catenin. This gene was also of particular interest due to it being the hub gene of an OLG enriched AD module from the co-methylation network analysis in Chapter 4 (**Appendix M**). In this module, increases of DNA methylation in methylation sites within the module were associated with increased disease risk. Here, we see that multiple methylation sites mapping to *CTNNA1* show hypomethylation and seem needed for the transition from USI to O4 (Stage 2 of differentiation). If opposing directions are seen in transition and disease, as we see here (**Figure 5.6**) this could indicate that increased DNA methylation of this gene in disease represents dysfunction of the decrease in methylation of the gene needed for maturation of OLGs.

The gene *PIP4K2A* also showed multiple methylation sites that were significantly differentially methylated during both stages of OLG differentiation. This gene contained the top-most differentially methylated OLG methylation site in the FTLD3-EWAS (Chapter 3), and was also a gene with high module membership in an FTLD-associated co-methylation module that was enriched for peroxisome-associated genes. The role of peroxisomes and associated signalling pathways have been found to be important in OPC proliferation and differentiation into myelinating OLGs [313], and once again possibly linking disease to OLG lineage changes.

5.3.2 Investigating the role of DNA methylation in gene expression within oligodendrocyte lineage genes in healthy brain tissue

In this section, we describe the analysis of correlations between DNA methylation and gene expressions in a large sample of control samples. The rationale behind this approach was to strengthen the link between DNA methylation changes and downstream gene expression by identifying CpGs that, in control tissue, exhibit a strong effect of DNA methylation on the expression levels of their corresponding genes. This analysis provides evidence to support the hypothesis that aberrantly differentially expressed genes in disease are driven, at least in part, by changes in DNA methylation.

Firstly, we examined which of the CpGs mapping to genes within the general OLG and OPC gene lists seem to show stronger effects on the expression levels of the corresponding genes. To do this we used a DNA methylation dataset for which we also had overlapping gene expression datasets - the AD bulk data described in **Chapter 3.** We subset these data for only control samples (N= 318), and carried out correlation analysis between DNA methylation and gene expression at all methylation sites in OLG/OPC genes (those described in **Appendix A**). Interestingly, the majority of these CpGs were not located within the promoter of the gene (66% and 56% for OLG and OPC, respectively), again illustrating the complexity of DNA methylation-gene expression dynamics, and highlighting the need for more research into the role of non-promoter region DNA methylation.

We looked at methylation sites within our EWAS analysis across diseases, and if any of these were significantly associated with gene expression in control brain tissue. Several disease associated OLG and OPC methylation sites were present (**Table 5.3**). Of the OPC genes, we found that two methylation sites mapping to *GALR1* promoter region (TSS1500) were significantly negatively correlated with gene expression in control data. This is in line with the model that DNA methylation at the promoter region is associated with a decrease in gene expression. The two methylation sites in *GALR1* were hypomethylated in disease (AD2 dataset), and we saw corresponding upregulation of this gene in the bulk AD RNA-sequencing dataset. *GALR1* codes for Galanin receptor 1, and expression changes of this gene have been observed during remyelination in MS [192,194], as well as the gene having been found to be overexpressed in AD[192]. We saw that methylation sites within *GALR1* were differentially methylated across multiple diseases (the FTLD-sorted, FTLD1, FTLD2 and AD EWAS analyses, **Chapter 3**). This finding adds strength to the model that aberrant DNA methylation at the promoter leads to aberrant upregulation of *GALR1* in disease. Another gene where a methylation site was significantly correlated with gene expression and was disease associated was *COL9A1.* This gene was differentially methylated in the FTLD1, FTLD2, FTLD-sorted and AD2 EWAS, and was downregulated in an OPC cluster in the AD snRNA-seq data.

Within our OLG genes, two methylation sites mapping to the promoter region of *HHIP,* which unexpectedly showed positive correlation with gene expression in control, were found across our AD EWAS analyses. We saw that in the bulk AD expression data, *HHIP* was downregulated. We had seen both hypo- and hypermethylation of the promoter region of this gene in AD, thus, the relationship between gene expression and methylation, and perturbations of this relationship in disease, remain unclear.

Table 5.3 Methylation sites significantly associated with gene expression in control tissue that are also present across disease EWAS analyses

CpG	Gene	Feature	EWAS	Correlation with gene expression in controls (P-value)	Altered gene expression in disease
OPC Genes					
cg03075065	*WFDC1*	Body	AD (Hypo)	-0.16 (0.029)	-
cg14273027	*KLHL1*	TSS200	AD (Hypo)	-0.23 (0.001)	-
cg17274742	*GPNMB*	1st Exon	AD (Hyper)	-0.45 (3.2E-11)	-
cg04525189	*VIPR2*	Body	AD1 (Hypo, top most differentially methylated methylation site in AD1 and meta-analysis of AD2 and AD2)	-0.22 (0.002)	-
cg05528293	*GALR1*	TSS1500	AD2 (Hypo)	-0.22 (0.002)	Upregulated in bulk AD, downregulated in OPC subcluster of snRNA-sequencing
cg07390210	*GALR1*	TSS1500	AD2 (Hypo)	-0.19 (0.009)	As above
cg07862930	*KCNJ16*	5'UTR	AD2 (Hypo)	-0.17 (0.021)	-
cg09961397	*COL9A1*	Body	AD2 (Hypo)	-0.15 (0.032)	Downregulated in OPC subcluster of snRNA-sequencing
cg05144285	*VIPR2*	Body	FTLD-sorted (Hyper)	-0.26 (0.0003)	-
OLG Genes					
cg03018256	*TTYH2*	Body	AD (Hypo)	0.21 (0.003)	Downregulated in bulk AD
cg07318204	*HHIP*	TSS1500	AD, AD2 (Both hypo)	0.18 (0.01)	Downregulated in bulk AD
cg10069691	*ST18*	5'UTR	AD (Hypo)	-0.16 (0.03)	Downregulated in bulk AD
cg17036418	*RFFL*	TSS1500	AD (Hyper)	-0.15 (0.03)	Downregulated in bulk AD
cg02524475	*HHIP*	TSS1500	AD1 (Hyper)	0.16 (0.03)	As above

cg04007987	*TTYH2*	Body	AD1 (Hyper)	0.21 (0.003)	As above
cg13882486	*LPAR1*	5'UTR	FTLD-sorted (Hyper)	-0.14 (0.04)	-
cg26347632	*CDC42EP1*	3'UTR	FTLD-sorted (Hyper)	0.18 (0.01)	-
cg04826663	*ANLN*	Body	FTLD1 (Hyper)	-0.16 (0.02)	-
cg07119871	*PRIMA1*	Body	FTLD1 (Hypo)	-0.21 (0.003)	-
cg07354124	*TCP11L2*	TSS1500	FTLD3 (Hypo)	0.20 (0.004)	-
cg17012863	*LIPA*	TSS200	FTLD3 (Hypo)	0.17 (0.018)	-

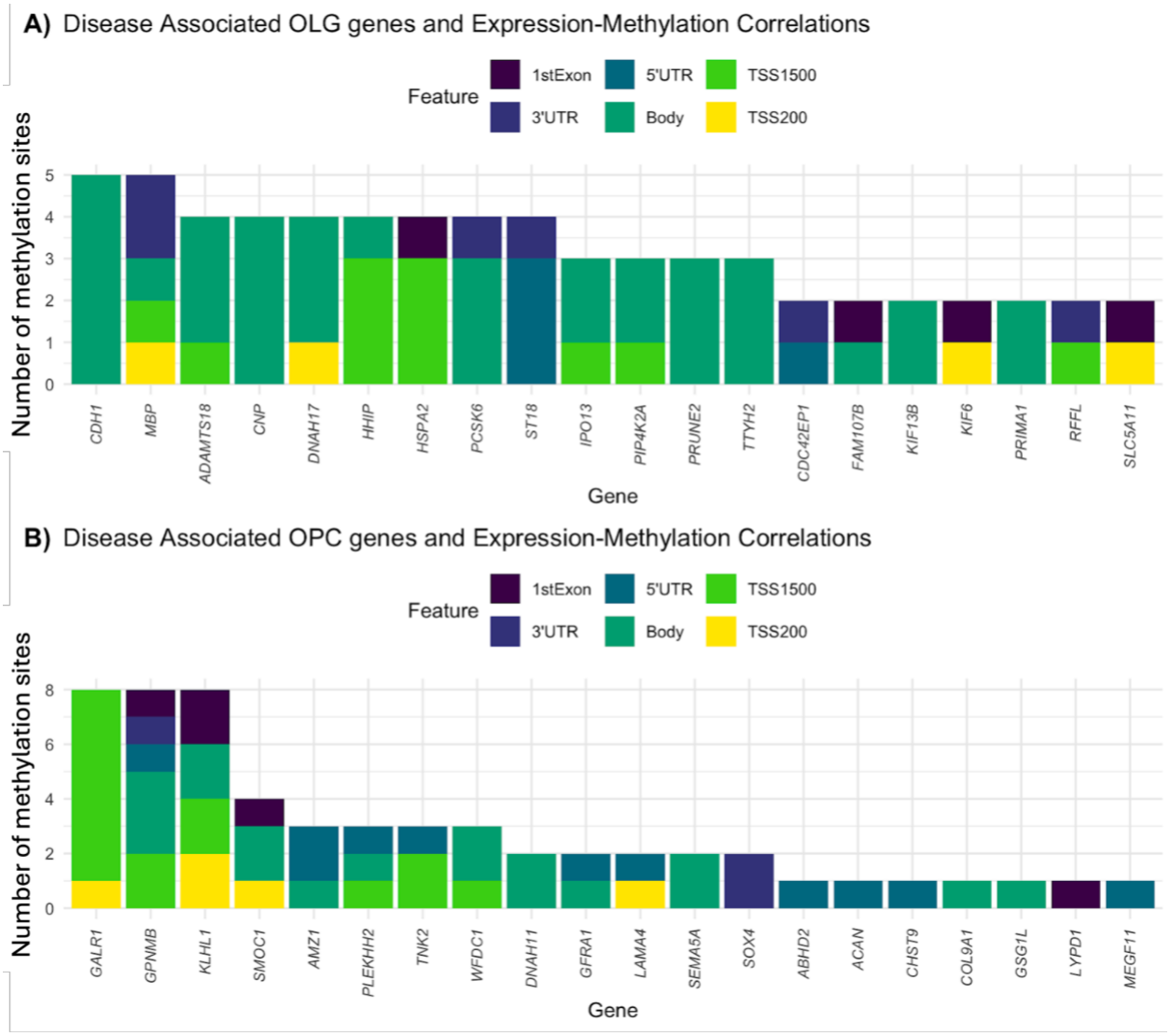
AD: Alzheimer's disease, FTLD: frontotemporal lobar degeneration, OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell, EWAS: epigenome-wide association study, Hypo: hypomethylated, Hyper: hypermethylated.

Next, we looked more generally at which genes within our list of curated disease-associated 'genes of interest' contained methylation sites with significant associations with gene expression. The aim here was to investigate which of the disease-associated genes appeared to have expression significantly influenced by DNA methylation. Within OLG genes of interest (curated for this Chapter), we saw several genes that showed correlation between gene expression and DNA methylation that had also been identified as being differentially expressed in disease, for example *MBP* and *PIP4K2A* (**Figure 5.8**). The finding that there is indeed strong correlation between DNA methylation and gene expression strengthens the likelihood that the differential methylation that we see in disease is associated with the observed differential expression in disease. In terms of OPC genes, we again highlight *GALR1,* which contained several methylation sites that were associated with gene expression, again providing evidence as to the importance of DNA methylation at this gene leading to aberrant expression of the gene in disease.

We also saw correlations between gene expression and DNA methylation in genes that we had identified as differentially methylated, but had not observed changes in gene expression in disease. For OLG genes, a notable example included *DNAH17,* which had differentially

methylated sites across all 7 disease EWAS analyses. In control tissue, it appears as though there are many methylation sites within this gene which are strongly correlated with gene expression (**Figure 5.8**), indicating that DNA methylation does play a strong role in expression of the gene, but that effect is likely perturbed in disease.

Figure 5.8 Distribution of differentially methylated sites showing high correlation with gene expression across disease-associated OLG and OPC genes



Bar plot showing the number of CpGs associated with A) OLG and B) OPC disease-associated genes. The x-axis represents genes, and the y-axis indicates the count of CpGs correlated with gene expression within each gene. Each bar is stacked by genomic features (e.g., 1stExon, 3'UTR, Body, TSS1500, TSS200). OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell.

5.4 Discussion

In this chapter, we describe the investigation of the relevance of genes identified as differentially methylated in disease, to the OLG life cycle, and to gene expression regulation in healthy brain tissue.

It was interesting to see that the majority of the most significantly differentially methylated positions were shared across the two stages of OLG differentiation we have investigated, but show opposing directions. This finding emphasises the importance of stage-specific DNA methylation effects. Furthermore, we found that several genes identified as differentially methylated in neurodegenerative diseases also displayed differential methylation throughout the OLG differentiation process. If the intricate changes in DNA methylation involved in genes relevant to OLG differentiation are disrupted, this could lead to an inability of the OLGs to pass through maturation stages, and lead to accumulation of more immature OPCs and decreased proportions of mature myelinating OLGs. Indeed, as described above and in the introduction, we see evidence of altered proportions of OPCs and OLGs across neurodegeneration. Our findings support the role of altered DNA methylation in this phenomenon. We investigated the directionality of methylation changes occurring in disease and in OLG differentiation. In some cases, we saw that the direction of effect was the same in disease and differentiation. It is possible that the same pathways that are involved in differentiation are also disease relevant, potentially pointing at an amplification or silencing of a particular pathway. Our findings could also indicate a compensatory response to disease. It could also be the case that, if more immature OLG lineage cells have methylation profiles more similar to earlier differentiation stages, their increases in disease would shift the methylation profiles to resemble less differentiated states. When the direction was reversed, i.e. observed hypomethylation of

methylation sites with increased DNA methylation through disease, pathological processes that interfere with normal differentiation programs could be implicated.

A gene of particular interest that we found was differentially methylated in disease and during differentiation was *CTNNA1.* This was the hub gene of an AD OLG enriched co-methylation network module described in Chapter 4. As described here, we saw that multiple methylation sites mapping to *CTNNA1* show hypomethylation in the transition from USI to O4 (Stage 2 of differentiation), but hypermethylation in the transition from NSC to USI (Stage 1 or differentiation). In our module of interest, we saw that increases in DNA methylation at this gene were associated with AD. *CTNNA1* was found to be upregulated in AD in single-cell RNA sequencing data, revealing the gene to be more highly expressed in mature OLGs than OPCs ([314]). *CTNNA1* codes, along with *CTNNA2* and *CTNNA3* (the latter of which is differentially methylated across neurodegenerative disease, Chapters 3 and 4), for α-catenins, which have the role of linking cadherins with the actin cytoskeleton [315], and have roles in maintaining the structure of tissue. As well as having structural roles, α-catenins, through interactions with β-catenins and Wnt signalling pathways (a recurring finding throughout this work), are thought to have roles in cell proliferation. Mice with altered Wnt/β-catenin signalling (which can be modulated through α-catenin [316]) have been found to exhibit delayed appearance of mature OLGs [317], and there is substantial evidence that the *Wnt*/β-catenin signalling pathway is vital in OPC differentiation and myelin formation [183]. It is therefore interesting that we see that this gene shows important DNA methylation modulation throughout OLG differentiation and is differentially methylated in disease. Interestingly, a significant disease and differentiation associated methylation site mapping to the OPC associated gene *FERMT1* also showed opposing directions of DNA methylation in disease and differentiation (**Figure 5.6**). In FTLD, a methylation site within the body of this gene was hypomethylated, whilst the same methylation site was

hypermethylated in Stage 2 of OLG differentiation in our iPSC model. This is of relevance due to the role of this gene having been found to interact with β-catenin, and provide activation of β-catenin transcriptional activity and the Wnt/β-catenin signalling pathway [308]. In FTLD, we found that *FERMT1* showed an increase in expression (Chapter 3). These findings potentially implicate DNA methylation changes as being important in regulating such pathways during OPC to OLG differentiation, which could be being disrupted in neurodegeneration.

Other genes showing highly significant DNA methylation changes throughout differentiation were *MOG* and *MBP*, that are key markers of mature OLGs and myelin production [318]. With methylation sites mapping to these genes, we showed changes both in neurodegeneration and in differentiation, supporting again a potential role for DNA methylation changes affecting proliferation and/or differentiation of OPCs in neurodegeneration.

To gain insight into the downstream consequences of DNA methylation patterns, we also investigated those genes from our OLG and OPC gene lists that showed strong and significant correlation with gene expression in control tissue. Although we had investigated changes in gene expression of genes which we identified as differentially methylated in previous chapters, we aimed here to investigate more direct consequences of DNA methylation on gene expression in control, to highlight how changes in disease may be functionally relevant, and when the relationship between DNA methylation and gene expression may be broken. Multiple genes showed significant correlations between DNA methylation and gene expression. Interestingly, we saw that many of the methylation sites associated with gene expression were not located within or near the transcription start site, but rather within the body of the gene or the 3'UTR. Throughout our investigation into disease-associated differentially methylated sites, many have not been within the transcription start sites. Whilst research into the role of DNA

methylation previously focused mostly on promoter region DNA methylation, this highlights the need to investigate changes across other regions.

Methylation sites within the gene *GALR1* were found to be associated with gene expression in control tissue and differentially methylated in disease, with the gene also showing aberrant gene expression in AD. Other genes which also contained sites showing both differential methylation in disease, correlation with gene expression in control data, and aberrant gene expression in disease included *COL9A1, TTYH2, HHIP, ST18* and *RFFL.* We also saw that, in healthy tissue, several genes contained multiple methylation sites associated where DNA methylation was associated with gene expression that we had not identified as being aberrantly expressed in disease, even though we had observed differential methylation across disease. This highlights the need for more cell-type specific investigations into DNA methylation changes, as any changes in gene expression in disease may have been diluted by noise from other cell types. It is also true that gene expression may be under the influence of other factors aside from DNA methylation, both in health and/or disease.

There are limitations to consider with the approaches we have used in this chapter. To investigate DNA methylation changes during OLG differentiation, we have utilised an iPSC derived model of OLGs, kindly provided through Dr James Evans and Prof Sonia Gandhi, Francis Crick Institute. Although the use of iPSC cells has been revolutionary in the study of neurodegeneration (and other fields), there are of course multiple caveats to their applicability to disease processes. iPSCs in isolation may not fully capture the complex interactions of multicellular environments present in diseases. This is particularly relevant to OLGs, whose role it is to wrap around neurons, and, without the presence of these neurons, the OLGs may not fully differentiate into myelinating OLGs. Furthermore, and of high relevance to the study of DNA methylation using iPSCs, although iPSCs are reprogrammed prior to differentiation, they may

retain epigenetic marks from the original cells, thus potentially altering the effects we can measure.

There were also caveats with our attempts to investigate gene expression changes associated with DNA methylation. In our first approach, simply investigating which DNA methylation sites showed significant correlation between gene expression and DNA methylation levels, we utilized a large (for DNA methylation datasets) control dataset. An important consideration here is that, although samples will be derived from the same patient, tissue heterogeneity might be a significant confounding factor. Brain tissue is composed of multiple cell types, and proportions of these cell types will vary between tissue samples obtained. Such differences, even intra donor, could mean that correlations between gene expression and DNA methylation may be disguised by noise. Indeed, we did see that, even when correlations were statistically significant, effect sizes (i.e. correlation coefficients) were relatively small in magnitude. As is the case with all of the DNA methylation datasets, with the analyses described in this thesis, sample size is a limitation, meaning that statistical power to infer subtle changes in DNA methylation and gene expression is limited, even though we have chosen the largest of our control datasets to analyse.

Further work is needed to complement the investigation of the consequences of disease-associated DNA methylation changes. In this analysis, we grouped together iPSC cells showing the AT53 mutation and the control lines (described Methods, Section 5.2). We saw, using singular value composition analysis, that there was no separation between these two groups, and we therefore took this approach to increase the power of detection of differentially methylated sites with a larger sample size. However, expanding the study to investigate these data to investigate if there are any, albeit potentially subtle, differences between DNA methylation changes throughout differentiation between the two conditions would be of interest.

It would also be interesting to explore the concept of DNA methylation leading to splicing changes throughout differentiation. In a mouse model investigating effects of ablation of *Dnmt1* on OPC differentiation, Moyon et al. found that dysregulation of DNA methylation during differentiation led to widespread changes in splicing events, caused by aberrant exon-skipping and intron retention[124]. We noted that several methylation sites undergoing changes during the two stages of differentiation investigated in IPSC derived cells were not located within the promoter region. This could be, in part, down to the alternative role for DNA methylation in the control of splicing throughout OPC differentiation, and would be worthy of follow up.

In work described in this chapter, we have attempted to investigate consequences of changes in DNA methylation in neurodegeneration that we have identified in previous chapters. We found that several genes containing differentially methylated sites in disease also underwent changes in DNA methylation in OLG lineage progression, highlighting a role for aberrant DNA methylation in disease leading to dysfunction of OPC differentiation and underpinning the need for further investigation into the role of DNA methylation changes affecting OPC differentiation in disease (discussed in **Chapter 1 Section 1.6.1**). We have also investigated a more direct causal relationship between gene expression and DNA methylation in genes which we have identified as being differentially methylated in previous chapters, adding strength to associations between aberrant DNA methylation and gene expression in disease.

Chapter 6 - Genetic and epigenetic insights into the role of *MOBP* across neurodegenerative diseases

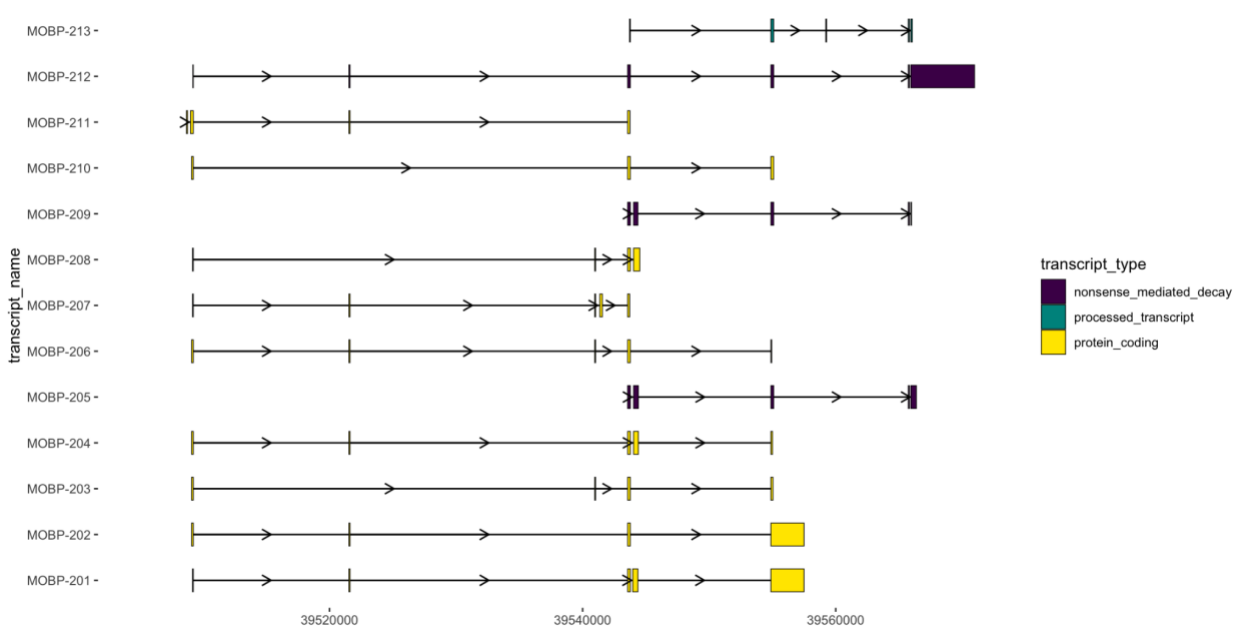
6.1 Introduction

MOBP, the gene encoding for myelin oligodendrocyte basic protein, is located on chromosome 3p22.1 (GRCh37 Chromosome 3: 39,508,689-39,570,970). This gene has several exons, and exhibits complex patterns of alternative splicing to produce distinct transcripts (**Figure 6.1**), some of which are predicted to undergo nonsense mediated decay (NMD) or lack open reading frames (ORF). There are three major protein isoforms of MOBP (RefSeq: isoform *a* [NP_001265251], isoform *b* [NP_001265252] and isoform *c* [NP_891980]), the most highly expressed of these in the brain is isoform *c* [129].

Although the exact function of MOBP is still unknown, it is the third most abundant myelin protein in the CNS [319], it is a myelin structural protein thought to contribute to stabilisation of myelin [319,320], and also to be involved in the morphological differentiation of OLGs [130]. It has been found that, during OLG differentiation in mice, *Mobp* mRNA was detected much earlier than MOBP protein, possibly indicating a form of translational repression [130]. Schäfer et al. investigated the role of MOBP through siRNA mediated knockdown of *MOBP* transcripts in primary oligodendrocytes, and observed a significant decrease in the cell surface area of MBP-positive oligodendrocytes compared to control siRNA-treated cells. Schäfer et al. further

investigated effects of MOBP knockdown, and found that overexpression of MOBP enhances the formation of myelin-like membranes, while MOBP knockdown significantly reduces their production by oligodendrocytes. Finally, Schäfer et al. investigated effects of MOBP overexpression on OLG processes - the long thin projections extending from the cell body which wrap around axons to form the myelin sheath [130]. Using Oli-neu cells, a simplified model for analyzing oligodendroglial morphology, the authors found MOBP overexpression led to a 3-fold increase in process length and a significant enhancement in process number and width.

Figure 6.1 Transcript structure of human *MOBP*



Rows represent unique *MOBP* transcripts described in Ensembl (*MOBP* Ensembl). Transcripts are color-coded based on their type. The exonic regions are shown as filled bars, coloured according to the transcript type. Introns are represented as connecting lines between the exonic ranges.

Genetic variation in the *MOBP* gene has been linked to multiple neurodegenerative diseases (**Table 6.1**), marking the relevance of the gene. SNPs in *MOBP* have been associated with disease risk, including PSP [76–79], corticobasal degeneration (CBD) [80], Alzheimer's Disease (AD) in apolipoprotein E-ε4 allele (*APOE*-ε4) carriers [81], frontotemporal lobar degeneration (FTLD)[321] and amyotrophic lateral sclerosis (ALS)[82], as well with disease duration and more severe white matter degeneration in behavioural variant of frontotemporal dementia (bvFTD)[84]. The SNP rs1768208 in particular is an intronic variant in *MOBP* (NC_000003.11:g.39523003T>C) that, through GWAS, was found to be associated with the risk of PSP[18,22] and CBD[19] , and has also been reported to be associated with white matter degradation and increased rates of decline in executive function in bvFTD[149]. The functional repercussions of such associations remain unclear. However, in human brain tissue, the risk allele T is also associated with increased expression of *MOBP* in PSP [22,150].

Table 6.1 Non-exhaustive summary of genetic disease-associated alterations in *MOBP*

Disease	Risk SNP (s)	Consequence	Reference (s)
FTD cohort (heterogenous)	rs1768208	- Disease risk - Reduced white matter integrity in midbrain	321
bvFTD	rs1768208	- Shorted median disease duration - Increased white matter neurodegeneration	84
PSP	rs1768208/multiple	- Disease risk - Increased *MOBP* expression - Increased tau threads	77 79,85 79,85 76–79 76–79
ALS	rs616147 rs631312	- Disease	82,322 82,322
CBD	rs1768208	- Disease risk	80
AD	rs538867	- Cognitive decline	323

FTD: Frontotemporal dementia, bvFTD: behavioural variant frontotemporal dementia, PSP: progressive Supranuclear Palsy, ALS: amyotrophic lateral sclerosis, CBD: corticobasal degeneration, AD: Alzheimer's Disease, SNP; single-nucleotide polymorphism.

Additionally, through investigations into DNA methylation patterns in MSA, Bettencourt et al. showed that in cerebellar white matter tissue, the promoter region of *MOBP* is the most differentially methylated region in MSA compared to controls [118]. CpGs mapping to this region were also differentially methylated in other brain regions, such as frontal and occipital white matter. Notably, as DNA methylation changes in *MOBP* were found even in brain regions very mildly affected by MSA pathology (e.g. occipital lobe) indicates that these may occur early in the disease and contribute to pathogenesis. In a subsequent study investigating downstream effects of altered DNA methylation at *MOBP*, Bettencourt et al. found that methylation status of *MOBP* in MSA is inversely correlated with *MOBP* expression levels (**Figure 6.2A**), indicating that the observed downregulation of this gene is likely driven by the hypermethylation of the promoter

region of this gene [129]. *MOBP* expression was also found to be decreased in OLGs extracted (through laser capture microdissection) from cerebellar white matter in MSA compared to healthy controls[223].

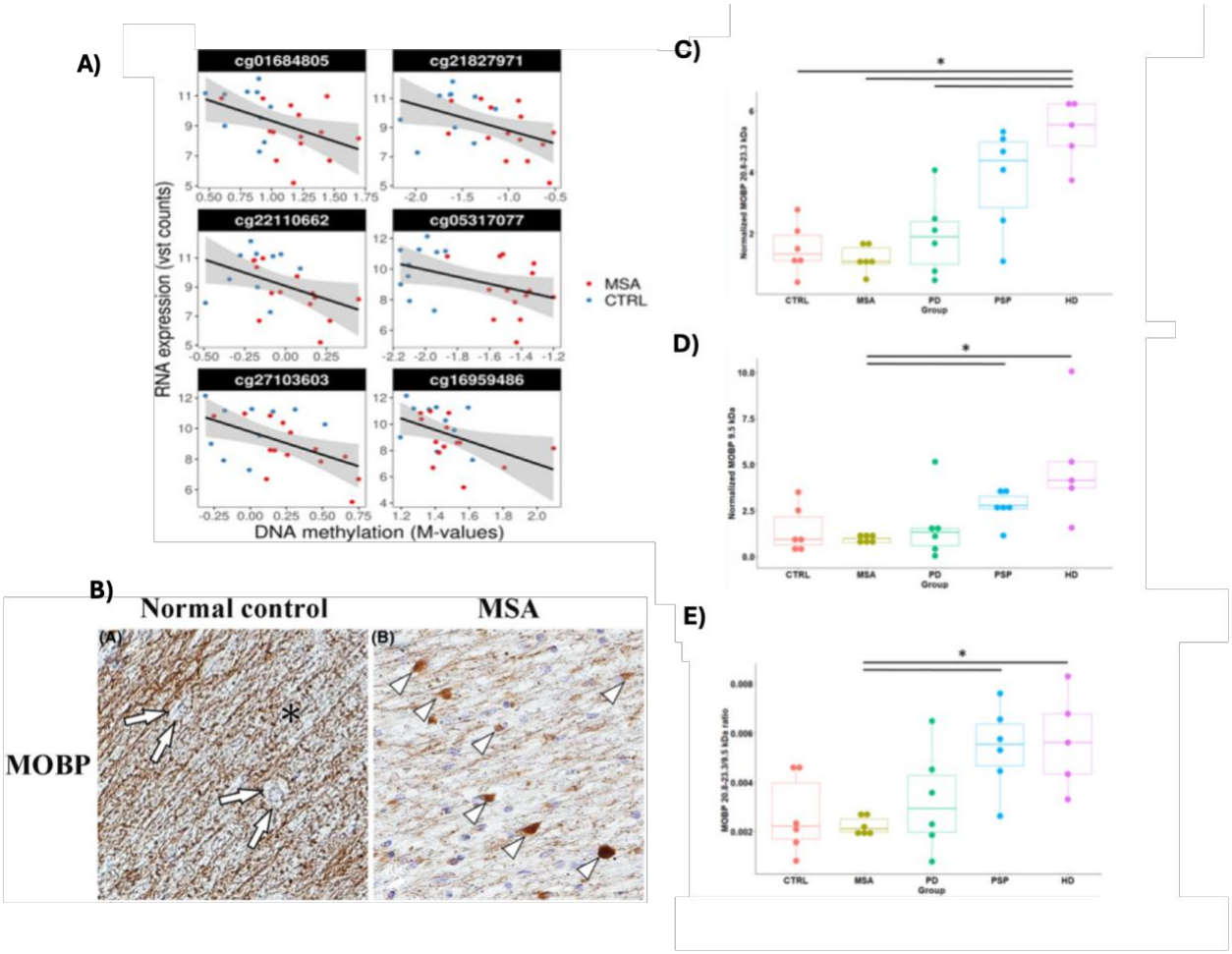
Protein expression of the three main isoforms of *MOBP* were also investigated in MSA, as well as in PSP, PD and Huntington's disease (HD). In MSA, it was found that there were no changes in any of the three isoforms, and no alteration in the ratio between the isoforms, compared to controls. However, it was notable that in PSP, there was both an increase in expression of the three isoforms (**Figure 6.2 B,C**), and an altered isoform ratio, with there being a increased ratio between levels of 20.8–23.3 kDa (isoforms *b* and *a*) over 9.5 kDa (isoform *c*) (**Figure 6.2 E**). Localisation of MOBP was also investigated In the cerebellar white matter in MSA and controls. In healthy controls, MOBP was detected in the myelin sheaths, while the cytoplasm of oligodendrocytes was immunonegative. In contrast, in MSA patients, GCIs (a hallmark of MSA pathology) showed strong immunopositivity for MOBP (**Figure 6.2 B**). Finally, protein–protein interactions of α-synuclein (as the pathological aggregating protein in MSA) and MOBP were investigated. Bettencourt et al. demonstrated that MOBP interacts with α-synuclein within GCIs in MSA cerebellar tissue.

Changes in MOBP protein expression and localization have also been implicated in other neurodegenerative diseases aside from MSA. Immunohistochemical analysis with anti-MOBP antibodies suggest that MOBP is sequestered into the core of Lewy bodies in PD and DLB [324], however, these findings were not replicated in subsequent studies, and should therefore be interpreted with caution.

Given the involvement of *MOBP* in common and rare neurodegenerative diseases, understanding shared mechanisms involving this gene in disease processes could give crucial

insight into common pathways that lead to neurodegeneration and provide targets for therapeutic intervention in a broader range of diseases. The aim of this work was to elucidate in as much detail as possible how this gene is dysregulated across several neurodegenerative diseases through investigations of genetic, epigenetic and transcriptomic data.

Figure 6.2: *MOBP* DNA Methylation, RNA Expression, and Immunohistochemistry key findings in MSA and Control Groups



A) Correlations of top differentially methylated methylation sites mapping to *MOBP* and gene expression in cerebellar white matter in MSA. B) Immunoreactivity of MOBP in healthy controls (N = 5) and MSA patients (N = 5): In controls, MOBP immunolabeling was restricted to myelin sheaths (asterisk), with no staining in the oligodendroglial cytoplasm (white arrows). In MSA patients, glial cytoplasmic inclusions (GCIs) showed strong MOBP immunopositivity (white arrowheads). C) Expression of 23.3 kDa (isoforms b and a) and D) 9.5 kDa (isoform c) bands, in CTRL, MSA, PD, PSP and HD. E) Analysis of the ratio between levels of 20.8–23.3 kDa (isoforms *b* and *a*) over 9.5 kDa (isoform *c*). *MOBP;* myelin-associated oligodendrocyte protein. MSA: multiple system atrophy, PSP:progressive supranuclear palsy, PD: Parkinson's Disease, HD: Huntington's Disease, CTRL: control. Adapted from Bettencourt et al[129]

6.2 Methods

6.2.1 Colocalization Analysis

6.2.1.i Datasets Used

To investigate the *MOBP* locus across neurodegenerative diseases, we carried out colocalization analysis. We first did this by using available GWAS summary statistics from multiple studies, as summarised/outlined in **Table 6.2**. GWAS summary statistics were kindly passed to us by Dr Jack Humphrey (Icahn School of Medicine at Mount Sinai), who had collected and processed the data.

Table 6.2. Description of the disease GWAS used to investigate genetic associations at *MOBP* across neurodegenerative diseases in this Chapter.

Disease	Sample Size	Reference
PSP	2,779 cases (2,595 neuropathologically-confirmed), 5,584 controls	[79]
FTD	4,685 cases, 15,308 controls	[325]
ALS	29,612 cases, 122,656 controls	[326]
LBD	2,981 cases, 4,391 controls	[327]
MSA	888 cases, 7,128 controls	[21]
AD	111,326 clinically diagnosed/'proxy' cases, 677,663 controls	[11]
CJD	4,110 cases, 13,569 controls	[328]
PD	37,700 cases, 18,600 proxy cases, 1,400,000 controls	[329]

PSP: progressive supranuclear palsy, FTD: frontotemporal dementia, ALS: amyotrophic lateral sclerosis, LBD: Lewy body dementia, MSA: multiple system atrophy, AD: Alzheimer's disease, CJD: Creutzfeldt–Jakob disease, PD: Parkinson's disease.

Summary statistics from methylation quantitative trait loci (mQTL) analysis were downloaded from https://mostafavilab.stat.ubc.ca/xqtl/ (N = 534). As described by Ng et al. [330], we use a

significance value of $p < 5x10^{-9}$ for SNP-CpG pairs, given the number of pairings between SNPs and CpGs (measured across 420,103 CpG sites). OLG specific expression quantitative trait loci (eQTL) summary statistics were made available by Bryois et al. [331] and downloaded from https://zenodo.org/records/7276971. For initial analyses, we used a region comprised between 1MB upstream and downstream of the coordinates *of MOBP* (GRCh37 Chromosome 3: 39,508,689-39,570,970) to investigate association between disease QTLs, mQTLs and eQTLs.

6.2.1.ii Computational colocalization analysis and visualisation

We utilised the R package COLOC [332]. COLOC assesses whether two traits (e.g. mQTL or eQTL and a disease GWAS signal) share a common causal genetic variant within a specific genomic region using summary statistics. COLOC uses a Bayesian approach to estimate posterior probabilities of five hypothesis:

- H0 - No association with either trait
- H1 - Association with trait 1 only
- H2 - Association with trait 2 only
- H3 - Association with both traits, but under different causal variants
- H4 - Association with both traits, under the same causal variant - i.e. colocalization

Inputs required to COLOC are summary statistics including Effect sizes (β), standard errors (SE) and Minor allele frequencies (MAF). For consistency across analyses, we chose to use the MAFs from the most recent GWAS included in our analysis, e.g. the PSP GWAS published by Farrell et al. [79]. Also required as an input are prior probabilities for trait 1, trait 2 and colocalization. Prior probabilities represent assumptions about the likelihoods of certain hypotheses, p_1 and p_2 being the prior probabilities that a SNP is associated with traits 1 or 2 respectively, and p_{12} being the prior probability that a SNP is associated with both traits. For

COLOC, small prior probabilities are considered due to a small fraction of SNPs expected to be causal for any given trait (p_1 and p_2). The default values of prior probabilities for p_1, p_2 and p_{12} are p_1=1×10^{-4}, p_2 = 1×10^{-4} and p_{12} = 1×10^{-6}. However, these can be adjusted based on the analysis being undertaken. As there was prior evidence to suggest a strong genetic association between the *MOBP* locus and diseases (as described above), we increased the size of these prior probabilities in order to reflect existing knowledge. The prior probabilities we considered after such relaxation were p_1 and p_2 = 1 x10^{-3}, and p_{12} = 1x10^{-5}. Once prior probabilities have been decided upon, COLOC utilises Approximate Bayes Factor (ABF) computation to evaluate the evidence for association, where ABF for each SNP is defined as

$$ABF = \sqrt{1 - r} \times exp\left(\frac{Z^2}{2} \times r\right)$$

Where $Z = \frac{\hat{\beta}}{\sqrt{V}}$. r is the ratio of the variance of the prior and total variance ($r = W/(V + W)$). *W* as the prior variance of the effect size reflects prior beliefs about the expected size distribution of the data, and *V* is the observed variance derived from the data. This ABF formula is used to calculate evidence for each SNP, taking into account both prior expectations and observations within the data. *ABF*s are then combined to calculate likelihoods for the data under each hypothesis;

- H0: no association with either trait - *P* (*Data* | *H0*) = 1
- H1: Association with trait 1 only - *P* (*Data* | *H1*) = $\sum_i \quad ABF_{1i}$ (The sum of all the ABFs across all SNPs for trait 1)
- H2: Association with trait 2 only - *P* (*Data* | *H2*) = $\sum_i \quad ABF_{2i}$ (The sum of all the ABFs across all SNPs for trait 2)

- H3: Both traits are associated, but with distinct causal variants - *P* (*Data* | *H3*)

  = $\sum_i \quad ABF_{1i} \times ABF_{2j}$ (The joint likelihood across all SNP pairs (e.g *i* and *j*), but assuming different causal variants for each trait)

- H4: Both traits are associated with the same causal variant - *P* (*Data* | *H4*) =

  $\sum_i \quad ABF_{1i} \times ABF_{2i}$ (Assuming SNP *i* is causal for both traits, therefore products of ABFs for traits 1-N are summed across all SNPs (N).

Prior probabilities are then incorporated as:

$$P(H_k| \text{Data}) = P(Data | H_k) * P(H_k)$$

Where, *P* (H_k) are defined as *P (H0)*=1−(p_1+p_2−p_{12}), *P(H1)*=p_1, *P (H1)* = p_1, , *P (H2)*=p_2, *P (H3)*=p_1*p_2 and *P (H4)*=p_{12}.

Final posterior probabilities are then calculated by normalising these values to each other so that the posterior probabilities sum to 1 across all 5 hypotheses.

Colocalisation results were plotted using the package locuscompare [333]. For transcript plotting, we used the package ggtranscript [334].

6.2.2 Genotyping of *MOBP* rs1768208

At the time of this analysis, several SNPs within *MOBP* had been shown to have associations with neurodegenerative diseases. rs1768208 was described by several studies to be the potential lead SNP (**Table 5.1**), we therefore decided to investigate this SNP. Although more recent findings have indicated that this may not be the lead SNP [79], rs1768208 is known to be in very high LD with SNPs in the surrounding region, meaning that even if it is not the causal

variant, it can still act as a reliable proxy for investigating the effect of genetic variation on DNA methylation within this region. rs1768208 is an intronic variant at Chromosome 3:39,508,968 (GRCh37).

The region of interest was amplified by polymerase chain reaction (PCR) using flanking primers (forward primer 5'-TCC TCT CAA GCC TCA AAC TCT C'3'; reverse primer 5'-GGC AAC TCA GCC CAG AAA TTT G-3'), which were designed using Primer3 software[153]. A PCR mix was made of 7.5µl of Promega GoTaq® Green Master Mix, 2.25µl of forward primer (3 µM), 2.25µl of reverse primer (3 µM), 2µl template DNA (~100 ng) and 2µl dH_2O per sample. The following PCR conditions were used:

94°C for 5 minutes

10 cycles of:

- 95°C for 30 seconds
- 72°C for 60 seconds

25 cycles of:

- 95°C for 30 seconds
- 62°C for 30 seconds
- 72°C for 30 seconds

72°C for 5 minutes

Hold at 4°C

PCR amplification was verified in all cases by visualisation in agarose gel electrophoresis. PCR products were then purified using a "home-made" enzyme mix (1ml of "ExoSap-IT" was made of 50 µl of exonuclease I [Thermo Fisher Scientific, USA], 200 µl of Fast-AP alkaline phosphatase [Thermo Fisher Scientific, USA] and 750 µl of autoclaved dH_2O) and the following conditions:

37 °C for 30 minutes

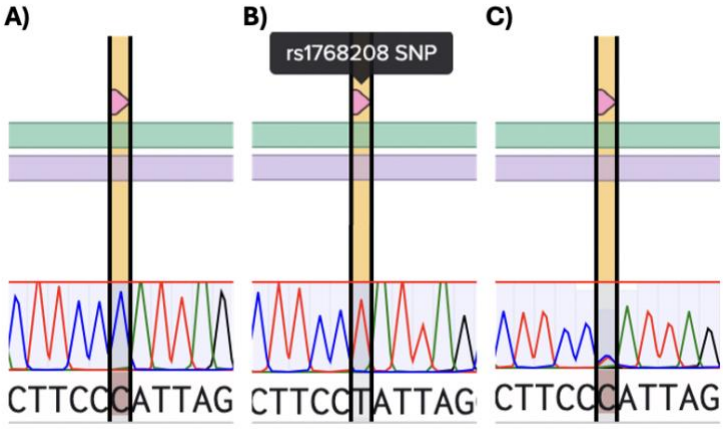
80 °C for 15 minutes

hold at 4 °C.

Sanger sequencing using the Forward primer (5'-TCC TCT CAA GCC TCA AAC TCT C'3') was performed by Source Biosciences. The sequence traces were visualised with Benchling ([Biology Software],2022)), and the rs1768208 genotypes determined as exemplified in **Figure 6.3.**

We genotyped DNA from MSA (*N* = 121), PD (*N* = 15) and PSP samples (*N* =15) as well as control samples (*N* = 89).

Figure 6.3. Possible genotypes at rs1769208 from Sanger sequencing results visualised with Benchling Software



A) Homozygous (CC). B) Homozygous (TT). C) Heterozygous (CT).

6.2.2 DNA methylation cohorts used for the genotyping of *MOBP* rs1768208

To investigate the effect of SNP rs1768208 on DNA methylation patterns, we used DNA methylation data previously generated by our group corresponding to overlapping samples that

were genotyped for this SNP (**Table 6.3**). Dr Megha Murthy carried out DNA methylation profiling on frontal lobe white matter from CTRL, MSA, PD, and PSP samples [116]. Dr Murthy also carried out loading, quality control and analysis of data, passing us the adjusted M-values which we then used to investigate effects of this SNP presence/absence of T alleles.

Table 6.3 Characteristics of the MSA, PD, and PSP frontal lobe white matter cohort

Disease	**Sample Number**	**Mean age (SD)**	**Sex**
CTRL	17	72.35 ± 4.47	8M/9F
MSA	17	67.71 ± 6.47	9M/8M
PD	17	68.06 ± 3.67	8M/9F
PSP	17	65.35 ± 3.75	9M/8F

Characteristics of samples used by Dr Megha Murthy to investigate DNA methylation changes in frontal lobe white matter tissue of CTRL, MSA, PD and PSP samples. *CTRL* controls, *MSA:* multiple system atrophy, *PD:* Parkinson's disease, *PSP:* progressive supranuclear palsy, SD: standard deviation, M; males, F; females.

6.3 Results

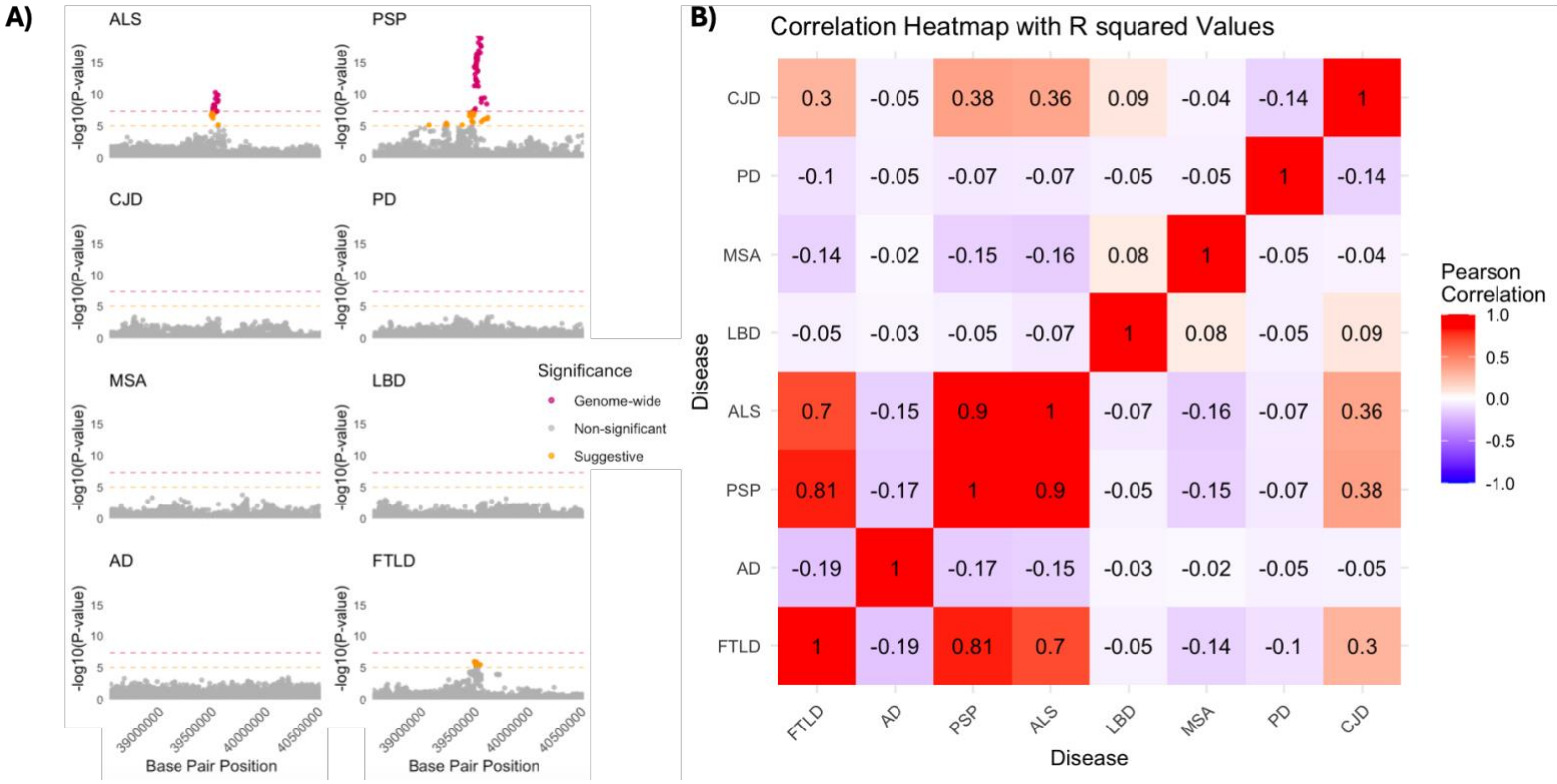
6.3.1 Genetic analyses of *MOBP*

6.3.1ii Neurodegenerative diseases GWAS colocalization with the *MOBP* locus

With the aim of gaining a preliminary understanding of genetic associations at *MOBP* across neurodegenerative diseases, we examined the *MOBP* locus and surrounding region, including 1MB upstream and downstream from the *MOBP* locus (as described in methods) using summary statistics from GWAS studies for the following diseases: AD, PSP, FTLD, ALS, LBD, PD, MSA and CJD (**Table 6.1**). Manhattan plots were utilized to assess the distribution and significance of SNPs across this region for each disease, with both genome-wide significance ($P < 5x10^{-8}$) and suggested significance ($P < 1x10^{-5}$) being taken into account (**Figure 6.4.A**). For ALS and PSP, there was a clear peak, with multiple SNPs passing both genome-wide and suggestive significance (**Figure 6.4 A**). For FTLD, there were several SNPs that passed the suggested significance threshold, but none that were genome-wide significant, possibly contributed to by small sample sizes and heterogeneous cohorts. No other disease GWAS showed SNPs passing significance threshold in this locus. Next we aimed to investigate genetic colocalization of associations within the *MOBP* region across distinct neurodegenerative diseases. Colocalization analysis is an approach used to determine whether two diseases share the same risk causing genetic variant, in this case within the *MOBP* locus. We carried out a simple correlation analysis using summary statistics significance values for SNPs within the *MOBP* locus and surrounding region (Figure 6.4). Using this approach, we aimed to identify which or if any of the diseases showed a shared genetic signature within this region. The result

of a correlation gives an R^2 value which was used to assess how correlated *p*-values are. If two GWAS show strong correlation, it is likely that there is/are shared risk causing variant/s. We found that ALS, FTLD and PSP all showed strong correlations (R^2 ranging between 0.7 and 0.9) pairwise (**Figure 6.4 B**). There was moderate correlation between CJD and ALS, FTLD and PSP (ranging between 0.3 and 0.38) (**Figure 6.4 B**) indicating that there may be some shared genetic signature, but which appears to be weaker than between ALS, FTLD and PSP. There was very little correlation between other diseases (**Figure 6.4 B**).

Figure 6.4 Correlation heatmap of GWAS p-values across neurodegenerative diseases.



A) Manhattan plots of GWAS results for MOBP (+/- 1MB) locus across neurodegenerative diseases: ALS, PSP, CJD, PD, MSA, LBD, AD, and FTLD. Each plot shows the -log10(P-value) of variants within the *MOBP* locus on chromosome 3. Dashed pink line represents the genome-wide significance threshold ($p > 5 \times 10^{-8}$), dashed orange line indicates the suggestive significance threshold ($p < 1 \times 10^{-5}$). SNPs are coloured based on significance: **pink** for genome-wide significant, **orange** for suggestive significant, and **grey** for non-significant. B) Heatmap visualising the Pearson correlation coefficients (r) between -log10-transformed GWAS p-values across neurodegenerative diseases at *MOBP*. Each cell represents the correlation between two diseases based on their GWAS summary statistics at this locus. ALS; amyotrophic lateral sclerosis, PSP; progressive supranuclear palsy, CJD; Creutzfeldt-Jakob disease, PD; Parkinson's disease, MSA; multiple system atrophy, LBD; Lewy body dementia, AD; Alzheimer's disease, and FTLD; frontotemporal lobar degeneration, GWAS; Genome-wide association study

Given that these three diseases showed significant genetic associations at this locus, we prioritised investigation of PSP, ALS and FTLD from this point on. We looked at which SNPs were the 'top' SNPs (i.e. most significant) in each of these three disease GWAS (**Table 6.4**).

Notably, the top four SNPs from the ALS and PSP GWAS were the same (albeit in a slightly different ranking), strengthening the finding of a shared genetic contribution to these diseases. While the top SNPs associated with PSP and ALS also show moderate association with FTD, their rankings are notably lower (e.g., rs631312 ranked 168 in FTD vs. 1 in PSP and ALS). The association of the top SNP in the FTLD GWAS at *MOBP*, rs2018725, which ranked lower in PSP and ALS, although in the PSP dataset, the association of this SNP was still genome-wide significant. This could indicate the possibility of a different causal variant driving the association in FTLD, but within a similar region.

Table 6.4 Top SNPs in PSP, ALS and FTLD GWAS data at *MOBP*

SNP	PSP p-value[79]	PSP rank	ALS p-value [82]	ALS rank	FTLD p-value[325]	FTLD rank
rs631312	4.60E-20	1	5.24E-11	1	0.0001613	168
rs1768208	1.10E-19	2	1.80E-10	3	0.0005135	173
rs616147	1.13E-19	3	1.32E-10	2	0.0003654	172
rs1768190	5.51E-19	4	5.03E-10	4	6.01E-05	65
rs545397	1.34E-18	5	1.59E-09	7	0.000275	171
rs1708104	2.11E-18	7	9.06E-10	5	0.0001368	165
rs2018725	4.83E-15	43	1.62E-07	53	1.12E-06	1
rs13081054	4.98E-15	44	1.67E-07	61	1.12E-06	1
rs12495185	5.01E-15	45	1.66E-07	59	1.12E-06	1
rs11129832	5.03E-15	46	1.66E-07	58	1.12E-06	1
rs1009966	5.33E-15	48	1.61E-07	52	1.12E-06	1

For each SNP, the table lists the p-values and ranks across all three datasets. SNPs that appear in the top five of multiple datasets are consolidated into a single row. PSP; progressive supranuclear palsy, ALS; amyotrophic lateral sclerosis, FTD; frontotemporal dementia, SNP; single nucleotide polymorphism.

Next, we aimed to investigate in more detail the relationship of the genetics at *MOBP* between FTLD, ALS and PSP, given that we had seen preliminary strong correlation of significant SNPs between these diseases. We utilised the package COLOC, which allows for the analysis of colocalization between genetic traits by estimating the posterior probability that a single causal variant is shared between traits, several probabilities are provided; PP.H0: Probability that neither disease is associated with any SNPs at *MOBP*, PP.H1: Only the first disease is associated, PP.H2: Only the second disease is associated. PP.H3: Both diseases are associated, but with different causal variants. PP.H4: Both diseases are associated and share the same causal variant. We found that between ALS and PSP, there was a very high probability (99.99%) that these diseases share the same causal variant at *MOBP* (**Table 6.5**). Between ALS and FTD, and PSP and FTD, there was a lower probability, 54% and 46%, respectively. For the ALS and FTD, and PSP and FTD, there was also a 45% and 53% probability, respectively, for PP.H3. Colocalization results for these comparisons are visualised in **Figure 6.5 A**.

Table 6.5. Colocalization analysis between the PSP, ALS and FTLD GWAS at the *MOBP* locus using the package COLOC.

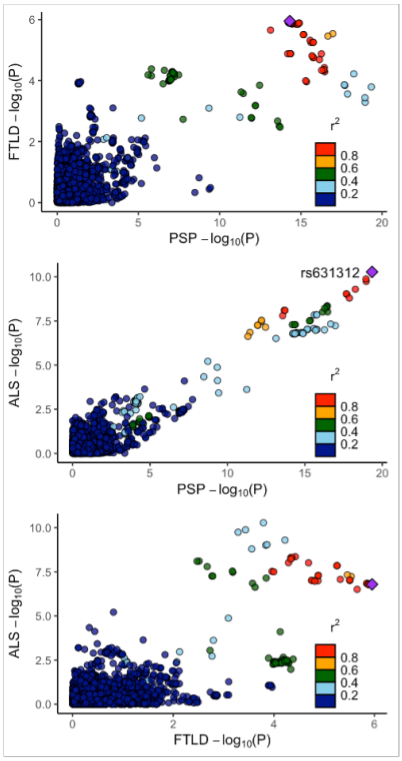
	PSP-ALS Colocalization	**ALS-FTLD Colocalization**	**PSP-FTD Colocalization**
PP.H0.abf	~0 %	~0 %	~0 %
PP.H1.abf	~0 %	1 %	1 %
PP.H2.abf	~0 %	~0 %	~0 %
PP.H3.abf	~0 %	45 %	53 %

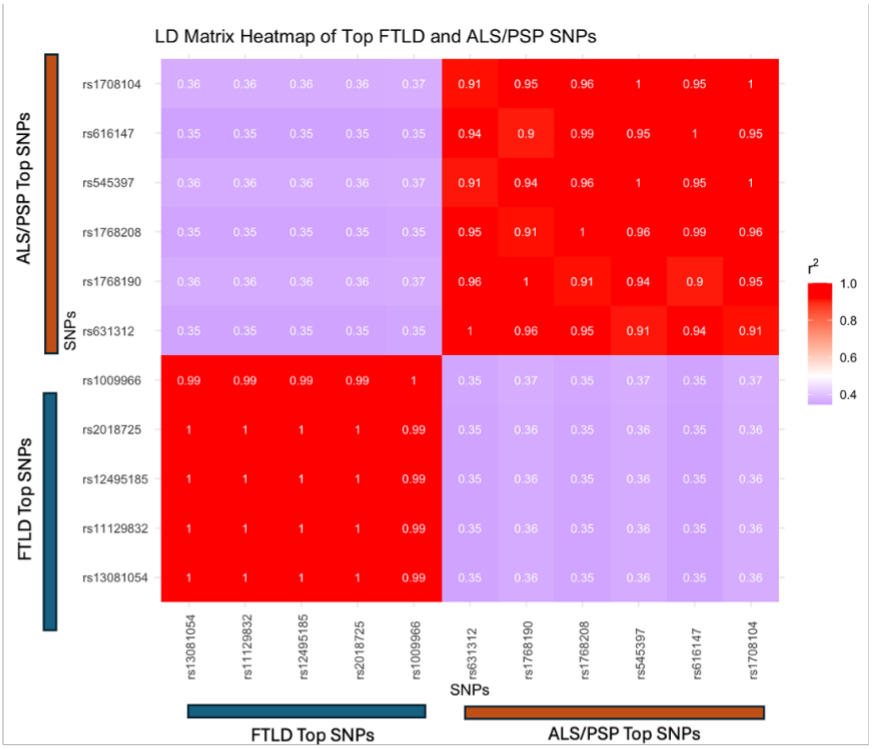
PP.H4.abf	~ 99.99 %	54 %	46 %

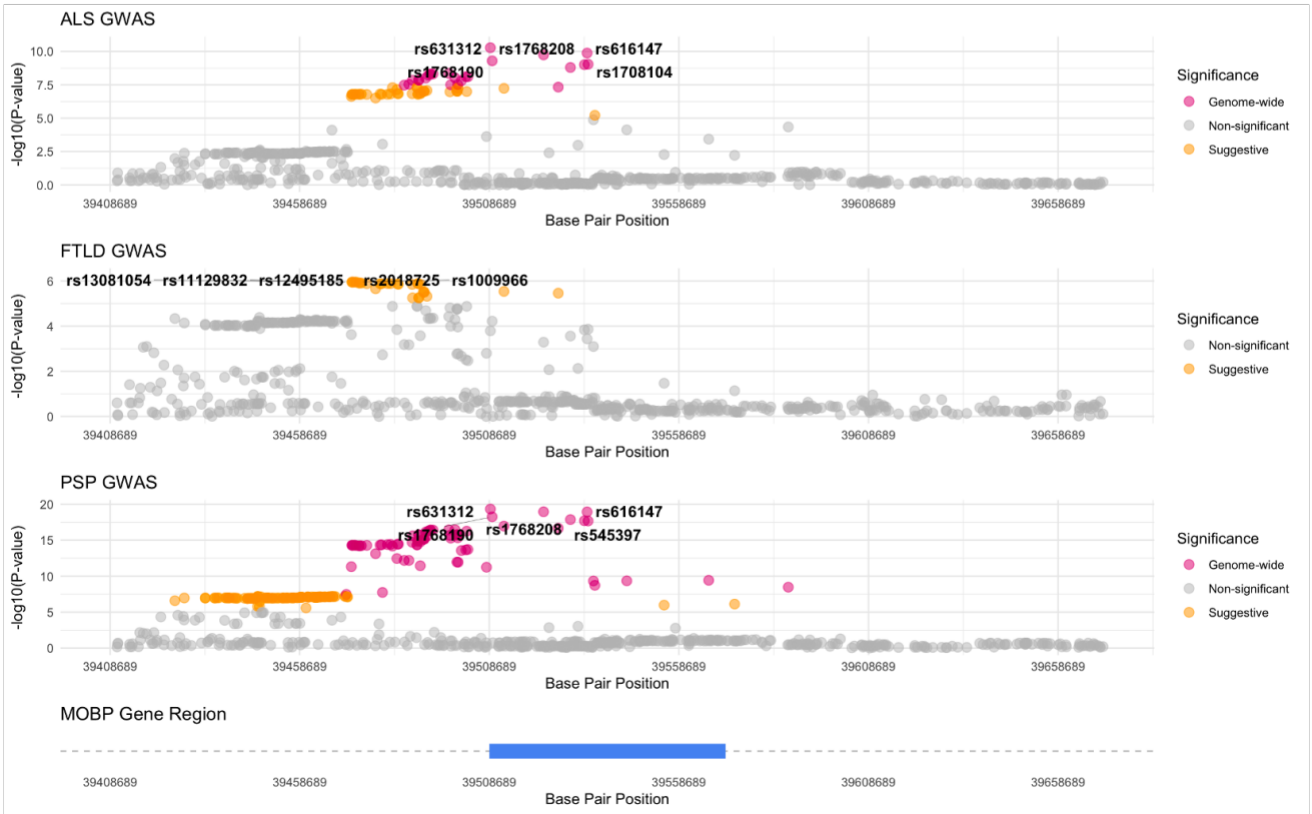
Table displaying results from COLOC between PSP, ALS and FTLD. The probability of the following scenarios are displayed for each combination - PP.H0: Probability that neither disease is associated with any SNPs at *MOBP*, PP.H1: Only the first disease is associated, PP.H2: Only the second disease is associated. PP.H3: Both diseases are associated, but with different causal variants. PP.H4**:** Both diseases are associated and share the same causal variant.For PP.H1.abf and PP.H2.abf in each column, the 'first' disease described refers to the first in each column, e.g. for column 'PSP-ALS Colocalization', PP.H1.abf refers to the probability that only PSP is associated with genetic variation at *MOBP.* PSP: Progressive Supranuclear Palsy, ALS: Amyotrophic Lateral Sclerosis, FTLD: Frontotemporal Lobar Degeneration, ABF; approximate bayes factor, PP; posterior probability.

To further understand the underlying genetic architecture of the potential shared association, we also analysed linkage disequilibrium (LD) of our top SNPs in the three diseases. LD is a phenomenon that occurs due to non-random segregation of alleles as a result of close-proximity [335]. LD can be measured through calculating an R^2 correlation coefficient between two loci, this allows quantification of how well a genotype at one locus predicts the genotype at another locus with values close to 1 indicating strong LD. We investigated LD at the top 5 *MOBP* SNPs within the ALS, FTD and PSP GWAS. Interestingly, although the top 5 associated SNPs for PSP and ALS had very high LD (R^2 between 0.91 and 1), indicating that they are within a haplotype block, these top PSP/ALS SNPs had low LD with the top FTLD SNPs (R^2 between 0.35 and 0.37) (**Figure 6.5.B**). This indicates that FTLD risk variants at the *MOBP* locus may reside in a distinct haplotype block. Visually inspecting the positions within the *MOBP* region of the top SNPs for the three datasets, we show that although the top SNPs within the ALS and PSP overlap with the start and gene body of *MOBP* (**Figure 6.5.C**), the (suggestively) significant FTLD SNPs sit upstream of the coding region gene. Together with the results from COLOC, these results support disease-specific causal variants at *MOBP* for FTD versus ALS/PSP.

Figure 6.5 Colocalization analysis between the PSP, ALS and FTD GWAS at the *MOBP* locus using the package COLOC







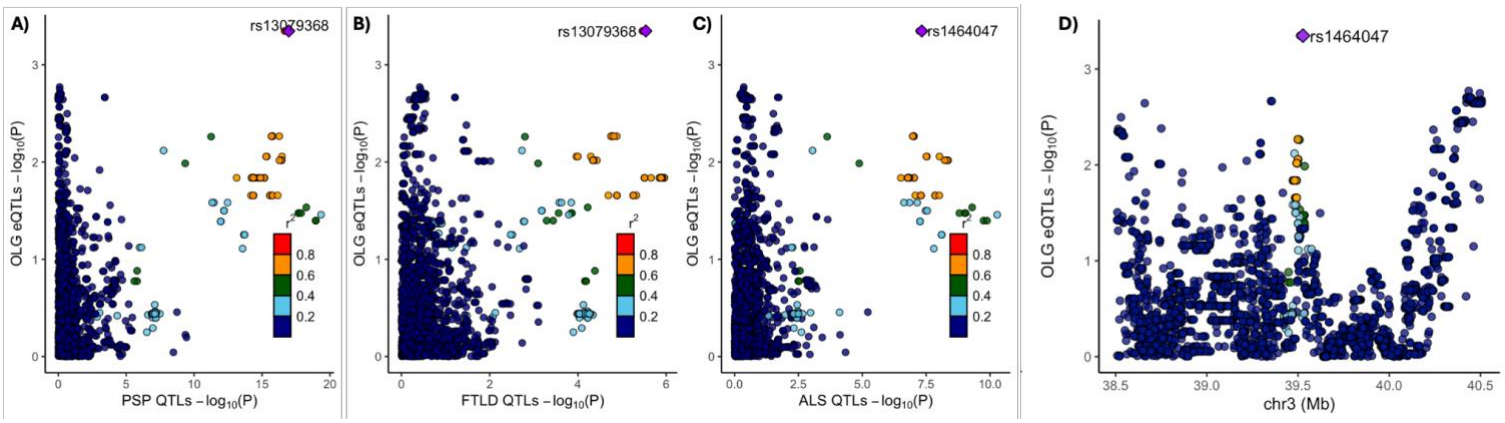
(A-C) Pairwise colocalization scatter plots showing the association between GWAS -log10(p-values) for (A) FTD and PSP, (B) ALS and PSP, and (C) ALS and FTD. The points represent individual SNPs, coloured based on their linkage disequilibrium (R^2) with the lead SNP in the region. (D) Manhattan plots displaying the -log10(p-values) for SNPs associated with PSP, FTLD, and ALS in the *MOBP* locus on chromosome 3. The x-axis represents the genomic location, while the y-axis shows the strength of the association. Genome-wide and suggestive SNPs are annotated and highlighted in pink and orange respectively. The region displayed is a smaller region around the coding region *MOBP:* 39408689 -39670970, to aid visualisation of location of genetic variants. Below the manhattan plots is a plot showing the coding region of *MOBP:* 39505000 - 39571000. GWAS: genome-wide association study, SNP: single-nucleotide polymorphism, PSP: progressive supranuclear palsy, FTLD: frontotemporal lobar degeneration, ALS: amyotrophic lateral sclerosis, *MOBP:* myelin-associated oligodendrocyte protein.

6.3.1ii Colocalization between GWAS and eQTLs at the *MOBP* locus

Genetic variants can have multiple phenotypic effects, depending on multiple factors including the type of mutation and location within a gene. Effects include but are not limited to: protein sequence changes, splicing variation, histone modifications, chromatin accessibility, and, of particular interest to us, DNA methylation and gene expression. Expression QTL data enables us to investigate genomic loci that are directly associated with expression of a certain gene. Using publicly available eQTL data, we investigated colocalisation between OLG specific eQTLs [331] (derived from single-nuclei RNA sequencing data) and disease SNPs at *MOBP*, to see whether our disease associated SNPs across PSP, FTLD and ALS were also found to modulate expression of the gene. Again, we used the package COLOC to estimate posterior probabilities for the following scenarios for each of these three diseases; PP.H0.abf: probability that neither disease or gene expression trait is associated with any SNPs at *MOBP*, PP.H1.abf: only the disease is associated with SNPs at *MOBP*, PP.H2: only gene expression of *MOBP* is associated with SNPs within *MOBP*. PP.H3.abf: both disease and gene expression of *MOBP*

are associated with genetic variation, but with different causal variants, and PP.H4.abf: both the disease and *MOBP* expression are associated with genetic variation in the region, and they share the same causal variant. For PSP, ALS and FTLD respectively, there was a 57%, 50% and 72% probability that the disease associated SNPs and eQTLs at *MOBP* are associated with the same causal variant (**Table 6.6, Figure 6.6**). It was interesting that the dataset that had the highest probability that the causal variant was shared was the FTLD data, we previously saw that the most significant FTLD SNPs sit upstream of the *MOBP* transcription start site (therefore potentially within promoter/enhancer regions), whilst the top ALS/PSP SNPs lie (in general), more across the gene body.

Figure 6.6 Visualisation of colocalisation of eQTLs at *MOBP* and PSP, FTLD and ALS SNPs at *MOBP.*



OLG eQTLs with disease-associated QTLs for A) PSP, B) FTLD and C) ALS across *MOBP* D) Manhattan plot displaying OLG eQTL significance (−log10(P)) along chromosome 3 (chr3). SNPs are coloured according to their LD R^2 with the lead SNP.

Table 6.6. Colocalisation analysis between the PSP, ALS and FTLD GWAS with OLG eQTL data at the *MOBP* locus using the package COLOC.

	PSP-eQTL Colocalisation	ALS-eQTL Colocalisation	FTLD-eQTL Colocalisation
PP.H0.abf	0.0 %	0.0 %	0.1%
PP.H1.abf	16.0 %	18.8 %	10.3 %
PP.H2.abf	0.0 %	0.0 %	0.2 %
PP.H3.abf	26.7 %	31.4 %	17.3 %
PP.H4.abf	57.3 %	49.8 %	72.0 %

Table displaying results from COLOC between PSP, ALS and FTLD QTLs and OLG eQTLs. The probability of the following scenarios are displayed for each combination, **PP.H0.abf**: Probability that neither disease or gene expression trait is associated with any SNPs at *MOBP*, **PP**.**H1.abf**: Only the disease (PSP, ALS or FTLD, depending on column) is associated with SNPs at *MOBP*, **PP.H2**: Only gene expression of *MOBP* is associated with SNPs within *MOBP*. **PP.H3.abf:** Both disease (PSP, ALS or FTLD depending on column) and gene expression of *MOBP* are associated with genetic variation, but with different causal variants. **PP.H4.abf:** Both the disease and *MOBP* expression are associated with genetic variation in the region, and they share the same causal variant. PSP, progressive supranuclear palsy; FTLD, frontotemporal lobar degeneration; ALS, amyotrophic lateral sclerosis.

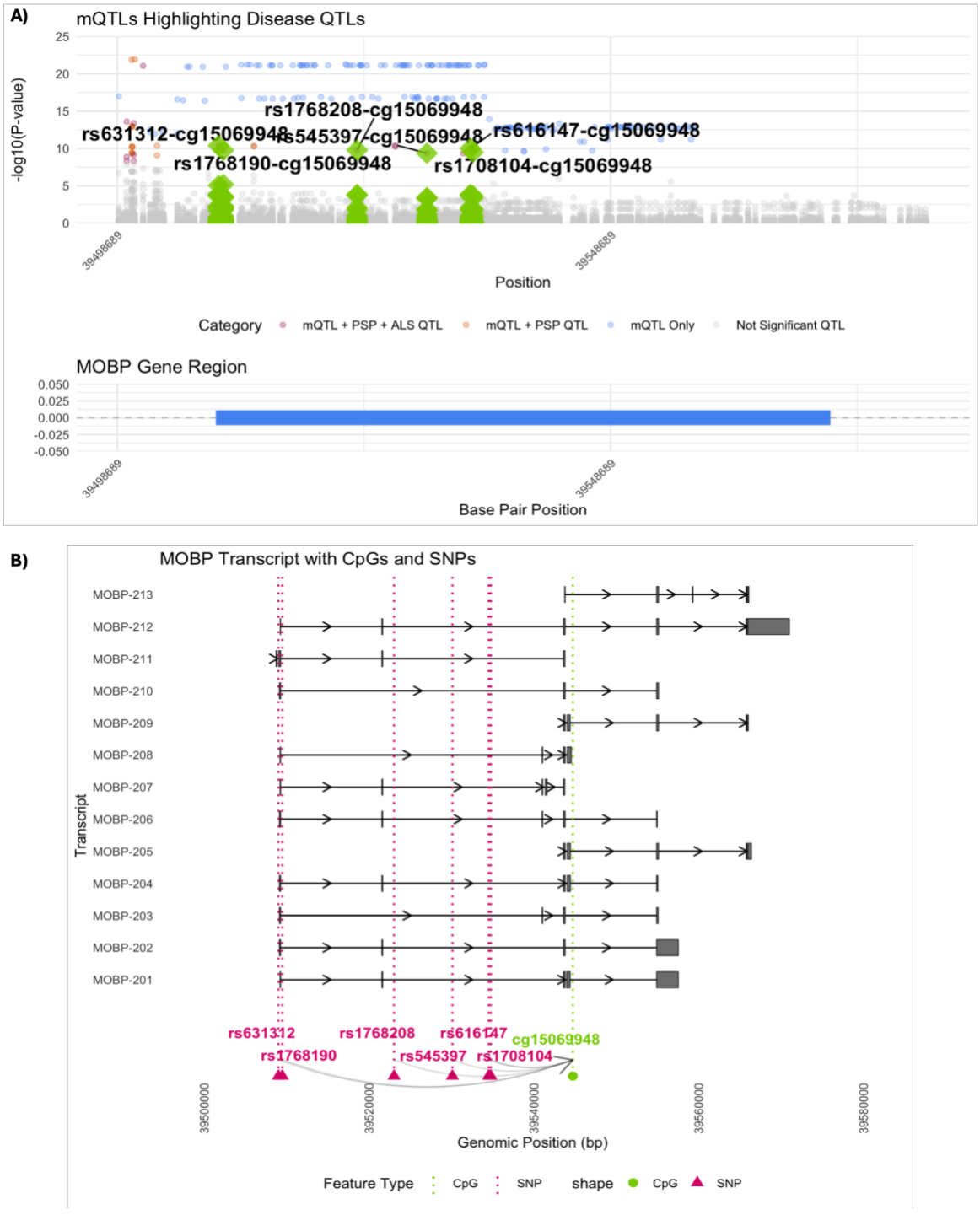
6.3.1.iii Analysis of disease GWAS and mQTLs at the *MOBP* locus

Next, we used mQTL data [330] to investigate if any of the disease associated SNPs at *MOBP* colocalized with mQTLs. We plotted mQTLs within the region spanning the *MOBP* locus and found that, in general, the most significant mQTLs that were also disease associated SNPs ($P < 5 \times 10^{-9}$ and $P < 5 \times 10^{-8}$ respectively for mQTLs and disease associated SNPs) lay upstream of coding region of the gene (**Appendix N**). However, these disease associated SNPs were not

the same SNPs as the SNPs of interest that we had identified as being shared between PSP and ALS and being the most significantly associated with disease.

Given that we had also identified a highly significant locus of genetic variation that was shared between ALS and PSP within the gene body, we therefore decided to investigate this region further. We examined mQTLs at the top 6 associated SNPs from the ALS and PSP GWAS. We found that these SNPs were also significant ($P < 5x10^{-9}$) mQTLs for one CpG - cg15069948, which we will refer to hence as the 'top CpG' (**Table 6.7, Figure 6.7.A**). We then sought to examine the manner of apparent regulation between the SNPs and the top CpG in terms of distance. We expected the associated CpG to be in close proximity to the region of genetic variation, but found that the apparent effect of genetic variation on methylation here spanned several Mb downstream of the SNPs (distances from 10kb to over 35kb). It was interesting to note that the CpG that is the significant mQTL for these SNPs lies within a region of the gene at the start site for multiple *MOBP* transcripts (**Figure 6.7.B**). This observation could implicate the presence of an alternative promoter/transcription start site at this region. We also looked at the effect of the SNPs at the level of methylation at this CpG. For all of the top 6 SNPs and the top CpG, there was an associated decrease in level of methylation (delta betas ranging between -0.265 and -0.278), indicating that the presence of the disease associated allele leads to a decrease in DNA methylation at this region. Interestingly, as previously described in Chapter 1, it is known that a decrease in DNA methylation at a promoter region of a gene is often associated with an increase in gene expression levels. If the genomic region containing this CpG does indeed act as an alternative promoter, we would therefore expect to see an increase in *MOBP* expression with the risk allele in these diseases. Farrell et al. recently reported that the risk SNP rs631312 was indeed associated with an increase in expression of *MOBP*, which is one of the top SNPs that is also an mQTL for this top CpG.

Figure 6.7 Position of mQTLs and disease significant SNPs at *MOBP*



A) Manhattan plot surrounding *MOBP* coding region which contains the “top SNPs” associated with ALS and PSP. Highlighted CpG-SNP pairs (e.g., rs1762808-cg15069948) are annotated with their identifiers. The bottom panel shows the *MOBP* gene region. CpG pairs are highlighted if the SNP is within our “top SNPs” associated with ALS and PSP. **B)** Schematic of *MOBP* transcripts with highlighted CpG sites and SNPs of interest. The diagram depicts *MOBP* transcript variants (MOBP-201 to MOBP-213), annotated with CpGs (green circles) and SNPs (dark pink triangles). Blocks indicate exons, whilst lines indicate intronic regions. Key SNP-CpG pairs of interest are shown with vertical dashed lines aligned to their genomic position. Only the 6 most significant SNPs from the ALS and PSP GWAS are included. ALS: amyotrophic lateral sclerosis, PSP: progressive supranuclear palsy, SNP: single-nucleotide polymorphism,

Table 6.7. SNP-CpG Associations in the *MOBP* Region.

SNP	SNP Position	A1	A2	CpG	CpG Position	Methylation mean difference	CpG P-value	PSP P-value	ALS P-value
rs1708104	Chr3:39534742	T	C	cg15069948	Chr3:39544697	-0.265	3.3E-10	2.1E-18	9.1E-10
rs1768190	Chr3:39509440	T	C	cg15069948	Chr3:39544697	-0.269	1.8E-10	5.5E-19	5.0E-10
rs1768208	Chr3:39523003	T	C	cg15069948	Chr3:39544697	-0.270	1.6E-10	1.1E-19	1.8E-10
rs545397	Chr3:39530083	T	C	cg15069948	Chr3:39544697	-0.264	4.3E-10	1.3E-18	1.6E-09
rs616147	Chr3:39534481	A	G	cg15069948	Chr3:39544697	-0.273	1.0E-10	1.1E-19	1.3E-10
rs631312	Chr3:39508968	G	A	cg15069948	Chr3:39544697	-0.278	4.0E-11	4.6E-20	5.2E-11

Table summarising key SNP-CpG associations identified in the MOBP region, highlighting their genomic positions, alleles, effect sizes, and statistical significance in mQTL and disease-specific analyses (PSP and ALS). SNP; The single nucleotide polymorphism identifier. Position: Genomic position of the SNP (GRCh37). A1/A2: Reference allele (A1) and alternative allele (A2). CpG: Identifier of the associated CpG site. CpG positions: Genomic position of the CpG site (GRCh37). Beta; Effect size of the SNP on CpG methylation levels in the mQTL analysis. P_value; p-value for the association between the SNP and CpG in the mQTL analysis. PSP_p_value; p-value for the association of the SNP with PSP, ALS_p_value; p-value for the association of the SNP with ALS.

Table 6.8 Colocalization of cg15069948, cg07405330 and cg25226092 mQTLs with ALS QTLs, PSP QTLs and *MOBP* eQTLs

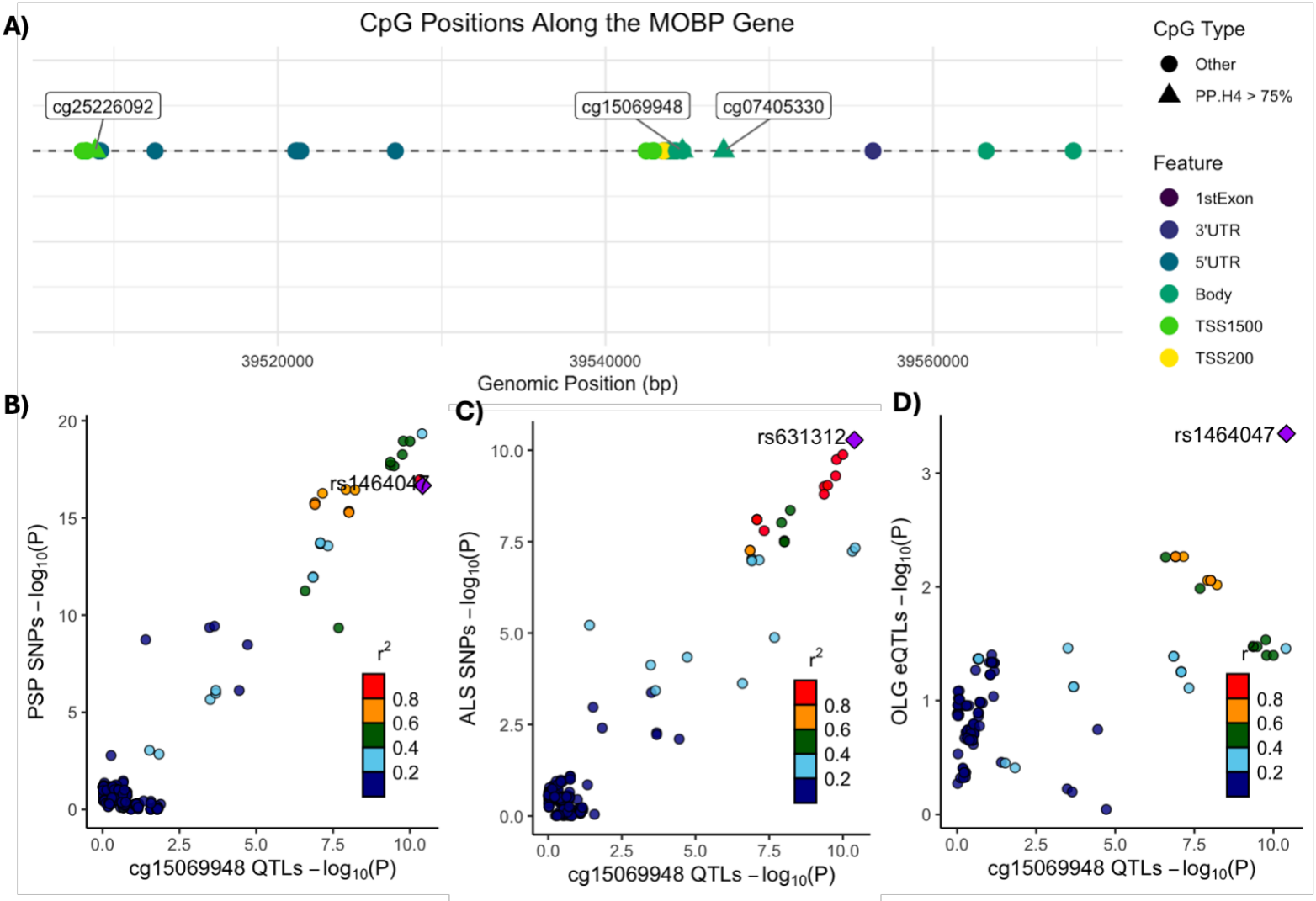
CpG	PP.H4.abf ALS	PP.H4.abf PSP	PP.H4.abf eQTL	Feature
cg15069948	99%	99%	62%	Body
cg07405330	90%	91%	4.2%	Body
cg25226092	89%	90%	7.04%	TSS1500

Table displaying results from COLOC between PSP, ALS and eQTLs and mQTLs of interest. For each combination, only **PP.H4.abf is shown ,** the posterior probability that both the disease/expression and DNA methylation are associated with genetic variation in the region, and they share the same causal variant. PSP, progressive supranuclear palsy; ALS, amyotrophic lateral sclerosis, mQTL; methylation quantitative trait loci, eQTL; expression quantitative trait loci, TSS1500; 200-1500bp from transcription start site.

To further establish shared causal variants, we performed colocalization analyses between mQTLs and PSP and ALS GWAS summary statistics. Remarkably, the same three CpGs showed strong colocalization (PP.H4 > 0.75) in both analyses (**Table 6.8, Figure 6.8 B,C)**. Among these, cg15069948 exhibited the highest posterior probability (PP.H4 = 0.99) for both PSP and ALS. This indicates a high likelihood that the mQTL and disease-associated SNP share the same causal variant, suggesting that the same genetic variant drives changes in DNA methylation and disease phenotype. Finally, we analysed colocalization between mQTLs and eQTLs at *MOBP* (**Table 6.8, Figure 5.8 D)**. The CpG that showed highest colocalization with the eQTL data was cg15069948, for which the PP.H4 was 62% (**Table 6.8**). Although this is a

moderate probability, it could indicate that genetic variation that is associated with disease (described above), could be affecting both DNA methylation and expression.

Figure 6.8 Position of CPGs showing high colocalization with PSP and ALS GWAS across *MOBP*



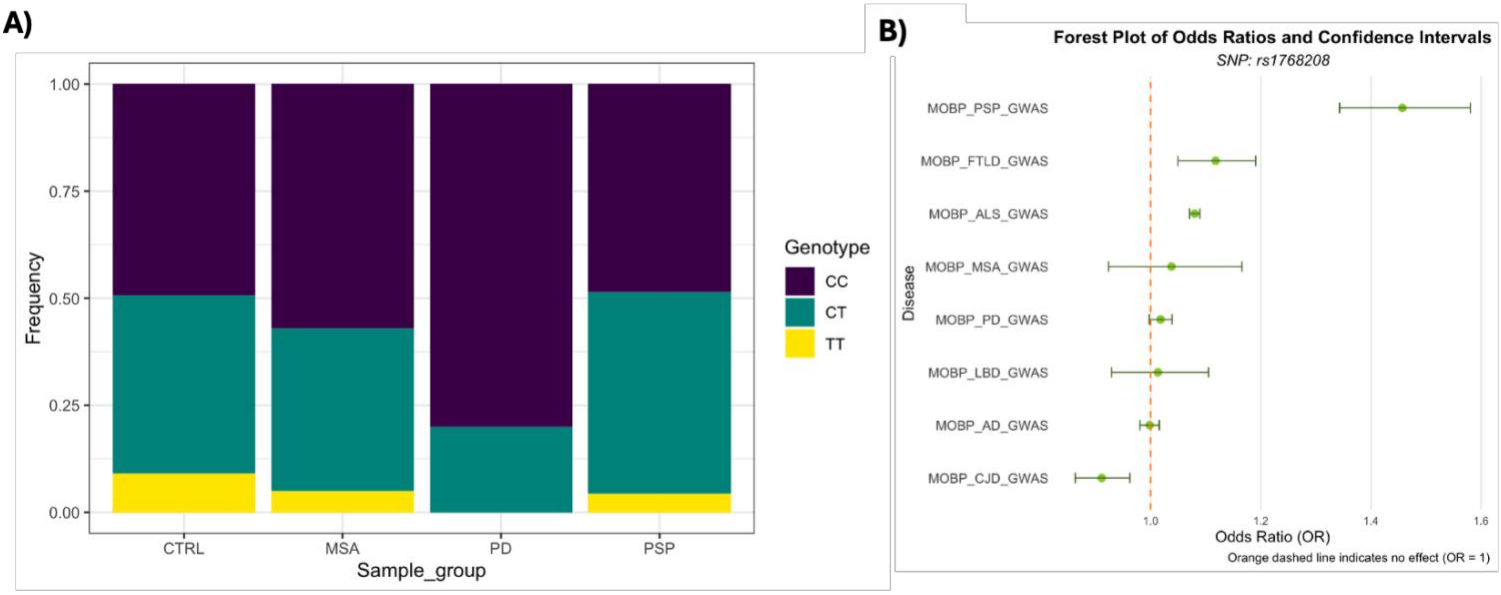
A) Schematic displaying the positions of CpG sites along the *MOBP* gene on the x-axis (genomic positions in base pairs) (GRCh37). Each CpG is represented by a point, with colours corresponding to their genomic features: 1stExon, 3'UTR, 5'UTR, Body, TSS1500, and TSS200. Points with a triangle shape represent CpGs with a posterior probability PP.H4 > 75% (those CpGs which show strong colocalization with disease-associated SNPs), while circles denote all other CpGs in the *MOBP* gene. B) Colocalization of cg15069948 QTLs with PSP SNPs. C) Colocalization of

cg15069948 QTLs with ALS SNPs. D) Colocalization of cg15069948 QTLs with OLG eQTLs. *MOBP:* myelin associated oligodendrocyte protein, QTL: quantitative trait loci, SNP: single-nucleotide polymorphism, PSP: progressive supranuclear palsy, ALS: amyotrophic lateral sclerosis.

6.3.3 Analysis of rs1768208 in pathologically confirmed MSA, PD and PSP cases and controls

To complement the analysis of genetic variation at *MOBP* and any associated DNA methylation variation across neurodegenerative diseases, we undertook genotyping of the SNP rs1768208 to investigate the frequency of the risk allele T in pathologically confirmed cases and controls from the Queen Square Brain Bank for neurological disorders (QSBB), for which we had DNA samples and in some cases associated DNA methylation profiles available (SNP selected as discussed in Methods, Section 6.2) (CTRL = 89, MSA = 121, PD = 15, PSP = 68). In our genotyping analysis of these movement disorders, we found no significant changes in allele frequency in MSA or PSP compared to controls (**Table 6.9)**. We did observe a significantly different allele frequency in the PD cohort, with the risk “T” allele being associated with a lower risk of PD (OR = 0.247, OR p = 0.047) (**Table 6.9**), however these results should be interpreted with caution due the very small sample size. The main purpose of generating these genotyping results was to enable us to investigate any effects of genotype on DNA methylation.

Figure 6.9. Genotype distribution of rs1768208 in QSBB samples and associated odds ratios in GWAS data from several neurodegenerative diseases



(A) Stacked bar plot showing the genotype distribution (CC, CT, TT) of rs1768208 across sample groups, including controls (CTRL), MSA, PD, and PSP QSBB cohorts (CTRL *N* = 89, MSA *N* = 121, PD *N* = 15, PSP *N* = 68) . (B) Forest plot displaying the odds ratios (ORs) and 95% confidence intervals (CIs) for rs1768208 in *MOBP* from PSP, FTLD, ALS, MSA, PD, LBD, AD, and CJD GWAS. The orange dashed line at OR = 1 indicates no effect, with points to the right representing increased disease risk associated with the T allele. ALS; amyotrophic lateral sclerosis, PSP; progressive supranuclear palsy, CJD; Creutzfeldt-Jakob disease, PD; Parkinson's disease, MSA: multiple system atrophy, LBD: Lewy body dementia, AD: Alzheimer's disease, FTLD: frontotemporal lobar degeneration, GWAS: Genome-wide association study, QSBB: Queen Square Brain Bank, OR: Odds Ratio, CI: confidence interval.

Table 6.9: Genotypes of SNP rs1768208 from a cohort of MSA, PD, PSP and control samples

Sample Group	Total Alleles	Frequency C	Frequency T	OR	OR P-value
CTRL	178	0.70	0.30	-	-
MSA	242	0.76	0.24	0.738	0.327

PD	30	0.90	0.10	0.247	0.047
PSP	136	0.72	0.28	1.037	1

Summary of allele counts, frequencies, and odds ratios (OR) for a genetic variant (rs1768208) across different sample groups. Sample Group: Groups analysed, including controls (CTRL) and disease cohorts (FTLD, MSA, PD, PSP). Total Alleles: Combined total of "C" and "T" alleles in each group, Frequency C: Proportion of the "C" allele within the total alleles. Frequency T: Proportion of the "T" allele within the total alleles, OR: Odds ratio comparing allele frequencies between each disease group and controls, OR P-value:Statistical significance of the odds ratio, FTLD: frontotemporal lobar degeneration, MSA: multiple system atrophy, PD: Parkinson's disease, PSP: progressive supranuclear palsy, CTRL: control.

Our group had previously produced genome-wide DNA methylation profiles of frontal lobe white matter for a subset of the cases genotyped across these three diseases subsequently genotyped, but had not looked in detail at the *MOBP* region. Therefore, we first investigated DNA methylation patterns across *MOBP* for these three diseases (regardless of genotype). With analysis of the data (preprocessing carried out by Dr Megha Murthy) [116], we observed several methylation sites differentially methylated between controls and disease. In MSA, 9/28 CpGs mapping to *MOBP* were nominally significantly differentially methylated ($p < 0.05$), 7 of which were in the promoter region of the gene (5'UTR, TSS1500 or TSS200), aligning with previous work carried out by Dr Bettencourt (described above in section 6.1), who described increased DNA methylation at the promoter region of the gene [118] (**Table 6.11**). Of the *MOBP* CpGs, 6/28 were nominally significantly differentially methylated in PD, again primarily within the promoter region. In PSP, only 3 methylation sites showed nominal significant differential methylation at *MOBP* (at the 5'UTR and 1st Exon) (**Table 6.11**).

Table 6.11. Differential methylation of methylation sites mapping to *MOBP* in frontal lobe white matter tissue in MSA, PD and PSP

		MSA vs Control		PD vs Control		PSP vs Control	
CpG	**Feature**	**Delta M**		**Delta M**	**P Value**	**Delta M**	**P Value**
cg23605644	TSS200	0.504	0.023	0.281		0.417	0.077
cg27272723	TSS1500	0.171		0.245	0.048	0.235	0.093
cg01684805	TSS1500	-0.259		-0.177		-0.261	0.054
cg21827971	TSS1500	0.256		0.221		0.114	0.268
cg22110662	TSS1500	0.339	0.008	0.304	0.011	0.190	0.154
cg21696316	5'UTR	0.109		-0.081		-0.169	
cg04442806	5'UTR	0.558		0.521	0.035	0.605	0.031
cg03054684	5'UTR	0.633		0.384		0.553	
cg14968361	5'UTR	0.811		0.629	0.014	0.691	0.017
cg05633900	1stExon	0.485		0.377		0.449	
cg07878407	1stExon	0.522	0.017	0.428	0.038	0.479	0.041

CpGs showing significant (P < 0.05) differential methylation in MSA, PD or PSP vs CTRL within *MOBP.* CTRL: Control, MSA: multiple system atrophy, PD: Parkinson's disease, PSP: progressive supranuclear palsy.

Next, we examined whether any CpGs within *MOBP* showed distinct patterns in samples with distinct rs1768208 genotypes. Given our modest sample size, we divided samples into "Positive" and "Negative", indicating the presence of the disease-associated T allele (i.e. either TT or TC genotypes for Positive, and CC for Negative). We found that none of the CpGs showed different patterns of DNA methylation when all samples (disease and controls) were grouped together (data not shown). There were three methylation sites which showed nominally significant differential methylation when stratified by disease and genotype (**Table 6.12)**. One methylation site, cg14968361, located in the 5'UTR, showed differential methylation both in PD and CTRLs, but with differing direction of effect - in CTRL samples, the presence of the T allele led to a decrease in DNA methylation, whilst in PD, an increase in DNA methylation was observed. This was one of 3 CpGs to be differentially methylated in MSA, PD and PSP. cg23603305, which is located in the promoter region of the gene (TSS1500), showed significant decrease in methylation in MSA T-positive cases only, and interestingly, was not one of the methylation sites to show differential methylation in any of the disease vs. CTRL analyses (**Table 6.11**). One CpGmethylation site was differentially methylated between genotype groups in PSP, cg15069948, was the same CpG identified above as being an mQTL with strong association with PSP and ALS, and having a moderate association with eQTLs in *MOBP* (**Table 6.12)**. Furthermore, the direction of effect - a decrease in DNA methylation being associated with the presence of the SNP T-risk allele, is in accordance with that described above. The finding of cg15069948 as linked to PSP through two distinct analyses highlights its biological relevance to linked genetic and epigenetic modulation of *MOBP.*

Table 6.12 Differential methylation of methylation sites mapping to *MOBP* in frontal lobe white matter tissue in MSA, PD and PSP stratified by genotype at rs1768208

		CTRL (Positive vs. Negative)		MSA Positive vs. Negative)		PD Positive vs. Negative)		PSP (Positive vs. Negative)	
Name	**feature**	**Delta M**	**P value**	**Delta M**	**P Value**	**Delta M**	**P Value**	**Delta M**	**P Value**
cg14968361	5'UTR	-0.4613	0.0488	0.1160	0.5876	0.5987	0.0282	-0.3717	0.0786
cg15069948	Body	-0.2093	0.0698	-0.0153	0.8851	0.1435	0.2794	-0.2942	0.0058
cg23603305	TSS1500	-0.1627	0.1868	-0.2300	0.0466	0.1468	0.3034	-0.0651	0.5576

Significantly differentially methylated methylation sites between carriers of the SNP rs1768208 stratified by genotype. Positive indicates individuals homozygous (TT) or heterozygous (CT) for the risk allele T at rs1768208, Negative indicates individuals homozygous (CC) at rs1768208. CTRL: Control, MSA: Multiple System Atrophy, PD: Parkinson's Disease, PSP: Progressive Supranuclear Palsy.

6.4 Discussion

Here, we have investigated genetic sequence and DNA methylation variation at the *MOBP* locus across neurodegenerative diseases, highlighting both shared and distinct mechanisms. Using GWAS summary statistics, we identified significant genetic associations at the *MOBP* locus with ALS, PSP, and FTLD, with ALS and PSP showing particularly strong overlap in top SNPs and high colocalization probabilities, suggesting shared causal variants between the two diseases. In contrast, FTLD-associated variants demonstrated weaker genetic correlation and distinct localisation upstream of the *MOBP* transcription start site, indicative of a distinct disease mechanism. mQTL analyses highlighted cg15069948 as a significant mQTL for shared ALS and PSP top SNPs, linking genetic variation to altered DNA methylation. These findings suggest that while ALS and PSP may share a common risk mechanism at *MOBP*, FTLD likely involves unique genetic and regulatory contributions within the same locus, underscoring the nuanced interplay of genetics and epigenetics in neurodegenerative disease risk at this locus.

The first thing we considered was the underlying proteinaceous inclusion of each disease, and whether this could offer an explanation as to the pattern of *MOBP* association across neurodegenerative diseases. ALS and PSP both have pathological links to tau, with PSP being classed as a primary tauopathy. Although ALS is classically defined by the presence of TDP-43 pathology, there is some evidence for altered tau metabolism in the disease [336]. We also saw a suggestively significant signal at the *MOBP* locus in FTLD, in a mixed FTLD-TDP and FTLD-tau cohort. Indeed, it has been reported that the T risk allele in PSP is associated with an increase in semi-quantitative scores of tau thread pathology and scores for phosphorylated tau-immunoreactive coiled bodies [337]. It is possible that the increased occurrence of the risk allele T in PSP cohorts is linked to tau pathology which is not replicated in α-synucleinopathies such as PD, MSA and LBD. However in the AD GWAS we saw no genetic signal at the *MOBP* locus,

thus, if the association was mediated through tau, we would expect to have seen a signal in the AD as another tauopathy. *MOBP* has been genetically linked to AD, but that link is thought to be mediated through the *APOE* genotype [81]. *APOE,* the strongest genetic risk factor for AD, has recently been linked to the function of OLGs in AD. Blanchard et al. investigated the effect of *APOE* genotype on post-mortem human OLGs, and found that OLGs derived from *APOE3/4* and *APOE4/4 (APOE4* being the AD risk associated allele) exhibited downregulation of key myelin-associated genes, regardless of AD status [70]. Furthermore, altered APOE function was associated with changes in cholesterol deposition in OLGs, which in turn led to a reduction in myelin [70]. The fact that genetic association of *MOBP* was only detected in *APOE4* carriers could lead to a hypothesis that the presence of genetic variation at *MOBP* exacerbates effects of OLG dysfunction, and that without the aberrant cholesterol function already associated with OLGs and APOE status, genetic variation may not have the same pathological impact. It would be interesting to investigate if there was any genetic association in AD *APOE4* carriers that was shared between PSP and ALS.

In a similar vein, it is possible that the observed potential genetic mediation of *MOBP* effects via *APOE* status could extend to other neurodegenerative diseases, offering some explanation for the patterns observed across diseases. All of these diseases are highly polygenic, involving contributions from numerous genetic variants. It is plausible that variants at *MOBP*, in combination with other disease-specific variants, could act synergistically to influence disease risk and progression. This interplay might help explain why *MOBP* genetic associations appear more pronounced in some diseases, such as PSP and ALS, compared to others like MSA and AD, where different genetic landscapes dominate.

We also considered whether the association was linked to the amount of OLG pathology reported in these diseases, as one of the main pathological hallmarks of PSP is cytoplasmic

inclusions in OLG presenting with coiled bodies [81,338338]. However, although there is evidence for OLG pathology in ALS (discussed in Chapter 1), loss of motor neurons is the major pathological hallmark. Furthermore, in MSA, the primary cell type affected is OLGs, and one would expect that if it was the case that the genetic association was linked to degree of OLG pathology, we would see a stronger genetic signal at *MOBP.*

It is also worth noting that whilst in PSP, there is an observed increase in expression of *MOBP* associated with the risk T allele, in MSA, *MOBP* mRNA levels were shown to be downregulated[87,129]. It is therefore possible that, although the gene is important across neurodegenerative disease pathology, its involvement in pathogenic mechanisms may be distinct between diseases. Furthermore, whilst in MSA, significant correlations between *MOBP* gene expression and DNA methylation at the promoter region have been observed[118,129], no similar correlations were found in a PSP cohort [85]. This further underscores the complexity of interactions between DNA methylation, genetic variants and gene expression, and demonstrates the need for deeper investigation to elucidate such potentially disease specific effects.

Whilst our analysis into *MOBP* across neurodegenerative diseases has provided important insights, there are of course some limitations to discuss. As with any analysis utilising publicly available data, we rely upon quality and completeness of datasets. Heterogeneity, mixed diagnosis and misclassification can all lead to GWAS results overestimating the importance of genetic signals that may not be disease relevant and underestimating important genetic loci. There are also limitations in our colocalization analysis, for example in the use of the COLOC package. Importantly, COLOC has the assumption of a single causal variant - i.e. that each region contains only one causal variant per trait. If this is not true, results may be misleading. SuSiE (Sum of Single Effects) is a computational approach that enables COLOC to include

multiple causal variants [339]. This could be interesting to apply this to the eQTL colocalization analysis, where we did not see strong colocalization, even though genetic variation at *MOBP* at SNPs discussed has been associated with altered expression of *MOBP* [79,85]

There is much promising future work to be carried out to continue this research. Farrell et al[79]. identify SNP rs631312 as the lead SNP driving association between *MOBP* and PSP, further work on fine mapping to identify if the same causal SNP is responsible across distinct diseases would be important. Detailed investigations into the relationship between genetics, DNA methylation and gene expression is also warranted. In ALS and PSP, we observe genetic associations at *MOBP* that are linked to changes in both gene expression and DNA methylation. However, additional computational modeling and functional studies are needed to clarify the causal directionality and potential synergy of these effects. For example, while the traditional assumption is that DNA methylation changes drive gene expression changes, it is also possible that these two processes are independent or influenced by a shared upstream mechanism. Elucidating these relationships could provide deeper insights into the complex regulatory dynamics at play in neurodegenerative diseases. It would also be interesting to explore the potential for a distinct alternative promoter region close to the region of the CpG that showed differential methylation with distinct genotypes in PSP (cg15069948), especially considering the finding of altered ratio of protein isoforms present in PSP compared to controls and other NDs [129]. Investigating transcript structure would also be important here. Although we looked at gene expression and genetics at *MOBP* through eQTL/disease QTL colocalization, it would be useful to use alternative transcript analysis, for example long-read sequencing which better allows for analysis of alternative splicing patterns or unannotated isoforms. Future work should also be the analysis of the neurodegenerative disorder corticobasal degeneration (CBD). CBD has been identified to share genetic associations at *MOBP* with PSP [80,340]. CBD is also a primary

tauopathy with significant white matter/OLG pathology [341]. Investigating colocalization between CBD associated loci at *MOBP* and other diseases, as we have carried out in this chapter between ALS, PSP and FTLD, would be interesting. Integrating DNA methylation and gene expression data with CBD genetics at *MOBP* could offer further elucidation as to why we see association at *MOBP* across some neurodegenerative diseases but not others.

In this work we have carried out analysis to investigate genetics, DNA methylation and transcriptomics at *MOBP* across neurodegenerative diseases, in order to further understanding of the involvement of this gene in pathology. By identifying both shared and distinct mechanisms, we have found that there appears to be a shared locus of genetic association within *MOBP* between ALS and PSP, which is also associated with DNA methylation. This work has provided key insights into the importance of *MOBP* in neurodegeneration and established a foundation for future research into this gene.

Chapter 7 - General discussion and conclusions

7.1 Overview

In this thesis we have analysed multiple DNA methylation and gene expression datasets, to attempt to obtain insights into the role of DNA methylation affecting OPCs and OLGs across neurodegenerative diseases. We have additionally used genetic data to understand the role of *MOBP,* as a gene which is highly relevant to OLGs and which has previously been associated with aberrant DNA methylation in neurodegeneration [118,129].

In Chapter 3, we report on EWAS analyses across numerous DNA methylation datasets and, where applicable, meta-analyses to allow us to report on robust findings from the EWAS analysis and to increase statistical power for additional identification of genes of interest. We used both DNA methylation datasets derived from bulk and sorted nuclei isolated from brain tissue and have found that several genes show consistent dysregulated DNA methylation patterns across FTLD and AD, many of which also show disturbance at the gene expression level. Some of these genes have previously been linked to altered myelin function/OLG lineage biology and/or disease mechanism. Here, we have provided evidence as to the potential role of DNA methylation contributing to such changes.

In Chapter 4, we report on a complementary analysis approach to further investigate DNA methylation perturbations in neurodegenerative disease relating to OLGs/OPCs. To reach that

goal, using a systems biology approach based on WGCNA, we have constructed co-methylation networks and investigated modules associated with neurodegeneration and enriched for OLG/OPC genes. We found several modules of interest and have investigated these for biological function, both through GO enrichment analysis and by detailed investigation of top genes in these modules. We also investigated preservation of these modules across distinct neurodegenerative diseases and brain regions. We observed generally high preservation of OLG/OPC enriched disease associated modules. This suggests that dysregulation of DNA methylation in OLGs may be common across the diseases we have analysed. Several genes, including by way of example *PIP4K2A,* that we identified in the EWAS analyses in Chapter 2 were found to be present in disease associated modules across FTLD. This finding implicates DNA methylation changes in the functioning of peroxisomes as relevant in neurodegeneration, which was strengthened through functional enrichment and identification of additional peroxisome related genes. We have identified other pathways of interest, including the *Wnt*/beta-catenin signalling pathway in OLGs/OPCs, known to be important in OPC differentiation, which warrants further investigation. We also identified co-methylation modules with hub genes that have high relevance to OLGs, including *MYRF* and *MOG*, which are known to be crucial modulators of myelination and OLG function, indicating that regulation of such crucial genes is modified in disease via changes in DNA methylation.

In Chapter 5, we have reinforced findings reported in Chapters 3 and 4 by leveraging an iPSC OLG cell differentiation model to investigate OLG/OPC genes which show distinct DNA methylation profiles throughout the OLG life cycle. In particular the gene *CTNNA1*, which was found as a hub gene in a co-methylation module associated with AD, showed significant and large changes in DNA methylation throughout OLG differentiation. As well as being of biological interest, this highlights the benefit of using multiple and complementary datasets and approaches in identifying disease relevant changes.

In Chapter 6, we report on a focused investigation into the genetic and epigenetic associations of the *MOBP* gene with neurodegeneration across a larger range of diseases. We chose to investigate this gene because of its genetic association with multiple neurodegenerative diseases and the previous findings within our group of disease-associated aberrant DNA methylation at the promoter region of the gene. Despite multiple lines of evidence pointing towards this gene as being important in neurodegeneration, studies involving the gene are limited. Chapter 6 sets out our attempt to understand how, in different neurodegenerative diseases, genetics and DNA methylation may be influential within the gene locus to lead to aberrant gene expression. We have demonstrated that there is strong genetic colocalisation of disease risk within this gene between ALS and PSP. We have also shown that such disease associated genetic variability is also influencing DNA methylation changes which are likely linked to altered *MOBP* expression, which has been observed in PSP [79] and MSA [342]. We have also demonstrated that, although *MOBP* is linked to multiple diseases, it appears as though underlying molecular alterations are distinct between diseases. These results highlight the complexity of risk associated SNPs at this gene locus, warranting further functional studies to dissect its role in neurodegeneration.

Overall we provide new evidence as of the importance of DNA methylation in the regulation of OLG/OPC genes across neurodegenerative diseases.

7.2 Issues arising

7.2.1 Implications of the findings

In work described throughout this thesis, we aimed to identify genes that are differentially methylated in disease and to investigate these genes in terms of OLG differentiation and gene expression changes. Through this, we have contributed to the field of study through identification of genes that may play a role in neurodegenerative disease pathogenesis and linked together findings from multiomics in order to identify biological pathways that play a role in disease, showing that DNA methylation is an important and still overlooked area in this context. There is, of course, much additional work to be done from this thesis in investigating further genes identified in terms of function (discussed further in Section 7.2.3). We have, in this work, identified genes that (as far as we know) have not previously been linked to DNA methylation changes, but have previously been shown to be important in disease pathogenesis, providing a potential functional role for DNA methylation and OLG lineage genes. We have also uncovered genes showing aberrant DNA methylation that, again to the best of our knowledge, have not previously been linked to neurodegenerative disease. Newly identified genes may represent hitherto unrecognised contributors to disease pathogenesis. In identifying both previously identified and novel disease candidates, we have broadened the understanding of OLGs in neurodegeneration.

As well as identifying disease associated genes and pathways to contribute to general understanding of disease pathogenesis, the study of DNA methylation in disease is important from a therapeutic perspective. As DNA methylation changes are reversible, this means that they are an attractive target for disease modifying treatments. Although not currently used as

therapeutic intervention strategies, several approaches have emerged to edit DNA methylation in specific genomic sites, see reviews by Kantor, Murthy, [104,343–345]. These approaches include the use of CRISPR-Cas9 fused to DNA methyltransferases which allow site specific manipulation of DNA methylation patterns [343]. In PD, elevated levels of *SNCA* are a hallmark, although normal expression levels of this gene are vital for normal cell physiology. Kantor et al. demonstrated that through CRISPR-Cas9 mediated targeting, mRNA and protein expression levels of *SNCA* were decreased, which increased cell function in an iPSC model. These examples provide promising evidence as to the efficacy of such approaches. It is possible that, with further investigations and biological validation of targets, genes such as *CTNNA1,* which we have found to be important in the differentiation of OLGs and dysregulated in disease, could be targeted to restore functionality of OLGs in neurodegeneration.

We have looked at DNA methylation changes occurring within brain tissue due to their relevance in understanding neurodegenerative disease associated mechanisms. Although beyond the scope of the work descibed in this thesis, DNA methylation changes are also known to be important in blood, hence their potential to be developed into blood-based biomarkers. Indeed, recent studies have demonstrated that there is significant correlation at some CpG sites between blood and brain [346]. Blood based biomarkers offer minimally invasive approaches for disease detection, as well as disease monitoring. Early detection of neurodegenerative diseases is crucial as it allows timely intervention. By the time patients present with clinical symptoms of neurodegeneration, significant and irreversible damage has already occurred. Therefore, timely detection and monitoring of disease is crucial. Although in this thesis we have mainly looked at changes consistent across neurodegenerative diseases, blood based biomarkers that allow clinicians to distinguish between distinct diseases will empower tailored treatments to match underlying pathologies more effectively.

7.2.2 General Limitations

Throughout this thesis, we have described limitations of the approaches used in each chapter. However, there are general considerations to discuss from the work.

A critical factor to consider in any research conducted into DNA methylation in disease in postmortem brain tissue is cause or effect. Whilst we and others have demonstrated the presence of aberrant DNA methylation across neurodegeneration, the determination of whether such changes are causing disease or simply a consequence of disease processes remains a challenge. We have investigated DNA methylation and gene expression data, and often have found that genes showing differential methylation also exhibit changes at the gene expression level. However, investigating the relationship between these two data types is challenging due to small sample sizes of overlapping data, and confounding factors such as tissue heterogeneity and cell type specificity. In Chapter 5, we report having used a large control dataset to investigate which OLG/OPC relevant genes show high correlation with gene expression, and determine for which of the disease associated genes this is also true. In our analysis of DNA methylation changes across multiple brain regions in AD, we have found consistent disease associated DNA methylation signatures across brain regions affected at different stages of disease progression; the entorhinal cortex and hippocampus are known to exhibit changes early in AD, whilst the dorsolateral prefrontal cortex is thought to be affected later on in disease. The finding of aberrant DNA methylation across both early and late stages of pathology suggests causal alterations contributing to disease rather than being the secondary effects of neurodegeneration. Although it is still possible that some of the changes we have observed may be a consequence of disease mechanisms rather than contributing themselves to disease, these findings are still of interest as they may reflect compensatory mechanisms and/or point to possible druggable targets; for example, decreased methylation of a DNA damage gene,

important in OLGs, could reflect the need for upregulation of damage response genes in response to increased levels of oxidative stress. Where genes important for OLG differentiation processes have been identified, it could be the case that OPCs are attempting to compensate for damage caused to mature OLGs by increasing in proliferation and/or differentiation. The research carried out for this thesis shows that multiple genes critical for OPC differentiation, including those implicated in the Wnt/beta-catenin pathway, show altered DNA methylation, which could lead to dysregulation of the differentiation process.

A significant challenge in studying neurodegenerative diseases is heterogeneity of cohorts. In Chapters 3 and 4 we have discussed DNA methylation changes across AD and FTLD. FTLD is an umbrella term for a clinically and pathologically heterogeneous group of diseases. We investigated changes across FTLD, as well as investigating more subtype/disease specific alterations, e.g. changes in FTLD-tau (importantly PSP) and FTLD-TDP cohorts. Although we found that multiple genes, including *CTNNA3, SCD* and *DNA17* showed dysregulation across these datasets, it is possible that we have missed more subtle subtype/disease specific effects. It is also true that we have not explored, aside from in our analysis of network analysis across brain regions in AD (Chapter 4), differences in DNA methylation between brain regions, which would be worthy of further investigation.

Another challenge is the potential misdiagnosis and/or degree of co-pathologies of the samples we have utilised. This is particularly applicable to Chapter 4, where we have leveraged multiple GWAS from a broad range of diseases. Neurodegenerative diseases are clinically heterogeneous, but also share many clinical features such as cognitive decline and changes in behaviour. There is also a complex relationship between clinical diagnosis and post-mortem pathology. By way of example, whilst PD and PSP share motor features, underlying pathologies are highly distinct. Importantly, PSP is a primary tauopathy, whilst PD is characterised by

accumulation of α-synuclein. Confounding this is the presence of mixed pathologies. Where cases have been misdiagnosed, this may lead to weakening of DNA methylation and genetic signals. Many of the cases utilised in the production of the GWAS analysed in Chapter 6 are pathologically diagnosed, which is considered the gold standard diagnosis, however not all cohorts are composed solely of such cases. Exciting new approaches such as machine learning and artificial intelligence are now being considered to mitigate issues of misdiagnosis and to more accurately detect and quantify neuropathological changes [347,348]. The advent of such techniques should enable clearer diagnosis and aid the identification of more specific disease effects.

The work written up in this thesis, aside from the use of the iPSC OLG differentiation cell model, relies wholly on the use of post-mortem brain tissue. Although the use of such data is invaluable in the study of neurodegenerative disease, its use presents several challenges. With any analysis using post-mortem tissue, there will be a post-mortem interval; the time between death of the subject and preservation of the tissue. Longer post-mortem delays may lead to DNA degradation and affect the detection of DNA methylation signals. We have attempted, where possible, to control for post-mortem intervals in our statistical analyses so as to reduce the possibility of finding differentially methylated sites which are artifacts of post-mortem degradation rather than as a cause of disease modifying effects. However, alterations in DNA methylation changes that may occur after death can hinder the detection of differentially methylated genes. Another caveat to the use of post-mortem tissue in the study of differential methylation in disease is that DNA methylation changes in post-mortem tissue only capture the end stage of disease, they do not capture the dynamic changes within tissue and disease progression. It is possible that end stage disease effects such as neuronal cell death and tissue inflammation will obscure additional disease relevant changes in the DNA methylation profile of OLGs. Altered cellular composition will also add to this; in advanced stages of

neurodegeneration, neuronal cell loss will change the composition of the tissue. This is particularly relevant where we have used bulk DNA methylation and RNA-sequencing data. Shifts in composition may lead to detection of DNA methylation changes that are more reflective of altered cellular proportions rather than disease causing modifications. We have attempted to mitigate this possibility through the use deconvolution methods to estimate cell type proportions and statistically adjusted for this as well as the use of brain-nuclei sorted DNA methylation datasets and single-nuclei RNA sequencing datasets, which should be less affected by cellular composition alterations, and through the use of our analysis of the iPSC model of OLG differentiation.

7.2.3 Future research directions

7.2.3.i Validation

Throughout this thesis, we have observed changes in OLG/OPC relevant genes across multiple DNA methylation datasets. To mitigate the detection of artifacts as a result of batch effects and data preparation, we have validated findings across multiple datasets relating to each disease analysed. However, to solidify findings, it would be beneficial to cross-check findings across a larger number of disease datasets, and carry out further meta-analyses to confirm reproducibility. As the use of single cell/single nuclei and sorted brain-nuclei DNA methylation datasets emerges and/or continues to grow, in the neurodegeneration field, it will be crucial to prioritise their investigation as a gold standard for achieving cell-type specificity in epigenetic studies.

7.2.3.ii Other effects of DNA methylation

We have attempted to investigate functional repercussions of changes in DNA methylation in neurodegeneration through the use of complementary gene expression datasets. However, it is known that changes in gene expression are not the only consequence of DNA methylation alterations. It would be interesting to explore the role of DNA methylation in alternative splicing of gene transcripts. Alternative splicing is a post-transcriptional modification through which a single gene results in the production of multiple transcript variants, and therefore diversity in the protein isoforms present. DNA methylation has been shown to be an important factor in the control of alternative splicing [349]. Mechanisms described include modulation of splicing through targeting of MeCP2 to included exons; ablation of MeCP2 was found to lead to aberrant splicing effects [350]. This was found to occur when there were intragenic DNA methylation marks, highlighting the importance of investigating DNA methylation changes outside of the promoter region. It has also been found that binding of CTCF resulting in inclusion of exons is mediated by changes in DNA methylation [351]. It is therefore possible that some of the DNA methylation changes that we have observed within the body of genes that do not appear, in our data, to modulate gene expression, are contributing to alternative splicing. Alternative splicing is an important phenomenon in neurodegenerative diseases. For example, aberrations in differential splicing of the *MAPT* pre-mRNA that cause changes in the balance of the different isoforms of the protein tau contributes to the pathogenesis of AD and other tauopathies [352]. Additionally, in Chapter 6, we have described how altered protein isoforms of the gene *MOBP* are detected in PSP [129], and that genetic and DNA methylation changes are located, within *MOBP,* around a site that could be acting as an alternative promoter region or is important in determining transcript usage, thus further investigation into this line of research is warranted. It is also worth noting that we have only investigated 5mC DNA methylation, and have not investigated DNA

hydroxymethylation (5hmC), an oxidative derivative of DNA methylation. As mentioned in Chapter 1, 5hmC is known to be enriched in the human brain compared to peripheral tissue, and studies in mice have defined a role for 5hmC in OLG fate determination. Although beyond the scope of this thesis, further investigation into this particular form of DNA methylation is warranted.

7.2.3.iv Extension to other diseases

In Chapters 3, 4 and 5 of this thesis, we have focused on my research into altered DNA methylation affecting OLG/OPC genes into the dementias AD and FTLD. In Chapter 5, we have looked at the *MOBP* locus in a wider range of diseases, including in MSA and PD. To gain a more holistic view of the role of DNA methylation in neurodegeneration, it would be useful to examine whether the effects we have seen consistently across diseases in Chapters 3 and 4, for example dysregulation of genes such as *CTNNA3, DNAH17* and *SCD,* extend to other neurodegenerative diseases or are found only in the diseases we have investigated. It would be interesting to gain more insight into how DNA methylation changes relate to underlying pathological processes that are hallmarks of each disease. MSA and PD are both α-synucleinopathies, and the work by Murthy et al. (2024) shows they share more similarities in alterations of DNA methylation between themselves than with PSP (a tauopathy). However, we have not explored this aspect in much detail in this thesis. Although we did examine dysregulation across tauopathies and TDP-43 proteinopathies, it is possible that dysregulation of DNA methylation in OLG/OPC relevant genes would be distinct.

Multiple sclerosis (MS), would be an intriguing disease to consider alongside the findings set out in this thesis. MS is a demyelinating disease where the loss of OLGs and OPCs to autoimmune attack is central to pathology [353]. Are epigenetic changes in these cells similar to those that

occur in the diseases we have analysed, representing general dysregulation and damage/death of cells? Or are different mechanisms in play?

7.3 Final remarks

Neurodegenerative diseases inflict an extremely high cost on society, both economically and socially. They also have devastating effects on individuals and their families. It is estimated that by 2050, the incidence of neurodegenerative disease will rise from 57.4 million cases in 2019 to 152.8 million cases globally [354]. This emphasises the importance of continued research into understanding the pathogenesis of the diseases in order to develop disease modifying therapies for care and intervention. Compared to fields such as cancer epigenetics, the study of epigenetic changes in neurodegeneration is less evolved, necessitating further progress. Similarly, research into OLG lineage cells in neurodegeneration has also been neglected. In this work, we have demonstrated the importance of dysregulated DNA methylation affecting OLGs in neurodegeneration, adding to the body of literature investigating epigenetics in neurodegenerative disease pathology. I hope that the work we have carried out for this thesis will contribute to the understanding of these devastating diseases so that a cure for them can be found.

8 Bibliography

1. Hou, Y. *et al.* Ageing as a risk factor for neurodegenerative disease. *Nat. Rev. Neurol.* **15**, 565–581 (2019).
2. Wilson, D. M., 3rd *et al.* Hallmarks of neurodegenerative diseases. *Cell* **186**, 693–714 (2023).
3. Gleichman, A. J. & Carmichael, S. T. Glia in neurodegeneration: Drivers of disease or along for the ride? *Neurobiol. Dis.* **142**, 104957 (2020).
4. Fu, H., Hardy, J. & Duff, K. E. Selective vulnerability in neurodegenerative diseases. *Nat. Neurosci.* **21**, 1350–1358 (2018).
5. Hardy, J. The genetic causes of neurodegenerative diseases. *J. Alzheimers. Dis.* **3**, 109–116 (2001).
6. Pandya, V. A. & Patani, R. Region-specific vulnerability in neurodegeneration: lessons from normal ageing. *Ageing Res. Rev.* **67**, 101311 (2021).
7. Erkkinen, M. G., Kim, M. O. & Geschwind, M. D. Clinical Neurology and Epidemiology of the Major Neurodegenerative Diseases. *Cold Spring Harbor Perspectives in Biology* **10**, (2018).
8. 2021 Alzheimer's disease facts and figures. *Alzheimers. Dement.* **17**, 327–406 (2021).
9. Zhang, J. *et al.* Recent advances in Alzheimer's disease: Mechanisms, clinical trials and new drug development strategies. *Signal Transduct. Target. Ther.* **9**, 211 (2024).
10. Maccioni, R. B., Muñoz, J. P. & Barbeito, L. The molecular bases of Alzheimer's disease and other neurodegenerative disorders. *Arch. Med. Res.* **32**, 367–381 (2001).
11. Bellenguez, C. *et al.* New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* **54**, 412–436 (2022).
12. Weggen, S. & Beher, D. Molecular consequences of amyloid precursor protein and

presenilin mutations causing autosomal-dominant Alzheimer's disease. *Alzheimers. Res. Ther.* **4**, 9 (2012).

13. Grossman, M. *et al.* Frontotemporal lobar degeneration. *Nat. Rev. Dis. Primers* **9**, 40 (2023).
14. Goldman, J. S. *et al.* Comparison of family histories in FTLD subtypes and related tauopathies. *Neurology* **65**, 1817–1819 (2005).
15. Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).
16. Steele, J. C., Richardson, J. C. & Olszewski, J. *Progressive Supranuclear Palsy: A Heterogeneous Degeneration Involving the Brain Stem, Basal Ganglia and Cerebellum with Vertical Gaze and Pseudobulbar Palsy, Nuchal Dystonia and Dementia*. (2014).
17. Komori, T. Tau-positive dial Inclusions in Progressive Supranuclear Palsy, Corticobasal Degeneration and Pick's Disease. *Brain Pathol.* **9**, 663–663 (1999).
18. Lashley, T., Rohrer, J. D., Mead, S. & Revesz, T. Review: An update on clinical, genetic and pathological aspects of frontotemporal lobar degenerations. *Neuropathol. Appl. Neurobiol.* **41**, 858–881 (2015).
19. Jellinger, K. A. Neuropathology of multiple system atrophy: New thoughts about pathogenesis. *Mov. Disord.* **29**, 1720–1741 (2014).
20. Poewe, W. *et al.* Multiple system atrophy. *Nat. Rev. Dis. Primers* **8**, 56 (2022).
21. Chia, R. *et al.* Genome sequence analyses identify novel risk loci for multiple system atrophy. *Neuron* **112**, 2142–2156.e5 (2024).
22. Prusiner, S. B. Evidence for alpha-synuclein prions causing multiple system atrophy in humans with parkinsonism. *Proc. Natl Acad. Sci. USA* **112**, E5308–E5317 (2015).
23. Bernis, M. E. *et al.* Prion-like propagation of human brain-derived alpha-synuclein in transgenic mice expressing human wild-type alpha-synuclein. *Acta Neuropathol. Commun.* **3**, 75 (2015).

24. Poewe, W. *et al.* Parkinson disease. *Nat. Rev. Dis. Primers* **3**, 17013 (2017).
25. Bandres-Ciga, S., Diez-Fairen, M., Kim, J. J. & Singleton, A. B. Genetics of Parkinson's disease: An introspection of its journey towards precision medicine. *Neurobiol. Dis.* **137**, 104782 (2020).
26. Angot, E., Steiner, J. A., Hansen, C., Li, J.-Y. & Brundin, P. Are synucleinopathies prion-like disorders? *Lancet Neurol.* **9**, 1128–1138 (2010).
27. Schapira, A. H. V. Mitochondrial dysfunction in Parkinson's disease. *Cell Death Differ.* **14**, 1261–1266 (2007).
28. Moehle, M. S. & West, A. B. M1 and M2 immune activation in Parkinson's Disease: Foe and ally? *Neuroscience* **302**, 59–73 (2015).
29. Sveinbjornsdottir, S. The clinical symptoms of Parkinson's disease. *J. Neurochem.* **139 Suppl 1**, 318–324 (2016).
30. Gibson, L. L. *et al.* Risk of dementia in Parkinson's disease: A Systematic review and meta-analysis. *Mov. Disord.* **39**, 1697–1709 (2024).
31. Masrori, P. & Van Damme, P. Amyotrophic lateral sclerosis: a clinical review. *Eur. J. Neurol.* **27**, 1918–1929 (2020).
32. Suk, T. R. & Rousseaux, M. W. C. The role of TDP-43 mislocalization in amyotrophic lateral sclerosis. *Mol. Neurodegener.* **15**, 45 (2020).
33. Abramzon, Y. A., Fratta, P., Traynor, B. J. & Chia, R. The overlapping genetics of amyotrophic lateral sclerosis and frontotemporal dementia. *Front. Neurosci.* **14**, 42 (2020).
34. Hofmann, J. W., Seeley, W. W. & Huang, E. J. RNA binding proteins and the pathogenesis of frontotemporal lobar degeneration. *Annu. Rev. Pathol.* **14**, 469–495 (2019).
35. Butti, Z. & Patten, S. A. RNA dysregulation in amyotrophic lateral sclerosis. *Front. Genet.* **9**, 712 (2018).
36. Mead, R. J., Shan, N., Reiser, H. J., Marshall, F. & Shaw, P. J. Amyotrophic lateral sclerosis: a neurodegenerative disorder poised for successful therapeutic translation. *Nat.*

Rev. Drug Discov. **22**, 185–212 (2023).

37. Pelvig, D. P., Pakkenberg, H., Stark, A. K. & Pakkenberg, B. Neocortical glial cell numbers in human brains. *Neurobiol. Aging* **29**, 1754–1762 (2008).
38. Bradl, M. & Lassmann, H. Oligodendrocytes: biology and pathology. doi:10.1007/s00401-009-0601-5.
39. Stadelmann, C., Timmler, S., Barrantes-Freer, A. & Simons, M. Myelin in the central nervous system: Structure, function, and pathology. *Physiological Reviews* **99**, 1381–1431 (2019).
40. Compston, A. & Coles, A. Multiple sclerosis. *Lancet* **372**, 1502–1517 (2008).
41. Ettle, B., Schlachetzki, J. C. M. & Winkler, J. Oligodendroglia and Myelin in Neurodegenerative Diseases: More Than Just Bystanders? *Molecular Neurobiology 2015 53:5* **53**, 3046–3062 (2015).
42. Valério-Gomes, B., Guimarães, D. M., Szczupak, D. & Lent, R. The absolute number of oligodendrocytes in the adult mouse brain. *Front. Neuroanat.* **12**, 90–90 (2018).
43. Timsit, S. *et al.* Oligodendrocytes originate in a restricted zone of the embryonic ventral neural tube defined by DM-20 mRNA expression. *J. Neurosci.* **15**, 1012–1024 (1995).
44. Vallstedt, A., Klos, J. M. & Ericson, J. Multiple dorsoventral origins of oligodendrocyte generation in the spinal cord and hindbrain. *Neuron* **45**, 55–67 (2005).
45. Warf, B. C., Fok-Seang, J. & Miller, R. H. Evidence for the ventral origin of oligodendrocyte precursors in the rat spinal cord. *J. Neurosci.* **11**, 2477–2488 (1991).
46. Kuhn, S., Gritti, L., Crooks, D. & Dombrowski, Y. Oligodendrocytes in Development, Myelin Generation and Beyond. *Cells* **8**, (2019).
47. Mr, D., A, Polito, Jm, L. & R, Reynolds. NG2-expressing glial progenitor cells: an abundant and widespread population of cycling cells in the adult rat CNS. *Molecular and cellular neurosciences* **24**, 476–488 (2003).
48. Falcão, A. M. *et al.* Disease-specific oligodendrocyte lineage cells arise in multiple

sclerosis. *Nature Medicine 2018 24:12* **24**, 1837–1844 (2018).

49. Kirby, L. & Castelo-Branco, G. Crossing boundaries: Interplay between the immune system and oligodendrocyte lineage cells. *Seminars in cell & developmental biology* **116**, 45–52 (2021).
50. Narine, M. & Colognato, H. Current Insights Into Oligodendrocyte Metabolism and Its Power to Sculpt the Myelin Landscape. *Frontiers in cellular neuroscience* **16**, (2022).
51. Fodder, K., de Silva, R., Warner, T. T. & Bettencourt, C. The contribution of DNA methylation to the (dys)function of oligodendroglia in neurodegeneration. *acta neuropathol commun* **11**, (2023).
52. de Faria, O. *et al.* Periods of synchronized myelin changes shape brain function and plasticity. *Nature Neuroscience 2021 24:11* **24**, 1508–1521 (2021).
53. Papuć, E. & Rejdak, K. The role of myelin damage in Alzheimer's disease pathology. *Archives of Medical Science : AMS* **16**, 345–345 (2020).
54. Braak, H. & Braak, E. Development of Alzheimer-related neurofibrillary changes in the neocortex inversely recapitulates cortical myelogenesis. *Acta Neuropathologica 1996 92:2* **92**, 197–201 (1996).
55. Brickman, A. M. Contemplating Alzheimer's disease and the contribution of white matter hyperintensities. *Curr. Neurol. Neurosci. Rep.* **13**, 415 (2013).
56. Brickman, A. M. *et al.* Regional White Matter Hyperintensity Volume, Not Hippocampal Atrophy, Predicts Incident Alzheimer Disease in the Community. *Arch. Neurol.* **69**, 1621–1627 (2012).
57. Nasrabady, S. E., Rizvi, B., Goldman, J. E. & Brickman, A. M. White matter changes in Alzheimer's disease: a focus on myelin and oligodendrocytes. *Acta Neuropathologica Communications* **6**, 22–22 (2018).
58. Dc, D. *et al.* Association of Amyloid Pathology With Myelin Alteration in Preclinical Alzheimer Disease. *JAMA Neurol.* **74**, 41–49 (2017).

59. Depp, C. *et al.* Myelin dysfunction drives amyloid-β deposition in models of Alzheimer's disease. *Nature* **618**, 349–357 (2023).
60. A, Matsuo *et al.* Myelin degeneration in multiple system atrophy detected by unique antibodies. *The American journal of pathology* **153**, 735–744 (1998).
61. T, Zhou *et al.* Implications of white matter damage in amyotrophic lateral sclerosis (Review). *Mol. Med. Rep.* **16**, 4379–4392 (2017).
62. Whitwell, J. L. *et al.* Clinical Correlates of White Matter Tract Degeneration in Progressive Supranuclear Palsy. *Arch. Neurol.* **68**, 753–760 (2011).
63. F, Caso *et al.* Progression of white matter damage in progressive supranuclear palsy with predominant parkinsonism. *Parkinsonism Relat. Disord.* **49**, 95–99 (2018).
64. Kaji, S., Maki, T., Ishimoto, T., Yamakado, H. & Takahashi, R. Insights into the pathogenesis of multiple system atrophy: focus on glial cytoplasmic inclusions. *Translational Neurodegeneration 2020 9:1* **9**, 1–15 (2020).
65. May, V. E. L. *et al.* α-Synuclein impairs oligodendrocyte progenitor maturation in multiple system atrophy. *Neurobiology of aging* **35**, 2357–2368 (2014).
66. Ahmed, Z., Asi, Y. T., Lees, A. J., Revesz, T. & Holton, J. L. Identification and quantification of oligodendrocyte precursor cells in multiple system atrophy, progressive supranuclear palsy and Parkinson's disease. *Brain Pathol.* **23**, 263–273 (2013).
67. G, Behrendt *et al.* Dynamic changes in myelin aberrations and oligodendrocyte generation in chronic amyloidosis in mice and men. *Glia* **61**, 273–286 (2013).
68. Je, S. *et al.* White matter lesions in an unselected cohort of the elderly: astrocytic, microglial and oligodendrocyte precursor cell responses. *Neuropathology and applied neurobiology* **33**, 410–419 (2007).
69. Gagyi, E. *et al.* Decreased oligodendrocyte nuclear diameter in Alzheimer's disease and Lewy body dementia. *Brain Pathol.* **22**, 803–810 (2012).
70. Blanchard, J. W. *et al.* APOE4 impairs myelination via cholesterol dysregulation in

oligodendrocytes. *Nature 2022* 1–11 (2022).

71. Hu, X. *et al.* 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 (2011).
72. Kunkle, B. W. *et al.* Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
73. Lambert, J. C. *et al.* Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat. Genet.* **45**, 1452–1458 (2013).
74. De Rossi, P. *et al.* Predominant expression of Alzheimer's disease-associated BIN1 in mature oligodendrocytes and localization to white matter tracts. *Molecular Neurodegeneration* **11**, (2016).
75. Chapuis, J. *et al.* Increased expression of BIN1 mediates Alzheimer genetic risk by modulating tau pathology. *Molecular Psychiatry* **16**, 1225–1234 (2013).
76. Chen, J. A. *et al.* Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. *Mol. Neurodegener.* **13**, (2018).
77. Gu, H. *et al.* Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nature genetics* **43**, 699–705 (2011).
78. Sanchez-Contreras, M. Y. *et al.* Replication of progressive supranuclear palsy genome-wide association study identifies SLCO1A2 and DUSP10 as new susceptibility loci. *Mol. Neurodegener.* **13**, 1–10 (2018).
79. Farrell, K. *et al.* Genetic, transcriptomic, histological, and biochemical analysis of progressive supranuclear palsy implicates glial activation and novel risk genes. *Nat. Commun.* **15**, 7880 (2024).
80. Kouri, N. *et al.* Genome-wide association study of corticobasal degeneration identifies risk

variants shared with progressive supranuclear palsy. *Nat. Commun.* **6**, 7247 (2015).

81. Liu, Q.-Y. *et al.* An exploratory study on STX6, MOBP, MAPT, and EIF2AK3 and late-onset Alzheimer's disease. *Neurobiol. Aging* **34**, 1519.e13–7 (2013).
82. W, van Rheenen *et al.* Genome-wide association analyses identify new risk variants and the genetic architecture of amyotrophic lateral sclerosis. *Nat. Genet.* **48**, 1043–1048 (2016).
83. Siokas, V. *et al.* Myelin-associated oligodendrocyte basic protein rs616147 polymorphism as a risk factor for Parkinson's disease. *Acta Neurol. Scand.* **145**, 223–228 (2022).
84. Irwin, D. J. *et al.* Myelin oligodendrocyte basic protein and prognosis in behavioral-variant frontotemporal dementia. *Neurology* **83**, 502–509 (2014).
85. Allen, M. *et al.* Gene expression, methylation and neuropathology correlations at progressive supranuclear palsy risk loci. *Acta Neuropathol.* **132**, 197–211 (2016).
86. Allen, M. *et al.* Conserved brain myelination networks are altered in Alzheimer's and other neurodegenerative diseases. *Alzheimer's & Dementia* **14**, 352–366 (2018).
87. Piras, I. S. *et al.* Transcriptional profiling of multiple system atrophy cerebellar tissue highlights differences between the parkinsonian and cerebellar sub-types of the disease. *Acta Neuropathol. Commun.* **8**, 1–20 (2020).
88. Hasan, R. *et al.* Transcriptomic analysis of frontotemporal lobar degeneration with TDP-43 pathology reveals cellular alterations across multiple brain regions. *Acta neuropathologica* **143**, (2022).
89. Moore, L. D., Le, T. & Fan, G. DNA methylation and its basic function. *Neuropsychopharmacology* Preprint at https://doi.org/10.1038/npp.2012.112 (2013).
90. Greenberg, M. V. C. & Bourc'his, D. The diverse roles of DNA methylation in mammalian development and disease. *Nat. Rev. Mol. Cell Biol.* **20**, 590–607 (2019).
91. Lou, S. *et al.* Whole-genome bisulfite sequencing of multiple individuals reveals complementary roles of promoter and gene body methylation in transcriptional regulation. *Genome Biol.* **15**, 1–21 (2014).

92. Pai, A. A., Bell, J. T., Marioni, J. C., Pritchard, J. K. & Gilad, Y. A Genome-Wide Study of DNA Methylation Patterns and Gene Expression Levels in Multiple Human and Chimpanzee Tissues. *PLoS Genet.* **7**, 1001316–1001316 (2011).
93. Vanderkraats, N. D., Hiken, J. F., Decker, K. F. & Edwards, J. R. Discovering high-resolution patterns of differential DNA methylation that correlate with gene expression changes. *Nucleic Acids Res.* **41**, 6816–6827 (2013).
94. Jeong, H. *et al.* Evolution of DNA methylation in the human brain. *Nat. Commun.* **12**, (2021).
95. Guo, J. U., Su, Y., Zhong, C., Ming, G. L. & Song, H. Hydroxylation of 5-methylcytosine by TET1 promotes active DNA demethylation in the adult brain. *Cell* **145**, 423–434 (2011).
96. Zhang, M. *et al.* Ten-eleven translocation 1 mediated-DNA hydroxymethylation is required for myelination and remyelination in the mouse brain. *Nature Communications 2021 12:1* **12**, 1–21 (2021).
97. Globisch, D. *et al.* Tissue Distribution of 5-Hydroxymethylcytosine and Search for Active Demethylation Intermediates. *PLoS ONE* **5**, 15367–15367 (2010).
98. Kozlenkov, A. *et al.* A unique role for DNA (hydroxy)methylation in epigenetic regulation of human inhibitory neurons. *Science Advances* **4**, (2018).
99. Münzel, M. *et al.* Quantification of the Sixth DNA Base Hydroxymethylcytosine in the Brain. *Angewandte Chemie International Edition* **49**, 5375–5377 (2010).
100. Moyon, S. & Casaccia, P. DNA methylation in oligodendroglial cells during developmental myelination and in disease. (2017) doi:10.1080/23262133.2016.1270381.
101. Moyon, S. *et al.* Efficient Remyelination Requires DNA Methylation. *eNeuro* **4**, (2017).
102. Moyon, S. *et al.* TET1-mediated DNA hydroxy-methylation regulates adult remyelination. *bioRxiv* 819995–819995 (2019).
103. Appleby-Mallinder, C. *et al.* TDP43 proteinopathy is associated with aberrant DNA methylation in human amyotrophic lateral sclerosis. *Neuropathology and Applied*

Neurobiology **47**, 61–72 (2021).

104. M, Murthy, Yy, C., Jl, H. & C, Bettencourt. Neurodegenerative movement disorders: An epigenetics perspective and promise for the future. *Neuropathol. Appl. Neurobiol.* (2021) doi:10.1111/NAN.12757.

105. Wen, K. X. *et al.* The Role of DNA Methylation and Histone Modifications in Neurodegenerative Diseases: A Systematic Review. *PLOS ONE* **11**, e0167201–e0167201 (2016).

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

107. Lunnon, K. *et al.* Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nat. Neurosci.* **17**, 1164–1170 (2014).

108. Smith, R. G. *et al.* Elevated DNA methylation across a 48-kb region spanning the HOXA gene cluster is associated with Alzheimer's disease neuropathology. *Alzheimers. Dement.* **14**, 1580–1588 (2018).

109. Watson, C. T. *et al.* Genome-wide DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease. *Genome Med.* **8**, (2016).

110. Shireby, G. *et al.* DNA methylation signatures of Alzheimer's disease neuropathology in the cortex are primarily driven by variation in non-neuronal cell-types. *bioRxivorg* 2022.03.15.484508–2022.03.15.484508 (2022).

111. Smith, A. R., Wheildon, G. & Lunnon, K. Invited Review - A 5-year update on epigenome-wide association studies of DNA modifications in Alzheimer's disease: progress, practicalities and promise. *Neuropathology and applied neurobiology* **46**, 641–653 (2020).

112. Smith, R. G. *et al.* A meta-analysis of epigenome-wide association studies in Alzheimer's disease highlights novel differentially methylated loci across cortex. *Nature Communications 2021 12:1* **12**, 1–13 (2021).

113. Zhang, L. *et al.* Epigenome-wide meta-analysis of DNA methylation differences in prefrontal cortex implicates the immune processes in Alzheimer's disease. *Nature Communications 2020 11:1* **11**, 1–13 (2020).

114. Masliah, E., Dumaop, W., Galasko, D. & Desplats, P. Distinctive patterns of DNA methylation associated with Parkinson disease: Identification of concordant epigenetic changes in brain and peripheral blood leukocytes. *Epigenetics* **8**, 1030–1038 (2013).

115. Kaut, O., Schmitt, I. & Wüllner, U. Genome-scale methylation analysis of Parkinson's disease patients' brains reveals DNA hypomethylation and increased mRNA expression of cytochrome P450 2E1. *Neurogenetics* **13**, 87–91 (2012).

116. Murthy, M. *et al.* DNA methylation patterns in the frontal lobe white matter of multiple system atrophy, Parkinson's disease, and progressive supranuclear palsy: a cross-comparative investigation. *Acta Neuropathol.* **148**, 4 (2024).

117. Weber, A. *et al.* Epigenome-wide DNA methylation profiling in Progressive Supranuclear Palsy reveals major changes at DLX1. *Nat. Commun.* (2018) doi:10.1038/s41467-018-05325-y.

118. Bettencourt, C. *et al.* White matter DNA methylation profiling reveals deregulation of HIP1, LMAN2, MOBP, and other loci in multiple system atrophy. *Acta Neuropathologica* **139**, 135–135 (2020).

119. Fodder, K. *et al.* Brain DNA methylomic analysis of frontotemporal lobar degeneration reveals OTUD4 in shared dysregulated signatures across pathological subtypes. *Acta Neuropathologica 2023* **1**, 1–19 (2023).

120. Horvath, S. *et al.* Huntington's disease accelerates epigenetic aging of human brain and disrupts DNA methylation levels. *Aging (Albany NY)* **8**, 1485–1512 (2016).

121. De Souza, R. A. G. *et al.* DNA methylation profiling in human Huntington's disease brain. *Hum. Mol. Genet.* **25**, 2013–2030 (2016).

122. Gasparoni, G. *et al.* DNA methylation analysis on purified neurons and glia dissects age

and Alzheimer's disease-specific changes in the human cortex. *Epigenetics Chromatin* **11**, (2018).

123. Tiane, A. *et al.* From OPC to Oligodendrocyte: An Epigenetic Journey. *Cells* **8**, (2019).
124. Moyon, S. *et al.* Functional characterization of DNA methylation in the oligodendrocyte lineage. *Cell reports* **15**, 748–748 (2016).
125. Pruvost, M. & Moyon, S. Oligodendroglial Epigenetics, from Lineage Specification to Activity-Dependent Myelination. *Life 2021, Vol. 11, Page 62* **11**, 62–62 (2021).
126. A, Tiane *et al.* DNA methylation regulates the expression of the negative transcriptional regulators ID2 and ID4 during OPC differentiation. *Cell. Mol. Life Sci.* (2021) doi:10.1007/S00018-021-03927-2.
127. Zhao, X. *et al.* Dynamics of ten-eleven translocation hydroxylase family proteins and 5-hydroxymethylcytosine in oligodendrocyte differentiation. *Glia* **62**, 914–926 (2014).
128. Sams, E. C. Oligodendrocytes in the aging brain. *Neuronal Signal.* **5**, (2021).
129. Bettencourt, C. *et al.* MOBP and HIP1 in multiple system atrophy: New α-synuclein partners in glial cytoplasmic inclusions implicated in the disease pathogenesis. *Neuropathol. Appl. Neurobiol.* **47**, 640–640 (2021).
130. Schäfer, I., Müller, C., Luhmann, H. J. & White, R. MOBP levels are regulated by Fyn kinase and affect the morphological differentiation of oligodendrocytes. *J. Cell Sci.* **129**, 930–942 (2016).
131. Yu, L. *et al.* Association of Brain DNA Methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 With Pathological Diagnosis of Alzheimer Disease. *JAMA Neurol.* **72**, 15–24 (2015).
132. Horvath, S. DNA methylation age of human tissues and cell types. *Genome Biol.* **14**, 1–20 (2013).
133. Horvath, S. & Ritz, B. R. Increased epigenetic age and granulocyte counts in the blood of Parkinson's disease patients. *Aging (Albany NY)* **7**, 1130–1142 (2015).

134. Picillo, M. *et al.* Parkinsonism due to A53E α-synuclein gene mutation: Clinical, genetic, epigenetic, and biochemical features. *Mov. Disord.* **33**, 1950–1955 (2018).
135. Siming, S., Aixiao, L., Jiadong, L., Candy, W. & Patrizia, C.-B. Epigenetic memory loss in aging oligodendrocytes in the corpus callosum. *Neurobiology of aging* **29**, 452–452 (2008).
136. Murthy, M. *et al.* Epigenetic age acceleration is associated with oligodendrocyte proportions in MSA and control brain tissue. *bioRxivorg* 2022.07.20.500795–2022.07.20.500795 (2022).
137. Singh, N. *et al.* COMPREHENSIVE INVITED REVIEW Brain Iron Homeostasis: From Molecular Mechanisms To Clinical Significance and Therapeutic Opportunities. doi:10.1089/ars.2012.4931.
138. Xu, H. *et al.* New progress on the role of Glia in iron metabolism and iron-induced degeneration of dopamine neurons in Parkinson's disease. *Front. Mol. Neurosci.* **10**, 455 (2017).
139. Zecca, L., Youdim, M. B. H., Riederer, P., Connor, J. R. & Crichton, R. R. Iron, brain ageing and neurodegenerative disorders. *Nature Reviews Neuroscience 2004 5:11* **5**, 863–873 (2004).
140. Connor, J. R. & Menzies, S. L. *Relationship of Iron to Oligodendrocytes and Myelination*. vol. 17 (1996).
141. Bettencourt, C. *et al.* Gene co-expression networks shed light into diseases of brain iron accumulation. *Neurobiol. Dis.* **87**, 59–59 (2016).
142. Heidari, M. *et al.* Brain iron accumulation affects myelin-related molecular systems implicated in a rare neurogenetic disease family with neuropsychiatric features. *Molecular Psychiatry* **21**, 1599–1607 (2016).
143. Back, S. A., Gan, X., Li, Y., Rosenberg, P. A. & Volpe, J. J. Maturation-Dependent Vulnerability of Oligodendrocytes to Oxidative Stress-Induced Death Caused by Glutathione Depletion. *J. Neurosci.* **18**, 6241–6253 (1998).

144. Fern, R. & Möller, T. Rapid Ischemic Cell Death in Immature Oligodendrocytes: A Fatal Glutamate Release Feedback Loop. *J. Neurosci.* **20**, 34–42 (2000).

145. Bartzokis, G. Age-related myelin breakdown: a developmental model of cognitive decline and Alzheimer's disease. *Neurobiol. Aging* **25**, 5–18 (2004).

146. French, H. M., Reid, M., Mamontov, P., Simmons, R. A. & Grinspan, J. B. Oxidative stress disrupts oligodendrocyte maturation. *J. Neurosci. Res.* **87**, 3076–3087 (2009).

147. Spaas, J. *et al.* Oxidative stress and impaired oligodendrocyte precursor cell differentiation in neurological disorders. *Cellular and Molecular Life Sciences* vol. 78 4615–4637 Preprint at https://doi.org/10.1007/s00018-021-03802-0 (2021).

148. Madugundu, G. S., Cadet, J. & Wagner, J. R. Hydroxyl-radical-induced oxidation of 5-methylcytosine in isolated and cellular DNA. *Nucleic Acids Res.* **42**, 7450–7460 (2014).

149. Menden, K. *et al.* A multi-omics dataset for the analysis of Frontotemporal Dementia genetic subtypes. *bioRxiv* (2020) doi:10.1101/2020.12.01.405894.

150. Semick, S. A. *et al.* Integrated DNA methylation and gene expression profiling across multiple brain regions implicate novel genes in Alzheimer's disease. *Acta Neuropathol.* **137**, 557–569 (2019).

151. Bizet, M. *et al.* Improving Infinium MethylationEPIC data processing: re-annotation of enhancers and long noncoding RNA genes and benchmarking of normalization methods. *Epigenetics* **17**, 2434–2454 (2022).

152. Aryee, M. J. *et al.* Minfi: A flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* (2014) doi:10.1093/bioinformatics/btu049.

153. Schalkwyk, L. C., Pidsley, R. & Wong, C. C. Y. wateRmelon: Illumina 450 methylation array normalization and metrics. *R package version 1.2.2* Preprint at (2013).

154. Tian, Y. *et al.* ChAMP: Updated methylation analysis pipeline for Illumina BeadChips. *Bioinformatics* (2017) doi:10.1093/bioinformatics/btx513.

155. Du, P. *et al.* Comparison of Beta-value and M-value methods for quantifying methylation levels by microarray analysis. *BMC Bioinformatics* **11**, (2010).

156. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).

157. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The SVA package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* (2012) doi:10.1093/bioinformatics/bts034.

158. Mathys, H. *et al.* Single-cell transcriptomic analysis of Alzheimer's disease. *Nature* doi:10.1038/s41586-019-1195-2.

159. Piras, I. S. *et al.* A review and meta-analysis of gene expression profiles in suicide. *European neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology* **56**, (2022).

160. Smith, A. R. *et al.* A cross-brain regions study of ANK1 DNA methylation in different neurodegenerative diseases. *Neurobiol. Aging* **74**, 70–76 (2019).

161. Houseman, E. A. *et al.* DNA methylation arrays as surrogate measures of cell mixture distribution. *BMC Bioinformatics* **13**, 86 (2012).

162. Balduzzi, S., Rücker, G. & Schwarzer, G. How to perform a meta-analysis with R: a practical tutorial. *Evid. Based. Ment. Health* **22**, 153–160 (2019).

163. Gene Ontology Consortium *et al.* The Gene Ontology knowledgebase in 2023. *Genetics* **224**, (2023).

164. Ashburner, M. *et al.* Gene Ontology: tool for the unification of biology. *Nature Genetics* **25**, 25–29 (2000).

165. Yu, G., Wang, L.-G., Han, Y. & He, Q.-Y. clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. *OMICS : a Journal of Integrative Biology* **16**, 284 (2012).

166. Cao, Y.-F. *et al.* Targeting USP10 induces degradation of oncogenic ANLN in esophageal

squamous cell carcinoma. *Cell Death Differ.* **30**, 527–543 (2023).

167. Pandey, S. *et al.* Disease-associated oligodendrocyte responses across neurodegenerative diseases. *Cell Rep.* **40**, 111189 (2022).

168. Sojka, D. R. *et al.* Functional redundancy of HSPA1, HSPA2 and other HSPA proteins in non-small cell lung carcinoma (NSCLC); an implication for NSCLC treatment. *Scientific Reports* **9**, 1–15 (2019).

169. Teeple, E. *et al.* Single nuclei sequencing of human putamen oligodendrocytes reveals altered heterogeneity and disease-associated changes in Parkinson's disease and Multiple System Atrophy. *bioRxiv* (2021) doi:10.1101/2021.05.06.442967.

170. Panitch, R. *et al.* Integrative brain transcriptome analysis links complement component 4 and HSPA2 to the APOE ε2 protective effect in Alzheimer disease. *Mol. Psychiatry* **26**, 6054–6064 (2021).

171. Hu, A. *et al.* PIP4K2A regulates intracellular cholesterol transport through modulating PI(4,5)P2 homeostasis. *J. Lipid Res.* **59**, 507–514 (2018).

172. McKenzie, A. T. *et al.* Multiscale network modeling of oligodendrocytes reveals molecular components of myelin dysregulation in Alzheimer's disease. *Mol. Neurodegener.* **12**, 82 (2017).

173. Ertekin-Taner, N. *et al.* Linkage of plasma Abeta42 to a quantitative locus on chromosome 10 in late-onset Alzheimer's disease pedigrees. *Science* **290**, 2303–2304 (2000).

174. Ertekin-Taner, N. *et al.* Fine mapping of the alpha-T catenin gene to a quantitative trait locus on chromosome 10 in late-onset Alzheimer's disease pedigrees. *Hum. Mol. Genet.* **12**, 3133–3143 (2003).

175. Martin, E. R. *et al.* Interaction between the alpha-T catenin gene (VR22) and APOE in Alzheimer's disease. *J. Med. Genet.* **42**, 787–792 (2005).

176. Busby, V. *et al.* α-T-catenin is expressed in human brain and interacts with the Wnt signaling pathway but is not responsible for linkage to chromosome 10 in Alzheimer's

disease. *NeuroMolecular Medicine* **5**, 133–146 (2004).

177. Vilariño-Güell, C. *et al.* Exome sequencing in multiple sclerosis families identifies 12 candidate genes and nominates biological pathways for the genesis of disease. *PLOS Genetics* **15**, e1008180 (2019).

178. Perrone, F., Cacace, R., van der Zee, J. & Van Broeckhoven, C. Emerging genetic complexity and rare genetic variants in neurodegenerative brain diseases. *Genome Medicine* **13**, 59 (2021).

179. Hamilton, L. K. *et al.* Stearoyl-CoA Desaturase inhibition reverses immune, synaptic and cognitive impairments in an Alzheimer's disease mouse model. *Nat. Commun.* **13**, 2061 (2022).

180. Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat. Rev. Genet.* **13**, 484–492 (2012).

181. Ruan, X. *et al.* The Research Progress in Physiological and Pathological Functions of TRAF4. *Frontiers in oncology* **12**, (2022).

182. Blaise, S. *et al.* In Vivo Evidence That TRAF4 Is Required for Central Nervous System Myelin Homeostasis. *PLoS ONE* **7**, e30917 (2012).

183. Dai, Z.-M. *et al.* Stage-specific regulation of oligodendrocyte development by Wnt/β-catenin signaling. *J. Neurosci.* **34**, 8467–8473 (2014).

184. Li, Q. S. & De Muynck, L. Differentially expressed genes in Alzheimer's disease highlighting the roles of microglia genes including OLR1 and astrocyte gene CDK2AP1. *Brain Behav. Immun. Health* **13**, 100227 (2021).

185. Liu, L.-Q., Feng, L.-F., Nan, C.-R. & Zhao, Z.-M. CREB3L1 and PTN expressions correlate with prognosis of brain glioma patients. *Bioscience Reports* **38**, BSR20170100 (2018).

186. Rowlands, D. *et al.* Aggrecan Directs Extracellular Matrix-Mediated Neuronal Plasticity. *The Journal of Neuroscience* **38**, 10102 (2018).

187. Keskin, T., Avsar, O., Eliacik, S. & Uysal Tan, F. Investigation of the relationship between

gene VNTR polymorphism and Alzheimer's disease in Turkish population. *Nucleosides Nucleotides Nucleic Acids* **43**, 1129–1138 (2024).

188. Ramírez VT, Ramos-Fernández E, Henríquez JP, Lorenzo A, Inestrosa NC. Wnt-5a/Frizzled9 Receptor Signaling through the Gαo-Gβγ Complex Regulates Dendritic Spine Formation. *Journal of Biological Chemistry* **291**, 19092–19107 (2016).

189. Braccioli, L., Vervoort, S. J., Puma, G., Nijboer, C. H. & Coffer, P. J. SOX4 inhibits oligodendrocyte differentiation of embryonic neural stem cells in vitro by inducing Hes5 expression. *Stem Cell Res* **33**, 110–119 (2018).

190. Mahajan, K. & Mahajan, N. P. ACK1/TNK2 Tyrosine Kinase: Molecular Signaling and Evolving Role in Cancers. *Oncogene* **34**, 4162 (2014).

191. Farlow, J. L. *et al.* Whole-Exome Sequencing in Familial Parkinson Disease. *JAMA Neurol* **73**, 68–75 (2016).

192. Counts, S. E., Perez, S. E., Ginsberg, S. D. & Mufson, E. J. Neuroprotective Role for Galanin in Alzheimer's Disease. *EXS* **102**, 143 (2010).

193. Lyubetska, H., Zhang, L., Kong, J. & Vrontakis, M. An elevated level of circulating galanin promotes developmental expression of myelin basic protein in the mouse brain. *Neuroscience* **284**, 581–589 (2015).

194. Zhang, L., Yu, W., Schroedter, I., Kong, J. & Vrontakis, M. Galanin Transgenic Mice with Elevated Circulating Galanin Levels Alleviate Demyelination in a Cuprizone-Induced MS Mouse Model. *PLOS ONE* **7**, e33901 (2012).

195. Maheu, M. *et al.* MicroRNA regulation of central glial cell line-derived neurotrophic factor (GDNF) signalling in depression. *Translational Psychiatry* **5**, e511–e511 (2015).

196. Strelau, J. & Unsicker, K. GDNF family members and their receptors: expression and functions in two oligodendroglial cell lines representing distinct stages of oligodendroglial development. *Glia* **26**, (1999).

197. Identification of risk genes for Alzheimer's disease by gene embedding. *Cell Genomics* **2**,

100162 (2022).

198. Genome-wide association study identifies susceptibility loci of brain atrophy to NFIA and ST18 in Alzheimer's disease. *Neurobiology of Aging* **102**, 200.e1–200.e11 (2021).

199. Zhen, Y. *et al.* Protocadherin 15 suppresses oligodendrocyte progenitor cell proliferation and promotes motility through distinct signalling pathways. *Commun. Biol.* **5**, 511 (2022).

200. Huang, W. *et al.* Origins and Proliferative States of Human Oligodendrocyte Precursor Cells. *Cell* **182**, 594–608.e11 (2020).

201. Ranjan, M. & Hudson, L. D. Regulation of tyrosine phosphorylation and protein tyrosine phosphatases during oligodendrocyte differentiation. *Mol Cell Neurosci* **7**, 404–418 (1996).

202. Terrone, G. *et al.* A further contribution to the delineation of epileptic phenotype in PACS2-related syndrome. *Seizure* **79**, 53–55 (2020).

203. Sadick, J. S. *et al.* Astrocytes and oligodendrocytes undergo subtype-specific transcriptional changes in Alzheimer's disease. *Neuron* **110**, 1788–1805.e10 (2022).

204. Celarain, N. & Tomas-Roig, J. Aberrant DNA methylation profile exacerbates inflammation and neurodegeneration in multiple sclerosis patients. *J. Neuroinflammation* **17**, 21 (2020).

205. Su, L., Chen, S., Zheng, C., Wei, H. & Song, X. Meta-analysis of gene expression and identification of biological regulatory mechanisms in Alzheimer's disease. *Front. Neurosci.* **13**, 633 (2019).

206. Ponnusamy, M. *et al.* Loss of forebrain BIN1 attenuates hippocampal pathology and neuroinflammation in a tauopathy model. *Brain* **146**, 1561–1579 (2023).

207. Gajewska, K. A., Lescesen, H., Ramialison, M. & Jans, D. A. Nuclear transporter Importin-13 plays a key role in the oxidative stress transcriptional response. *Nature Communications* **12**, 1–13 (2021).

208. Hampton, D. W. *et al.* HspB5 Activates a Neuroprotective Glial Cell Response in Experimental Tauopathy. *Front. Neurosci.* **14**, 545943 (2020).

209. Upregulation of alphaB-crystallin expression in the substantia nigra of patients with

Parkinson's disease. *Neurobiology of Aging* **36**, 1686–1691 (2015).

210. Website. https://www.jns-journal.com/article/0022-510X(93)90135-L/fulltext.

211. Maitre, M. *et al.* Myelin in Alzheimer's disease: culprit or bystander? *Acta Neuropathol. Commun.* **11**, 56 (2023).

212. Boggs, J. M. Myelin basic protein: a multifunctional protein. *Cellular and molecular life sciences : CMLS* **63**, (2006).

213. Stadelmann, C., Timmler, S., Barrantes-Freer, A. & Simons, M. Myelin in the Central Nervous System: Structure, Function, and Pathology. *Physiological Reviews* (2019) doi:10.1152/physrev.00031.2018.

214. Zhu, Q. *et al.* Genetic evidence that Nkx2.2 and Pdgfra are major determinants of the timing of oligodendrocyte differentiation in the developing CNS. *Development (Cambridge, England)* **141**, 548 (2014).

215. Miyashita, A. *et al.* Genetic association of CTNNA3 with late-onset Alzheimer's disease in females. *Hum. Mol. Genet.* **16**, 2854–2869 (2007).

216. Fancy, S. P. J. *et al.* Dysregulation of the Wnt pathway inhibits timely myelination and remyelination in the mammalian CNS. *Genes Dev.* **23**, 1571–1585 (2009).

217. Zacarías-Fluck, M. F. *et al.* The Wnt signaling receptor Fzd9 is essential for Myc-driven tumorigenesis in pancreatic islets. *Life Science Alliance* **4**, e201900490 (2021).

218. Jia, L., Piña-Crespo, J. & Li, Y. Restoring Wnt/β-catenin signaling is a promising therapeutic strategy for Alzheimer's disease. *Mol Brain* **12**, 104 (2019).

219. García-Velázquez, L. & Arias, C. The emerging role of Wnt signaling dysregulation in the understanding and modification of age-associated diseases. *Ageing Res Rev* **37**, 135–145 (2017).

220. Flowers, M. T. & Ntambi, J. M. Role of stearoyl-coenzyme A desaturase in regulating lipid metabolism. *Curr. Opin. Lipidol.* **19**, 248–256 (2008).

221. Cunnane, S. C. *et al.* Plasma and brain fatty acid profiles in mild cognitive impairment and

Alzheimer's disease. *J. Alzheimers. Dis.* **29**, 691–697 (2012).

222. Fraser, T., Tayler, H. & Love, S. Fatty acid composition of frontal, temporal and parietal neocortex in the normal human brain and in Alzheimer's disease. *Neurochem. Res.* **35**, 503–513 (2010).

223. Piras, I. S. *et al.* Transcriptional profiling of multiple system atrophy cerebellar tissue highlights differences between the parkinsonian and cerebellar sub-types of the disease. *Acta Neuropathologica Communications* **8**, 76 (2020).

224. Balusu, S., Praschberger, R., Lauwers, E., De Strooper, B. & Verstreken, P. Neurodegeneration cell per cell. *Neuron* **111**, 767–786 (2023).

225. Bednarski, T. *et al.* Stearoyl-CoA desaturase 1 deficiency reduces lipid accumulation in the heart by activating lipolysis independently of peroxisome proliferator-activated receptor α. *Biochim Biophys Acta* **1861**, 2029–2037 (2016).

226. Quinonoid dihydropteridine reductase, a tetrahydrobiopterin-recycling enzyme, contributes to 5-hydroxytryptamine-associated platelet aggregation in mice. *Journal of Pharmacological Sciences* **150**, 173–179 (2022).

227. Vasquez-Vivar, J., Shi, Z. & Tan, S. Tetrahydrobiopterin in Cell Function and Death Mechanisms. *Antioxid Redox Signal* **37**, 171–183 (2022).

228. Asea, A. A. A. & Brown, I. R. *Heat Shock Proteins and the Brain: Implications for Neurodegenerative Diseases and Neuroprotection*. (Springer Science & Business Media, 2008).

229. Dabir, D. V., Trojanowski, J. Q., Richter-Landsberg, C., Lee, V. M.-Y. & Forman, M. S. Expression of the Small Heat-Shock Protein αB-Crystallin in Tauopathies with Glial Pathology. *The American Journal of Pathology* **164**, 155 (2004).

230. Zhan, X. *et al.* Myelin basic protein associates with AβPP, Aβ1-42, and amyloid plaques in cortex of Alzheimer's disease brain. *J Alzheimers Dis* **44**, 1213–1229 (2015).

231. Shayevitch, R., Askayo, D., Keydar, I. & Ast, G. The importance of DNA methylation of

exons on alternative splicing. *RNA* **24**, 1351 (2018).

232. Lev Maor, G., Yearim, A. & Ast, G. The alternative role of DNA methylation in splicing regulation. *Trends Genet.* **31**, 274–280 (2015).

233. Langfelder, P. & Horvath, S. WGCNA: An R package for weighted correlation network analysis. *BMC Bioinformatics* (2008) doi:10.1186/1471-2105-9-559.

234. Zhang, B. *et al.* Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell* **153**, 707 (2013).

235. Chen, Y., Li, Z., Ge, X., Lv, H. & Geng, Z. Identification of novel hub genes for Alzheimer's disease associated with the hippocampus using WGCNA and differential gene analysis. *Front. Neurosci.* **18**, 1359631 (2024).

236. Horvath, S. *et al.* Aging effects on DNA methylation modules in human brain and blood tissue. *Genome Biology* **13**, 1–18 (2012).

237. Botía, J. A. *et al.* An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. *BMC Syst. Biol.* (2017) doi:10.1186/s12918-017-0420-6.

238. Langfelder, P., Luo, R., Oldham, M. C. & Horvath, S. Is My Network Module Preserved and Reproducible? *PLoS Comput. Biol.* **7**, 1001057–1001057 (2011).

239. Skene, N. G. & Grant, S. G. N. Identification of Vulnerable Cell Types in Major Brain Disorders Using Single Cell Transcriptomes and Expression Weighted Cell Type Enrichment. *Front. Neurosci.* **0**, 16–16 (2016).

240. Zeisel, A. *et al.* Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science* **347**, 1138–1142 (2015).

241. Xu, J. *et al.* Inhibition of proliferation by knockdown of transmembrane (TMEM) 168 in glioblastoma cells via suppression of Wnt/β-catenin pathway. *Oncol. Res.* **27**, 819–826 (2019).

242. Jana, A. & Pahan, K. Sphingolipids in multiple sclerosis. *Neuromolecular medicine* **12**, 351

(2010).

243. Dill, H., Linder, B., Fehr, A. & Fischer, U. Intronic miR-26b controls neuronal differentiation by repressing its host transcript, ctdsp2. *Genes & development* **26**, (2012).

244. Miyazaki, M. & Esser, K. A. REDD2 is enriched in skeletal muscle and inhibits mTOR signaling in response to leucine and stretch. *Am. J. Physiol. Cell Physiol.* **296**, C583–92 (2009).

245. Tyler, W. A. *et al.* Activation of the mammalian target of rapamycin (mTOR) is essential for oligodendrocyte differentiation. *J. Neurosci.* **29**, 6367–6378 (2009).

246. Grupe, A. *et al.* Evidence for novel susceptibility genes for late-onset Alzheimer's disease from a genome-wide association study of putative functional variants. *Hum. Mol. Genet.* **16**, 865–873 (2007).

247. Xia, Z. *et al.* A putative Alzheimer's disease risk allele in PCK1 influences brain atrophy in multiple sclerosis. *PLoS One* **5**, e14169 (2010).

248. Millward, C. A. *et al.* Phosphoenolpyruvate carboxykinase (Pck1) helps regulate the triglyceride/fatty acid cycle and development of insulin resistance in mice. *J. Lipid Res.* **51**, 1452–1463 (2010).

249. Rodriguez-Hernandez, A., Mayo, M., Jauregui, L. & Patel, P. Autosomal dominant GDAP1 mutation with severe phenotype and respiratory involvement: A case report. *Front. Neurol.* **13**, 905725 (2022).

250. Sipione, S., Monyror, J., Galleguillos, D., Steinberg, N. & Kadam, V. Gangliosides in the Brain: Physiology, Pathophysiology and Therapeutic Applications. *Front. Neurosci.* **14**, 572965 (2020).

251. Kister, A. & Kister, I. Overview of myelin, major myelin lipids, and myelin-associated proteins. *Frontiers in Chemistry* **10**, 1041961 (2023).

252. Nowack, L., Teschers, C. S., Albrecht, S. & Gilmour, R. Oligodendroglial glycolipids in (Re)myelination: implications for multiple sclerosis research. *Nat. Prod. Rep.* **38**, 890–904

(2021).

253. Huber, N., Guimaraes, S., Schrader, M., Suter, U. & Niemann, A. Charcot-Marie-Tooth disease-associated mutants of GDAP1 dissociate its roles in peroxisomal and mitochondrial fission. *EMBO Rep.* **14**, 545–552 (2013).

254. Qu, X. *et al.* Stabilization of dynamic microtubules by mDia1 drives Tau-dependent Aβ1-42 synaptotoxicity. *J. Cell Biol.* **216**, 3161–3178 (2017).

255. Agís-Balboa, R. C. *et al.* Formin 2 links neuropsychiatric phenotypes at young age to an increased risk for dementia. *EMBO J.* **36**, 2815–2828 (2017).

256. Monzo, P. *et al.* Adaptive mechanoproperties mediated by the formin FMN1 characterize glioblastoma fitness for invasion. *Dev. Cell* **56**, 2841–2855.e8 (2021).

257. Website. https://www.jneurosci.org/content/29/15/4794.

258. Peschl, P., Bradl, M., Höftberger, R., Berger, T. & Reindl, M. Myelin oligodendrocyte glycoprotein: Deciphering a target in inflammatory demyelinating diseases. *Front. Immunol.* **8**, 529 (2017).

259. Tiane, A. *et al.* From methylation to myelination: epigenomic and transcriptomic profiling of chronic inactive demyelinated multiple sclerosis lesions. *Acta Neuropathol.* **146**, 283–299 (2023).

260. Segawa, K. *et al.* A sublethal ATP11A mutation associated with neurological deterioration causes aberrant phosphatidylcholine flipping in plasma membranes. *J. Clin. Invest.* **131**, (2021).

261. Bharadwaj, P. & Martins, R. N. PRKAG2 gene expression is elevated and its protein levels are associated with increased amyloid-β accumulation in the Alzheimer's disease brain. *J. Alzheimers. Dis.* **74**, 441–448 (2020).

262. Lin, X., Deng, F. Y., Lu, X. & Lei, S. F. Susceptibility genes for multiple sclerosis identified in a gene-based genome-wide association study. *J. Clin. Neurol.* **11**, 311–318 (2015).

263. Shepard, C. J., Cline, S. G., Hinds, D., Jahanbakhsh, S. & Prokop, J. W. Breakdown of

multiple sclerosis genetics to identify an integrated disease network and potential variant mechanisms. *Physiol. Genomics* **51**, 562–577 (2019).

264. Koenning, M. *et al.* Myelin gene regulatory factor is required for maintenance of myelin and mature oligodendrocyte identity in the adult CNS. *J. Neurosci.* **32**, 12528–12542 (2012).

265. Sanders, A. A. W. M. *et al.* KIAA0556 is a novel ciliary basal body component mutated in Joubert syndrome. *Genome Biol.* **16**, 293 (2015).

266. Romani, M., Micalizzi, A. & Valente, E. M. Joubert syndrome: congenital cerebellar ataxia with the molar tooth. *Lancet Neurol.* **12**, 894–905 (2013).

267. Lasagna-Reeves, C. A. *et al.* Reduction of Nuak1 decreases tau and reverses phenotypes in a tauopathy mouse model. *Neuron* **92**, 407–418 (2016).

268. Cavallin, M. *et al.* WDR81 mutations cause extreme microcephaly and impair mitotic progression in human fibroblasts and Drosophila neural stem cells. *Brain* **140**, 2597–2609 (2017).

269. Chomyk, A. M. *et al.* DNA methylation in demyelinated multiple sclerosis hippocampus. *Sci. Rep.* **7**, 8696 (2017).

270. Cavallin, M. *et al.* Recurrent RTTN mutation leading to severe microcephaly, polymicrogyria and growth restriction. *Eur. J. Med. Genet.* **61**, 755–758 (2018).

271. Malla, S. *et al.* Molecular profiling of frontal and occipital subcortical white matter hyperintensities in Alzheimer's disease. *Neuroscience* (2024).

272. Liu, Q. *et al.* Cerebellum-enriched protein INPP5A contributes to selective neuropathology in mouse model of spinocerebellar ataxias type 17. *Nat. Commun.* **11**, 1101 (2020).

273. Kumari, A. *et al.* Temporal Cortex Microarray Analysis Revealed Impaired Ribosomal Biogenesis and Hyperactivity of the Glutamatergic System: An Early Signature of Asymptomatic Alzheimer's Disease. *Frontiers in Neuroscience* **16**, 966877 (2022).

274. Konki, M. *et al.* Peripheral blood DNA methylation differences in twin pairs discordant for Alzheimer's disease. *Clin. Epigenetics* **11**, 130 (2019).

275. Andrés-Benito, P., Moreno, J., Aso, E., Povedano, M. & Ferrer, I. Amyotrophic lateral sclerosis, gene deregulation in the anterior horn of the spinal cord and frontal cortex area 8: implications in frontotemporal lobar degeneration. *Aging (Albany NY)* **9**, 823–851 (2017).
276. Logie, L. *et al.* Rab-GTPase binding effector protein 2 (RABEP2) is a primed substrate for Glycogen Synthase kinase-3 (GSK3). *Sci. Rep.* **7**, 17682 (2017).
277. Lei, P., Ayton, S., Bush, A. I. & Adlard, P. A. GSK-3 in neurodegenerative diseases. *Int. J. Alzheimers. Dis.* **2011**, 189246 (2011).
278. Poitelon, Y., Kopec, A. M. & Belin, S. Myelin fat facts: An overview of lipids and fatty acid metabolism. *Cells* **9**, 812 (2020).
279. Ye, B. *et al.* ZNF143 in Chromatin Looping and Gene Regulation. *Front. Genet.* **11**, 522878 (2020).
280. Pandey, S. *et al.* Disease-associated oligodendrocyte responses across neurodegenerative diseases. *Cell reports* **40**, (2022).
281. Bernstein, A. I. *et al.* 5-Hydroxymethylation-associated epigenetic modifiers of Alzheimer's disease modulate Tau-induced neurotoxicity. *Human Molecular Genetics* **25**, 2437 (2016).
282. Benito, C. *et al.* Cannabinoid CB2 receptors and fatty acid amide hydrolase are selectively overexpressed in neuritic plaque-associated glia in Alzheimer's disease brains. *J Neurosci* **23**, 11136–11141 (2003).
283. Giussani, P., Prinetti, A. & Tringali, C. The role of Sphingolipids in myelination and myelin stability and their involvement in childhood and adult demyelinating disorders. *J. Neurochem.* **156**, 403–414 (2021).
284. Marian, O. C. *et al.* Disrupted myelin lipid metabolism differentiates frontotemporal dementia caused by GRN and C9orf72 gene mutations. *Acta Neuropathol. Commun.* **11**, 52 (2023).
285. Brugarolas, J. *et al.* Regulation of mTOR function in response to hypoxia by REDD1 and the TSC1/TSC2 tumor suppressor complex. *Genes Dev* **18**, 2893–2904 (2004).

286. Lebrun-Julien, F. *et al.* Balanced mTORC1 Activity in Oligodendrocytes Is Required for Accurate CNS Myelination. *The Journal of Neuroscience* **34**, 8432 (2014).
287. Cuesta, A. *et al.* The gene encoding ganglioside-induced differentiation-associated protein 1 is mutated in axonal Charcot-Marie-Tooth type 4A disease. *Nat. Genet.* **30**, 22–25 (2002).
288. Niemann, A. *et al.* The Gdap1 knockout mouse mechanistically links redox control to Charcot-Marie-Tooth disease. *Brain* **137**, 668–682 (2014).
289. Semikasev, E., Ahlemeyer, B., Acker, T., Schänzer, A. & Baumgart-Vogt, E. Rise and fall of peroxisomes during Alzheimer´s disease: a pilot study in human brains. *Acta Neuropathol. Commun.* **11**, 80 (2023).
290. Dietschy, J. M. Central nervous system: cholesterol turnover, brain development and neurodegeneration. *Biol. Chem.* **390**, 287–293 (2009).
291. Björkhem, I., Leoni, V. & Meaney, S. Genetic connections between neurological disorders and cholesterol metabolism. *J. Lipid Res.* **51**, 2489–2503 (2010).
292. Berghoff, S. A., Spieth, L. & Saher, G. Local cholesterol metabolism orchestrates remyelination. *Trends Neurosci* **45**, 272–283 (2022).
293. Jo, D. S. & Cho, D.-H. Peroxisomal dysfunction in neurodegenerative diseases. *Archives of Pharmacal Research* **42**, 393–406 (2019).
294. Drummond, E. *et al.* Phosphorylated tau interactome in the human Alzheimer's disease brain. *Brain* **143**, 2803 (2020).
295. Casanova, R. *et al.* High dimensional classification of structural MRI Alzheimer's disease data based on large scale regularization. *Front. Neuroinform.* **5**, 22 (2011).
296. Huang, H., Zhou, F., Zhou, S. & Qiu, M. MYRF: A mysterious membrane-bound transcription factor involved in myelin development and human diseases. *Neurosci. Bull.* **37**, 881–884 (2021).
297. Hornig, J. *et al.* The transcription factors Sox10 and Myrf define an essential regulatory network module in differentiating oligodendrocytes. *PLoS Genet.* **9**, e1003907 (2013).

298. Cioffi, F., Adam, R. H. I., Bansal, R. & Broersen, K. A review of oxidative stress products and related genes in early Alzheimer's disease. *J. Alzheimers. Dis.* **83**, 977–1001 (2021).
299. Orobets, K. S. & Karamyshev, A. L. Amyloid precursor protein and Alzheimer's disease. *Int. J. Mol. Sci.* **24**, (2023).
300. Du, Q., Luu, P.-L., Stirzaker, C. & Clark, S. J. Methyl-CpG-binding domain proteins: readers of the epigenome. *Epigenomics* **7**, 1051–1073 (2015).
301. Cerneckis, J., Cai, H. & Shi, Y. Induced pluripotent stem cells (iPSCs): molecular mechanisms of induction and applications. *Signal Transduct. Target. Ther.* **9**, 112 (2024).
302. Lin, G., Mela, A., Guilfoyle, E. M. & Goldman, J. E. Neonatal and adult O4(+) oligodendrocyte lineage cells display different growth factor responses and different gene expression patterns. *J. Neurosci. Res.* **87**, 3390–3402 (2009).
303. De Kleijn, K. M. A., Zuure, W. A., Peijnenborg, J., Heuvelmans, J. M. & Martens, G. J. M. Reappraisal of human HOG and MO3.13 cell lines as a model to study oligodendrocyte functioning. *Cells* **8**, 1096 (2019).
304. Barbarese, E. *et al.* Expression and localization of myelin basic protein in oligodendrocytes and transfected fibroblasts. *J. Neurochem.* **51**, 1737–1745 (1988).
305. Atrophin protein RERE positively regulates Notch targets in the developing vertebrate spinal cord. http://dx.doi.org/10.1111/jnc.13969 doi:10.1111/jnc.13969.
306. Linneberg, C., Harboe, M. & Laursen, L. S. Axo-Glia Interaction Preceding CNS Myelination Is Regulated by Bidirectional Eph-Ephrin Signaling. *ASN NEURO* **7**, 1759091415602859 (2015).
307. Miyamoto, T., Furusawa, C. & Kaneko, K. Pluripotency, differentiation, and reprogramming: A gene expression dynamics model with epigenetic feedback regulation. *PLoS Comput. Biol.* **11**, e1004476 (2015).
308. Liu, C.-C. *et al.* FERMT1 mediates epithelial-mesenchymal transition to promote colon cancer metastasis via modulation of β-catenin transcriptional activity. *Oncogene* **36**, 1779–

1792 (2017).

309. Yu, S., Han, R. & Gan, R. The Wnt/β-catenin signalling pathway in Haematological Neoplasms. *Biomark. Res.* **10**, 74 (2022).

310. Saxena, S., Purohit, A., Varney, M. L., Hayashi, Y. & Singh, R. K. Semaphorin-5A maintains epithelial phenotype of malignant pancreatic cancer cells. *BMC Cancer* **18**, 1283 (2018).

311. Chang, B. *et al.* Epsin is required for Dishevelled stability and Wnt signalling activation in colon cancer development. *Nat. Commun.* **6**, 6380 (2015).

312. Qi, Y. B., Xu, Z., Shen, S., Wang, Z. & Wang, Z. MYRF: A unique transmembrane transcription factor- from proteolytic self-processing to its multifaceted roles in animal development. *Bioessays* **46**, e2300209 (2024).

313. Bernardo, A., Bianchi, D., Magnaghi, V. & Minghetti, L. Peroxisome proliferator-activated receptor-gamma agonists promote differentiation and antioxidant defenses of oligodendrocyte progenitor cells. *J. Neuropathol. Exp. Neurol.* **68**, 797–808 (2009).

314. Single cell type - CTNNA1 - The Human Protein Atlas. https://www.proteinatlas.org/ENSG00000044115-CTNNA1/single+cell.

315. Drees, F., Pokutta, S., Yamada, S., Nelson, W. J. & Weis, W. I. Alpha-catenin is a molecular switch that binds E-cadherin-beta-catenin and regulates actin-filament assembly. *Cell* **123**, 903–915 (2005).

316. Giannini, A. L., Vivanco, M. d. & Kypta, R. M. alpha-catenin inhibits beta-catenin signaling by preventing formation of a beta-catenin*T-cell factor*DNA complex. *J. Biol. Chem.* **275**, 21883–21888 (2000).

317. Feigenson, K., Reid, M., See, J., Crenshaw, E. B., 3rd & Grinspan, J. B. Wnt signaling is sufficient to perturb oligodendrocyte maturation. *Mol. Cell. Neurosci.* **42**, 255–265 (2009).

318. Aggarwal, S., Yurlova, L. & Simons, M. Central nervous system myelin: structure, synthesis and assembly. *Trends Cell Biol.* **21**, 585–593 (2011).

319. Montague, P., McCallion, A. S., Davies, R. W. & Griffiths, I. R. Myelin-Associated Oligodendrocytic Basic Protein: A Family of Abundant CNS Myelin Proteins in Search of a Function. *Dev. Neurosci.* **28**, 479–487 (2006).

320. Gould, R. M., Freund, C. M., Palmer, F. & Feinstein, D. L. Messenger RNAs Located in Myelin Sheath Assembly Sites. *Journal of Neurochemistry* **75**, 1834–1844 (2000).

321. McMillan, C. T. *et al.* Genetic and neuroanatomic associations in sporadic frontotemporal lobar degeneration. *Neurobiol. Aging* **35**, 1473–1482 (2014).

322. Benyamin, B. *et al.* Cross-ethnic meta-analysis identifies association of the GPX3-TNIP1 locus with amyotrophic lateral sclerosis. *Nat. Commun.* **8**, (2017).

323. Sherva, R. *et al.* Genome-wide association study of rate of cognitive decline in Alzheimer's disease patients identifies novel genes and pathways. *Alzheimers. Dement.* **16**, 1134–1145 (2020).

324. Kon, T. *et al.* Immunoreactivity of myelin-associated oligodendrocytic basic protein in Lewy bodies. *Neuropathology* **39**, 279–285 (2019).

325. Manzoni, C. *et al.* Genome-wide analyses reveal a potential role for the MAPT, MOBP, and APOE loci in sporadic frontotemporal dementia. *Am. J. Hum. Genet.* **111**, 1316–1329 (2024).

326. van Rheenen, W. *et al.* Common and rare variant association analyses in Amyotrophic Lateral Sclerosis identify 15 risk loci with distinct genetic architectures and neuron-specific biology. *Research Square* (2021) doi:10.21203/rs.3.rs-322430/v1.

327. Chia, R. *et al.* Genome sequencing analysis identifies new loci associated with Lewy body dementia and provides insights into its genetic architecture. *Nat. Genet.* **53**, 294–303 (2021).

328. Jones, E. *et al.* Identification of novel risk loci and causal insights for sporadic Creutzfeldt-Jakob disease: a genome-wide association study. *Lancet Neurol.* **19**, 840–848 (2020).

329. Nalls, M. A. *et al.* Identification of novel risk loci, causal insights, and heritable risk for

Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* **18**, 1091–1102 (2019).

330. Ng, B. *et al.* An xQTL map integrates the genetic architecture of the human brain's transcriptome and epigenome. *Nat. Neurosci.* **20**, 1418–1426 (2017).
331. Bryois, J. *et al.* Cell-type-specific cis-eQTLs in eight human brain cell types identify novel risk genes for psychiatric and neurological disorders. *Nat. Neurosci.* **25**, 1104–1112 (2022).
332. Giambartolomei, C. *et al.* Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet.* **10**, e1004383 (2014).
333. Liu, B., Gloudemans, M. J., Rao, A. S., Ingelsson, E. & Montgomery, S. B. Abundant associations with gene expression complicate GWAS follow-up. *Nat. Genet.* **51**, 768–769 (2019).
334. Website. https://academic.oup.com/bioinformatics/article/38/15/3844/6617821.
335. Slatkin, M. Linkage disequilibrium--understanding the evolutionary past and mapping the medical future. *Nat. Rev. Genet.* **9**, 477–485 (2008).
336. Strong, M. J., Donison, N. S. & Volkening, K. Alterations in tau metabolism in ALS and ALS-FTSD. *Front. Neurol.* **11**, 598907 (2020).
337. Kouri, N. *et al.* Latent trait modeling of tau neuropathology in progressive supranuclear palsy. *Acta Neuropathologica* **141**, 667–667 (2021).
338. Höglinger, G. U. *et al.* Clinical diagnosis of progressive supranuclear palsy: The movement disorder society criteria. *Mov. Disord.* (2017) doi:10.1002/mds.26987.
339. Wallace, C. A more accurate method for colocalisation analysis allowing for multiple causal variants. *PLoS Genet* **17**, e1009440 (2021).
340. Yokoyama, J. S. *et al.* Shared genetic risk between corticobasal degeneration, progressive supranuclear palsy, and frontotemporal dementia. *Acta neuropathologica* **133**, (2017).
341. Constantinides, V. C., Paraskevas, G. P., Paraskevas, P. G., Stefanis, L. & Kapaki, E. Corticobasal degeneration and corticobasal syndrome: A review. *Clinical Parkinsonism &*

Related Disorders **1**, 66 (2019).

342. Piras, I. S. *et al.* Transcriptional profiling of multiple system atrophy cerebellar tissue highlights differences between the parkinsonian and cerebellar sub-types of the disease. *Acta Neuropathol. Commun.* **8**, (2020).

343. Kantor, B. *et al.* Downregulation of SNCA Expression by Targeted Editing of DNA Methylation: A Potential Strategy for Precision Therapy in PD. *Molecular Therapy* **26**, 2638–2638 (2018).

344. Sapozhnikov, D. M. & Szyf, M. Unraveling the functional role of DNA demethylation at specific promoters by targeted steric blockage of DNA methyltransferase with CRISPR/dCas9. *Nature Communications 2021 12:1* **12**, 1–26 (2021).

345. Smith, J. *et al.* Locus-Specific DNA Methylation Editing in Melanoma Cell Lines Using a CRISPR-Based System. *Cancers* **13**, (2021).

346. Mendonça, V., Soares-Lima, S. C. & Moreira, M. A. M. Exploring cross-tissue DNA methylation patterns: blood-brain CpGs as potential neurodegenerative disease biomarkers. *Commun. Biol.* **7**, 904 (2024).

347. Signaevsky, M. *et al.* Artificial intelligence in neuropathology: deep learning-based assessment of tauopathy. *Lab. Invest.* **99**, 1019–1029 (2019).

348. Vizcarra, J. C. *et al.* Toward a generalizable machine learning workflow for neurodegenerative disease staging with focus on neurofibrillary tangles. *Acta Neuropathol. Commun.* **11**, 202 (2023).

349. Linker, S. M. *et al.* Combined single-cell profiling of expression and DNA methylation reveals splicing regulation and heterogeneity. *Genome Biol.* **20**, 30 (2019).

350. Maunakea, A. K., Chepelev, I., Cui, K. & Zhao, K. Intragenic DNA methylation modulates alternative splicing by recruiting MeCP2 to promote exon recognition. *Cell Res.* **23**, 1256–1269 (2013).

351. Shukla, S. *et al.* CTCF-promoted RNA polymerase II pausing links DNA methylation to

splicing. *Nature* **479**, 74–79 (2011).

352. Buchholz, S. & Zempel, H. The six brain-specific TAU isoforms and their role in Alzheimer's disease and related neurodegenerative dementia syndromes. *Alzheimers. Dement.* **20**, 3606–3628 (2024).

353. Huang, W.-J., Chen, W.-W. & Zhang, X. Multiple sclerosis: Pathology, diagnosis and treatments. *Exp. Ther. Med.* **13**, 3163–3166 (2017).

354. GBD 2019 Dementia Forecasting Collaborators. Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *Lancet Public Health* **7**, e105–e125 (2022).

9. Appendix

Appendix A: OPC/OLG gene lists

Gene list	Gene names
OPC genes	*COL20A1, GPR17, NR0B1, BEST3, HAS2, CACNG5, B3GNT7, NEU4, COL9A1, STK32A, MEGF11, MYT1, GPNMB, SAPCD2, MROH9, RUFY4, SPSB4, BCHE, KLRC4-KLRK1, SMOC1, SOX4, SPC25, CHST9, LHFPL3, SNX22, WSCD1, FZD9, TSPAN11, XYLT1, AMZ1, ATP2C2, FERMT1, CREB3L1, KLHL1, USP24, ACAN, NTN1, HRASLS, ASIC4, VIPR2, WFDC1, BLM, CSPG4, LRRN1, PLEKHH2, GSG1L, VSX1, TRAF4, ASCL1, PCDH15, SEMA3E, CHST8, OLIG2, KCNJ16, EPN2, DNAH11, TNK2, GALR1, CCDC146, NKAIN4, SEMA5A, ABHD2, PDGFRA, FABP7, LYPD1, NLGN3, TMEM255A, LAMA4, GFRA1, LIMD1, BCAN, SPRY4, ZDHHC14, GALNT13, CSPG5, NOS1, CCDC50, NAV1, PRKG2, RAB3IP, SULF2, NLGN4Y, OLIG1, PRTG, ZNF462, SCRG1, DCC, CALCRL, BRINP3, KIF13A, MAP3K1, SEZ6L, CRISPLD2, SORCS3, TEK, TACC2, CSMD2, THBS4, S100A16, RHOC, NLGN4X, AFAP1L2, ALK, PXDN, PEAK1, SCN9A, ADAMTS17, CTTNBP2, DPYSL3, TMEM100, ETV5, C3orf70, SOX13, TIMP4, LRRC4C, STK32B, HIP1R, FBLN2, PHLDA1, GPSM2, C1QL1, PYGO1, TMEM163, TNKS, TTC16, TAOK3, MATN2, GLCCI1, WWP2, GRIK4, CTSO, SLC22A3, INHBA, TGFA, PCOLCE2, CRISPLD1, SETD5, KANK4, BTBD17, ALDH1A3, SUSD5, DSEL, C1QL2, ONECUT1, TPCN2, C1orf106, NME9, GPC2, UPF3A, MMP2, ANGPTL2, EMILIN3, PDPN, COL4A3, GALNT3, TMC2, SOX3, COL4A4, MIDN, CDCA7L, FAM184B, UGDH, B3GAT2, OXTR, KCNG4, TMEM196, NSUN7, ASIC1, PCMTD2, DLL3, KLHL7, WDR11, INSC, ISG15, EBF4, DBX2, CHAD, GLDC, DHRS12, SIM2, MAMLD1, SLC25A27, ANGPT2, HFM1, RCBTB1, PRSS48, ANGEL1, 44630, INTS12, QPRT, KANSL1L, ZNF730, TLL1, ADAMTS7, SENP7, SLC5A9, SLC40A1, TMEM41A, KIF18B, G0S2, TMEM229B, STAT5B, HIBADH, MEX3A, FARP2, LMOD1, HMX1, VAX2, ASB5, SPRED3, EMC10, NUPL2, NAT16, TRA2A, NUP160, BMP8B, GDF6, PIGN, RP1-27O5.3, PPA2, GPR82, NBPF12, FUS, RAB5A, CXorf57, SLC8B1, PPP2R3B, JADE3, SLC15A4, PYCR1, MGARP, CA1, FHDC1, GSTCD, YIF1A, TM6SF2, RP11-403P17.5, LRCH3, CLEC12A, TMEM38B, ARHGEF38, FAM89A, MMP3, BMP2, OPTC, PCCA, SPATA9, SELV, CCKAR, LOR, BRWD3, MAB21L1, PDE6D, KCTD14, ZFP37, EXOG, GPR146, DLK1, DUSP9, ZNF107, SMARCC1, SLC35F5, TRMT10B, TMEM167A, ZNF790, RAB9A, CCDC90B, CHD1L, SLC29A2, SFPQ, KLRG1, NBPF3, AQP5, PIGF, CAPN7, RDH11, ARHGEF35, GALNT8, SHD, HELZ2, FBXL4, PARP12, GLIPR1L2, GSX1, ZNF439, KRBOX4, ZNF677, EPC2, APOA1BP, LECT1, CACNG1, ZNF577, ZNF469, SLC25A15, ZNF300, CST1, AC003006.7, VPS54, MRPS28, TP73, TRIM16, FZD10, DCAF4L2, MYCN, C9orf50, TK2, CNTD2, HCAR1, ZNF85, CTHRC1, TIGD7, QRSL1, ZFP69, COX6A2, MAP3K7CL, METTL3, ATP6V1B1, CD1D, ZMAT5, GDNF, PSG9, ZNF578, FGFBP3, RBM4B, C6orf10, ZNF880, C5orf15, EEA1, MST1R, TMED7, TMPRSS11D, TRIM73, SLC26A5, HPS5, SLC43A3, CLEC18A, PABPC5, OPN1SW, ARMCX6, TMEM39A, ZNF772, LEFTY2, RPS27L, PCDHB8, PZP, LAMB4, S100G, USP41, AMDHD1, PLN, ZNF850, SYNPO2L, SGOL1, LIX1L, MSMP, POLG2, PLEKHG7, COMMD3, PCDHB1, ZKSCAN7, GNRHR, BATF2, CD300LD, ZNF705D, PRSS42, CDH15, SLC17A1, SCAMP3, SFTPA1, WEE1, TAL2, IRX2, HSD17B10, ZNF649, ASF1B, KLRC2, MNX1, CLDN22, DMP1, TMCO5A, NECAP2, ZNF729, RNASEK-C17orf49, PDZD3, C1orf158, DAZL, DEFB136, GOLGA6L2, GOLGA6L9, BPY2, POU5F1B, CRISP1, CTD-2278I10.6, UMOD, BOP1, FAM217A, SNAI1, HLA-G, RFX6, NOTO, SLC17A4, FMR1NB, TPPP2, TBC1D28, HLA-DMA, POTEJ, IRX1, CTSD, NUTM2A, ZNF181, ITPRIPL1, CYP2A6, CLEC18C, C8A, BARHL2, FAM111B, PRR15L, PRDM14, CRISP2, LRRC52, TAS2R8, AMBP, MMP10, BPIFB3, GCM1, PBK, CTSE, BCKDHA, XAGE5, EPS8L1, TM4SF5, CCT8L2, MMP13, FNIP1, C1orf94, COMMD3-BMI1, PRRG2, RP11-166N6.3, VSIG8, C2orf16, CDCP2, CYP2F1, LUZP4, ZNF607, PHOSPHO2, RFPL4B, DMRTB1*
OLG genes	*ST18, CNDP1, PLP1, MAG, OPALIN, MOBP, FOLH1, MYRF, NKX6-2, SLC5A11, TMEM235, ANLN, CARNS1, GPR62, RP11-432B6.3, KLK6, LDB3, MOG, GJB1, ENPP2, HHIP, LRP2, GPR37, ERMN, CPB1, TF, CD22,*

	CERCAM, FAM107B, CDK18, CAPN3, SH3TC2, NINJ2, TMEM125, TP53TG5, SLC31A2, SCD, CLDND1, CLCA4, TMEM98, ADAMTS14, TMEM144, ADAMTS18, PLA2G16, PALM2, KIF6, CTNNA3, LRRC63, S1PR5, CNTN2, PLEKHH1, TMEM165, POLR2F, TMTC4, MAN2A1, C21orf91, BOK, PCSK6, DNAH17, CHRM5, ABCA2, LAMP2, ZDHHC9, GJC2, CBR1, QDPR, MBP, KIF13B, PPP1R14A, TMEM63A, TJAP1, FRYL, MID1IP1, CCP110, PSEN1, PXK, TMC7, SUN2, CLMN, HSPA2, LPGAT1, SLC44A1, MAL, HAPLN2, PRIMA1, CRYAB, TCP11L2, IP6K3, MOSPD2, PDE1C, DYSF, CDC42EP1, PRUNE2, CDH19, ATG4C, ASPA, PIP4K2A, PLEKHG3, LPAR1, RFFL, SPATA22, ITGA2, LIPA, RASGRP3, SEPP1, CA2, ARHGAP23, 44808, CNP, ADAMTS4, HSD17B3, SPINT2, EFHD1, PACS2, MAP6D1, NDE1, PLCL1, PIEZO2, TPRN, UGT8, TTYH2, VRK2, IPO13, AMER2, CPNE2, ARHGAP1, GALNT6, VWA1, GLTP, CDH1, AFMID, NEK3, KCNJ2, SLC25A13, IFIT3, FA2H, RYBP, ICOSLG, LIPE, NT5DC1, GOLGA7, PRR5L, DICER1, TPPP3, MYLIP, TFEB, SYNJ2, LARP6, RNF220, PDE8A, CALD1, CDKN1C, GPIHBP1, GNAI1, SLC45A3, SEMA4D, TSPAN15, ABHD17B, NXPE3, ZDHHC20, SLCO3A1, FAM102A, DIP2B, ITCH, HIPK2, YPEL2, KLHL4, CCDC122, KLHL32, AATK, 44811, RDX, TMTC2, FAM124A, UNC5C, RHOG, NPC1, APOLD1, TMCC2, DEPTOR, EVA1C, STXBP3, DAAM2, SIK3, CPOX, PLEKHB1, RHOU, ZNF397, FAM222A, SLCO1A2, CUEDC1, CDK19, GREB1L, HBS1L, TMEM123, MVB12B, ANKIB1, KDM6A, TBC1D2, WDR20, NENF, LRRC1, SLC24A2, PTBP2, RNASE1, MOB3B, FGF1, TARSL2, PRKCQ, DLG1, PIK3C2B, SLC12A2, SHROOM4, MYO1E, CD55, BIN1, LAMP1, ARHGAP21, ELOVL1, DFNB31, ZNF708, DOCK5, TRIM2, DPYSL5, SH3GL3, CHD7, LSS, SORT1, EXOC6B, NFASC, BAIAP2L2, SLC22A15, COL4A5, FAM13C, DUSP16, SHC4, COLGALT2, FUT8, CHADL, PHLDB1, MAP4K4, MYO1D, C12orf76, C10orf90, ARRDC2, TULP4, DNAJC6, ZNF565, TUBA1A, FAM69C, HS3ST5, HPN, HDAC11, PSME4, GREM1, ADA, ACSS2, RNH1, TRIM59, FCHO1, NSMCE2, PTRF, CFL2, ANAPC5, SLC6A9, C4orf48, TRPM6, KEL, TRAPPC10, PAQR4, FAT4, ANKRD18A, LZTS2, FRK, DOHH
Stringent OPC genes	*COL20A1, GPR17, NR0B1, BEST3, HAS2, CACNG5, B3GNT7, NEU4, COL9A1, STK32A, MEGF11, MYT1, GPNMB, SAPCD2, MROH9, RUFY4, SPSB4, BCHE, KLRC4-KLRK1, SMOC1, SOX4, SPC25, CHST9, LHFPL3, SNX22, WSCD1, FZD9, TSPAN11, XYLT1, AMZ1, ATP2C2, FERMT1, CREB3L1, KLHL1, USP24, ACAN, NTN1, HRASLS, ASIC4, VIPR2, WFDC1, BLM, CSPG4, LRRN1, PLEKHH2, GSG1L, VSX1, TRAF4, ASCL1, PCDH15, SEMA3E, CHST8, OLIG2, KCNJ16, EPN2, DNAH11, TNK2, GALR1, CCDC146, NKAIN4, SEMA5A, ABHD2, PDGFRA, FABP7, LYPD1, NLGN3, TMEM255A, LAMA4, GFRA1*
Stringent OLG genes	*ST18, CNDP1, PLP1, MAG, OPALIN, MOBP, FOLH1, MYRF, NKX6-2, SLC5A11, TMEM235, ANLN, CARNS1, GPR62, RP11-432B6.3, KLK6, LDB3, MOG, GJB1, ENPP2, HHIP, LRP2, GPR37, ERMN, CPB1, TF, CD22, CERCAM, FAM107B, CDK18, CAPN3, SH3TC2, NINJ2, TMEM125, TP53TG5, SLC31A2, SCD, CLDND1, CLCA4, TMEM98, ADAMTS14, TMEM144, ADAMTS18, PLA2G16, PALM2, KIF6, CTNNA3, LRRC63, S1PR5, CNTN2, PLEKHH1, TMEM165, POLR2F, TMTC4, MAN2A1, C21orf91, BOK, PCSK6, DNAH17, CHRM5, ABCA2, LAMP2, ZDHHC9, GJC2, CBR1, QDPR, MBP, KIF13B, PPP1R14A, TMEM63A, TJAP1, FRYL, MID1IP1, CCP110, PSEN1, PXK, TMC7, SUN2, CLMN, HSPA2, LPGAT1, SLC44A1, MAL, HAPLN2, PRIMA1, CRYAB, TCP11L2, IP6K3, MOSPD2, PDE1C, DYSF, CDC42EP1, PRUNE2, CDH19, ATG4C, ASPA, PIP4K2A, PLEKHG3, LPAR1, RFFL, SPATA22, ITGA2, LIPA, RASGRP3, SEPP1, CA2, ARHGAP23, 44808, CNP, ADAMTS4, HSD17B3, SPINT2, EFHD1, PACS2, MAP6D1, NDE1, PLCL1, PIEZO2, TPRN, UGT8, TTYH2, VRK2, IPO13, AMER2, CPNE2, ARHGAP1, GALNT6, VWA1, GLTP, CDH1*

OLG/OPC gene lists used throughout this thesis derived from single-nuclei RNA sequencing data [158] and processed by Dr Piras [159]. For OLG and OPC genes, genes were included as described in Section 2.3 (FDR p-value < 0.05 and ratio ≥ 1.76) . For stringent OPC and OLG gene lists, a filtering strategy based on the expression of that gene in OPCs/OLGs compared to other cell types was applied as described in Section 2.3 (FDR p-value < 0.05 and ratio ≥ 3). OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell, FDR: false-discovery rate.

Appendix B: Results from the FTLD1, FTLD2 and FTLD3 EWAS analyses

Probe	logFC FTLDvsCTRL	P.Value FTLDvsCTRL	Gene	Feature
FTLD1 OPC				
cg19594305	-0.7389965	0.00087542	*CHST8*	1stExon
cg06998031	0.78900016	0.00113898	*XYLT1*	Body
cg12415224	-0.8027056	0.00182561	*XYLT1*	Body
cg10004990	0.64903964	0.00196173	*VSX1*	Body
cg00425944	0.68081313	0.00250536	*ACAN*	Body
cg23889875	1.30016875	0.00258153	*SEMA3E*	1stExon
cg17414248	0.70405921	0.00387719	*ACAN*	5'UTR
cg26836413	1.59923	0.00395594	*PDGFRA*	Body
cg14029669	-1.3434532	0.00405219	*WFDC1*	Body
cg15484988	1.431929	0.00435901	*GALR1*	TSS1500
cg11197909	-0.6619404	0.00570855	*TSPAN11*	3'UTR
cg08719965	0.487121	0.00583543	*GPNMB*	Body
cg01951671	-0.6939151	0.00733734	*KLHL1*	Body
cg02111748	0.827545	0.00755314	*PLEKHH2*	Body
cg00429706	1.085571	0.00761221	*CCDC146*	TSS200
cg23713090	0.77171699	0.00777569	*COL9A1*	Body
cg14487304	0.476144	0.00812086	*SEMA5A*	5'UTR
cg02230373	0.4214282	0.0083862	*AMZ1*	Body
cg01223423	-0.6472784	0.00860494	*SEMA5A*	5'UTR
cg16270670	0.8222554	0.00868788	*KLHL1*	Body
cg16421850	0.5250777	0.0087429	*FERMT1*	Body
cg04848502	0.8004356	0.00972661	*TNK2*	TSS1500
cg03502002	-0.5744363	0.009892	*GALR1*	1stExon
FTLD2 OPC				
cg05602183	-0.5082425	0.00013019	*TRAF4*	Body
cg14487304	-0.757614	0.00073355	*SEMA5A*	5'UTR

cg04645817	-0.8479956	0.0008377	*ABHD2*	Body
cg11052958	-0.6365886	0.00125161	*DNAH11*	Body
cg17448156	-0.4248624	0.00130194	*ABHD2*	5'UTR
cg27607412	0.8937851	0.00134947	*CREB3L1*	Body
cg16601153	-0.5647103	0.00146811	*SMOC1*	Body
cg02954590	-0.7668536	0.00160999	*STK32A*	TSS1500
cg13497866	-0.6142556	0.00186501	*SEMA5A*	5'UTR
cg20762419	-1.135096	0.00186811	*MEGF11*	Body
cg23032674	-0.4398372	0.00198863	*COL9A1*	Body
cg06668073	-0.4274583	0.00240955	*FERMT1*	1stExon
cg15665400	-0.9595823	0.00247184	*LYPD1*	TSS200
cg17860366	-0.90838	0.00273243	*ABHD2*	5'UTR
cg06418646	-1.056016	0.00314078	*EPN2*	TSS200
cg21211413	-0.4256513	0.00354986	*ACAN*	5'UTR
cg27198485	-0.4061729	0.00454962	*CHST8*	5'UTR
cg26187005	-0.6901525	0.00456194	*WSCD1*	TSS1500
cg26047151	0.5693594	0.00467557	*ACAN*	Body
cg22593533	-0.5630169	0.00492138	*OLIG2*	TSS1500
cg06170053	2.127599	0.00517061	*SOX4*	3'UTR
cg02303520	0.3518848	0.00564889	*BLM*	Body
cg01994513	-0.540371	0.00570537	*BLM*	5'UTR
cg11671308	0.69326392	0.00605581	*MEGF11*	Body
cg03885119	-0.5416982	0.00639067	*GFRA1*	Body
cg20181251	-0.4413067	0.00643226	*CREB3L1*	TSS200
cg06997545	0.7462881	0.00714068	*XYLT1*	Body
cg09414426	-0.4871463	0.00733413	*LAMA4*	Body
cg12200950	0.32829403	0.00757922	*CHST8*	5'UTR
cg10397941	2.092119	0.00824359	*CHST9*	TSS200
cg12828294	0.505654	0.00848264	*HAS2*	5'UTR
cg23178192	-0.291465	0.00862625	*GSG1L*	TSS1500

cg17371049	0.4503356	0.00951015	*MEGF11*	Body
cg04232246	0.4804495	0.0095852	*MYT1*	Body
cg04462652	-0.4119489	0.00986454	*LYPD1*	TSS1500
cg10414198	0.3282102	0.00988817	*CREB3L1*	Body
FTLD3 OPC				
cg24825027	0.3720794	0.00073627	*CREB3L1*	TSS1500
cg05855618	-1.689192	0.00078838	*SEMA3E*	1stExon
cg02884053	0.9646468	0.00081628	*VIPR2*	Body
cg18716096	-0.811466	0.00107871	*TNK2*	TSS200
cg11842610	-0.4905241	0.00110194	*GFRA1*	TSS1500
cg13890706	-0.6040917	0.00139539	*GFRA1*	1stExon
cg18438300	0.5409309	0.00207465	*FZD9*	TSS1500
cg17110364	0.7396864	0.00281672	*PLEKHH2*	Body
cg17912112	0.5591033	0.0032162	*LRRN1*	Body
cg14547067	-1.381547	0.00356013	*CREB3L1*	Body
cg18340535	0.5504516	0.00432008	*WSCD1*	Body
cg08927739	-0.3029243	0.00477042	*ASCL1*	TSS1500
cg18443629	0.4921888	0.00590208	*GALR1*	TSS1500
cg25351565	-0.4358636	0.006367	*GFRA1*	TSS1500
cg19818218	0.5541061	0.00729548	*CCDC146*	5'UTR
cg09576074	0.9484115	0.00739336	*WFDC1*	Body
cg25483854	0.4251287	0.00880668	*GFRA1*	Body
cg16375358	0.4456595	0.00880839	*CACNG5*	TSS200
cg20738665	0.4833503	0.00895735	*NKAIN4*	Body
cg04555779	-0.4110074	0.00933346	*LAMA4*	5'UTR
FTLD1 OLG				
cg13010326	1.11572342	6.74E-05	*ANLN*	Body
cg21204870	0.8332753	0.00010403	*PALM2*	Body
cg25560173	-0.896103	0.00033562	*AMER2*	TSS1500

cg10370144	-0.8376445	0.00041853	*PIP4K2A*	Body
cg14506332	0.6867308	0.00072852	*CCP110*	ExonBnd
cg10042572	1.225336	0.00080398	*CTNNA3*	Body
cg16585827	2.114521	0.00154328	*LRP2*	TSS200
cg17849530	0.7074948	0.00280321	*FAM107B*	1stExon
cg18733967	0.5734606	0.00289283	*DNAH17*	Body
cg08817786	1.51848	0.0033508	*EFHD1*	TSS200
cg06592333	-0.6724416	0.0034582	*GPR37*	Body
cg05297352	-0.6387336	0.00393163	*FAM107B*	5'UTR
cg10746396	-0.8199421	0.00422079	*PCSK6*	Body
cg07980015	-0.7254877	0.00432458	*TMTC4*	Body
cg00803762	0.68349949	0.0051869	*CLCA4*	Body
cg04489786	-0.5912267	0.00547777	*LPGAT1*	TSS1500
cg08285862	0.56423214	0.00626887	*CDH1*	Body
cg06868415	0.5095751	0.00633513	*CTNNA3*	Body
cg16864731	0.6362055	0.00649491	*CTNNA3*	Body
cg23659216	0.4137947	0.00700419	*CDH1*	Body
cg00093095	0.6037456	0.00701231	*DNAH17*	Body
cg20037072	-0.7869802	0.00702374	*PIEZO2*	Body
cg25114586	0.3865834	0.00759183	*PACS2*	Body
cg08866897	0.4171995	0.00761641	*MYRF*	Body
cg10503635	2.064732	0.00771467	*MOG*	3'UTR
cg14625975	0.6614531	0.00797631	*ABCA2*	Body
cg07119871	-0.5983757	0.00825134	*PRIMA1*	Body
cg01806713	0.7326345	0.00868674	*MAN2A1*	Body
cg19813868	-1.487371	0.00887614	*QDPR*	TSS1500
cg08220149	-0.517982	0.00889706	*NDE1*	Body
cg26467952	0.5017531	0.00908322	*PRUNE2*	Body
cg18628493	-0.4682805	0.00909713	*TMTC4*	5'UTR
cg24375244	0.4657584	0.00914539	*PCSK6*	Body

cg02956019	-0.8176598	0.00937005	*PCSK6*	Body
cg02835742	1.17569117	0.00949579	*KIF13B*	TSS1500
cg04826663	1.263206	0.00961232	*ANLN*	Body
cg03440556	-0.9276868	0.00988155	*SCD*	Body
FTLD2 OLG				
cg08407007	-0.7350882	4.27E-05	*HSPA2*	5'UTR
cg08909938	0.7958662	0.00013333	*NINJ2*	Body
cg09547698	-0.5485639	0.00033158	*PALM2*	5'UTR
cg11598403	-0.8537922	0.00038872	*MBP*	Body
cg13572782	-0.6652821	0.00052119	*MBP*	Body
cg13543436	0.7015383	0.00095792	*PRUNE2*	Body
cg01786715	-0.9495888	0.00109873	*HSPA2*	1stExon
cg24807354	-0.4595228	0.00115413	*S1PR5*	TSS1500
cg23225103	0.7864026	0.00144214	*CNTN2*	Body
cg06138643	-0.6374172	0.00173103	*DYSF*	Body
cg14371817	-0.9695061	0.00189402	*PLEKHH1*	5'UTR
cg06400428	-2.216374	0.00202769	*SCD*	Body
cg23288103	-0.4925922	0.00203698	*BOK*	Body
cg09025960	-0.5625344	0.00219234	*TMEM63A*	TSS1500
cg12600692	0.6305011	0.00220556	*PLCL1*	Body
cg18563812	-0.926239	0.00234096	*ST18*	Body
cg06304841	-0.9663024	0.00244937	*SPATA22*	TSS1500
cg20162626	-0.9852278	0.00303445	*ENPP2*	Body
cg10662093	-0.3749633	0.00337498	*PPP1R14A*	TSS200
cg14272175	-0.4969135	0.00358419	*PXK*	Body
cg01560493	-0.438741	0.00368364	*MAN2A1*	TSS1500
cg13850625	-0.809915	0.00390843	*TMEM144*	5'UTR
cg10771968	-0.7090675	0.00471617	*CPB1*	Body
cg27379573	-0.5508829	0.00483076	*TMEM235*	5'UTR
cg22413063	0.3766988	0.00499656	*CAPN3*	5'UTR

cg16864731	-0.5226198	0.00549206	*CTNNA3*	Body
cg06912990	-0.6718896	0.00560266	*TMEM144*	1stExon
cg09032863	-0.4716535	0.00565476	*IPO13*	5'UTR
cg12249207	-0.4153343	0.00575364	*C21orf91*	TSS200
cg00954566	-0.853223	0.00585419	*SCD*	TSS1500
cg14102267	-0.68168	0.00616531	*PALM2*	TSS1500
cg12695986	-1.482131	0.0065382	*ITGA2*	TSS200
cg19298588	-0.7513747	0.00661768	*ADAMTS18*	Body
cg18191162	-0.3398265	0.00672503	*KIF6*	TSS200
cg22621867	-0.8039745	0.00692401	*GPR62*	1stExon
cg06424168	0.4166857	0.00724177	*TMC7*	TSS1500
cg20316614	-0.5270099	0.00724912	*TMEM63A*	5'UTR
cg13515395	-0.3199654	0.00766233	*MBP*	Body
cg15375239	-1.094902	0.00802145	*SPINT2*	5'UTR
cg07309124	-0.683337	0.00811736	*CTNNA3*	Body
cg21042919	-0.4805335	0.0082647	*ARHGAP23*	Body
cg23105820	-0.3400106	0.00850634	*MOG*	1stExon
cg00936790	-0.7049655	0.00875874	*KIF13B*	Body
cg13534698	0.6088663	0.00878147	*NDE1*	5'UTR
cg17667454	-0.6937434	0.0088171	*KIF6*	TSS1500
cg03663576	-0.5714469	0.00890287	*ENPP2*	Body
cg00204249	0.3173641	0.00925061	*DNAH17*	Body
cg17604758	-0.6062349	0.00930714	*ANLN*	Body
cg16596039	0.2951763	0.00941052	*PIEZO2*	Body
cg02906707	0.5066916	0.00965292	*PCSK6*	Body
cg26559661	0.2923501	0.0097141	*GALNT6*	Body
cg06784867	-0.3487582	0.00976167	*CNTN2*	Body
cg14849071	-0.5209993	0.00978822	*TMEM63A*	TSS200
cg20481032	-0.5521592	0.00980069	*PRIMA1*	Body
FTLD3 OLG				

cg11965880	-2.679896	1.31E-05	*PIP4K2A*	TSS200
cg03558769	0.9217398	4.98E-05	*BOK*	TSS1500
cg07354124	-0.4845015	8.22E-05	*TCP11L2*	TSS1500
cg05523056	-0.9266833	0.00026453	*ADAMTS18*	Body
cg23109129	-1.902458	0.00054699	*HHIP*	Body
cg16004911	-4.905445	0.00068653	*FRYL*	TSS1500
cg01870140	-1.346707	0.00069201	*TMTC4*	TSS200
cg03142956	-0.5795641	0.00114831	*CNTN2*	1stExon
cg23053525	0.4573756	0.00144186	*DNAH17*	Body
cg21815781	-0.6416225	0.00179734	*PLEKHH1*	TSS200
cg26371957	0.9821989	0.0024596	*NINJ2*	Body
cg00477086	0.609925	0.00299513	*DYSF*	Body
cg10734940	-0.7988895	0.00300492	*PSEN1*	Body
cg14272175	-0.5981389	0.00315265	*PXK*	Body
cg17012863	-1.384354	0.00366199	*LIPA*	TSS200
cg14245120	0.7782098	0.00413889	*KIF6*	Body
cg27026695	0.5063622	0.0045363	*HSPA2*	TSS1500
cg20264966	-0.9257232	0.0046845	*CTNNA3*	Body
cg06252382	-1.456909	0.00578693	*SLC31A2*	Body
cg00148325	0.3315188	0.00593041	*TTYH2*	TSS200
cg07230380	-1.601482	0.00600129	*SCD*	TSS1500
cg09444531	0.5188627	0.0070737	*ENPP2*	Body
cg08645980	0.7751442	0.00723884	*LRP2*	3'UTR
cg17275844	-0.8360219	0.00754782	*FRYL*	3'UTR
cg19756789	-0.5243171	0.00755741	*RASGRP3*	1stExon
cg18568335	0.5080226	0.00760835	*PACS2*	Body
cg04934643	0.557157	0.00763676	*PACS2*	Body
cg25272394	0.663744	0.00770489	*CLCA4*	Body
cg26787256	0.5185892	0.00808321	*DYSF*	Body
cg18501001	0.4494832	0.00947437	*DNAH17*	Body

cg14875449	-0.5759124	0.00984583	*PXK*	1stExon

Differentially methylated sites mapping to OLG and OPC genes from the FTLD1, FTLD2 and FTLD3 EWAS. FTLD: frontotemporal lobar degeneration, OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell.

Appendix C: Differential expression of genes identified in the FTLD1, 2 and 3 EWAS analyses

Gene	Probe	logFC_FTLDvsCTRL_DNAmeth	Feature	logFC_Expression	P.Value_Expression	Expression	Methylationstatus
FTLD1_OPC							
ACAN	cg00425944	0.681	Body	0.462	0.000	Up	Hyper
ACAN	cg17414248	0.704	5'UTR	0.462	0.000	Up	Hyper
AMZ1	cg02230373	0.421	Body	-0.280	0.002	Down	Hyper
CCDC146	cg00429706	1.086	TSS200	0.144	0.026	Up	Hyper
GPNMB	cg08719965	0.487	Body	0.430	0.013	Up	Hyper
SEMA3E	cg23889875	1.300	1stExon	-0.432	0.042	Down	Hyper
TNK2	cg04848502	0.800	TSS1500	-0.148	0.029	Down	Hyper
VSX1	cg10004990	0.649	Body	0.404	0.001	Up	Hyper
WFDC1	cg14029669	-1.343	Body	-0.306	0.003	Down	Hypo
FTLD1_OLG							
AMER2	cg25560173	-0.896	TSS1500	0.268	0.004	Up	Hypo
CDH1	cg08285862	0.564	Body	0.442	0.017	Up	Hyper
CDH1	cg23659216	0.414	Body	0.442	0.017	Up	Hyper
FAM107B	cg17849530	0.707	1stExon	0.437	0.001	Up	Hyper
FAM107B	cg05297352	-0.639	5'UTR	0.437	0.001	Up	Hypo
GPR37	cg06592333	-0.672	Body	0.360	0.015	Up	Hypo
LPGAT1	cg04489786	-0.591	TSS1500	-0.239	0.003	Down	Hypo
LRP2	cg16585827	2.115	TSS200	0.511	0.006	Up	Hyper
MAN2A1	cg01806713	0.733	Body	0.307	0.002	Up	Hyper
MYRF	cg08866897	0.417	Body	0.359	0.033	Up	Hyper
PIEZO2	cg20037072	-0.787	Body	0.385	0.013	Up	Hypo
PIP4K2A	cg10370144	-0.838	Body	0.259	0.017	Up	Hypo
PRIMA1	cg07119871	-0.598	Body	0.601	0.000	Up	Hypo
QDPR	cg19813868	-1.487	TSS1500	0.315	0.003	Up	Hypo

FTLD2_OPC							
ACAN	cg26047151	0.569	Body	0.512	0.030	Up	Hyper
ACAN	cg21211413	-0.426	5'UTR	0.512	0.030	Up	Hypo
CHST8	cg12200950	0.328	5'UTR	-0.660	0.002	Down	Hyper
CHST8	cg27198485	-0.406	5'UTR	-0.660	0.002	Down	Hypo
CHST9	cg10397941	2.092	TSS200	0.513	0.012	Up	Hyper
DNAH11	cg11052958	-0.637	Body	1.086	0.006	Up	Hypo
FERMT1	cg06668073	-0.427	1stExon	0.556	0.008	Up	Hypo
GSG1L	cg23178192	-0.291	TSS1500	0.530	0.004	Up	Hypo
MEGF11	cg11671308	0.693	Body	0.208	0.032	Up	Hyper
MEGF11	cg17371049	0.450	Body	0.208	0.032	Up	Hyper
MEGF11	cg20762419	-1.135	Body	0.208	0.032	Up	Hypo
SMOC1	cg16601153	-0.565	Body	0.698	0.011	Up	Hypo
SOX4	cg06170053	2.128	3'UTR	0.372	0.012	Up	Hyper
WSCD1	cg26187005	-0.690	TSS1500	0.551	0.002	Up	Hypo
FTLD2_OLG							
CPB1	cg10771968	-0.709	Body	-1.363	0.010	Down	Hypo
SPINT2	cg15375239	-1.095	5'UTR	-0.805	0.001	Down	Hypo

Differentially expressed genes to which methylation sites mapped to in the FTLD1, FTLD2 and FTLD3 EWAS. FTLD: frontotemporal lobar degeneration, OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell.

Appendix D: Results from the FTLD1, FTLD2 and FTLD3 meta-analysis

CpG	TE.random	P-value random	Chromosome	Gene	Position
OPC Genes					
cg18716096	-0.503519734	0.00056605	3	*TNK2*	TSS200
cg18443629	0.334404672	0.000927	18	*GALR1*	TSS1500
cg06580318	-0.291271963	0.001135707	2	*SPC25*	TSS200
cg09368832	0.412167423	0.001742224	17	*NTN1*	3'UTR
cg25617725	-0.293112611	0.00180402	10	*GFRA1*	TSS1500
cg17154602	0.362747343	0.004265685	18	*GALR1*	TSS1500
cg21690489	-0.230764176	0.004589552	11	*CREB3L1*	Body
cg20872937	-0.644242889	0.004833112	18	*GALR1*	TSS200
cg14454477	0.453642849	0.005796448	2	*PLEKHH2*	Body
cg27462405	-0.342714813	0.007437921	5	*SEMA5A*	5'UTR
cg06271623	0.518119015	0.007710976	2	*RUFY4*	Body
cg20095233	-0.305913628	0.009561602	10	*GFRA1*	TSS1500
OLG Genes					
cg12600692	0.528939091	1.78E-05	2	*PLCL1*	Body
cg14849071	-0.55108699	1.95E-04	1	*TMEM63A*	TSS200
cg14272175	-0.400489864	2.59E-04	3	*PXK*	Body
cg07119871	-0.486667472	3.58E-04	14	*PRIMA1*	Body
cg27026695	0.386164257	6.26E-04	14	*HSPA2*	TSS1500
cg12571928	-0.412468723	7.38E-04	2	*BOK*	Body
cg17122437	0.365888198	7.74E-04	10	*FAM107B*	1stExon
cg01806713	0.48626362	0.001524677	5	*MAN2A1*	Body
cg04905644	0.319038166	0.001706192	8	*KIF13B*	Body
cg26421140	0.203980251	0.002135411	14	*HSPA2*	TSS1500
cg06998038	0.308756865	0.002415002	14	*PLEKHG3*	Body
cg01786715	-0.539645912	0.002556728	14	*HSPA2*	1stExon

cg24807354	-0.310620376	0.00264963	19	*S1PR5*	TSS1500
cg25139187	-0.335516953	0.003287719	1	*HAPLN2*	Body
cg12162778	-0.397636053	0.003836848	14	*PSEN1*	5'UTR
cg14775560	-0.718265541	0.004025309	11	*FOLH1*	1stExon
cg02396982	0.272952911	0.004763075	16	*NDE1*	Body
cg08187425	-0.260382685	0.005338688	2	*EFHD1*	TSS1500
cg16719517	-0.67921887	0.005718725	21	*CBR1*	1stExon
cg08589981	0.58294705	0.005806083	10	*PIP4K2A*	Body
cg00807430	-0.25141277	0.006169178	5	*MAN2A1*	1stExon
cg02960853	-0.299005552	0.006226633	7	*GPR37*	1stExon
cg13993734	0.359257057	0.007142274	19	*MAG*	Body
cg16747785	0.360384725	0.007368041	14	*CLMN*	Body
cg11311579	-0.440685595	0.008169193	10	*SCD*	TSS200
cg19098785	0.291765283	0.008428158	1	*CDK18*	5'UTR
cg26677958	0.262819459	0.008618372	14	*PACS2*	Body
cg01560493	-0.238625822	0.008720767	5	*MAN2A1*	TSS1500
cg12768447	0.49044515	0.009079316	10	*CTNNA3*	TSS1500
cg16559695	-0.29379934	0.009378443	14	*PACS2*	3'UTR
cg15375239	-0.604902952	0.009685521	19	*SPINT2*	5'UTR

Differentially methylated sites mapping to OLG and OPC genes from the FTLD1, FTLD2 and FTLD3 meta-analysis. FTLD: frontotemporal lobar degeneration, OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell.

Appendix E: Results from FTLD-sorted EWAS

Probe	Gene	Feature	Delta M	P.Value
OPC Genes				
cg20588045	*PCDH15*	TSS200	0.70951951	0.0001668
cg12092351	*GSG1L*	Body	0.72406599	0.00029508
cg26321126	*MEGF11*	Body	0.99596103	0.00076029
cg17783815	*NTN1*	Body	0.66983169	0.0008812
cg14789080	*KLHL1*	Body	0.58409469	0.00111138
cg03570006	*ATP2C2*	Body	1.02343102	0.00163563
cg07229946	*ABHD2*	5'UTR	-0.738612	0.0019155
cg08543143	*CHST8*	Body	1.28546765	0.00259315
cg21191541	*TNK2*	5'UTR	0.86932414	0.00308807
cg21950367	*KLRC4-KLRK1*	5'UTR	0.66088237	0.00314037
cg01439383	*AMZ1*	5'UTR	0.64505164	0.0033725
cg25617725	*GFRA1*	TSS1500	0.58920422	0.00472367
cg09699193	*SEMA5A*	5'UTR	0.63812851	0.0051067
cg14029669	*WFDC1*	Body	-0.8371062	0.00530147
cg05144285	*VIPR2*	Body	0.88131434	0.00538402
cg00556126	*WFDC1*	TSS200	0.71607837	0.00550451
cg08909835	*USP24*	Body	0.86308607	0.00575563
cg26577658	*ATP2C2*	Body	-0.6434043	0.00612552
cg20919596	*AMZ1*	TSS1500	0.48907194	0.00649895
cg05711251	*AMZ1*	Body	0.5759589	0.00703424

cg06433369	*MYT1*	Body	0.59301831	0.00705627
cg12169338	*WSCD1*	Body	0.63339923	0.00712321
cg13944537	*MEGF11*	5'UTR	-2.5569339	0.00739029
cg00545918	*COL20A1*	Body	0.46559438	0.00743353
cg23865698	*WFDC1*	TSS1500	0.52310609	0.00760983
cg26066180	*COL20A1*	TSS1500	-1.1485145	0.00815615
cg17110364	*PLEKHH2*	Body	0.49579913	0.00846157
cg00452694	*HRASLS*	ExonBnd	0.81104928	0.00848034
cg14006448	*COL9A1*	Body	0.93771172	0.0090853
cg17922749	*GSG1L*	Body	0.50341686	0.00915629
cg25849086	*USP24*	3'UTR	0.57171346	0.00948906
cg03483267	*FERMT1*	5'UTR	-0.8672395	0.00957876
cg21609684	*CSPG4*	1stExon	0.51772177	0.00973855
cg10276400	*XYLT1*	Body	0.7813862	0.00978865
cg01661350	*KLHL1*	1stExon	-0.5425529	0.00999528
OLG Genes				
cg10415442	*ST18*	5'UTR	-1.2198579	0.00038771
cg20723425	*DNAH17*	Body	0.60867088	0.00040227
cg16176654	*FAM107B*	Body	0.68918859	0.00071179
cg21696316	*MOBP*	5'UTR	1.51241846	0.00075903
cg14465747	*PDE1C*	TSS1500	0.7378028	0.00095886
cg00932128	*HSPA2*	TSS1500	-0.8732706	0.00134025
cg19113053	*KIF6*	Body	0.72414173	0.00150166

cg05784996	*PIEZO2*	Body	0.73389475	0.00174207
cg06028605	*SLC5A11*	5'UTR	-0.5124864	0.00177054
cg20198768	*MOG*	3'UTR	-0.7047336	0.00178449
cg12600692	*PLCL1*	Body	-0.9061647	0.00182993
cg07068376	*IPO13*	Body	0.59792189	0.00188306
cg05778847	*PPP1R14A*	Body	-0.9922623	0.00207578
cg26347632	*CDC42EP1*	3'UTR	1.08269643	0.00213249
cg02691058	*ST18*	3'UTR	0.62380579	0.00218387
cg05304658	*ADAMTS18*	Body	0.49142331	0.00230198
cg08808571	*QDPR*	Body	0.64548631	0.00237343
cg11552370	*NINJ2*	Body	0.64867318	0.00242588
cg15783427	*MBP*	Body	0.55201912	0.002526
cg14940461	*KIF6*	Body	0.64658781	0.00280886
cg19962468	*SH3TC2*	1stExon	0.55026426	0.00281129
cg17275551	*KIF6*	Body	0.51788159	0.00309269
cg04906616	*POLR2F*	Body	0.7699259	0.00328913
cg13882486	*LPAR1*	5'UTR	0.65557069	0.00344196
cg00702126	*VWA1*	TSS1500	-1.4712543	0.00376084
cg21481658	*RFFL*	Body	0.70184717	0.00431881
cg03072665	*BOK*	Body	-1.7055425	0.00440333
cg24400204	*PRUNE2*	Body	-0.4936703	0.00450896
cg21816234	*CTNNA3*	Body	0.43443761	0.00476436
cg09281830	*KIF13B*	Body	0.55474027	0.00477658

cg11139063	*TMEM165*	Body	-0.6742209	0.0049246
cg18419549	*GALNT6*	5'UTR	0.52117047	0.00544792
cg22447679	*TMEM63A*	Body	0.56385719	0.00580976
cg16544209	*PRUNE2*	TSS200	-2.1972918	0.00691746
cg07649988	*SCD*	TSS200	0.63222166	0.00748193
cg13735453	*CLMN*	Body	0.54556979	0.00756862
cg07063370	*POLR2F*	Body	-0.7402299	0.00772039
cg25678491	*TMTC4*	TSS200	-1.1461034	0.00808232
cg03760081	*FAM107B*	5'UTR	0.51443107	0.00815369
cg24855781	*CTNNA3*	Body	0.51771535	0.00828858
cg05264408	*PCSK6*	Body	0.52810523	0.0083788
cg08899203	*RASGRP3*	Body	0.57062988	0.00880356
cg07099998	*SLC5A11*	TSS1500	0.67899381	0.0088213
cg19040750	*CDK18*	5'UTR	0.61710267	0.00883616
cg15651293	*PRIMA1*	Body	0.44730736	0.00924092
cg25499397	*GPR62*	1stExon	0.55918941	0.00936414
cg11027058	*DYSF*	5'UTR	0.58612938	0.00945425
cg03641375	*FAM107B*	5'UTR	-0.7418409	0.00974516
cg14399369	*VRK2*	TSS200	-0.5633325	0.00981632
cg25432316	*PLEKHH1*	Body	0.58575896	0.00990618

Differentially methylated OLG and OPC methylation sites from brain-nuclei sorted FTLD EWAS reaching our chosen significance threshold ($p < 0.01$). FTLD: frontotemporal lobar degeneration, OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell.

Appendix F: Results from the AD EWAS analysis

Probe	Gene	logFC	P.Value	adj.P.Val	Feature
OPC Genes					
cg20692569	*FZD9*	-0.116	0.0000	0.0150	1stExon
cg00916179	*PLEKHH2*	-0.084	0.0001	0.0429	3'UTR
cg17640485	*TNK2*	0.101	0.0001	0.0488	Body
cg22274825	*SOX4*	-0.141	0.0001	0.0505	TSS200
cg19178642	*ABHD2*	0.094	0.0001	0.0671	5'UTR
cg23865698	*WFDC1*	0.047	0.0009	0.1546	TSS1500
cg17650274	*VSX1*	-0.106	0.0013	0.1796	TSS1500
cg26338723	*GFRA1*	-0.135	0.0015	0.1920	Body
cg14273027	*KLHL1*	-0.086	0.0022	0.2220	TSS200
cg13016408	*LYPD1*	-0.090	0.0030	0.2503	5'UTR
cg02631468	*VSX1*	-0.088	0.0031	0.2548	TSS200
cg15484988	*GALR1*	0.137	0.0039	0.2793	TSS1500
cg17274742	*GPNMB*	0.083	0.0040	0.2803	1stExon
cg03075065	*WFDC1*	-0.061	0.0041	0.2830	Body
cg11970679	*DNAH11*	-0.077	0.0046	0.2950	Body
cg18285681	*AMZ1*	0.037	0.0056	0.3154	5'UTR
cg03175653	*GALR1*	0.035	0.0061	0.3263	TSS1500
cg20473723	*FERMT1*	0.182	0.0064	0.3308	TSS1500
cg17158750	*COL20A1*	0.053	0.0071	0.3420	TSS200
cg11835978	*SEMA5A*	0.188	0.0071	0.3430	3'UTR
cg17792108	*CACNG5*	0.053	0.0073	0.3467	TSS1500
cg25012274	*NTN1*	0.065	0.0075	0.3488	Body
cg24182468	*VIPR2*	-0.055	0.0078	0.3518	Body
cg17742418	*EPN2*	0.041	0.0078	0.3534	5'UTR
cg16465663	*MEGF11*	0.099	0.0089	0.3666	Body
cg10979891	*SMOC1*	0.065	0.0092	0.3716	TSS1500

cg04986899	*XYLT1*	-0.065	0.0100	0.3815	Body
OLG Genes					
cg01941881	*PACS2*	-0.10263	0.00001	0.02123	Body
cg16179521	*PCSK6*	0.07483	0.00003	0.03306	Body
cg10464462	*RFFL*	0.08819	0.00008	0.04923	TSS1500
cg02363711	*TMEM144*	-0.10500	0.00008	0.05098	Body
cg12461930	*GJC2*	0.08076	0.00016	0.07292	5'UTR
cg05706061	*SLC31A2*	-0.07092	0.00032	0.09882	TSS1500
cg20704602	*MOG*	-0.11282	0.00036	0.10616	3'UTR
cg00479347	*PACS2*	-0.08688	0.00057	0.12836	Body
cg02534163	*ENPP2*	-0.16252	0.00079	0.14757	1stExon
cg01994290	*PLEKHG3*	0.04983	0.00082	0.14934	TSS1500
cg10772086	*TF*	-0.06499	0.00100	0.16047	TSS1500
cg24765079	*CDH1*	-0.08609	0.00103	0.16224	Body
cg11950778	*UGT8*	-0.05604	0.00108	0.16655	TSS1500
cg07637837	*MBP*	0.10646	0.00117	0.17174	5'UTR
cg15035133	*CDC42EP1*	0.05920	0.00159	0.19437	Body
cg09627520	*PXK*	-0.12641	0.00168	0.19888	3'UTR
cg25677261	*PCSK6*	0.07272	0.00168	0.19888	Body
cg17655614	*CDH1*	-0.04394	0.00168	0.19912	TSS1500
cg20123637	*TF*	-0.09411	0.00182	0.20600	Body
cg22908922	*PACS2*	0.07143	0.00183	0.20663	Body
cg01644592	*MOG*	-0.07124	0.00186	0.20749	3'UTR
cg03065165	*GJC2*	0.03886	0.00187	0.20812	TSS200
cg15552051	*PPP1R14A*	-0.11182	0.00196	0.21220	Body
cg02428538	*SLC5A11*	0.11445	0.00225	0.22397	TSS1500
cg16419054	*PACS2*	-0.03944	0.00235	0.22819	Body
cg15069948	*MOBP*	-0.11550	0.00277	0.24372	Body
cg24402667	*ARHGAP23*	-0.04957	0.00298	0.25112	TSS1500
cg09317371	*CPNE2*	-0.03429	0.00329	0.26052	Body

cg04105230	*PCSK6*	0.07950	0.00333	0.26148	Body
cg03018256	*TTYH2*	-0.05343	0.00386	0.27777	Body
cg22810875	*NINJ2*	-0.04092	0.00400	0.28112	3'UTR
cg03147503	*MOG*	-0.06631	0.00462	0.29622	3'UTR
cg21927177	*PLEKHG3*	0.03513	0.00471	0.29832	5'UTR
cg07318204	*HHIP*	-0.07353	0.00499	0.30415	TSS1500
cg19178876	*MBP*	-0.08628	0.00501	0.30432	3'UTR
cg03917584	*MAG*	-0.09676	0.00503	0.30460	Body
cg17036418	*RFFL*	0.09743	0.00505	0.30490	TSS1500
cg02475474	*MOG*	-0.04385	0.00506	0.30520	3'UTR
cg05279622	*MOG*	-0.06362	0.00523	0.30920	Body
cg14833933	*CAPN3*	-0.04203	0.00551	0.31483	1stExon
cg01279990	*ARHGAP23*	0.06418	0.00569	0.31854	Body
cg26341831	*TMEM63A*	0.04017	0.00642	0.33101	Body
cg07020596	*CTNNA3*	-0.07216	0.00667	0.33579	Body
cg23053525	*DNAH17*	0.07406	0.00681	0.33765	Body
cg25060035	*TTYH2*	-0.03932	0.00714	0.34391	Body
cg01611017	*CLMN*	0.04558	0.00715	0.34404	Body
cg09138865	*LDB3*	0.03119	0.00717	0.34442	TSS1500
cg12236045	*CTNNA3*	0.15215	0.00723	0.34549	Body
cg10069691	*ST18*	-0.05922	0.00729	0.34602	5'UTR
cg16563470	*CNP*	-0.05273	0.00739	0.34723	Body
cg24487969	*TMTC4*	-0.14258	0.00756	0.34924	Body
cg10944735	*DYSF*	0.07196	0.00796	0.35488	TSS200
cg08220149	*NDE1*	-0.08923	0.00849	0.36281	Body
cg03238797	*ADAMTS18*	-0.15741	0.00858	0.36414	1stExon
cg14875449	*PXK*	-0.08785	0.00876	0.36574	1stExon
cg23761616	*TMEM144*	-0.05979	0.00897	0.36855	TSS200
cg09662369	*NDE1*	0.12975	0.00903	0.36959	TSS200
cg06570930	*PPP1R14A*	-0.07620	0.00934	0.37349	Body

cg22681074	*GJC2*	0.03656	0.00948	0.37520	5'UTR
cg06880930	*CPNE2*	-0.08852	0.00962	0.37707	Body

Differentially methylated methylation sites (P < 0.01) mapping to OLG/OPC genes in the bulk AD data. AD: Alzheimer's disease, OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell.

Appendix G: Differentially expressed genes identified in the AD EWAS

Gene	CpG	Delta M CpG Methylation	P.Value CpG Methylation	Feature	logFC Gene Expression	P.Value Gene expression
OPC genes						
LYPD1	cg13016408	-0.0904	0.0030	5'UTR	-0.1242	0.0084
GALR1	cg15484988	0.1372	0.0039	TSS1500	0.1248	0.0265
DNAH11	cg11970679	-0.0766	0.0046	Body	-0.1989	0.0263
GALR1	cg03175653	0.0354	0.0061	TSS1500	0.1248	0.0265
EPN2	cg17742418	0.0406	0.0078	5'UTR	-0.0438	0.0429
FZD9	cg20692569	-0.1155	4.1842e-06	1stExon	-0.1513	0.0042
SOX4	cg22274825	-0.1412	7.9940e-05	TSS200	-0.1144	0.0064
OLG Genes						
PACS2	cg01941881	-0.1026	1.2991e-05	Body	-0.0402	0.0055
TF	cg20123637	-0.0941	0.0018	Body	-0.1637	0.0027
NDE1	cg08220149	-0.0892	0.0085	Body	-0.1870	5.4491e-06
CPNE2	cg06880930	-0.0885	0.0096	Body	-0.0835	0.0031
PACS2	cg00479347	-0.0869	0.0006	Body	-0.0402	0.0055
CDH1	cg24765079	-0.0861	0.0010	Body	-0.1417	0.0260
HHIP	cg07318204	-0.0735	0.0050	TSS1500	-0.1476	0.0154
CTNNA3	cg07020596	-0.0722	0.0067	Body	-0.1575	0.0094
SLC31A2	cg05706061	-0.0709	0.0003	TSS1500	-0.1224	0.0153
TF	cg10772086	-0.0650	0.0010	TSS1500	-0.1637	0.0027
ST18	cg10069691	-0.0592	0.0073	5'UTR	-0.1714	0.0018
UGT8	cg11950778	-0.0560	0.0011	TSS1500	-0.2265	0.0004
TTYH2	cg03018256	-0.0534	0.0039	Body	-0.1032	0.0224
ARHGAP23	cg24402667	-0.0496	0.0030	TSS1500	-0.0753	0.0007
CDH1	cg17655614	-0.0439	0.0017	TSS1500	-0.1417	0.0260
NINJ2	cg22810875	-0.0409	0.0040	3'UTR	-0.1455	0.0115
PACS2	cg16419054	-0.0394	0.0023	Body	-0.0402	0.0055

TTYH2	cg25060035	-0.0393	0.0071	Body	-0.1032	0.0224
CPNE2	cg09317371	-0.0343	0.0033	Body	-0.0835	0.0031
PLEKHG3	cg21927177	0.0351	0.0047	5'UTR	-0.0884	0.0364
GJC2	cg22681074	0.0366	0.0095	5'UTR	-0.0930	0.0240
GJC2	cg03065165	0.0389	0.0019	TSS200	-0.0930	0.0240
TMEM63A	cg26341831	0.0402	0.0064	Body	-0.1343	0.0058
CLMN	cg01611017	0.0456	0.0071	Body	-0.1970	8.7250e-05
PLEKHG3	cg01994290	0.0498	0.0008	TSS1500	-0.0884	0.0364
CDC42EP1	cg15035133	0.0592	0.0016	Body	-0.1310	0.0018
ARHGAP23	cg01279990	0.0642	0.0057	Body	-0.0753	0.0007
PACS2	cg22908922	0.0714	0.0018	Body	-0.0402	0.0055
DYSF	cg10944735	0.0720	0.0080	TSS200	-0.1119	0.0049
PCSK6	cg25677261	0.0727	0.0017	Body	-0.1734	0.0006
DNAH17	cg23053525	0.0741	0.0068	Body	-0.1475	0.0283
PCSK6	cg16179521	0.0748	3.3244e-05	Body	-0.1734	0.0006
PCSK6	cg04105230	0.0795	0.0033	Body	-0.1734	0.0006
GJC2	cg12461930	0.0808	0.0002	5'UTR	-0.0930	0.0240
RFFL	cg10464462	0.0882	7.7249e-05	TSS1500	-0.1056	0.0003
RFFL	cg17036418	0.0974	0.0050	TSS1500	-0.1056	0.0003
SLC5A11	cg02428538	0.1145	0.0022	TSS1500	-0.2456	0.0033
NDE1	cg09662369	0.1297	0.0090	TSS200	-0.1870	5.4491e-06
CTNNA3	cg12236045	0.1522	0.0072	Body	-0.1575	0.0094

Significantly differentially methylated probes mapping to genes that are significantly differentially expressed in bulk AD data.CpG methylation and gene expression of significantly differentially methylated ($P < 0.01$) which also map to genes which are significantly differentially expressed ($P < 0.05$). AD: Alzheimer's disease, OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell.

Appendix H: Results from the AD1 and AD2 EWAS

Probe	Gene	Feature	Delta M	P.Value
AD1 OPC Genes				
cg04525189	*VIPR2*	Body	-0.428006919	0.000381785
cg20791593	*NEU4*	TSS1500	0.381533044	0.000714713
cg04913160	*ABHD2*	1stExon	0.339511987	0.000733702
cg14173476	*EPN2*	5'UTR	0.305605811	0.000836407
cg15299832	*OLIG2*	TSS200	0.347481408	0.002155247
cg18493027	*SMOC1*	5'UTR	-0.275376224	0.002548964
cg16933181	*CHST9*	TSS200	0.321088356	0.002715211
cg07099991	*WFDC1*	Body	0.298544835	0.002810256
cg15042332	*FABP7*	5'UTR	0.441169	0.002967258
cg22736323	*PDGFRA*	1stExon	-0.362727447	0.004345708
cg25164490	*SPSB4*	Body	0.42217454	0.004398757
cg06039355	*GFRA1*	5'UTR	0.342250606	0.004403473
cg20588045	*PCDH15*	TSS200	0.303864393	0.004534342
cg12162377	*B3GNT7*	Body	-0.25843262	0.004575852
cg00164894	*USP24*	Body	0.239656027	0.005062574
cg27418204	*TNK2*	Body	0.276391066	0.005907791
cg10071275	*MYT1*	5'UTR	0.306768601	0.006231633
cg25857018	*BLM*	TSS200	0.279950974	0.006317593
cg25156843	*ABHD2*	5'UTR	0.275076175	0.006364144
cg22235407	*MEGF11*	Body	0.313078601	0.007522292
cg11959156	*MYT1*	Body	0.401388784	0.007621494
cg10354232	*LYPD1*	Body	0.409098519	0.008187175
cg07575407	*MEGF11*	5'UTR	0.350284376	0.008365171
AD1 OLG Genes				
cg16985259	*IPO13*	5'UTR	0.354612413	3.17E-05
cg24402667	*ARHGAP23*	TSS1500	0.516080314	0.0001973

cg22238209	*MAG*	Body	-0.597199878	0.000394012
cg13311357	*FRYL*	TSS200	0.359260178	0.000576944
cg15923042	*QDPR*	Body	0.352396179	0.001385939
cg04007987	*TTYH2*	Body	0.414274168	0.001670194
cg00524271	*PLCL1*	TSS1500	0.337585692	0.002183969
cg10251094	*DNAH17*	Body	0.353749432	0.002193613
cg23394510	*GLTP*	Body	0.397719217	0.002269928
cg23428192	*TMTC4*	Body	-0.435264872	0.003238146
cg25296646	*POLR2F*	5'UTR	0.289887521	0.003254817
cg23687677	*CNP*	3'UTR	0.317024493	0.003480274
cg09345954	*PLCL1*	TSS200	0.247360311	0.003813256
cg10586087	*SH3TC2*	3'UTR	-0.328883746	0.003960395
cg10415442	*ST18*	5'UTR	-0.303589435	0.004223364
cg03312124	*RFFL*	5'UTR	0.346823273	0.005186746
cg18200741	*RASGRP3*	Body	-0.334863863	0.005487429
cg03005055	*HSPA2*	TSS1500	0.364003109	0.00581851
cg06439124	*TMEM144*	5'UTR	0.297153891	0.006660309
cg00401471	*SPINT2*	3'UTR	0.377712287	0.007231842
cg05314420	*MAL*	TSS1500	0.342248873	0.007947175
cg16759204	*TJAP1*	5'UTR	-0.325167341	0.00823156
cg02524475	*HHIP*	TSS1500	0.377314281	0.008757836
cg11311579	*SCD*	TSS200	0.216390763	0.009053459
cg06102777	*PLA2G16*	TSS200	0.340298193	0.009267348
cg19295451	*PCSK6*	Body	0.280125747	0.009514906
cg24446429	*MBP*	Body	0.261574601	0.009619992
AD2 OPC genes				
cg07390210	*GALR1*	TSS1500	-0.98560592	0.000758257
cg23046475	*EPN2*	5'UTR	-0.974270892	0.000892543
cg25592910	*PCDH15*	TSS200	-0.959383843	0.000952641
cg27201457	*WSCD1*	Body	-1.036821236	0.001031079

cg10632899	*DNAH11*	Body	0.900855588	0.001809256
cg19713057	*HRASLS*	TSS1500	-0.960449292	0.00272003
cg10122698	*GALR1*	TSS200	-0.937963537	0.003013186
cg06742101	*SEMA3E*	1stExon	1.188576122	0.003157185
cg25734490	*ASCL1*	TSS1500	-0.695167666	0.003586539
cg09961397	*COL9A1*	Body	-1.186319397	0.003935182
cg17414248	*ACAN*	5'UTR	0.881974956	0.004374363
cg06859463	*CREB3L1*	Body	-0.67109949	0.004663787
cg00588575	*SAPCD2*	Body	-1.706791291	0.005283239
cg21822822	*LAMA4*	Body	0.661557799	0.005342867
cg16532510	*DNAH11*	Body	0.609776098	0.00543077
cg16862641	*COL9A1*	Body	-0.903574381	0.005447197
cg19594305	*CHST8*	1stExon	-0.659322522	0.005742579
cg27341926	*GFRA1*	1stExon	-0.604803204	0.005907863
cg08637302	*WSCD1*	Body	0.699561112	0.006346086
cg03808835	*PCDH15*	5'UTR	-0.846864275	0.006853095
cg07805424	*USP24*	Body	-0.667860868	0.006928677
cg26836413	*PDGFRA*	Body	-0.608543895	0.008007674
cg07862930	*KCNJ16*	5'UTR	-0.640420146	0.008330074
cg24279131	*DNAH11*	Body	-0.756721556	0.008607698
cg12499584	*USP24*	TSS200	-0.822399385	0.009102444
cg03165383	*GALR1*	3'UTR	-0.874968608	0.009231183
cg25156843	*ABHD2*	5'UTR	-0.648799861	0.009424758
cg05528293	*GALR1*	TSS1500	-1.038056455	0.009679931
cg08625851	*SOX4*	TSS200	-1.064461663	0.009879825
AD2 OLG genes				
cg15545878	*CRYAB*	TSS200	-1.3757029	0.00018849
cg09627520	*PXK*	3'UTR	-1.7361527	0.00018921
cg07527273	*MBP*	5'UTR	1.09957218	0.00056493
cg10841593	*TCP11L2*	TSS1500	-0.8845859	0.00133706

cg02510091	*MAN2A1*	TSS1500	-0.8893519	0.0014006
cg00281717	*PALM2*	TSS200	-0.9548064	0.00259212
cg17478193	*LRP2*	Body	-0.6947888	0.0028
cg15428653	*ABCA2*	TSS1500	-0.8886078	0.00290592
cg23588204	*LPGAT1*	5'UTR	-0.9153007	0.00313023
cg07318204	*HHIP*	TSS1500	-0.6908225	0.00318705
cg01253534	*LRP2*	ExonBnd	1.02698315	0.00344745
cg16400871	*LDB3*	Body	0.71432482	0.00387244
cg05069012	*UGT8*	TSS1500	-0.7972486	0.00398548
cg01655548	*DNAH17*	Body	0.84779563	0.00408599
cg13173405	*PSEN1*	TSS200	-1.3085912	0.00449504
cg17187287	*UGT8*	TSS200	-0.6528107	0.00501436
cg16176654	*FAM107B*	Body	-0.7916488	0.00501683
cg08505473	*PRIMA1*	TSS1500	-0.7174529	0.00502496
cg14864972	*PIP4K2A*	Body	-0.5869494	0.00585676
cg19239924	*TCP11L2*	5'UTR	-0.6326963	0.00592009
cg17037859	*DYSF*	Body	1.23216535	0.00649162
cg24402667	*ARHGAP23*	TSS1500	-0.7385276	0.00663073
cg27333693	*PIP4K2A*	Body	-0.6181074	0.00681462
cg07230380	*SCD*	TSS1500	-0.9844734	0.00697924
cg01140770	*MYRF*	Body	0.63138307	0.00697964
cg16231452	*TCP11L2*	TSS200	-0.7120043	0.00703719
cg16354502	*DYSF*	Body	0.64928926	0.00705941
cg23676734	*DNAH17*	Body	0.59021489	0.00728303
cg10182697	*NKX6-2*	TSS1500	-0.7146255	0.00768955
cg05565537	*GALNT6*	Body	0.81063644	0.0080416
cg27553637	*CDC42EP1*	5'UTR	-0.7239523	0.0080421
cg13914324	*PRUNE2*	Body	-0.5912214	0.00810189
cg13215003	*CTNNA3*	Body	0.78146135	0.00820934
cg05347732	*DYSF*	Body	0.6853443	0.00829565

cg05941521	*NINJ2*	Body	0.68954763	0.00911718
cg21099653	*CDH1*	Body	0.72571087	0.0097135
cg11598403	*MBP*	Body	-0.8059314	0.00987361
cg03588861	*PIEZO2*	Body	0.63280071	0.00993767

AD: Alzheimer's disease, OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell.

Appendix I: Genes differentially expressed in snRNA-sequencing OLG/OPC clusters identified in the AD1 and AD2 EWAS

Gene	P_value	avg_log2FC	Subcluster
OPC Genes			
ABHD2	0.00082	0.18531	Oli_0
TNK2	0.00701	0.18906	Oli_0
NTN1	0.04483	0.67801	Oli_0
OLIG2	0.00797	0.40552	Oli_1
USP24	0.01083	0.44474	Oli_3
SEMA5A	0.04577	0.61578	Oli_3
GFRA1	0.04638	4.29681	Oli_5
MEGF11	0.00002	0.24337	Opc_0
GSG1L	0.00004	0.37619	Opc_0
SMOC1	0.00042	0.18076	Opc_0
CHST9	0.00072	0.44280	Opc_0
XYLT1	0.00247	-0.15182	Opc_0
OLIG2	0.00640	0.18199	Opc_0
SEMA5A	0.01293	-0.15586	Opc_0
GALR1	0.01445	-0.44809	Opc_0
PDGFRA	0.00048	0.46213	Opc_1
LAMA4	0.00065	0.44792	Opc_1
COL9A1	0.00738	0.24335	Opc_1
GFRA1	0.01023	0.43260	Opc_1
XYLT1	0.01066	-0.20521	Opc_1
SMOC1	0.01619	0.17890	Opc_1
GFRA1	0.01713	-2.48011	Opc_2
OLG Genes			
CNP	0.00000	-0.20383	Oli_0
DNAH17	0.00000	-0.62638	Oli_1
SCD	0.00003	-0.15493	Oli_1
PACS2	0.00386	0.22217	Oli_1
QDPR	0.00690	-0.12480	Oli_1
RASGRP3	0.02113	0.21127	Oli_1
QDPR	0.00000	1.35254	Oli_3
MBP	0.00000	-0.51347	Oli_3
ST18	0.00000	-0.52480	Oli_3
POLR2F	0.00000	-0.88393	Oli_3

Gene	P_value	avg_log2FC	Subcluster
GLTP	0.00005	0.90301	Oli_3
PLEKHG3	0.01127	-0.87469	Oli_3
PACS2	0.00072	0.61158	Oli_5
CNP	0.04330	0.17969	Opc_1
MBP	0.02454	-1.85002	Opc_2
POLR2F	0.03585	-1.47701	Opc_2

AD: Alzheimer's disease, OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell.

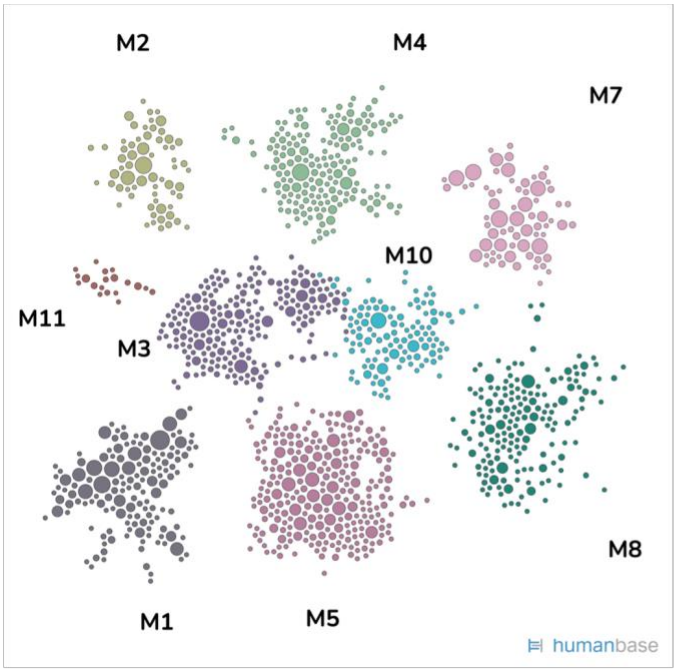
Appendix J: Results from the AD1 and AD2 meta-analysis

CpG	Gene	Feature	TE.fixed	pval.fixed
AD Meta OPC				
cg04525189	*VIPR2*	Body	-0.433840	0.000003
cg27201457	*WSCD1*	Body	-0.629704	0.000027
cg07429743	*SEMA5A*	Body	-0.380999	0.000543
cg20791593	*NEU4*	TSS1500	0.318704	0.000597
cg07099991	*WFDC1*	Body	0.301686	0.000622
cg12162377	*B3GNT7*	Body	-0.268130	0.000653
cg18493027	*SMOC1*	5'UTR	-0.261915	0.000668
cg22143285	*KCNJ16*	TSS1500	0.388692	0.000712
cg04913160	*ABHD2*	1stExon	0.273336	0.000829
cg25857018	*BLM*	TSS200	0.281460	0.000830
cg23190972	*SEMA5A*	Body	-0.231266	0.001070
cg07390210	*GALR1*	TSS1500	-0.451657	0.001289
cg03082589	*GSG1L*	Body	-0.365783	0.001289
cg14173476	*EPN2*	5'UTR	0.238721	0.001693
cg10525432	*SEMA5A*	5'UTR	-0.295811	0.002777
cg00164894	*USP24*	Body	0.211763	0.003125
cg26806527	*XYLT1*	Body	0.356987	0.003155
cg15299832	*OLIG2*	TSS200	0.283836	0.003531
cg21841583	*ABHD2*	TSS1500	-0.297851	0.003968
cg06039355	*GFRA1*	5'UTR	0.298516	0.004065
cg16933181	*CHST9*	TSS200	0.251742	0.004376
cg08083599	*COL20A1*	Body	-0.244581	0.005113
cg18080303	*LAMA4*	TSS1500	-0.283774	0.005449
cg23550826	*PDGFRA*	1stExon	-0.282623	0.006101
cg27418204	*TNK2*	Body	0.233487	0.006208
cg07575407	*MEGF11*	5'UTR	0.316550	0.006286

cg19441674	*NTN1*	3'UTR	0.238670	0.008127
cg04487479	*COL9A1*	Body	-0.307709	0.008211
cg20349377	*KLHL1*	TSS1500	-0.565467	0.008402
cg22335692	*CHST8*	Body	0.251177	0.008874
cg16713743	*OLIG2*	TSS1500	-0.309166	0.009277
cg10764907	*VIPR2*	Body	0.198061	0.010000
AD Meta OLG				
cg16985259	*IPO13*	5'UTR	0.3477004	0.0000004
cg23676734	*DNAH17*	Body	0.3419555	0.0003198
cg24446429	*MBP*	Body	0.2873290	0.0006537
cg15923042	*QDPR*	Body	0.3051573	0.0007392
cg13311357	*FRYL*	TSS200	0.2752738	0.0011142
cg00618155	*ADAMTS14*	Body	-0.3081429	0.0013589
cg23687677	*CNP*	3'UTR	0.2560790	0.0039307
cg10251094	*DNAH17*	Body	0.2814563	0.0041784
cg07865517	*PLEKHG3*	TSS1500	-0.3304465	0.0045710
cg23394510	*GLTP*	Body	0.3003015	0.0058522
cg11311579	*SCD*	TSS200	0.2024740	0.0061537
cg12131510	*ADAMTS14*	Body	-0.2458166	0.0073264
cg12529006	*LIPA*	5'UTR	-0.2231279	0.0077497
cg25296646	*POLR2F*	5'UTR	0.2172582	0.0080987
cg07421682	*PCSK6*	TSS1500	-0.3993335	0.0081129
cg18200741	*RASGRP3*	Body	-0.2678405	0.0093903
cg11028291	*PACS2*	Body	0.2117862	0.0097648
cg22238209	*MAG*	Body	-0.3331616	0.0097802
cg10415442	*ST18*	5'UTR	-0.2235869	0.0099360

AD: Alzheimer's disease, OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell.

Appendix K: Example output of functional enrichment using HumanBase



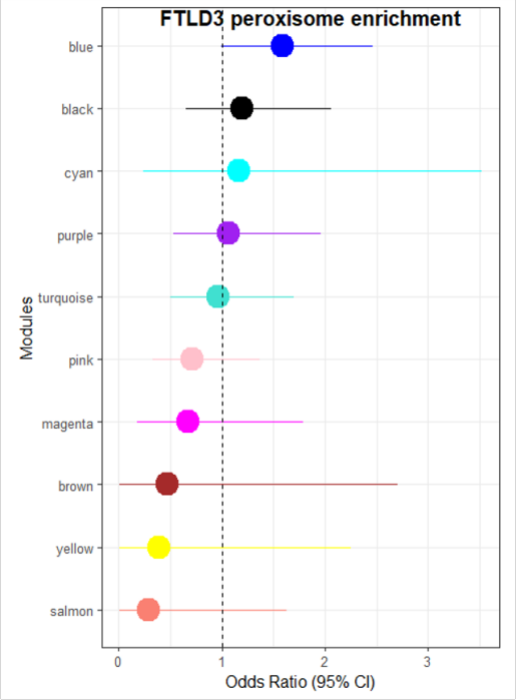
MODULE	TOP TERMS (Max 10)	Q VAL	GENES	TERMS
M4	glucosylceramide metabolic process	0.00939435	176	110
	cholesterol transport	0.01385383		
	regulation of protein localization to cell periphery	0.01620434		
	ERAD pathway	0.01721997		
	glycosylceramide metabolic process	0.01741401		
	positive regulation of JNK cascade	0.01875036		
	sterol transport	0.01875036		
	histone H2A monoubiquitination	0.01875036		
	regulation of non-motile cilium assembly	0.01938635		
	intracellular cholesterol transport	0.02078830		
M5	regulation of transcription from RNA polymerase II promoter in response to hypoxia	0.00328027	253	128
	response to tumor necrosis factor	0.00430893		
	negative regulation of cell junction assembly	0.00442388		
	negative regulation of wound healing, spreading of epidermal cells	0.00511500		
	cellular response to tumor necrosis factor	0.00877669		
	regulation of wound healing, spreading of epidermal cells	0.00877669		
	positive regulation of transcription from RNA polymerase II promoter in response to hypoxia	0.00939435		
	regulation of myosin-light-chain-phosphatase activity	0.01070162		
	positive regulation of transcription from RNA polymerase II promoter involved in cellular response to chemical stimulus	0.01236495		
	negative regulation of focal adhesion assembly	0.01721997		
M7	retrograde protein transport, ER to cytosol	0.01648328	70	24
	endoplasmic reticulum to cytosol transport	0.01648328		
	peroxisomal membrane transport	0.01704610		
	protein targeting to peroxisome	0.01720343		
	peroxisomal transport	0.01720343		
	establishment of protein localization to peroxisome	0.01720343		
	protein localization to peroxisome	0.01720343		
	DNA-templated transcription, initiation	0.01741401		
	positive regulation of translation	0.01754146		
	positive regulation of intrinsic apoptotic signaling pathway	0.02261341		

MODULE	TOP TERMS (Max 10)	Q VAL	GENES	TERMS
M1	protein localization to chromosome, centromeric region	0.00045519	143	93
	protein localization to chromosome	0.00045519		
	regulation of DNA biosynthetic process	0.00045519		
	positive regulation of DNA biosynthetic process	0.00045519		
	metaphase plate congression	0.00148879		
	RNA-dependent DNA biosynthetic process	0.00252874		
	regulation of telomerase activity	0.00252874		
	regulation of telomere maintenance via telomerase	0.00323814		
	mitotic metaphase plate congression	0.00328027		
	establishment of chromosome localization	0.00328027		
M2	fatty acid derivative biosynthetic process	0.00328027	64	30
	cytoplasmic translation	0.00328027		
	leukotriene biosynthetic process	0.00513129		
	activation of cysteine-type endopeptidase activity involved in apoptotic process	0.00727685		
	leukotriene metabolic process	0.00882202		
	fatty acid derivative metabolic process	0.01003430		
	positive regulation of cysteine-type endopeptidase activity involved in apoptotic process	0.01518368		
	icosanoid biosynthetic process	0.01666360		
	signal transduction by p53 class mediator	0.01835811		
	positive regulation of endopeptidase activity	0.01944054		
M3	ionotropic glutamate receptor signaling pathway	0.00045519	213	121
	glutamate receptor signaling pathway	0.00094478		
	cellular response to vascular endothelial growth factor stimulus	0.00511500		
	branch elongation involved in ureteric bud branching	0.00556474		
	regulation of endothelial cell proliferation	0.00750017		
	negative regulation of vasculature development	0.00846880		
	collagen biosynthetic process	0.00882202		
	regulation of clathrin-dependent endocytosis	0.00917729		
	ureteric bud elongation	0.01041398		
	endothelial cell proliferation	0.01084339		

MODULE	TOP TERMS (Max 10)	Q VAL	GENES	TERMS
M8	positive regulation of cortisol secretion	0.00328027	171	80
	positive regulation of glucocorticoid secretion	0.00328027		
	positive regulation of corticosteroid hormone secretion	0.00430893		
	cortisol secretion	0.00487544		
	regulation of cortisol secretion	0.00487544		
	positive regulation of steroid hormone secretion	0.00694491		
	regulation of glucocorticoid secretion	0.00877669		
	negative regulation of interleukin-6 production	0.00926554		
	glucocorticoid secretion	0.00945139		
	regulation of corticosteroid hormone secretion	0.01137526		
M10	positive regulation of lipopolysaccharide-mediated signaling pathway	0.00871811	116	133
	blood vessel endothelial cell migration	0.00877669		
	positive regulation of astrocyte differentiation	0.00877669		
	heparan sulfate proteoglycan biosynthetic process, polysaccharide chain biosynthetic process	0.00877669		
	cell migration involved in sprouting angiogenesis	0.00928734		
	glycosaminoglycan biosynthetic process	0.01041398		
	regulation of endothelial cell migration	0.01070162		
	positive regulation of amyloid-beta formation	0.01168345		
	positive regulation of cytokine secretion	0.01257908		
	positive regulation of amyloid precursor protein catabolic process	0.01421299		
M11	activation of MAPKKK activity	0.00094478	19	9
	positive regulation of p38MAPK cascade	0.00328027		
	activation of MAPKK activity	0.00431392		
	regulation of p38MAPK cascade	0.00487544		
	p38MAPK cascade	0.00846880		
	positive regulation of JNK cascade	0.01741401		
	positive regulation of stress-activated MAPK cascade	0.01938635		
	positive regulation of stress-activated protein kinase signaling cascade	0.01938635		
	positive regulation of MAP kinase activity	0.01938635		

Example of output from functional enrichment analysis with HumanBase of the FTLD-sorted black module.

Appendix L: Gene list enrichment analysis of peroxisome related genes with FTLD3 co-methylation network modules



Appendix M: Curated set of genes of interest for Chapter 5

Reason for inclusion	Gene (s)	Disease
FTLD Bulk Networks		
Hub gene FTLD1- skyblue	*TMEM168*	FTLD
Hub gene FTLD1-tan	*C5orf51*	
Hub gene FTLD1-white	*CTDSP2*	
Hub gene FTLD2 orangered	*ZNF696*	FTLD
Hub gene FTLD3 cyan	*DDIT4L*	FTLD (PSP)
Hub gene FTLD3 salmon	*PCK1*	FTLD (PSP)
AD Bulk Networks		
Hub gene FTLD3 blue	*GDAP1*	FTLD (PSP)
Hub gene AD DLPFC greenyellow	*MOG*	AD
Hub gene AD DLPFC purple		AD
Hub gene AD ERC tan (also differentially methylated (FTLD1 EWAS, AD2 EWAS) and expressed (FTLD1-expression)	*MYRF*	AD
Hub gene AD HIPPO grey60	*MGC14436*	AD
Hub gene AD HIPPO greenyellow		AD
Consistently present in disease associated AD signature (Chapter 3.X)	*ADARB2, FAM107B, QDPR, RASGRF2, UNC5C, CAV1, SLC5A11, ATP11A, INPP5A*	
Sorted FTLD Networks		
Hub gene FTLD-sorted midnightblue	*ZNF292*	FTLD
Hub gene FTLD-sorted brown	*PARM1*	FTLD
Hub gene FTLD-yellow	*ARHGAP15*	FTLD
Sorted AD Networks		
Hub gene AD1 brown	*CTNNA1*	AD

Hub gene AD1 blue	*ZNF143*	AD
Hub gene AD1 midnightblue	*METT10D*	AD
Hub gene AD2 blue	*FAAH*	AD
Common genes across FTLD and AD EWAS and meta-analyses		
OPC genes present in 3 or more analyses (Chapter 3. 3.ii)	*GFRA1, ABHD2, TNK2, WFDC1, CHST8, COL9A1, EPN2, FERMT1, GALR1, MEGF11, PLEKHH2, SEMA5A, VIPR1, WSCD1, XLYT1, ACAN, AMZ1, CREB3L1, DNAH11, KLHL1, LAMA4, LYPD1, MYT1, PCDH15, PDGFRA, SEMA3E, SMOC1, SOX4, USP24*	AD, FTLD
OLG genes present in 3 or more analyses (Chapter 3.3.ii)	*DNAH17, CTNNA3, SCD, DYSF, MBP, NINJ2, PCSK6, TMTC4, ADAMTS18, ARGHAP23, HHIP, HSPA2, MOG, PIEZO2, PRIMA1, PRUNE2, PXK, ST18, BOK, CDC42EP1, CDH1, ENPP2, FAM107B, GALNT6, IPO13, KIF13B, KIF6, LRP2, MAN2A1, NDE1, PACS2, PALM2, PIP4K2A, PLCL1, PLEKHH1, PPP1R14A, QDPR, RAGRP3, RFFL, TMEM144, TMEM63A, TTYH2*	AD, FTLD
OLG Genes from AD EWAS also showing differential expression patterns	*MBP, CTNNA3, CRYAB, FAM107B, TMEM144, DNAH17, ST18, QDPR, CNP, MBP, POLR2F*	AD
OPC genes from AD EWAS also showing differential expression patterns	*USP24, CREB3L1, SEMA3E, PCDH15, GALR1, SAPCD2, TNK2, PDGFRA, CHST9, OLIG2, MYT1, MEGF11, GFRA1, ABHD2*	AD
OLG Genes from FTLD EWAS also showing differential expression patterns	*LRP2, PIP4K2A, LPGAT1, CDH1, PIEZO2, MAN2A1, GPR37, QDPR, FAM107B, AMER2, PRIMA1*	FTLD
OPC genes from FTLD EWAS also showing differential expression patterns	*WFDC1, SEMA3E, SOX4, CHST9, GSG1L, SMOC1, FERMT1, DNAH11, ACAN, CHST8, MEGF11, TNK2, FERMT1, AMZ1, CCDC146, GPNMB*	FTLD

Curated set of genes, as well as diseases associated with, and reason for inclusion in list across Chapters in this thesis. FTLD: frontotemporal lobar degeneration, AD: Alzheimer's disease, OLG: oligodendrocyte, OPC: oligodendrocyte precursor cell.

Appendix N: Position of mQTLs that are also disease associated SNPs across FTLD, PSP and ALS