Genome-wide long-read transcriptomics and DNA methylation in *App*$^{NL\text{-}G\text{-}F}$ knock-in mouse model of Alzheimer's disease

Umran Yaman

University College London

Thesis submitted for the degree of Doctor of Philosophy

13 January 2025

Acknowledgements

I have been very fortunate to work with my supervisor, **Dr. Dervis Salih**, who consistently supported me, and was a source of motivation, every single day and throughout my time in UK as an MSc and PhD student in his lab. Dervis, thanks for guiding me and giving me the countless opportunities to learn and practise, and freedom to pursue research in the field I am most passionate about.

I cannot possibly thank **Prof. Sir. John Hardy** enough for his support, always, providing me with the countless opportunities. It has been an absolute privilege to work in your lab. Thanks very much for allowing me to explore and pursue research in the long-read and bioinformatics fields. John, thanks very much for always trusting in me and encouraging my growth. I am very hopeful and excited about the future ahead, and you and Dervis were solely supportive of every situation, it has meant the world to me.

I would like to also thank my secondary supervisor, **Dr Emil Gustavsson**, for his massive help and guidance towards technical aspects, and invaluable advice towards this PhD work. Emil, your suggestions have been incredibly helpful, making it much easier for me to navigate the complexities of the work. Your insights have been invaluable and greatly appreciated!

My thesis committee members, **Prof. Sao Bettencourt** and **Prof. Pietro Fratta**, have provided insights and suggestions that greatly contributed and helped me to explore many aspects of this work. I truly appreciate your contributions and suggestions.

Special thanks to **Dr Frances Wiseman**, being an incredible RDGT and collaborator. I appreciate your support throughout my PhD, and your teams' massive contributions to the long-read RNA sequencing work, along with **Dr Paige Mumford,** and **Dr Gareth Banks**.

I would like to thank **Hannah McPherson**, and **Jasmine Lee** from the long-read facility, for teaching, and helping with long-read, thanks for your patience and great contributions to this work.

I would like to thank **my funders, Ministry of Education in Turkiye**. This scholarship was a massive step forwards and hope, and changed my whole perspective and career track. I truly appreciate this enormous financial support covering both MSc and PhD education in the UK.

To **each member of Hardy lab**, I am constantly amazed by the incredible work you do, and always broadened my perspective. I have learned so much from you, and your support has been invaluable during every presentation, exam, and challenge along the way. I am grateful to have had worked with you. My colleague and also dear friend, **Naciye Magusali**, thanks for making the journey memorable and truly enjoyable so far! Hep yanımda ol!

To my amazing friends from my MSc days—**Cansu, Daniel, Ricardo, Morgan, Mayank**, and **Carol**—you made my first days in London unforgettable and this whole journey so much more fun. Thank you for all the laughs and for always keeping things enjoyable.

My dearests in London— **Chris** and my lovely **Luna**. Every single day you were a great source of happiness, and you were always by my side.

To **Büşra Canik**, **Gamze Gökmen, Gamze Köse** and **Ali Doğan**, my chosen family, you possibly can not imagine how much I longed to see you and how I look for every opportunity to visit Türkiye. Your trust and support have always motivated me in everything I do.

Finally, I would like to acknowledge my family, thank you for always believing in me, being great parents, and great friends. You have supported and motivated me all the time, everywhere, and with every choice I make. Sizi çok seviyorum.

Declaration

I, Ümran Yaman, hereby declare that the work presented in this PhD thesis is the result of my own independent research. Where information has been derived from other sources, it has been properly acknowledged and referenced within the thesis.

Genotyping of all mice was performed by Paige Mumford. Library preparation of long-read RNA-sequencing and the sequencing was performed by Hannah McPherson. Library preparation of long-read DNA sequencing, and the sequencing the samples was performed with the help of Jasmaine Lee.

Publications arising from this thesis:

Yaman, U., Banks, G., Gustavsson, E. K. *et al.* (2024). Long-read transcriptomic identification of synaptic adaptation to amyloid pathology in the App (NL-GF) knock-in mouse model of the earliest phase of Alzheimer's disease. *bioRxiv*, 2024-08.

Findings and content from the publication above, along with the supplementary material available via data access links, have been incorporated into this thesis.

Publications not directly related to this thesis:

Graham, A. C., Bellou, E., Harwood, J., **Yaman, U**., *et al.* (2025). Human longevity and Alzheimer's disease variants act via microglia and oligodendrocyte gene networks, Brain, awae339.

Gezegen, H., .. **Yaman, U**., *et al.* (2024). Exploratory blood biomarker patterns in a mixed dementia cohort. medRxiv, 2024.

Arber, C., Casey, J. M., Crawford, S., Rambarack, N., **Yaman, U.,** *et al.* (2024). Microglia contribute to the production of the amyloidogenic ABri peptide in familial British dementia. Acta Neuropathologica, 148, 65.

Deb, S. K., Kalra, D., Kubica, J., .. **Yaman, U.,** *et al.* (2024). The fifth international hackathon for developing computational cloud-based tools and resources for pan-structural variation and genomics. F1000Research, 13, 708.

Simmonds, E., Leonenko, G., **Yaman, U.,** *et al.* (2024). Chromosome X-wide association study in case-control studies of pathologically confirmed Alzheimer's disease in a European population. Translational Psychiatry, 14(1), 34.

Magusali, N., Graham, A. C., .. **Yaman, U.,** *et al.* (2021). A genetic link between risk for Alzheimer's disease and severe COVID-19 outcomes via the OAS1 gene. Brain, 144(8), 2403–2416.

Umran Yaman

13/01/2025

Abstract

Alzheimer's disease (AD) is a complex neurodegenerative disorder driven by genetic and epigenetic changes. Genome-wide association studies (GWAS) have identified a network of AD risk genes, primarily expressed in microglia. However, short-read sequencing often fails to capture transcript diversity, isoform-level alterations, and alternative splicing, limiting understanding of how genetic risk variants regulate gene expression and contribute to disease progression.

This thesis describes and presents the use of long-read RNA and DNA sequencing, which was performed as a part of this study, in the *App*$^{NL\text{-}G\text{-}F}$ knock-in mouse model to address these limitations. By integrating transcriptional and epigenetic analyses, differential alternative splicing, novel transcript isoforms, and genome-wide DNA methylation patterns were investigated to uncover mechanisms underlying AD.

Long-read RNA sequencing identified activation of microglial AD risk genes, such as *Trem2*, and novel isoforms in AD-associated genes. Isoform usage and alternative splicing were also observed in genes not previously linked to AD, including *Syngr1* and *Clta*, which influence synaptic function. These findings highlight the importance of profiling splicing and isoform alterations across diverse brain cell types, including microglia, neurons, and oligodendrocytes.

Genome-wide DNA methylation analysis revealed differentially methylated regions (DMRs) in AD-related genes such as *App* and *Mapt*. Many human AD risk genes and differentially spliced genes overlapped with DMRs. Strong correlations between methylation and gene expression were observed in *Capg* and *Csf1*, emphasizing the role of methylation in fine-tuning expression. Hypermethylated promoters were linked to RNA splicing and stem cell maintenance, while hypomethylated promoters were associated with immune activation and cell trafficking. DMRs extending into gene bodies and intergenic regions were enriched in inhibitory neurons, suggesting connections between synaptic regulation and microglial interactions during amyloid response.

This thesis demonstrates the utility of long-read sequencing in revealing AD mechanisms, providing insights into amyloid pathology and potential therapeutic targets.

Impact Statement

Alzheimer's disease (AD) is a progressive neurodegenerative disorder characterized by the accumulation of amyloid plaques and tau tangles, which lead to the gradual decline of cognitive abilities and a diminished quality of life. Despite extensive research into the disease, AD remains incurable, and effective treatments are still lacking. This thesis aims to fill critical gaps in our understanding of AD by investigating the molecular mechanisms driving disease progression, with a specific focus on gene expression regulation. Using a well-established AD mouse model that mimics the preclinical stages of AD, this work offers a comprehensive analysis of the early molecular changes in AD, providing vital insights into its onset and identifying potential therapeutic targets.

A major contribution of this research is the generation of high-quality, genome-wide transcriptomic and epigenetic datasets through the application of advanced long-read RNA and DNA sequencing technologies. These cutting-edge tools enabled the identification of genome-wide novel transcript isoforms, alternative splicing events, and methylation particularly in genes involved in immune response and synaptic function. Unlike short-read sequencing technologies, long-read RNA sequencing facilitated the detection of isoform usage patterns that were previously overlooked, advancing our understanding of how amyloid pathology disrupts gene regulation. These findings not only contribute to the molecular landscape of AD but also underscore the importance of isoform-level analyses in disease research, highlighting new pathways that could be targeted to understand and mitigate amyloid-driven neurodegeneration.

In addition to transcriptomics, genome-wide DNA methylation analysis shed light on the epigenetic regulation of AD. By focusing on specific neuronal subtypes, particularly inhibitory neurons, this study uncovers differential DNA methylation patterns that may be implicated in the cellular dysfunction characteristic of AD. These results suggest that epigenetic changes, such as DNA methylation, play a significant role in regulating gene expression and influencing how neurons respond to amyloid pathology. These insights deepen our understanding of how non-genetic factors, such as epigenetic

modifications, contribute to AD progression and highlight their potential as biomarkers or therapeutic targets.

A key aspect of this thesis is the emphasis on reproducibility and accessibility. Robust and reproducible pipelines for DNA methylation analysis and single-cell RNA sequencing re-analysis were developed to ensure that these methods can be adopted and extended by other researchers. Datasets and code generated from Chapters 2, 3, and 4 will be shared publicly on platforms such as bioRxiv and GitHub for open and accessible to the broader scientific community. A manuscript based on the findings from Chapter 2 has already been submitted for publication, contributing to the wider dissemination of important results obtained by long-read RNA sequencing.

In conclusion, this thesis makes significant strides in advancing our understanding of AD by identifying genome-wide novel transcript isoforms, alternative splicing events, and DNA methylation, in response to amyloid pathology. These findings provide fresh perspectives on the underlying molecular mechanisms of AD progression, and enhance our understanding of the disease's onset and advances the search for potential therapeutic targets. While further research is needed to translate these discoveries into clinical applications, the results presented in this work lay a strong foundation for the development of targeted therapeutic strategies, for example to control splicing in AD. Ultimately, this thesis aims to contribute to the growing body of knowledge on AD and support the advancement of more effective treatments for those affected by this devastating disease.

List of Figures

List of Tables

Abbreviations

Acronyms / Abbreviations

AD: Alzheimer's Disease

App$^{NL\text{-}G\text{-}F}$: A knock-in mouse model carrying humanized mutations in the amyloid precursor protein gene

Aß: Amyloid-beta

CpG: Cytosine-Phosphate-Guanine dinucleotide

DAM: Disease Associated Microglia

DEXSeq: Differential Exon Usage in RNA-seq

DESeq2: Differential Expression Analysis

DNA: Deoxyribonucleic Acid

DMR: Differentially Methylated Region

EWCE: Expression Weighted Cell Type Enrichment

GWAS: Genome-Wide Association Studies

GO: Gene Ontology

LOAD: Late Onset Alzheimer's Disease

ONT: Oxford Nanopore Technologies

RNA: Ribonucleic Acid

RNA-seq: RNA Sequencing

eQTL: Expression Quantitative Trait Loci

sQTL: Splicing Quantitative Trait Loci

SNP: Single Nucleotide Polymorphism

TSS: Transcription Start Site

WT: Wild Type

5mC: 5-methylcytosine

Chapter 1

1 Introduction

1.1 Genetics of Alzheimer's Disease

Alzheimer's disease (AD) is a leading neurodegenerative disorder, accounting for the majority of dementia cases worldwide - an estimated 50 million cases, a number projected to surpass 152 million by 2050 (Nichols et al., 2022). Late-onset Alzheimer's disease (LOAD) typically progresses through a long, asymptomatic phase marked by the gradual build-up of amyloid, followed by tau pathology. The pathological progression of AD occurs across distinct cellular states, including neurons, glia, oligodendrocytes, and the vascular system (De Strooper & Karran, 2016a).

In early-onset AD, the initial hallmark of AD, Aß plaques, are generated due to rare dominant gene variants in the amyloid precursor protein *(APP),* presenilin 1 *(PSEN1)* and presenilin 2 *(PSEN2)* which drive APP cleavage to amyloid-ß (Aß). The genetic pathways differ between early-onset and LOAD (Hartl et al., 2020; Karch & Goate, 2015; Suh et al., 2013). Genome-wide association studies (GWAS) have shone a light onto the DNA variants in the human population associated with LOAD, showing the involvement of complex alterations in the innate immune system and lipid metabolism in response to the hallmark pathological features of LOAD (Aß plaques and tau tangles)(Bellenguez et al., 2022; De Strooper and Karran, 2016a; Efthymiou and Goate, 2017; Felsky et al., 2019; Kunkle et al., 2019; Lambert et al., 2013; Salih et al., 2019).

Gene expression network analyses in LOAD indicate that differentially expressed genes may not directly link to Aß generation (Cruchaga et al., 2014; Edwards et al., 2019; Hong et al., 2016; Sala Frigerio et al., 2019; Salih et al., 2019). Instead, many risk are primarily expressed in microglial cells, forming a transcriptional network that influences AD pathology (Sala Frigerio et al., 2019; Salih et al., 2019; Sierksma et al., 2020).

Single nucleotide polymorphisms (SNPs) linked to AD are found in several genes, including apolipoprotein E (APOE), triggering eeceptor expressed on myeloid Cells 2

(*TREM2*), phosphatidylinositol binding clathrin assembly protein (*PICALM*), siglec-3 (*CD33*), ABI family member 3 (*ABI3*), phospholipase C gamma 2 (*PLCG2*), and transcription factor PU.1 (*SPI1*) (Efthymiou & Goate, 2017; K. L. Huang et al., 2017; W. Liu et al., 2020; Sims et al., 2017). Among these, *APOE-ε4* is the strongest genetic risk factor for LOAD (Corder et al., 1993). The *APOE-ε4* allele significantly increases the risk of developing LOAD and accelerates disease onset in a dose-dependent manner. It affects Aß metabolism by promoting aggregation and impairing clearance, thereby contributing to a more extensive Aβ deposition in the brain compared to the ε2 and ε3 alleles (Castellano et al., 2011; Liu et al., 2015). Additionally, *APOE-ε4* interacts with microglia, enhancing inflammatory responses that may further exacerbate AD pathology through immune dysregulation (Efthymiou & Goate, 2017; Pankiewicz et al., 2017; Salih et al., 2019; Weigand et al., 2021).

1.2 Transcriptomic and cellular changes in AD mouse models

Transcriptome studies using microarrays, bulk RNA-seq, and single-cell RNA-seq in AD mouse models have shown that amyloid plaques alone can drive the expression of microglial risk genes and induce microglial changes including proliferation and activation, indicating these AD risk genes and microglial adaptations may act at early stages of disease in humans to impact the onset of clinical symptoms (Matarin et al., 2015; Salih et al., 2019; Shireby et al., 2022). Overt cognitive deficits in humans typically lead to diagnosis when a substantial level of neuronal loss occurs that starts approximately twenty years before the onset of the disease (Bateman et al., 2012; Liang et al., 2013).

However, whether mouse models accurately replicate the complex features of human AD remains debated. AD mouse models do not exhibit the amyloid-dependent tau pathology that leads to neuronal degeneration (Lee et al., 2021; Leyns et al., 2019), raising questions about their relevance (Chen et al., 2023; Friedman et al., 2018; Holtzman et al., 2000). Amyloid-based mouse models, including $App^{NL\text{-}G\text{-}F}$, fail to develop tau pathology, likely due to the absence of human tau or the additional pathological triggers—such as aging, inflammation, or other cellular stressors—that drive tau hyperphosphorylation and neurofibrillary tangle formation in human AD (Dujardin et al., 2020; He et al., 2018). This absence of tau pathology limits the ability of amyloid models to fully replicate later-stage neurodegeneration. However, despite

these limitations, AD mouse models remain invaluable for dissecting disease mechanisms, particularly those that characterize early prodromal stages (Keren-Shaul et al., 2017; Sala Frigerio et al., 2019).

The *App*$^{NL-G-F}$ mouse model is a genetically engineered knock-in system that carries three familial AD mutations: Swedish (KM670/671NL), Iberian (I716F), and Arctic (E693G). Each mutation contributes to distinct aspects of Aβ pathology. The Swedish mutation enhances β-secretase cleavage, leading to elevated Aβ production; the Iberian mutation increases the Aβ42/Aβ40 ratio, favoring the accumulation of more aggregation-prone Aβ42 species; and the Arctic mutation accelerates Aβ aggregation, promoting protofibril formation and plaque deposition (Saito et al., 2014).

Unlike traditional transgenic models such as 5xFAD and *APP/PS1*, which rely on non-physiological *APP* overexpression, *App*$^{NL-G-F}$ preserves endogenous *APP* expression levels, providing a more biologically relevant system for studying amyloid pathology. To generate this model, the murine Aβ sequence was humanized, introducing the Swedish mutation into exon 16 and the Iberian and Arctic mutations into exon 17. These modifications were carried out using gene targeting in C57BL/6 embryonic stem (ES) cells, ensuring that regulatory elements and intronic architecture remained intact, thereby preserving normal APP transcription, splicing, and processing. Following gene targeting, heterozygous mutant mice were crossed with EIIa-Cre transgene to excise the pgk-neo selection cassette, eliminating any potential confounding effects from vector sequences (Saito et al., 2014). However, the Arctic mutation introduces an important caveat: it accelerates Aβ aggregation, yielding protofibrils with structural properties that differ from those in sporadic AD, where Aβ species are more heterogeneous and influenced by a broader range of pathological factors.

Despite the limitations, the *App*$^{NL-G-F}$ model provides a well-controlled system for studying amyloid-driven disease mechanisms, as it develops extensive Aβ plaque deposition without the artificial consequences of *APP* overexpression and its proteolytic fragments which are known to drive more aggressive amyloid pathology, potentially and introduce non-physiological effects that cab obscure downstream mechanisms (Oakley et al., 2006). In contrast to transgenic models, *App*$^{NL-G-F}$ more accurately reflects endogenous APP expression levels, allowing for the study amyloid

response in a more physiologically relevant context. Other knock-in models, such as *App*$^{NL-F}$ and *App*$^{NL-G}$, also lack tau pathology but exhibit differences in their Aβ42/Aβ40 ratio and aggregation dynamics, which influence the trajectory of disease progression.

While no single mouse model fully captures the complexity of human AD, the *App*$^{NL-G-F}$ model is a valuable system for investigating the early amyloid-driven mechanisms of AD. Unlike traditional transgenic models which rely on *APP* overexpression, knock-in models better preserve physiological APP processing, avoiding artificial overproduction of APP fragments that do not occur in human disease. This distinction is particularly relevant for studying crucial processes such as *APP* processing, amyloid-β plaque deposition, and microglial activation mediated by an increase in expression of a number of risk genes identified in human GWAS (Wood et al., 2022).

GWAS have pinpointed nearly 100 loci associated with AD (Bellenguez et al., 2022; Kunkle et al., 2019; Lambert et al., 2013). While a causal gene is directly responsible for the development of the disease, where mutations can alter the function of the protein product, a gene with an expression quantitative trait loci (eQTL) variant may impact gene regulation by alternating expression and/or splicing patterns (Fan et al., 2021; Raj & Blencowe, 2015; Takata et al., 2017; Wang et al., 2015). The identification of microglia-specific eQTLs provides insights into the gene regulatory mechanisms that may underlie AD pathology. In a study by Young et al. (2021), eQTLs were mapped to primary human microglia and revealing genetic variants that impact gene expression, which may influence AD pathology (Young et al., 2021).

In parallel to eQTLs, splicing quantitative trait loci (sQTLs) have emerged as crucial determinants of alternative splicing and transcript isoform diversity. The regulation of alternative splicing is essential for generating functional protein diversity. Alternative transcript isoforms and splice-forms are pivotal mechanisms in gene expression, and errors in splicing can lead to regulatory dysfunctions (García-Ruiz et al., 2023; Gruijs da Silva et al., 2022; Jaffrey & Wilkinson, 2018; D. Li et al., 2021; Mills et al., 2013; Reyes & Huber, 2018; Simone et al., 2021). Alternative splicing plays a crucial role in protein function, particularly in neurological diseases (Marques-Coelho et al., 2021; Tollervey et al., 2011), and selection of specific mRNA isoforms is instrumental in driving various cellular pathways and functions (Chen PC et al., 2022; Darnell, 2013; B. Raj & Blencowe, 2015). Recent studies have highlighted the role of sQTLs in

influencing splicing patterns, thereby impacting cellular processes that are vital for proper regulatory function (García-Ruiz et al., 2023; Kim-Hellmuth et al., 2020).

Raj et al. (2018) conducted integrative transcriptome analyses that implicated altered splicing as a significant factor in AD susceptibility. Their study identified many alternatively spliced genes in the brain, and they associated these splicing events with pathological features of AD, such as neuritic plaques and neurofibrillary tangles (Raj et al., 2018). Zhang et al. emphasized the regional variation of sQTLs in the human brain, identifying specific SNPs that create binding sites for splicing factors, thereby influencing the splicing of genes like microtubule-associated protein tau *(MAPT)* (Zhang et al., 2020). The generation of diverse tau isoforms through alternative splicing is essential for neuronal health, and dysregulation of this process can lead to the accumulation of toxic tau aggregates characteristic of AD.

Long-read RNA- and DNA-sequencing technologies are increasingly recognized for their potential to characterize the earliest molecular and cellular alterations in response to amyloid pathology in AD. These advanced sequencing methods provide unique advantages over traditional short-read sequencing, particularly in resolving complex genomic structures and transcriptomic profiles that are critical for understanding the pathogenesis of AD (Cogan et al., 2024). One of the significant benefits of long-read sequencing is its ability to capture full-length transcripts (Leung et al., 2021), which is essential for accurately assessing gene expression changes associated with amyloid pathology. For instance, Roeck et al. demonstrated that long-read cDNA sequencing could elucidate the complexities of ATP binding cassette subfamily A member 7 (*ABCA7*) transcript variations in early-onset AD, revealing alternative splicing events that might escape detection by short-read methods (De Roeck et al., 2017). This capability is crucial, as the expression of genes involved in amyloid processing and clearance can be significantly altered in the presence of Aβ aggregates, affecting the overall cellular response to amyloid pathology. In summary, understanding these splicing events is crucial, as they may contribute to the pathophysiological mechanisms underlying amyloid pathology.

1.3 Epigenetic changes in AD brain

AD results from the interplay of genetic and epigenetic factors, with the role of GWAS risk variants remains unclear. It is widely acknowledged that AD does not result from a single genetic mutation but rather from a multifaceted interaction between genetic predispositions and environmental influences. Within this complex network, epigenetic modifications, such as DNA methylation and histone modifications may play a critical role in regulating splicing or transcript or gene expression, thereby influencing disease progression(De Jager et al., 2014; Fodder et al., 2023; Klein et al., 2016; X. Liu et al., 2018; Lunnon et al., 2014; Qazi et al., 2018; Shireby et al., 2022; Yokoyama et al., 2017; L. Zhang et al., 2020).

DNA methylation, the addition of a methyl group to cytosine residues in DNA (5-methyl cytosine), particularly at CpG sites, has gathered attention as a potential biomarker for AD. Technologies such as bisulfite sequencing and methylation arrays (e.g., Illumina 450K or EPIC) have revealed methylation alterations occurring years before clinical onset, suggesting these epigenetic changes may contribute to AD pathogenesis(Altuna et al., 2019; Fodder et al., 2023; Gasparoni et al., 2018; Lardenoije et al., 2019; Lunnon et al., 2014; Shireby et al., 2022; R. G. Smith et al., 2021). Lunnon et al. (2014) demonstrated methylation changes at genomic regions such as *ANK1,* and associated with AD pathology. The impact of DNA methylation on gene expression in AD was further explored via a large-scale epigenome-wide association study (EWAS) with Illumina Human Methylation Array on brain tissue from AD patients (De Jager et al., 2014; Lang et al., 2022; Piras et al., 2023; Wang et al., 2023). Their findings indicated that differential methylation at key loci correlates with AD pathology, altering the expression of genes implicated in the disease and contributing to neurodegeneration. Consequently, DNA methylation, specifically at CpG sites is increasingly recognised as an important biomarker for AD, with alterations in methylation patterns detectable years before the clinical onset of dementia, as supported by several studies (Fransquet et al., 2020; Kobayashi et al., 2020; Madrid et al., 2018; Mitsumori et al., 2020; Roubroeks et al., 2020). Key examples include ankyrin 1 (*ANK1*) and rhomboid 5 homolog 2 (*RHBDF2*), which were found to be hypermethylated in AD brains, with *ANK1* showing a particularly strong association with AD pathology (De Jager et al., 2014; Lunnon et al., 2014; Sanchez-Mut et al., 2016; Semick et al., 2019). Additionally, bridging integrator 1 (*BIN1*) has been shown

to have methylation changes that may influence amyloid processing and immune responses within AD pathology (De Jager et al., 2014; Gasparoni et al., 2018; Semick et al., 2019; Shireby et al., 2022).

Specific histone modifications, such as trimethylation at lysine 4 (H3K4me3), lysine 27 (H3K27me3), or lysine 27 acethylation (H3K27ac), have been shown to correlate with gene expression changes in AD, indicating that chromatin state alterations may also be pivotal in the disease process (Marzi et al., 2018; Smith et al., 2021). Nott and colleagues (Nott et al., 2019), identified cell-type-specific cis-regulatory elements in AD by mapping regulatory regions associated with both AD and psychiatric disorders. They found that some AD-associated SNPs were primarily located within microglia-specific enhancers. Importantly, deleting a microglia-specific enhancer harboring an AD-risk variant via CRISPR-Cas9 led to the loss of *BIN1* expression specifically in microglia, without affecting neurons or astrocytes. This suggests that sequence polymorphisms in enhancer regions can disrupt regulatory elements, potentially altering histone modifications via transcription factor binding or chromatin accessibility, leading to gene expression changes seen in AD. In conclusion, the epigenetic landscape in AD is characterized by significant, cell-type-specific alterations in DNA modifications and chromatin states, underscoring the intricate regulatory mechanisms that drive the disease.

Altuna et al. identifies a DNA methylation signature in the human hippocampus linked to neurogenesis, reporting differentially methylated genes that were found among the top-ranked genes in previous AD methylome studies conducted on various brain regions, including the frontal and temporal cortices (Altuna et al., 2019). They validated some of the differentially methylated positions using bisulfite cloning sequencing, observing differential methylation in genes such as *RHOB*, which showed a significant correlation with tau burden. *RHOB* is involved in cell-cell interaction and adhesion, essential processes for maintaining synaptic integrity and communication (McNair et al., 2010; Vega et al., 2015). This suggests that the methylation changes observed in the hippocampus may reflect broader patterns relevant to AD pathology.

DNA methylation changes in AD also occur in a brain region-specific manner, reflecting the complex interplay between genetic risk factors, environmental influences, and amyloid pathology. Several studies have identified distinct DNA

methylation patterns in regions particularly vulnerable to amyloid plaque accumulation, emphasizing their role in disease progression. The middle temporal gyrus (MTG), heavily affected by amyloid deposition, shows significant DNA methylation changes linked to neurodegeneration and cognitive decline (Piras et al., 2023). The parahippocampal gyrus (PHG), another key region associated with neuritic amyloid plaques, exhibits specific methylation patterns indicative of early cognitive dysfunction (Wang et al., 2023). As a hub for memory and spatial navigation, these epigenetic changes may reflect its critical role during the early stages of AD pathology. In the later stages of AD, the prefrontal cortex (PFC)—essential for executive functions and higher-order cognition—shows pronounced methylation changes. These alterations strongly correlate with the burden of neuritic amyloid plaques, suggesting a progressive relationship between methylation, amyloid accumulation, and cognitive decline (De Jager et al., 2014).

Iwata et al. reported that aberrant DNA methylation on AD-related genes such as*,* *APP*, *MAPT,* and glycogen synthase kinase 3 beta (*GSK3B*), is associated with AD (Iwata et al., 2014). Roubroeks et al. further indicated that differential methylation at specific loci within the *HOXB6* gene correlates with cognitive dysfunction in AD patients (Roubroeks et al., 2020). In a comprehensive analysis, Semick et al. conducted an integrated study of DNA methylation and gene expression across multiple brain regions, implicating several novel genes contributing to AD. They found that differential methylation was enriched in genes associated with biological processes hypothesized to underlie AD pathology, such as cell adhesion and calcium ion homeostasis (Semick et al., 2019). This suggests that the methylation changes in these genes may play a role in the neurodegenerative processes characteristic of AD. Additionally, Watson et al. identified differentially methylated regions (DMRs) in the superior temporal gyrus, with the majority showing hypermethylation in AD cases compared to controls. This study emphasizes the complexity of methylation changes in AD and the potential for these DMRs to serve as biomarkers for the disease (Watson et al., 2016). Bakulski et al. also reported significant differences in DNA methylation between late-onset AD patients and cognitively normal controls, identifying specific genes that were differentially methylated in the frontal cortex, which is critical for cognitive functions (Bakulski et al., 2012).

The ability to detect these early methylation changes holds immense promise for the development of stage-specific biomarkers, which could improve the precision of diagnostic tools by identifying individuals at risk before clinical symptoms emerge. Targeting these epigenetic modifications could offer new avenues for therapeutic intervention, potentially slowing or even preventing the progression of Alzheimer's disease. Ultimately, this research could lead to more personalized approaches in both diagnosis and treatment, tailoring interventions to the molecular changes occurring at different stages of the disease.

1.4 Application of long-read DNA sequencing to genome-wide methylation analysis in AD mouse models

Long-read sequencing has potential to transform our understanding of the epigenetic landscape in AD. This technology enables the direct detection of DNA modifications without the need for bisulfite treatment, which can introduce biases and artifacts (Lüth et al., 2021). Long-read sequencing, thus, offers a more accurate and genome-wide view of the methylation landscape without the biases associated with bisulfite treatment (Sigurpalsdottir et al., 2024). For instance, a single PromethION flow cell can generate up to 290 billion base pairs of sequencing data—more than enough to detect subtle, yet significant (Snajder et al., 2023).

Beyond its role in epigenetics, long-read sequencing is transforming transcriptomic analysis, particularly in model systems where it provides unparalleled accuracy in isoform detection. One of the key advantages of model systems is the ability to work with biologically consistent experimental groups, allowing for robust identification of recurrent isoforms within the same disease or pathology context. This consistency not only reduces experimental variability but also enhances statistical power, leading to more reliable exon- and isoform-level expression measurements. In addition to facilitating the discovery of novel RNA isoforms and alternative splicing events, long-read sequencing enables cross-model validation, ensuring that observed transcriptomic changes are biologically meaningful and reproducible across different systems. Recent studies have demonstrated that long-read sequencing captures previously undetected transcriptomic complexity, revealing new isoforms and splicing patterns that may play critical roles in disease mechanisms (Gallo et al., 2024; Leung et al., 2021, 2023).

In this thesis, the *App*$^{NL\text{-}G\text{-}F}$ knock-in mouse model was used at 9 months of age, representing an early prodromal stage of AD. This model incorporates three familial AD mutations (Swedish, Iberian, and Arctic) within the endogenous *APP* gene, leading to robust amyloid-β pathology without the confounding effects of transgene overexpression. Compared to traditional transgenic models, which often rely on supraphysiological *APP* expression, *App*$^{NL\text{-}G\text{-}F}$ more accurately reflects physiological amyloid accumulation and its associated molecular changes. However, a key limitation of this model is its inability to fully recapitulate tau pathology, a defining feature of late-stage AD. Additionally, the Arctic (G) mutation presents another constraint: while it enhances Aβ aggregation, the resulting Aβ sequence and conformation differ from those observed in sporadic AD cases**.** Despite this, it remains a powerful system for investigating early amyloid-driven disease mechanisms, particularly in the context of transcriptional and epigenetic alterations that precede neurodegeneration.

We leveraged the long-read DNA sequencing to explore how adjacent CpG sites are methylated, investigating where methylation occurs in the genome within CpG sites. This fine-scale resolution of CpG methylation is correlated to the gene expression, to investigate whether epigenetic changes may potentially drive gene regulation and disease progression. We initially focused on characterising known and novel isoforms, differential RNA isoform usage and isoform switch, and finally alternative splicing in the *App*$^{NL\text{-}G\text{-}F}$ mouse model, examining how these changes may relate to amyloid pathology. In parallel, we performed genome-wide DNA methylation analysis to identify differentially methylated regions (DMRs) associated with amyloid accumulation. Finally, by integrating transcriptomic and epigenomic data, this study aims to uncover methylation patterns and potential novel regulatory mechanisms relevant to AD progression. Additionally, we compared these molecular alterations between mouse models and human AD brains, to identify conserved pathways and candidate genes that could contribute to disease mechanisms and serve as potential biomarkers.

1.5 Thesis aims

To investigate the molecular mechanisms underlying early AD by focusing on transcriptomic and epigenomic changes in response to amyloid pathology. Specifically, it seeks to:

1. Characterize differential RNA isoform usage and alternative splicing in genes using long-read RNA sequencing in the *App*$^{NL\text{-}G\text{-}F}$ mouse model, with an emphasis on understanding how these changes contribute to interactions between cell-types and cellular processes, including synaptic changes and microglial activation, during the early stages of amyloid pathology.
2. Identify novel transcript isoforms and annotate genome-wide, absent from current catalogues, revealing previously uncharacterized mechanisms of gene regulation that may be critical in AD progression.
3. Explore genome-wide DNA methylation patterns in the same mouse model using long-read DNA sequencing, identifying DMRs associated with amyloid pathology.
4. Investigate the relationship between DNA methylation and expression, and correlating these with changes in gene expression to uncover how epigenetic modifications may regulate transcriptomic changes in AD, potentially contributing to disease progression.
5. Compare the new transcriptomic and DNA methylation between mouse models and human AD brains to identify common pathways and candidate genes that could serve as biomarkers or therapeutic targets.

Chapter 2

2 Long-read transcriptomic identification of synaptic adaptation to amyloid pathology in the knock-in *App*$^{NL-G-F}$ mouse model of the earliest phase of Alzheimer's disease

2.1 Introduction

Alternative splicing is a mechanism by which a single gene can produce multiple distinct mRNA transcripts, achieved by the selective inclusion or exclusion of exons during the processing of pre-mRNA (Frankish et al., 2012; Mudge et al., 2011). The composition of regulatory sequences in the mature mRNA can influence various properties of the transcript, such as its stability, efficiency of translation, and localization within the cell. For transcripts that are translated into proteins, alternative splicing can lead to differences in protein domains, variation in expression and protein levels, altered cellular localization, or the complete loss of a functional protein, as seen in cases involving nonsense-mediated decay.

2.1.1 Aims and objectives

The aim of chapter 2 is to investigate the transcriptomic changes, particularly focusing on alternative splicing and isoform usage, in response to amyloid pathology in the *App*$^{NL-G-F}$ mouse model of AD. To elucidate how these molecular changes contribute to difference cellular processes during AD during the earliest phases of amyloid, following objectives were set:

1. Identify differentially expressed genes and isoforms in response to amyloid pathology and obtain expected enrichment of microglia-expressed AD risk genes. Perform cell type enrichment analysis to understand whether isoform-level expression is enriched in cell types other than immune cells.
2. Examine alternative splicing events and isoform usage genome-wide and uncover novel mechanisms of transcript regulation in response to amyloid plaques.

3. Perform functional annotation and pathway enrichment analysis to explore the biological significance of gene and isoform level analysis, highlighting their potential contributions to AD.

2.2 Methods

2.2.1 Animal genetics and experimental design

The *App*$^{NL\text{-}G\text{-}F}$ mouse model (Apptm3.1Tcs, MGI:5637817)(Saito et al., 2014) was maintained on a C57BL/6J genetic background as a heterozygous backcross. Cohorts of homozygous and WT controls were generated by heterozygous crosses. The study used App+/+ (WT allele homozygotes) male n=23, female n=21, and *App*$^{NL\text{-}G\text{-}F/NL\text{-}G\text{-}F}$ (mutant allele homozygotes) male n=15, female n=24. Animals were housed under specific pathogen-free (SPF) conditions in individually ventilated cages at the Mary Lyon Centre, MRC Harwell Institute, with 3-5 animals per cage. Mice had access to a cardboard tunnel with bedding material and wood chips (grade 4 aspen) and were provided with water and RM3 chow ad libitum. To generate WT and homozygous *App*$^{NL\text{-}G\text{-}F}$ animals, heterozygous *App*$^{NL\text{-}G\text{-}F}$ animals were bred in an intercross. Animals were euthanized under sodium pentobarbitone (Euthatal) anaesthesia by terminal perfusion with 1X phosphate-buffered saline (PBS) in accordance with the Animals (Scientific Procedures) Act 1986 (United Kingdom). The right hemisphere of the brain was separated and fixed in 10% buffered formal saline, while the left hemisphere was dissected into cortex, hippocampus, and cerebellum, snap-frozen on dry ice, and stored at -70°C.

A subset of these animals were used for long-read sequencing analysis *App*$^{+/+}$ (WT allele) male n=3, female n=3, and *App*$^{NL\text{-}G\text{-}F/NL\text{-}G\text{-}F}$ (homozygous for the mutant allele) male n=3, female n=3. A further subset of animals were used for biochemical amyloid-β quantification *App*$^{+/+}$ (WT allele) male n=6, female n=6, and *App*$^{NL\text{-}G\text{-}F/NL\text{-}G\text{-}F}$ (homozygous for the mutant allele) male n=6, female n=6. At the time of the experiment, no long-read bulk RNA-seq datasets were available for AD mouse models, particularly for the diverse types of analyses conducted in this study, such as transcript-level differential gene expression and isoform usage. Consequently, the n=6 sample size was selected based on established short-read bulk RNA-seq standards, where this number has been shown to provide sufficient power for detecting gene expression differences. Additionally, to mitigate potential confounding factors, groups

were sex-balanced, and samples were randomised across sequencing batches to control for gender and batch effects.

2.2.2 DNA extraction and genotyping

DNA was extracted from ear biopsy, isolated at postnatal day 14 using TaqMan Sample-to-SNP (Applied Biosystems). Mice were genotyped for the *App*$^{NL\text{-}G\text{-}F}$ allele using TaqMan multiplexed qPCR for the NL mutant and WT *App* alleles and *Dot1l* reference allele; using the following primers (forward 5′-GGAAGAGATCTCGGAAGTGAAGA-3′; reverse 5′-CAGTTTTTGATGGCGGACTTCAA-3′) and probes (5′-FAM-TGGATGCAGAATTCGGACATG-BHQ1-3′) for the *App* WT allele and the following forward 5′-CGGAAGAGATCTCGGAAGTGAATCT-3′; reverse 5′-ACCAGTTTTTGATGGTGGACTTCA-3′) and probes (5′- FAM-AGATGCAGAATTCAGACATGATTC-BHQ1-3′) for the *App*^{NL} mutant allele and *Dot1l* allele primers (forward 5′-TAGTTGGCATCCTTATGCTTCATC-3′; reverse 5′-GCCCCAGCACGACCATT-3′) and probe (5′- VIC-CCAGCTCTCAAGTCG-MGBNFQ-3′). G and F mutant and WT *App* alleles were genotyped using allelic discrimination assays; using the following primers for the *App* WT/G allele (forward 5′-CGATGATGGCGCCTTTGTTC-3′; reverse 5′-GTTGCCTCTTGCGCTTACAG-3′) and probes (*App*-WT allele 5′-TET-ACCCACATCTTCAGCAA-BHQ1-3′ and *App*-mutant (G) allele 5′- FAM-CCACATCTCCAGCAAA-BHQ1-3′) and the following primers for the *App* WT/F allele (forward 5′-GTGGGCGGCGTTGTCA-3′; reverse 5′-CGCCATGATGGATGGATGTGTA -3′) and probes (*App*-WT allele 5′- FAM-AGCAACCGTGATTGTCAT-BHQ1-3′ and *App*-mutant (F) allele 5′-TET-AGCAACCGTGTTTGTC-BHQ1-3′).

2.2.3 Tissue preparation

Following perfusion with PBS (pH 7.4) brain samples were dissected and divided along the sagittal midline with one hemisphere immersion fixed in 10% NBF (neutral, phosphate buffered formalin) for a minimum of 48 hours prior to tissue processing and paraffin embedding. Once dissected, tissue samples were dehydrated, cleared, and processed into paraffin wax using a Tissue-Tek VIP 6 AI (Sakura) tissue processor and finally embedded into paraffin wax blocks using a HistoCore Arcadia (Leica). The

remaining hemisphere was further dissected into the hippocampal and cortex regions which were then placed into cryotubes and immediately snap frozen in liquid nitrogen.

2.2.4 Immunohistochemistry of mouse brain

The formalin-fixed paraffin-embedded (FFPE) blocks were trimmed dorsally to give a coronal section of either the ventral or dorsal hippocampus. For each FFPE block, two 4 µm tissue sections (40 µm apart) were mounted onto SuperFrost Plus slides for subsequent staining and analysis. First, the tissue sections were deparaffinized and rehydrated in a Gemini AS Automated Stainer (Epredia) using xylene and a series of ethanol baths (100%, 95%, 85%) and washed with distilled water. The sections were then pretreated with 80% formic acid for 8 minutes. After pretreatment, a Ventana ULTRA automated stainer was used for the following; heat-induced epitope retrieval was performed for 16 minutes at 100^{o}C in Tris boric acid EDTA buffer (Ventana Medical Systems, 06414575001); endogenous peroxidases were quenched with Inhibitor D (DABMap™, Ventana Medical System); the primary antibody incubation was performed using biotinylated mouse monoclonal IgG1 antibodies against Aß (82E1, IBL, 0.2 µg/ml) for 8 hours at room temperature; chromogen visualization was achieved using Ventana DABMap™ kit (Ventana Medical Systems); the sections were counterstained with Haematoxylin II (4 minutes, Ventana Medical Systems) and treated with a bluing agent (4 minutes, Ventana Medical Systems) for clearer imaging. Stained slides were washed with a mild detergent and dehydrated/cleared in a Gemini AS Automated Stainer using a series of ethanol baths (85%, 95%, 100%) and xylene. Stained slides were then mounted with coverslips using a ClearVue Coverslipper (Epredia) before slide scanning with the Zeiss Axio Scan Z1 slide scanner.

Regions of interest (ROI) of cortex and hippocampus for each tissue section were selected from the scans using QuPath (Bankhead et al., 2017). Analysis of ROI images was conducted with an ImageJ macro (Abràmoff et al., 2004). Briefly, a standard threshold of pixel value was used to create a binarized mask (Binary Mask plugin) of all ROI images. Pixels above the threshold were identified as Aβ staining, while pixels below the threshold were identified as not containing Aβ staining. ImageJ quantified the percentage of the pixels within the ROI covered by the mask (Analyze Particles plugin), thus, positive for staining.

2.2.5 Tissue fractionation for amyloid-β MSD assay

Total cortical proteins were fractionated based on the method in Shankar *et al.* (2009) (Shankar et al., 2009). Total cortex was homogenized in three volumes of the weight of the heaviest sample of ice-cold Tris-buffered saline (TBS) (50 mM Tris-HCl pH 8.0) plus complete protease and phosphatase inhibitors (Roche). Homogenates were centrifuged at 186 000 *g* at 4°C for 30 min, and the resultant supernatant (the soluble TBS fraction) was stored at −80°C. The resultant pellet was homogenized in three volumes times the heaviest sample of 1% Triton™ X-100 in TBS plus complete protease and phosphatase inhibitors (Roche) and centrifuged at 186 000 *g* at 4°C for 30 minutes, and the resultant supernatant (the Triton soluble fraction) was stored at −80°C. The resultant pellet was homogenized in three volumes of 50 mM Tris-HCl buffer, pH 8.0, containing 5 M guanidine-HCl plus complete protease and phosphatase inhibitors (Roche). This resuspension (the guanidine HCl soluble fraction) was incubated at 4°C for a minimum of 14 hours with shaking and was stored at −80°C. Protein concentration was determined by Bradford assay (Bio-Rad).

Samples were then analysed by human amyloid-β 6E10 Triplex (Meso Scale Discovery) following the manufacturer's protocols. Briefly the TBS, Triton, and guanidine HCl soluble fractions were diluted into Diluent 35 (Meso Scale Discovery) and added to a precoated plate prior to addition of amyloid-β detection antibody and incubation overnight at 4°C (Meso Scale Discovery). After washing, Read Buffer (Meso Scale Discovery) was applied immediately prior to plate reading on a Meso Scale Discovery Sector Imager S600. Analyte data (pg/ml) was normalised to weight of brain region divided by total buffer homogenised in (mg/ml) to result in amount of amyloid-β analyte per brain weight (pg/mg).

2.2.6 Long-read RNA-sequencing and data pre-processing

Total RNA was extracted from tissue samples using the Monach Total RNA kit (NEB) and assessed for quality using RNA Integrity Number (RIN) values via Tapestation (Agilent). Libraries were prepared using the PCR-cDNA Barcoding kit (SQK-PCB109) from Oxford Nanopore Technologies, following the manufacturer's protocol. Briefly, 50 ng of total RNA were converted into cDNA through reverse transcription using Maxima H Minus Reserve Transcriptase. A strand-switching primer was added to guarantee the selection of full transcripts. cDNAs were amplified and barcoded with a unique

barcode through a 15 cycle PCR with 7 minutes of extension time. Six pools composed of two samples each (thus two individual mice pooled, one mutant and one WT in each pool) were created and purified using Ampure XP beads. After adapter ligation, each pool was sequenced in a PromethION flow cell (FLO-PRO002) for 72 hours. Flow cells were refuelled when the translocation speed was low, and washed once and reloaded with another library aliquot. Base calling was performed in real time with the "High accuracy base calling" mode using Guppy 4.3.4.

2.2.7 Pipeline

Quality control was performed with fastQC (Simon Andrews, 2020) on the raw sequencing data. Around 30 million reads for each sample passed the initial QC step and the estimated N50, statistical measure of average length of a set of sequence, was found to be ~1kb. The pipeline from Oxford Nanopore Technologies, IsoformSwitchAnalyzer, and PSI-Sigma tools were used to obtain transcript assembly, differential isoform usage and splicing detection, respectively.

2.2.8 Genomic alignment

Raw reads obtained from Oxford Nanopore Technology (ONT) flow cells were aligned to the latest release of the GENCODE reference genome (GRCm39) with Minimap2 (Li, 2018). The alignment parameters '-ax splice: hq -uf –secondary=no' were used for the splice-aware alignment. The latest GENCODE annotation (GRCm39) was used for the junction information. The bam files were sorted with samtools (Li et al., 2009).

2.2.9 Gene and isoform abundance estimation

Assembly (gtf) files of each sample were used as an input list to the *prepDE.py* function according to the manual of StringTie (Pertea et al., 2015, 2016). For the quantification step, all gtf files were collapsed using Stringtie --merge option. The merged gtf file was used for the quantification of the isoforms. StringTie -eB parameters were used to obtain read-count data.

2.2.10 Differential expression analysis with DESeq2

Gene and isoform level differential expression analysis were performed with DESeq2 (Anders et al., 2012; Love et al., 2014) with raw counts obtained from StringTie (Pertea

et al., 2016) as described in Pertea *et al.* (2016). Sex was included to the contrast design.

2.2.11 Exon and isoform usage via DEXSeq

Differential transcript usage (DTU) analysis between genotypes was performed using DEXSeq (Anders et al., 2012). For exon usage, the full experimental design formula used for this analysis was ~sample + exon + Sex:exon + Genotype:exon. To isolate the specific effects of genotype, the full model was compared to a reduced model (~sample + exon + Sex:exon). The statistical test for differential exon usage was carried out using the *testForDEU* function in DEXSeq, which compares the full model to the reduced model and identifies significant differences in exon usage attributed to genotype.

For isoform usage, a similar approach was used, with the focus on detecting differential isoform usage across genotypes. This was tested separately from exon usage. Isoform-level differences were assessed by comparing the usage of specific transcript isoforms between the two genotype groups, using the same model design framework to control for sex and sample effects. Statistical significance for differential isoform usage was also determined using the *testForDEU* function.

2.2.12 Functional annotation via two-stage tappAS analysis

We conducted additional detailed analysis of isoform diversity and functional consequences using tappAS (De La Fuente et al., 2020), a Java Application that integrates various functional analyses of isoforms. We used several existing tools implemented in tappAS for our analysis. To generate a functionally annotated gtf file, we used IsoAnnot$_{Lite}$ (https://isoannot.tappas.org/) with a SQANTI3 output (Tardaguila et al., 2018) file obtained from abovementioned merged gtf file from StringTie to provide input to tappAS Application to perform the DTU analysis(De La Fuente et al., 2020). The *sqanti3_qc.py* and *sqanti3_filter.py* functions were used with the merged gtf file from gffcompare (Pertea & Pertea, 2020). tappAS additionally outputs isoform usage and isoform switch, which was included in the GO annotations of the usage and switch analyses.

2.2.13 Isoform switch analysis and predicted functional consequences

We performed isoform switch analysis using IsoformSwitchAnalyzer (Vitting-Seerup et al., 2019). Briefly, we identified isoform switches using the merged transcript assembly from StringTie (Pertea et al., 2015), and generated isoform-level count data using *prepDE.py* function. Next, we checked whether the transcripts of interest contained Open Reading Frames (ORF/CDS) (Vitting-Seerup et al., 2014; Vitting-Seerup & Sandelin, 2017), as assesed by coding potential calculator version 2 (CPC2) (Y. J. Kang et al., 2017). We also evaluated whether the isoform switches resulted in any changes in protein domains using the protein family database (Pfam) (Mistry et al., 2021), idenfidied the presence of signal peptides and their cleavage sites using signal peptide prediction tool (SignalP) (Almagro Armenteros et al., 2019), and analysed intrinsically disordered regions using intrinsically disordered protein prediction tool version 2A (IUPred2A) (Mészáros et al., 2018). Finally, we performed Gene Ontology (GO) (Harris et al., 2004) and pathway enrichment analysis on the transcripts showing switch events. For isoform switch/usage overlap analysis with human AD data, short-read RNA-seq data from the Mayo Clinic, Mount Sinai Brain Bank (MSBB), and ROSMAP cohorts were incorporated (Marques-Coelho et al., 2021).

2.2.14 Alternative splicing analysis

We used PSI-Sigma for comprehensive splicing detection for long-read RNA-seq (Lin & Krainer, 2019). Compared to other alternative splicing analyses methods, PSI-Sigma has several advantages over short-read splicing detection tools such as covering important alternative splicing events, such as multiple-exon-skipping (MES) or more complex splicing events.

2.2.15 Data Availability

All data will be made available via GEO. The dataset for differential transcript usage comparisons across Mayo Clinic, Mount Sinai Brain Bank (MSBB), and ROSMAP cohorts (Marques-Coelho et al., 2021) is available at https://diegomscoelho.github.io/AD-IsoformSwitch/.

2.2.16 Software and Algorithms

Minimap2 (Li, 2018)	Long-read aligner	https://github.com/lh3/minimap2
Pychopper	Identifying, orienting and trimming full-length Nanopore cDNA reads	https://github.com/epi2me-labs/pychopper
StringTie (Pertea et al., 2015)	Transcriptome assembly	https://github.com/gpertea/stringtie
GffCompare (Pertea & Pertea, 2020)	Comparing and annotating transcriptome annotations	https://ccb.jhu.edu/ / stringtie/gffcompare.shtml
clusterProfiler (Wu et al., 2021; Yu et al., 2012)	Gene set enrichment analysis	https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
Snakemake (Köster & Rahmann, 2012)	Workflow management system that helps build reproducible and scalable data analyses.	https://snakemake.readthedocs.io/
Salmon (Patro et al., 2017)	Used for quantifying gene expression from RNA-seq data	https://salmon.readthedocs.io/.
DESeq2 (Love et al., 2014)	Statistical software package for differential expression analysis of RNA-seq data	https://bioconductor.org/packages/release/bioc/html/DESeq2.html
DEXSeq (Anders et al., 2012; Tardaguila et al., 2018)	Used for analysis of differential exon usage in RNA-seq data	https://bioconductor.org/packages/release/bioc/html/DEXSeq.html

CoExpNets (Botía et al., 2017)	Weighted Gene Co-expression Network Analysis with additional k-means clustering	https://github.com/juanbot/CoExpNets
SQANTI3 (Tardaguila et al., 2018)	Transcriptome quality control	https://github.com/ConesaLab/SQANTI
stageR (Van den Berge et al., 2017)	Statistical analysis of gene expression data	https://bioconductor.org/packages/release/bioc/html/StageR.html
tappAS (De La Fuente et al., 2020; Van den Berge et al., 2017)	Computational framework for functional transcriptomics analysis	https://App.tappas.org
PSI-Sigma (Lin & Krainer, 2019)	Splicing detection method	https://github.com/wososa/PSI-Sigma

2.3 Results

2.3.1 Gene-level expression changes confirm microglial proliferation and activation

To investigate the impact of amyloid pathology on transcription isoforms and splicing of gene expression, we performed long-read RNA-seq on total cortical RNA of *App*$^{NL-G-F}$ knock-in mice and littermate C57BL/6J controls at 9 months of age. Initially amyloid accumulation was confirmed and shown to be similar to previous studies (Benitez et al., 2021; Saito et al., 2014) (Fig. 2.1., Fig. 2.2.). We analysed the long-read RNA-seq to determine differential gene-level expression, usage of individual transcript isoforms, exon switching, alternative splicing, alternative polyadenylation sequences and changes to the protein coding sequence, using DESeq2 (Love et al., 2014), DEXSeq (Anders et al., 2012), PSI-Sigma (Lin & Krainer, 2019; Vitting-Seerup et al., 2019), IsoformSwitchAnalyzer (Vitting-Seerup et al., 2019), and tappAS which is a part of the Functional Iso-Transcriptomics (FIT) framework (De La Fuente et al., 2020).

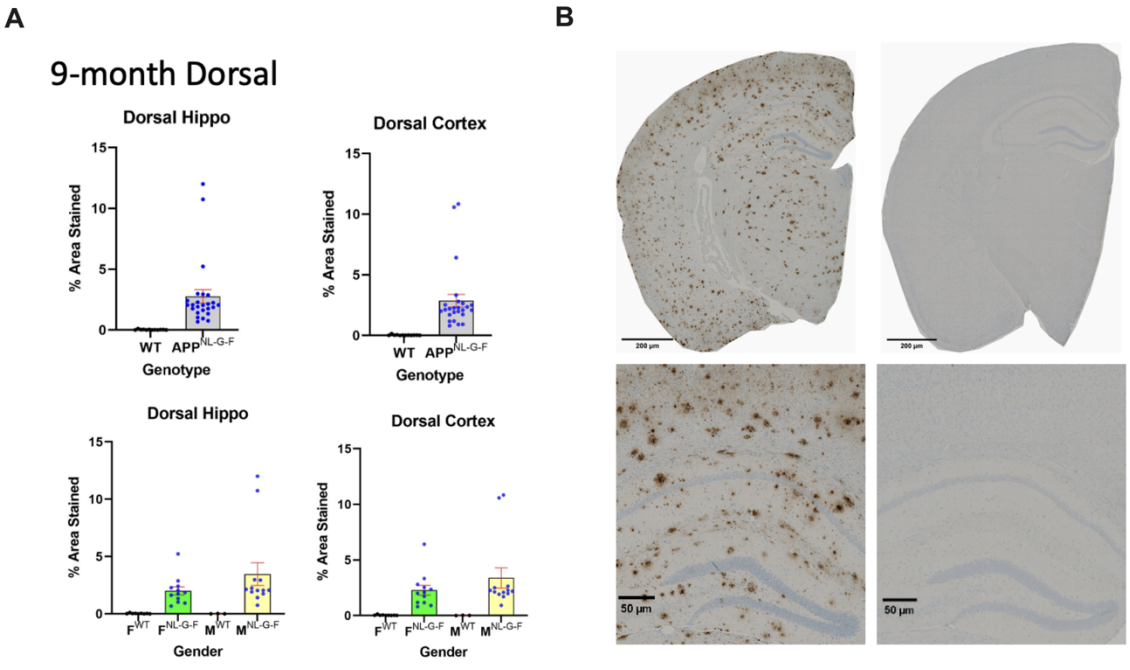

Figure 2-1 Aß plaque coverage via immunohistochemistry in *App*$^{NL-G-F}$ mice in dorsal hippocampus and cortex.

(A) This figure illustrates the Aß plaque coverage in both female and male *App*$^{NL-G-F}$ mice in dorsal hippocampus and cortex compared to WT controls. (B) Immunostained

brain sections from both *App*$^{NL-G-F}$ mice (left panels) and WT control mice (right panels) reveal Aβ plaque staining exclusively in *App*$^{NL-G-F}$ mice, highlighting the distinctive Aβ distribution.

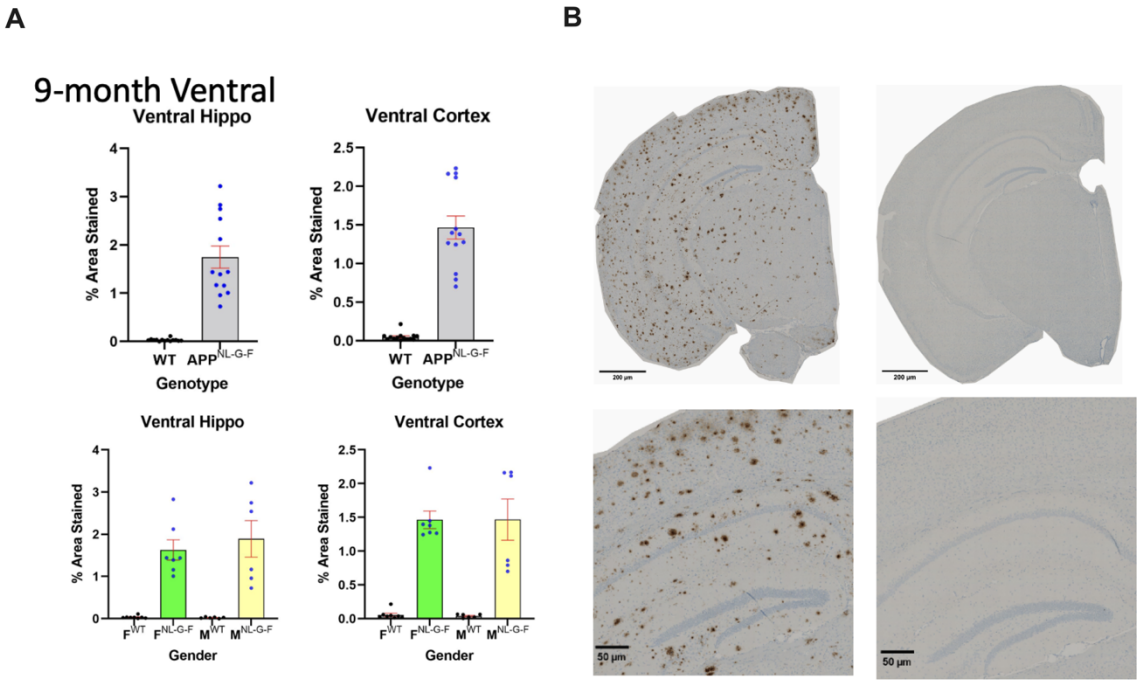

Figure 2-2 Aβ plaque coverage via immunohistochemistry in *App*$^{NL-G-F}$ mice mice in ventral hippocampus and cortex.

(A) This figure presents Aβ coverage measured by immunohistochemistry assay in ventral hippocampus and cortex, categorized by gender *App*$^{NL-G-F}$ mice compared to controls. (B) Immunostained brain sections in both *App*$^{NL-G-F}$ mice and control mice show Aβ staining specifically in *App*$^{NL-G-F}$ mice, emphasizing the unique Aβ distribution pattern in these mice.

Our differential gene-level expression analysis using DESeq2 confirmed the expected gene-level changes at the age of 9 months in the cortex, as seen in previous studies with short read RNA-seq using the same mouse model (Sala Frigerio et al., 2019), and other mouse models with amyloid plaques (Keren-Shaul et al., 2017; Matarin et al., 2015; Salih et al., 2019).

Our analysis identified 178 differentially expressed genes (176 upregulated and 2 downregulated, |log2FC| > 0.5 and FDR < 0.05) in *App*$^{NL-G-F}$ knock-in versus control mice (Fig. 2.3a; Supplementary Table 2.1.), showing a significant overlap with the amyloid-responsive microglia (ARM), also known as the disease-associated microglial

(DAM) gene cluster identified in previous literature using single-cell RNA-seq analysis (Fisher's exact test, p-value < 2.2e-16)(Keren-Shaul et al., 2017), including many well-known genes such as *Trem2*, *Tyrobp* and *Ctsd*, and orthologues of many AD risk genes. Biological annotation of the differentially expressed genes highlighted terms associated with the expected immune system associated processes in $App^{NL\text{-}G\text{-}F}$ mice (Fig. 2.3b), reflecting the proliferatiovn and activation of microglia. Network analysis to identify the co-expressed genes in response to amyloid produced a genetic network representing amyloid-responsive microglia, with *Spi1* and *Ctsd* as hub genes, which are proposed to drive the observed transcriptional response to amyloid (Fig. 2.3c; Supplementary Table 2.2.). This amyloid-responsive microglial network was similar to those described previously by our group by short-read RNA-seq and microarrays in other mouse models with amyloid plaques (Matarin et al., 2015; Salih et al., 2019).

We also investigated if changes in gene-expression were sex-dependent, but only a limited number of differentially expressed genes were observed to be dependent on sex at 9 months of age (6 differentially expressed genes; Supplementary Table 2.3.). These limited genes had no obvious enrichment for known biological annotations, and this indicated that sex-dependent changes in gene expression were limited at this relatively young age of 9 months.

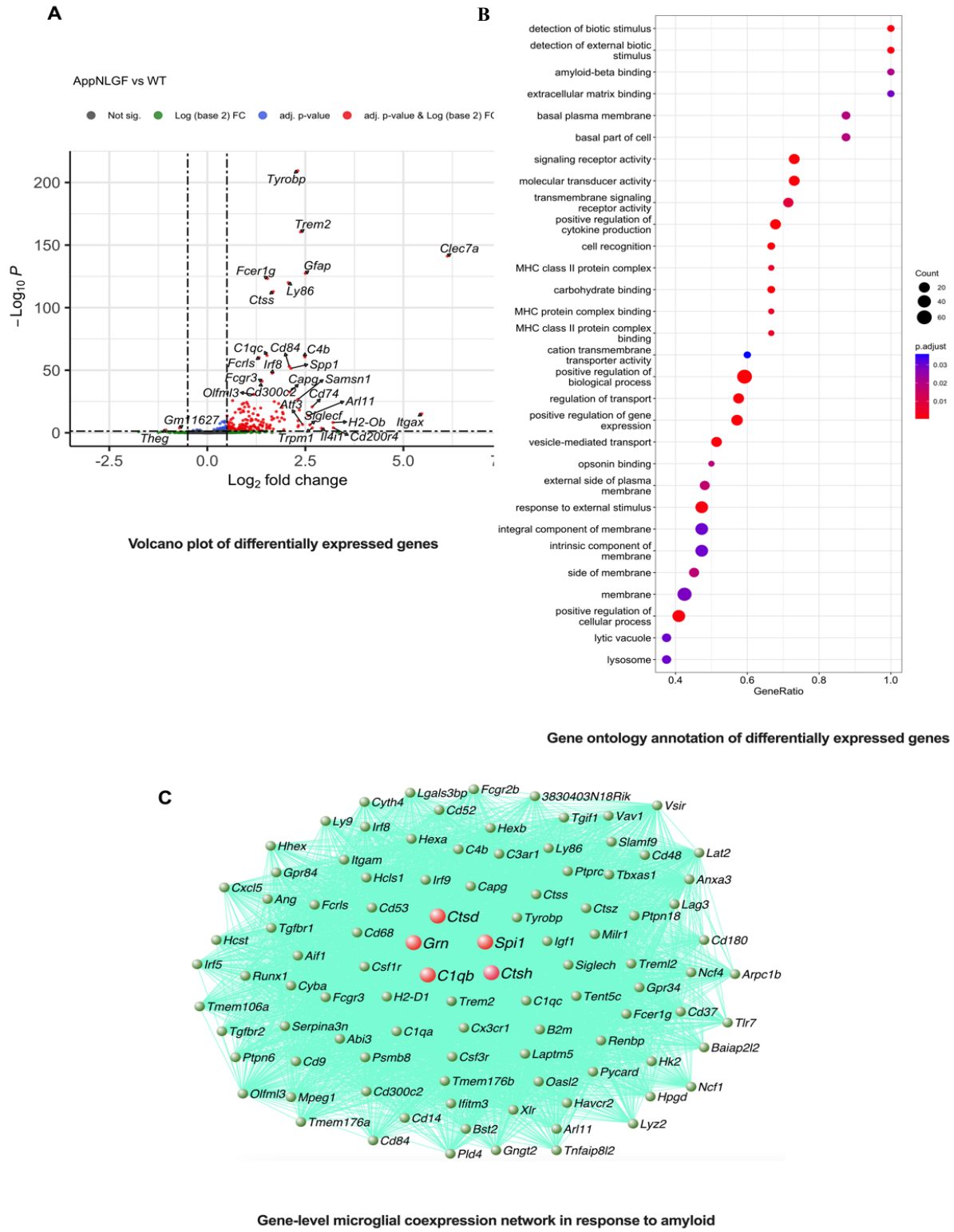

Figure 2-3 Gene-level differential expression analysis in *App*$^{NL-G-F}$ versus control mice at 9 months of age using long-read RNA-seq

(A) The differential expression analysis confirms the microglial response to amyloid plaques, with many genes associated with ARM/DAM (*e.g. Trem2, Tyrobp, Ctsd, C1qb*), assessed by DESeq2 (FDR < 0.05). The analysis was performed on a cohort

of n=6 mice for each genotype, with an equal representation of n=3 males and 3 females per genotype. (B) Differentially expressed genes are enriched for immune cell activation pathways, indicating a microglial response to amyloid plaques. (C) The coexpression analysis of gene-level expression using long-read RNA-seq reveals an overlap with the ARM/DAM response generated using short-read sequencing data (Keren-Shaul et al., 2017; Sala Frigerio et al., 2019).

2.3.2 Isoform-level differential expression changes refine myeloid cell function

A number of isoforms, such as *Ccl6-201*, *chemokine (C-C motif) ligand 6* (Fig. 2.4a; Supplementary Table 2.4.), displayed significant fold-expression changes at the transcript level as compared to whole gene expression level. Specific *Trem2* isoforms (*ENSMUST00000113237.3,* log2FC = 2.75, FDR = 3.41e-118 and *ENSMUST00000024791.14,* log2FC = 2.28, FDR = 6.41e-42), an *Apoe* isoform *(ENSMUST00000174064.8)*, and a *Gfap* isoform (*ENSMUST00000067444.9,* log2FC = 2.60, FDR = 4.38e-46), amongst others, were highly expressed in the cortex of $App^{NL\text{-}G\text{-}F}$ mice compared to WT controls.

This suggests that certain genes may exert effects via the transcript level, potentially driving specific microglial network functions, in response to the amyloid pathology (Fig 2.4a). Our isoform-level annotation via GO terms revealed changes in chemokine activity, reduced endopeptidase activity, and signalling receptor regulation (Fig. 2.4b). The hub isoforms represented by gene names in Fig. 2.5 are likely to be major cellular drivers at the isoform-scale, which are not evident at the gene level, suggesting specific isoforms of genes may contribute to the early pathological changes. Our findings highlight the significance of isoform-level changes, which offer new insights into subtle yet significant cellular alterations that may go unnoticed at the gene level. This observation emphasizes the importance of considering isoform-level dynamics to fully comprehend the molecular mechanisms underlying the response to pathology. Moreover, the enrichment of genes with altered expression within the broader innate immune response pathway suggests that these genes may contribute to functional differences in the orchestrated immune response to pathology. These findings underscore the complex interplay between isoform-level regulation and the broader gene-level enrichment within the innate immune response pathway.

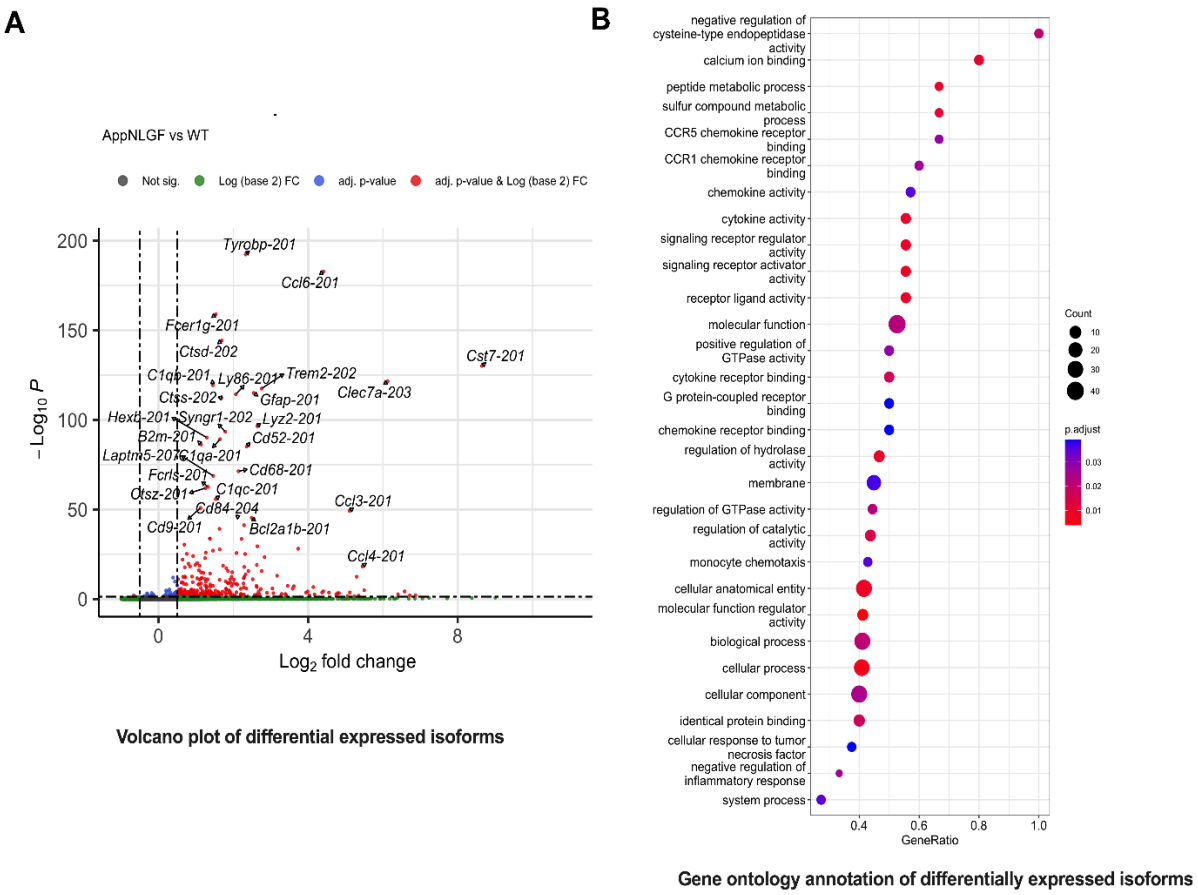

Figure 2-4 Transcript isoform-level expression analysis and enrichment of immune-related pathways.

(A) Isoform-level expression analysis revealed that several transcripts, such as *Tyrobp-201*, *Ccl6-201*, exhibited greater fold-expression level changes at the transcript-level compared to the gene-level, assessed by DESeq2 (FDR < 0.05). This suggests strong preference bias of specific transcript isoforms in response to amyloid pathology (n=6 per genotype). (B) Enrichment analysis showed prominent enrichment of cytokine activity, membrane transport and metabolism. GO enrichments highlight the importance of isoform-level changes in understanding the cellular response to pathology and underscore the interplay between isoform-level regulation and the broader gene-level enrichment within various pathways encompassing the innate immune response pathway and membrane trafficking.

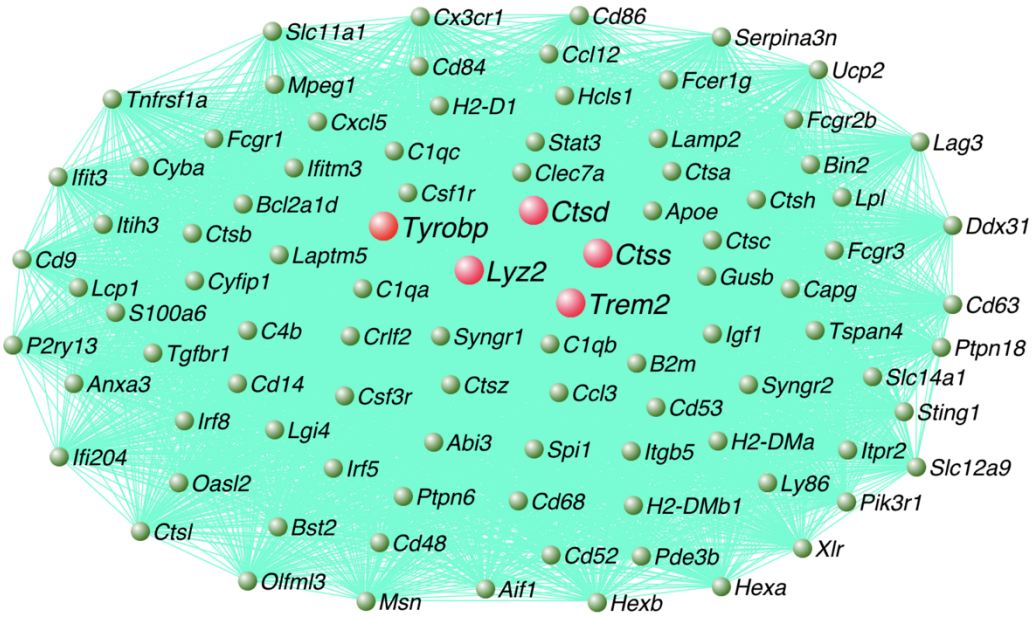

Figure 2-5 Microglial isoform-level coexpression network.

This figure depicts the coexpression network of isoforms that are differentially expressed in response to amyloid pathology. For simplicity, isoforms are labeled by their corresponding gene names and are predominantly expressed in microglial cells.

2.3.3 Novel transcript isoforms of familial AD genes and risk genes

The use of genome-wide long-read RNA-seq enabled the identification of several transcript variants originating from the canonical dementia causative and risk genes. Notably, novel transcripts of *Apoe, App, Mapt,* and *Oas1a* were discovered, that were previously absent in the Ensembl catalogue (Supplementary Table 2.5.). Among these, compared to the reference annotation in Fig. 2.6a, we note a novel *Apoe* anti-sense transcript represented by StringTie identifier MSTRG.40715.1 (Fig 2.6b). While the various transcripts of these dementia causative and risk genes likely play distinct roles under specific conditions (e.g. during development), the changes in the levels of these isoforms at 9 months of age were not dependent on Aß.

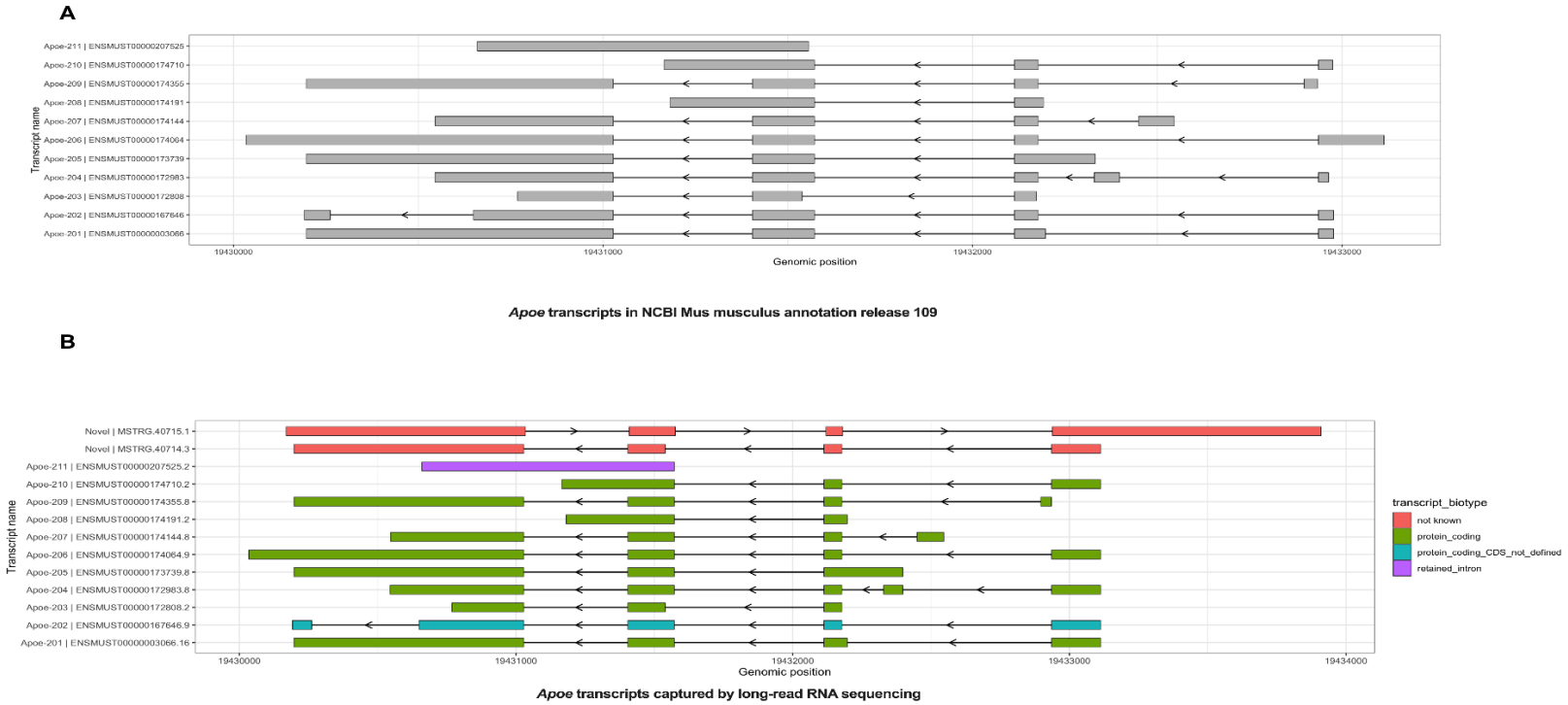

Figure 2-6 Integrative Genomic Viewer visualization of all detected *Apoe* transcripts.

(a) Canonical *Apoe* transcripts in reference annotation (NCBI Mus Musculus 109). The common *Apoe-201* isoform is shown at bottom. (b) The isoforms detected via long-read RNA-seq represents two novel transcript variants of the *Apoe* gene. In addition to the canonical transcripts, our analysis uncovered two novel transcripts (MSTRG.40715.1, MSTRG.40714.3), including an antisense transcript aligned in the reverse direction of the reference genome.

2.3.4 Detecting genome-wide splicing patterns, isoform switching, and isoform/exon usage using long-reads

We employed the DEXSeq analysis to investigate the intricate mechanisms underlying gene and transcript expression in *App*$^{NL-G-F}$ mice compared to controls. This method utilizes a negative binomial distribution to model the feature counts of each gene, such as exons, and incorporates Generalized Linear Models with interaction terms to capture the interplay between different features within the same gene (Anders et al., 2012; Love et al., 2018) (Supplementary Table 2.6). To gain further insights into the landscape of alternative splicing, we initially performed differential transcript usage (DTU) and differential exon usage (DEU) analyses, which are calculations based on isoform and exon fraction per sample. Changes in isoform usage provide additional clarity on instances where the proportional contribution of isoforms relative to the overall gene expression undergoes significant alterations between different genotypes (also known as isoform switching). Isoform switch detection and coding potential analysis, protein domain identification and external annotations of the novel and known isoforms have been performed via the IsoformSwitchAnalyzer tool (Vitting-Seerup et al., 2019). To further elucidate the functional ramifications of altered transcript/exon use and alternative splicing, and their potential effects on changes at the protein level, we employed the tappAS tool which is a part of the Functional Iso-transcriptomics Framework (De La Fuente et al., 2020), also accounting for differential polyadenylation usage of transcripts. In addition, we identified the genome-wide splicing profile using PSI-Sigma, a tool developed for detecting alternative splicing events in both short and long-read RNA-seq (Lin & Krainer, 2019). The combination of genome-wide gene- and isoform-level analyses, along with improved transcript detection and quantification, enabled us to validate and identify alternative splicing events while also predicting their potential functional consequences. Additionally, the preferential usage of specific transcript isoforms suggests that the selective expression of certain exons may play an important role in mRNA regulation, cellular trafficking, and potentially influence protein function by altering the protein sequence.

2.3.5 Selective transcript usage and exon usage

Our genome-wide analysis using *DEXSeq* and *tappAS* revealed differential isoform usage events in response to amyloid plaques (Fig. 2.7a, Supplementary Table 2.6, and 2.7), including several genes associated with the ARM/DAM phenotype, such as *Apoe* (*ENSMUST00000174064*, FDR= 2.59e-03), *Colony Stimulating Factor 1 (Csf1, ENSMUST00000120243*, FDR = 4.22e-03*), Synaptogyrin 1 (Syngr1, ENSMUST00000009728,* FDR = 6.17e-34), *Colony Stimulating Factor 2 Receptor Subunit Alpha (Csf2ra, ENSMUST00000235172,* FDR = 4.20e-03), *Cell Adhesion Molecule 2 (Cadm2, ENSMUST00000114548,* FDR = 7.00e-03), and *Insulin-like growth factor 1 (Igf1, ENSMUST00000121952,* FDR = 5.20e-05). Strikingly, the GO annotations associated with these genes exhibiting transcript preference were predominantly enriched in post-synapse organization, microtubule cytoskeleton, regulation of MAPK cascade, interferon-γ production, and intracellular signal transduction; shedding light on their role on synaptic pathways in response to amyloid pathology (Fig. 2.7b).

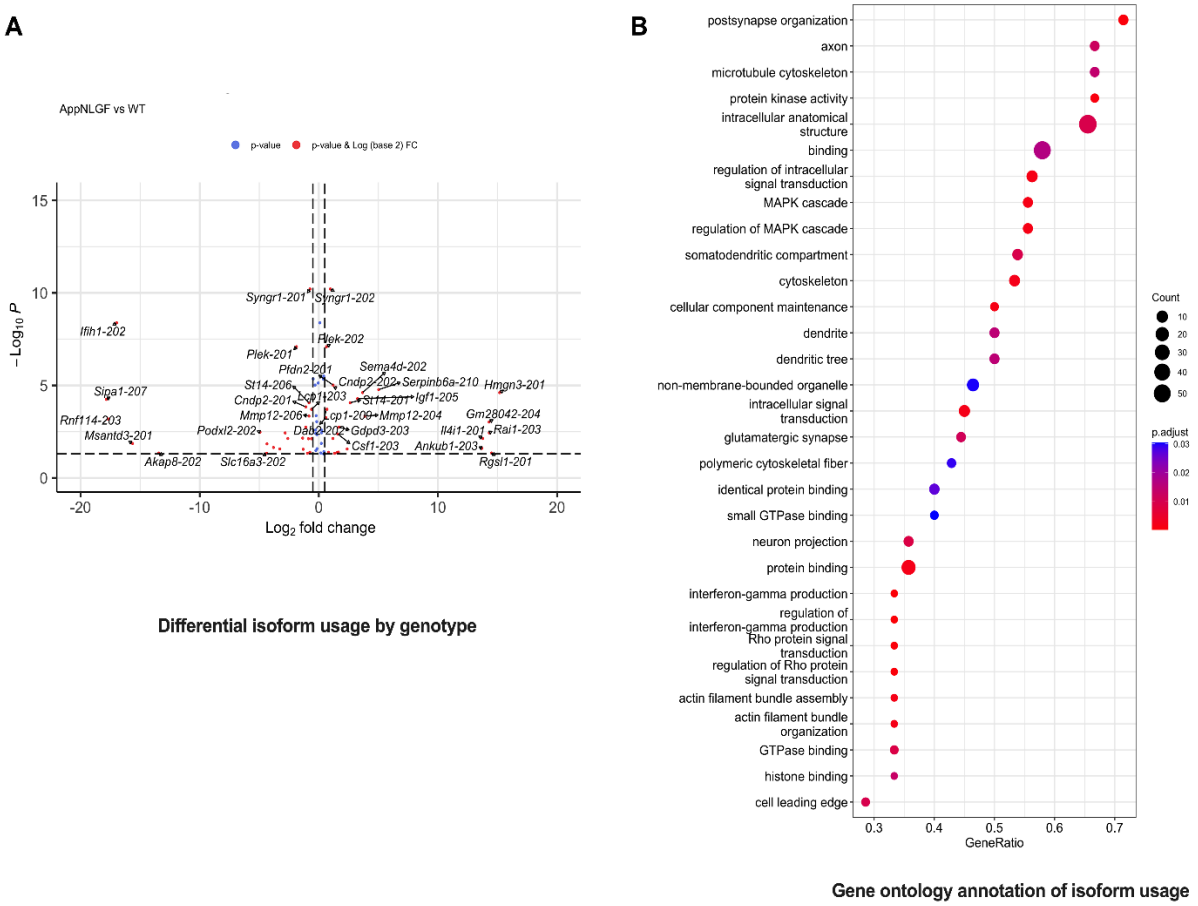

Figure 2-7 Differential transcript usage analysis of genes in *App*$^{NL-G-F}$ vs control mice at 9 months of age using long-read RNA-seq.

(a) DEXSeq isoform usage analysis identified a total of 70 transcripts that exhibit preferential usage of specific isoforms in response to amyloid plaques (n=6). (b) Differentially used transcript isoforms are enriched in postsynaptic organisation, regulation of intracellular signal transduction and the MAPK cascade. These findings diverge notably from the outcomes observed in both gene-level and isoform-level expression analyses.

The isoform switch results revealed several novel transcript variants of Capping Actin Protein (*Capg)* in *App*$^{NL-G-F}$ animals (denoted *MSTRG.38755.8*; Fig. 2.8a and 2.8b) (Supplementary Table 2.8.). There was a decreased in the usage of the *Capg* isoform *ENSMUST00000071044.13,* and increase in the usage of *ENSMUST000000114071.8* and *MSTRG.38755.8* compared to controls. These three isoforms may exert different functional consequences, as *MSTRG.38755.8* is a coding isoform that lacks a protein domain between amino acid positions 31-107 (Fig 2.8c), which could influence the functional role of *Capg.*

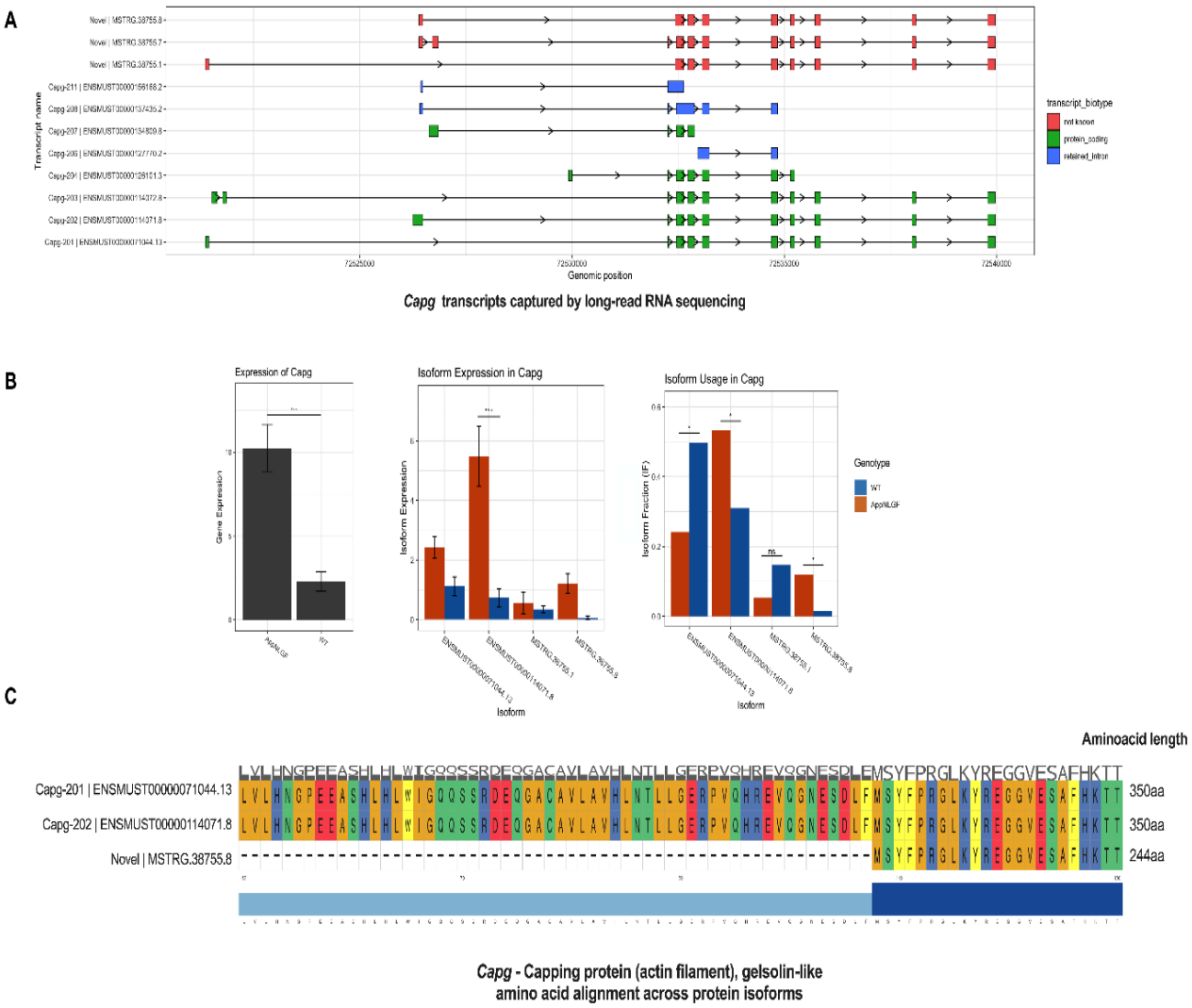

Figure 2-8 *Capg* showed isoform switching and amino acid changes in the predicted protein sequences in its novel transcript variant.

(A) The isoforms detected in the *Capg* gene. In addition to the canonical transcripts, three novel transcripts (MSTRG.38755.8, MSTRG.38755.7, MSTRG.38755.1) were found. (B) Gene-level expression of *Capg* is increased with amyloid pathology (left panel), while isoform expression indicates individual isoforms upregulated in response to amyloid pathology (middle panel). Isoform usage analysis with IsoformSwitchAnalyzer revealed an isoform-switch pattern in proportional transcript use between *ENSMUST00000071044.13, ENSMUST00000114071.8* and novel transcript variant *MSTRG.38755.8* (right panel) (n=6) (C) The novel coding transcript (*MSTRG.38755.8*) lacks a full domain between 31-107 amino acids of the canonical isoform *ENSMUST00000071044.13* (https://www.uniprot.org/uniprotkb/Q3TNN6).

Several other genes including amyloid-responsive microglial genes, *Gusb* (glucuronidase beta; Fig. 2.9a) and PTPRF interacting protein alpha 4 *(Ppfia4*; Fig. 2.9b) showed significant isoform switch events, including usage of the novel variant *MSTRG.2151.1* of *Ppfia4*. *Capg,* transmembrane protein with EGF-like and two follistatin-like domains 1 *(Tmeff1)* and occludin (*Ocln*) underwent switching of their polyadenylation sites in response to amyloid plaques (Fig. 2.9c). Biological annotation of genes showing isoform switching indicated enrichment of pathways involving development, cytoskeletal protein binding, and cell junction (Fig 2.10b).

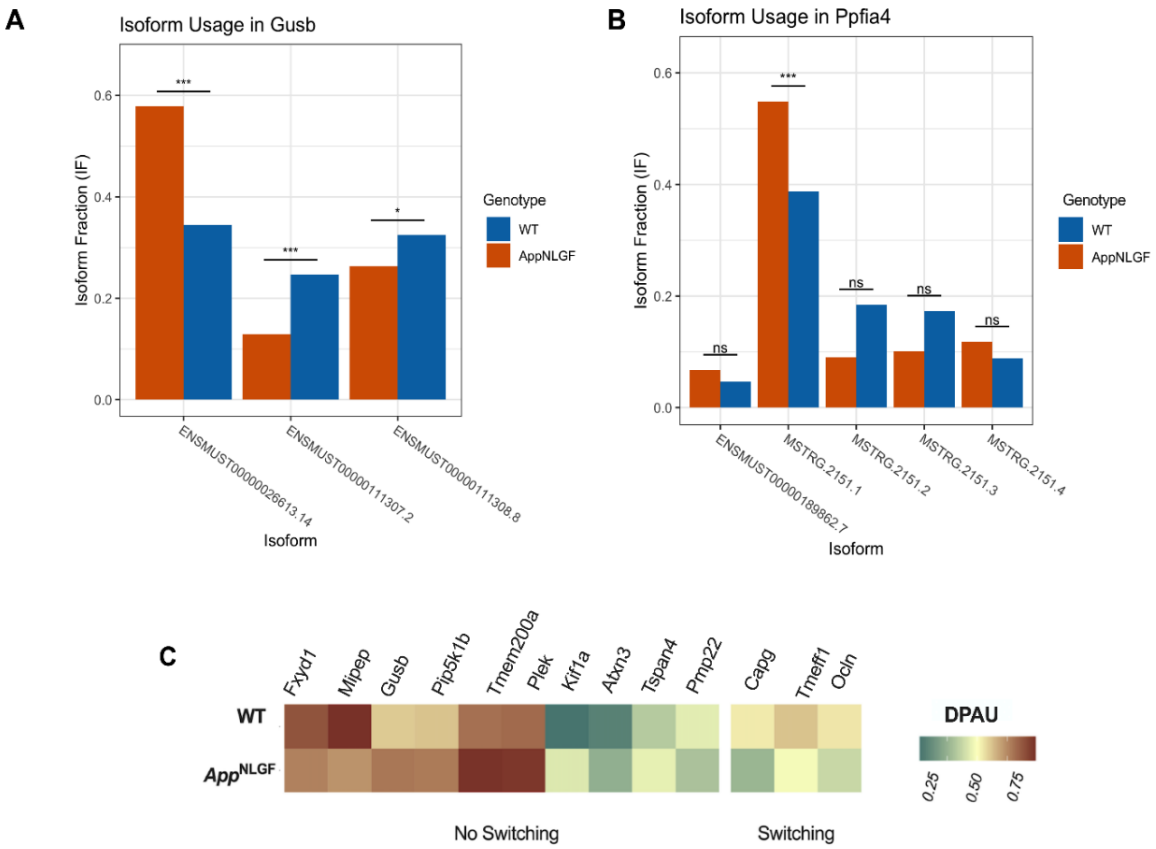

Figure 2-9 ARM/DAM gene *Gusb* and trans-synaptic signalling gene, *Ppfia4,* are amongst alternatively spliced genes showing transcript usage and isoform switching

(A) *Gusb* shows an isoform switch event in proportional transcript usage (FDR = 2.5e-06) and single exon skipping event on Exon 3 of ENSMUST00000026613.14 (FDR = 1.0e-03) and exon usage (FDR = 4.1e-02) in *App*$^{NL\text{-}G\text{-}F}$ mice. (B) *Ppfia4* shows an isoform switch event for MSTRG.2151.1 (FDR = 3.7e-07), and an exon skipping event on Exon 3 (FDR = 5.9e-04) which potentially increases overall isoform expression in response to the amyloid as found by IsoformSwitchAnalyzer. (C) tappAS analysis

shows altered polyadenylation sites within genes, including *Capg, Tmeff*, and *Ocln*, demonstrate a switch in polyadenylation patterns between $App^{NL\text{-}G\text{-}F}$ and control samples (n=6).

This observed change in differential alternative polyadenylation site usage is suggestive of alterations in mRNA stability, and RNA-protein binding indicating potential regulatory modifications at the post-transcriptional level (Edwalds-Gilbert et al., 1997; Kwan et al., 2008; Sheets et al., 1994).

Furthermore, our analysis discovered multiple instances of exon usage differences in key genes such as *Cathepsin D (Ctsd), Cathepsin B (Ctsb), Syngr1, BCL2 Related Protein A1 (Bcl2a1), App,* and *Clusterin (Clu)* in response to amyloid (Supplementary Table 2.9.). Particularly intriguing is the alternative splicing of *Clu*, which is known to modulate β-amyloid metabolism and/or deposition (Raj et al., 2018).

Changes in exon usage has the potential to lead to specific isoform preferences, as exemplified by the *Syngr1* gene. We observed an increased usage of Exon 7, Exon 9, and Exon 10 (Stage-wise testing, FDR < 0.05), concomitant with decreased usage of *Syngr1-201,* and increased usage of *Syngr1-202* isoforms in response to amyloid. This isoform switch (FDR = 1.21e-37) highlights the dynamic nature of alternative splicing and suggests a potential functional implication in the context of amyloid pathology.

CYFIP2 has shown decreased protein expression and is associated with memory loss in late-stage human AD (Tiwari et al., 2016), our data showed changes in this mouse model of early AD, where *Cyfip2-205* exhibited changes in usage (Stage-wise testing, FDR = 1.82e-05). This mechanistic insight might signify an early phase response to amyloid marked by preferential expression of specific mRNA isoforms, eventually culminating in the accumulation of tau tangles and transition to the later stages of the disease with neurodegeneration (Garcia-Esparcia et al., 2017; Ghosh et al., 2020; Ohno, 2018).

Marques-Coelho *et al.* (2021) identified differential isoform switch/usage in temporal and frontal lobes in healthy and AD adult individuals; mining this data we found several genes showing isoform usage changes in mice in response to amyloid also showed isoform usage changes in late-stage AD brain. This comparison of short-read and

long-read RNA-seq between human and mouse, and between late-stage human AD and early amyloid pathology may indicate conservation of response to AD-related pathology via mechanisms involving isoform switch and transcript usage in 42 common genes, including *Cathepsin D (Ctsd), Cathepsin A (Ctsa)* and clathrin light chain A (*Clta*) (Supplementary Table 2.10.)(Marques-Coelho et al., 2021).

2.3.6 Alternative splicing analysis in response to amyloid plaques serves to adapt cellular cytoskeleton and metabolism

To identify different alternative splicing events more directly, we used the PSI-Sigma software to classify splicing events into five standard types: single or multi-skipped exons (SES/MES, also known as cassette exons), alternative 5' and 3' splice sites (A5SS and A3SS), intron retention (IR), and mutually exclusive exons (MXE) (Lin & Krainer, 2019). We identified approximately one hundred genes (Supplementary Table 2.11) displaying one of these classes of alternative splicing in response to amyloid (Supplementary Table 14). Interestingly, we identified single/multi-exon-skipping events in several ARM/DAM genes, including *Ctsd, Ctsa,* glycosylated type I transmembrane glycoprotein *(Cd68),* glucuronidase beta *(Gusb),* tumor protein D52 (*Tpd52*), and cytokine receptor like factor 2 *(Crlf2).* Additionally, we found that *Csf2ra* and LYN proto-oncogene, Src family tyrosine kinase *(Lyn)* exhibited alternative 5' UTR splicing sites with amyloid pathology. Overall splicing increased with amyloid pathology, with a mean change in Percent Spliced-In (ΔPSI) of 1.76% for all significant splicing events ($p < 0.01$ and $|\Delta PSI| >= 5\%$) between $App^{NL\text{-}G\text{-}F}$ and WT samples. The genes exhibiting alternative splicing in response to amyloid were found to be enriched for functions primarily associated with endomembrane system organisation, actin filament binding, actin cytoskeleton reorganisations and kinase binding as shown in Fig. 2.10a.

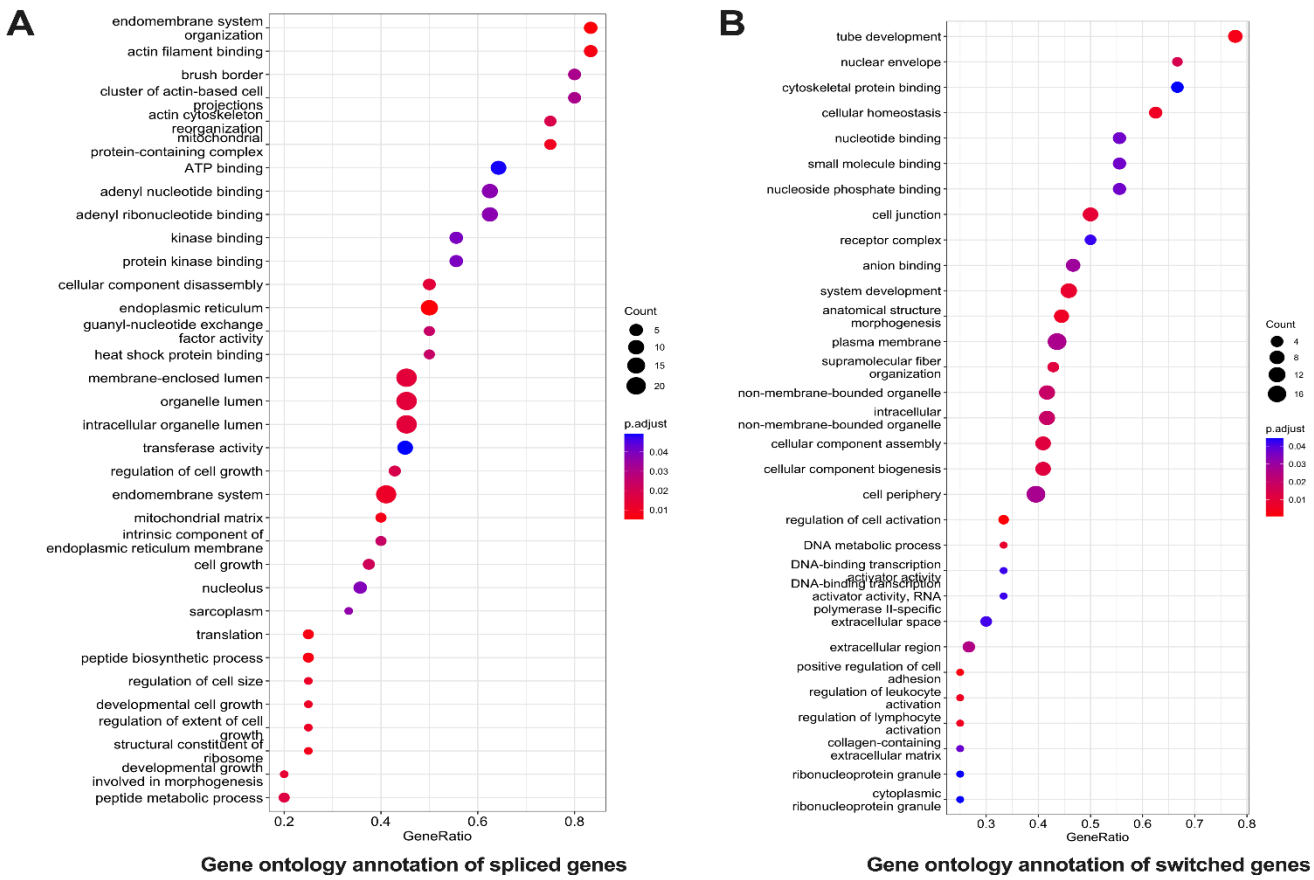

Figure 2-10 GO enrichment of genes undergoing splicing or isoform switching in response to amyloid pathology.

(A) GO annotations of spliced genes from PSI-Sigma analysis, highlighting pathways such as endomembrane system organization. (B) Isoform-switching genes found by IsoformSwitchAnalyzer enriched in functions like tube development and cellular homeostasis (n=6). Dot size represents gene count, and colour gradient indicates adjusted p-value (p.adjust), with darker colours showing higher significance. A threshold of 0.05 was applied for the adjusted p-value.

2.3.7 Transcript variants and splicing reveals changes to numerous cell-types including synaptic changes during early amyloid accumulation

Our detailed analyses on transcript usage, exon usage and alternative splicing offer novel insights into the functional implications of transcript isoform switching during microglial activation, synaptic adaptation and a series of other cell type changes in response to amyloid plaques in models of early AD. Importantly, these transcript isoform changes exhibit widespread enrichment across various cell types present in the brain as illustrated in Fig. 2.11, including T cells, border associated cells, astrocytes, oligodendrocytes, and both excitatory and inhibitory neurons (Botía et al., 2017; Cahoy et al., 2008; Miller et al., 2011; Zeisel et al., 2015). The enrichment analysis compares overlap directly between the gene inputs and marker sets using a simple probability model. Although, it is a limitation to represent isoforms via their gene names, a particularly intriguing aspect is the enrichment of genes with isoform-level changes associated with diverse synaptic pathways. This suggests concurrent alterations in synaptic physiology alongside the upregulation of AD risk genes occurring microglial proliferation and activation during the initial stages of disease progression, aligning with the earliest cognitive changes. These findings hold the potential to provide more precise isoform-specific indicators of disease stage for diagnostic tracking purposes, independent of the gene-level expression changes resulting from microglial cell proliferation/activation.

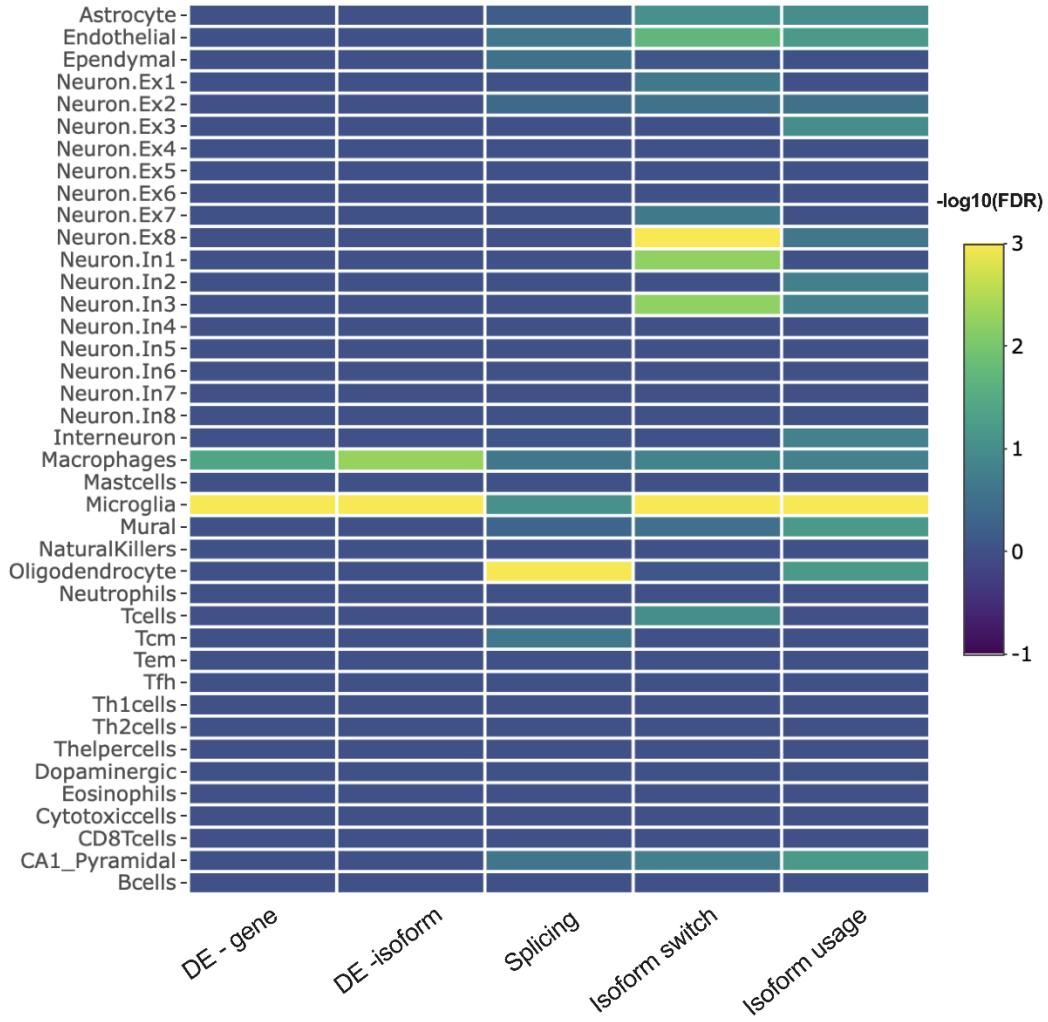

Figure 2-11 Heatmap of gene set overrepresentation analysis of gene and isoform level analyses.

Gene-level differential expression analysis reveals an enrichment of genes and isoforms within those expressed by microglia and macrophages, primarily attributed to the heightened proliferation and activation in response to amyloid accumulation. Genes showing splicing events were predominantly expressed by oligodendrocytes. Isoform switching mainly occurred in genes expressed by neuronal subtypes, endothelial cells, mural cells, and astrocytes. Isoform usage changes occurred in genes expressed by interneurons, mural cells, neuron subtypes, endothelial cells, and astrocytes (Cahoy et al., 2008; Zeisel et al., 2015).

We used EWCE to assess cell type-specific enrichment across our results, using the mouse reference dataset (Figure 2.12) (Zeisel et al., 2015) starting with 21,207 mouse genes, 16,482 remained after ortholog conversion (78%). EWCE accounts for expression bias (genes that are generally highly expressed across all cell types) by

using the bootstrap test. This provides a more accurate assessment of whether genes are specifically enriched in a particular cell type, not just highly expressed overall.

For differential gene expression, 140 hit genes were retained, and enrichment tests across 48 cell types at annotation level 2 identified cell types with significant enrichment ($q < 0.05$) (Fig 2.12.), suggesting these genes may influence cell type-specific expression patterns linked to disease. In the isoform expression analysis, 778 hit genes were identified, with microglial subtypes (Mg1 and Mg2), perivascular macrophages (Pvm1 and Pvm2), and vascular endothelial cells (Vend1 and Vend2) showing significant enrichment ($q < 0.05$), indicating that isoform expression may vary across specific cell types. For splicing, 362 hit genes were evaluated, and for isoform switch analysis, there were 115 hit genes, no significant enrichment was found, indicating a potential lack of cell specificity (Fig. 2.12). Similarly, in the isoform usage analysis, no significant enrichment was found for the 95 hit genes tested, implying that isoform usage may be broadly shared across cell types rather than restricted to specific populations.

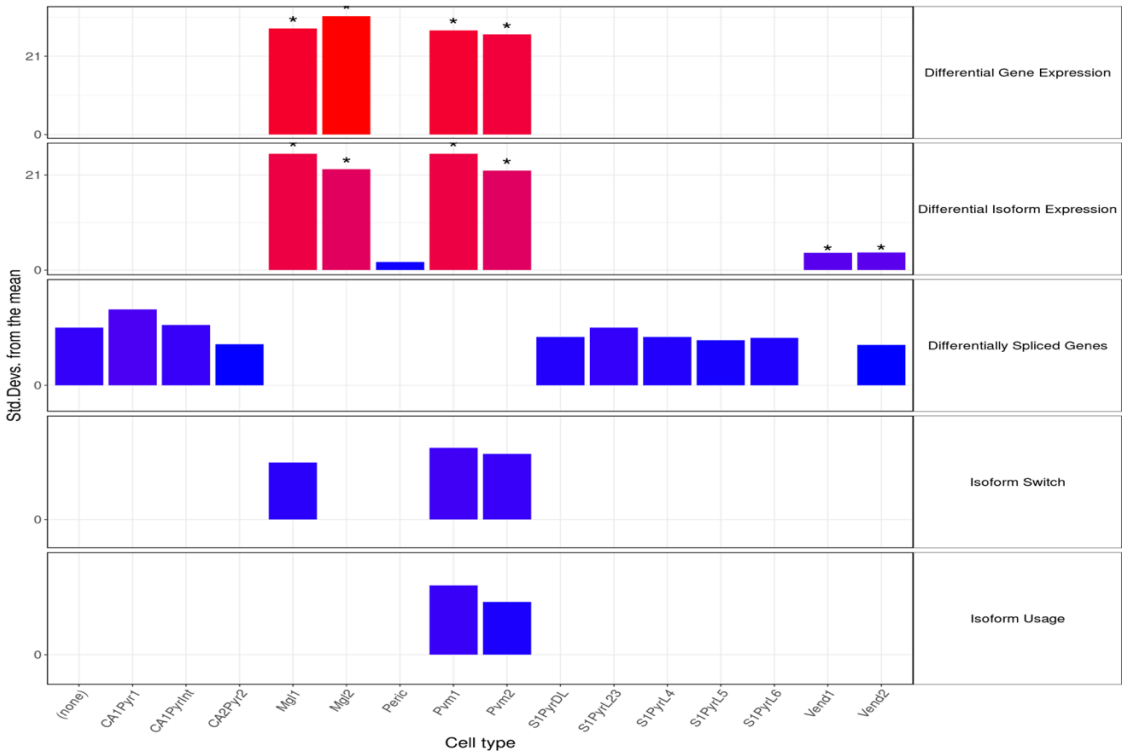

Figure 2-12 Enrichment analysis across gene and isoform level analysis.

Each panel row represents different types of analysis based on gene or isoform expression and splicing events. From top to bottom, the panels display differential

gene-level expression, where Microglia subtype-1(Mgl1) and Microglia subtype-2 (Mgl2) cell types showed significantly elevated expression changes, as indicated by red bars. Asterisks (*) mark significance ($q < 0.05$). In the differential isoform expression panel, similar patterns were seen for Mgl1 and Mgl2, with additional significance noted for the Vascular endothelial subtype-2 (Vend2) cell type. The differentially spliced genes panel showed a wider distribution of expression of these genes across several cell types, although not significant. The isoform switching genes revealed expression of these genes for a few cell types. The isoform usage panel showed expression of these genes by various cell types such as Vend2 and Perivascular microglia (Pvm), although no overall enrichment.

2.4 Discussion

In this study, we employed long-read RNA-seq in the *App*$^{NL\text{-}G\text{-}F}$ knock-in mouse models alongside WT controls to discern changes in transcript usage, splicing and novel transcript isoforms in response to amyloid. By adopting an isoform-centric approach, we provide valuable insights into the changes in alternative splicing events (Zhang et al., 2013). Previous gene-level expression analyses lack the sensitivity needed to detect possible changes at the transcript level, such as alterations in splicing. To overcome this limitation, we employed differential transcript usage analysis to identify additional shifts in gene expression in the *App*$^{NL\text{-}G\text{-}F}$ knock-in mouse model compared to controls. Our results reaffirm the significance and intricacy of known AD risk genes while also revealing novel splicing and potential regulatory patterns in these genes.

Moreover, our analysis uncovers previously unidentified splicing events and transcript isoforms that were not previously discernible through gene-level and isoform-level expression analysis, particularly of the genes integral to synaptic physiology. Our analysis identified differential isoform usage in genes linked to pre- and post-synaptic functions which might implicate an association with synaptic resilience in AD (King et al., 2023). We have comprehensively annotated the novel using external databases and prediction tools, identified gene and isoform level alterations contributing to a deeper understanding of protein diversity and specific patterns of isoform changes that contribute to the disease. Such insights may prove valuable in detecting the earliest stages of disease.

2.4.1 Differentially expressed genes and transcripts in long-read RNA-seq bulk experiments are mostly microglial

Our gene-level expression analysis, employing long-read RNA-sequencing, corroborated short-read data from other mouse models, highlighting induction of *Trem2, Tyrobp*, and other genes associated with the ARM/DAM microglia cell-state and AD risk. The utility of long-read RNA-seq in capturing gene-level expression changes underscored microglial proliferation and activation in this mouse model of amyloid accumulation. Importantly, the behavioral analysis of our collaborators demonstrated that these gene-level expression changes coincide with early cognitive alterations, suggesting their relevance in disease progression from early stages (Yaman et al., 2024). Notably, several ARM/DAM genes were found to be upregulated, further supporting the involvement of microglia in the response to amyloid pathology.

While gene-level changes provide insights into the amyloid-responsive genetic network expressed by microglia, our analysis reveals that specific microglial genes exhibit preferential selection of transcript isoforms in response to amyloid. Remarkably, these genes demonstrate notable disparities in fold-expression ranking and significance when compared at the gene-level versus the transcript level, as exemplified by genes such as *Igf1*. The differential usage of specific transcripts by these genes in the presence of amyloid plaques suggests their potential role in driving the cellular response to amyloid. It is worth noting that *IGF1*, one of the genes showing selective transcript usage, plays a crucial role as a mediator in the clearance and regulation of Aß in the brain, highlighting its relevance in the context of AD progression (Carro et al., 2002).

A further important example is the distinction between the mouse *Trem2* isoforms: we detect three isoforms of *Trem2*, the *ENSMUST00000024791.14* isoform, encoding for full-length receptor, and the *ENSMUST00000113237.3* isoform, lacking the transmembrane receptor, which are both induced by amyloid plaques in our analysis. In contrast, the *ENSMUST00000132340.2* isoform which represents an intron-retention event is not thought to produce protein and is not induced by amyloid. Focussing on specific isoforms is functionally important, the human *TREM2* isoform *ENST00000373113*, responsible for encoding the full-length transmembrane domain, along with the alternatively spliced isoforms *ENST00000373122* and

ENST00000338469, demonstrate a moderate increase in specific brain regions among individuals with AD (Moutinho et al., 2023). Alongside this, experimental findings from 7-month-old 5xFAD mouse model of Aß (Moutinho et al., 2023) expressing the human *TREM2* gene, compared to control mice (B6hT2 and 5xFADhT2), provide evidence that the alternatively spliced isoforms of *TREM2* undergo translation and secretion, leading to the formation of soluble TREM2 (sTREM2), which has been used as a biomarker for late pre-symptomatic and early symptomatic stages of AD (Suárez-Calvet et al., 2016) .

Beyond microglia, the most abundantly expressed *Gfap* isoform detected in our dataset, was also found to be major driver of the *Gfap* expression in a transgenic mouse model of tau pathology after long-read RNA-seq analysis (Leung et al., 2023), suggesting that the role of this isoform in the role of the astrocyte response in AD may warrant further investigation. Thus, identifying specific transcript isoforms provide new biological insights into mechanisms leading to AD.

2.4.2 Microglial activation, AD risk genes and alternative mechanisms linking to synapses

GWAS have transformed how we view AD, which led to a shift in focus towards the innate immune system. These studies have identified around 100 genes/loci variants associated with AD risk (Bellenguez et al., 2022; Kunkle et al., 2019; Lambert et al., 2013; Salih et al., 2019), many of which are expressed by microglia in the brain, implicating the innate immune system in AD pathogenesis (Lambert et al., 2013). The genetic network formed by these risk genes is specifically induced by amyloid pathology, and in large due to triggering of microglial proliferation and activation (Keren-Shaul et al., 2017). Intriguingly, this microglial risk network activation occurs early in disease progression, preceding neuronal death and cognitive decline (Hansen et al., 2018). Exploring how these risk genes/loci influence AD susceptibility encompasses exploring alternative mechanisms when variants do not alter the protein coding sequence of genes. Non-coding variants can modulate gene expression and splicing patterns (expression quantitative loci, eQTLs, and splicing quantitative loci, sQTLs), which may influence transcript abundance, specific transcript isoforms, subcellular localisation of RNA, and changes in the proteome to compound AD risk (T. Raj et al., 2018a; Simone et al., 2021; Westra et al., 2013). While short-read RNA-seq

is valuable for quantifying gene-level expression changes, it has limitations in capturing differential transcript usage, exon usage and alternative splicing. To address such limitations, we performed long-read RNA-seq of cortical RNA in the *App*$^{NL-G-F}$ mouse model of amyloid plaque accumulation compared with WT control samples. Analysis of gene expression was undertaken at 9 months of age in mice that had already developed an Aß associated spatial short-term memory deficit (Yaman et al., 2024). We identified novel transcripts from causal and risk genes associated with AD such as *Apoe*, *Ctsb*, *Grn*. We observed induction of microglial-expressed AD risk genes at the whole gene-level, but also discovered preferential transcript usage in microglial genes in response to amyloid plaques. Moreover, a significant number of the genes which exhibited differential exon usage and alternative splicing in response to amyloid, were enriched in functional pathways coordinating synaptic physiology and the interaction with diverse cell types in the brain encompassing T cells, astrocytes and oligodendrocytes. Understanding these novel changes at the transcript-level will provide new insights into microglial activation, synaptic adaptation and interaction between other cell types at early disease stages, potentially informing disease-stage and isoform-specific diagnostic and therapeutic strategies.

2.4.3 Transcript-level alterations reveal insights into the microglial response to amyloid

Accompanying the well-characterised induction of microglial-expressed AD risk genes at the gene-level in response to amyloid, we also revealed preferentially used novel isoforms, transcript usage, switching and splicing events in a series of microglial genes, including many ARM/DAM cell-state genes associated with amyloid pathology, independent from changes in cell number due to proliferating microglia with amyloid. Transcript isoform changes were seen in genes such as *Capg*, *Trem2*, *Ocln, Ctsd, Ctsb, Ctsa, Cd68, Gusb* and *Csf2ra*. Recent work has shown that changes in *CTSB* activity may contribute to AD (Wu et al., 2023). Thus, intricate control of transcript isoforms of these microglial genes within AD risk genetic pathways/networks is likely to tailor the functional microglial response to amyloid.

2.4.4 Conservation of isoform switches in amyloid pathology between AD mouse models and late-stage human AD brain

We identified transcript isoform switches and usage profiles, comparing these gene sets with isoform changes detected in human AD brains from the MAYO, MSBB, and ROSMAP cohorts (Marques-Coelho et al., 2021). Notably, 42 genes showing isoform usage or switches in response to amyloid pathology in the mouse model were also altered in human late-AD brains. These genes are involved in crucial pathways, including endocytic and vesicle trafficking (e.g. elastin microfibril interfacer 1 (*EMILIN1*), *CLTA*, adaptor-related protein complex 2 subunit alpha 1 (*AP2A1*), and DENN domain containing 2A (*DENND2A*)); neuroinflammation and microglial activation (e.g. apoptosis-inducing factor mitochondria-associated 2 (*AIFM2*), regulator of G-protein signaling 10 (*RGS10*), fractalkine receptor (*CX3CR1*), and interferon regulatory factor 9 (*IRF9*)); lysosomal function and protein degradation (e.g. *CTSD*, *CTSA*, and ATPase H+ transporting V1 subunit E1 (*ATP6V1E1*)); and cytoskeletal dynamics (e.g. leucine-rich repeats and immunoglobulin-like domains 1 (*LRIG1*), kinesin family member 1A (*KIF1A*), abl interactor 1 (*ABI1*), tocopherol transport protein (*TPGS1*), and rhotekin 2 (*RTKN2*)).

EMILIN1, an extracellular matrix glycoprotein, plays a critical role in forming anchoring filaments essential for proper lymphatic drainage. *EMILIN1*'s interaction with integrins involved in regulating lymphatic valve formation and maintenance, potentially influencing immune cell trafficking and inflammation in the brain. Given the established role of neuroinflammation in AD, changes in *EMILIN1* expression could exacerbate the inflammatory response associated with amyloid plaque deposition (Pivetta et al., 2016, 2022).

LRIG1, a membrane glycoprotein, negatively regulates receptor tyrosine kinases (RTKs), including the epidermal growth factor receptor (EGFR). *LRIG1* has been identified as potential AD risk gene in GWAS from *APOE*-stratified East Asian population ((S. Kang et al., 2021)). Overexpression of *LRIG1* has been shown to block neurotrophin signaling (Trinchero et al., 2017), and reduce dendritic growth in developing hippocampal neurons (Alsina et al., 2016).

CTSD, a lysosomal protease, plays a pivotal role in AD pathogenesis. It is involved in lysosomal function, protein degradation, and *APP* processing. *CTSD* is responsible for

degrading a significant portion of Aβ, particularly in lysosomes, which are key to maintaining brain Aβ proteostasis. Differential degradation rates of Aβ42 and Aβ40 suggest CTSD's involvement in regulating the Aβ42/40 ratio. Dysfunction in *CTSD* could thus contribute to Aβ accumulation, exacerbating AD progression (Grubman et al., 2019; Ouyang et al., 2023; Suire et al., 2020).

AIFM2, which encoding *ferroptosis suppressor protein 1 (FSP1),* plays a crucial role in protecting cells from ferroptosis by regeneratingmembrane-embedded antioxidants in healthy conditions (Xavier da Silva et al., 2023). It has a dual role; it can also induce caspase-activated cell death. Increased isoform-usage of *AIFM2* in both human AD and an AD mouse model may suggest the presence of an amyloid-specific specific isoform that contributes to the iron-dependent cell death or excessive apoptosis contributing to the disease. This highlights a potential novel mechanism in dysregulated apoptosis in earlier stage of the AD.

Key limitations exist in the comparison of isoform usage and switching events across species and platforms. Our study revealed isoform changes in mice using long-read sequencing from bulk whole cortex samples, while human data from the Mayo Clinic, MSBB, and ROSMAP cohorts relied on short-read RNA-seq from dorsolateral prefrontal and entorhinal cortices (Marques-Coelho et al., 2021). Our study was based on whole cortex in mouse model with amyloid pathology only, whereas human datasets analyzed focused on specific brain regions. These differences in brain regions and the stage of the disease pathology could also influence the results, in addition to species-specific differences. In addition, short-read sequencing offers limited resolution at the exon level, complicating precise mapping of exon usage and splicing events. Although we identified shared isoform-level changes in response to amyloid pathology between human and mouse models, the exact splicing events, such as which exons are differentially spliced, remain uncertain. Also, from bulk tissue it is difficult to attribute transcript splicing and usage changes to specific cell types, and it may be that certain transcript forms are expressed or induced by very specific cell types normally not expressing canonical versions of the gene in healthy, non-disease circumstances.

Furthermore, the timing of AD pathology differs between models. Human tissue from the aforementioned cohorts represents late-stage AD, including tau pathology and

underrepresentation of neurons lost, and the influence of ageing, whereas the mouse models we used capture earlier amyloid pathology stages (without tau pathology and neuronal death). Therefore, some isoform changes in the mouse model may not fully correspond to early-stage human AD pathology, complicating direct comparisons across disease stages.

The statistical approaches also introduce variability: isoform switch analysis in short-read human data differs from that of long-read sequencing in mice. These methodological differences may impact the detection of isoforms and their annotations (novel/canonical), and the transcript models also complicates cross-species comparisons. Future studies employing long-read sequencing in human AD tissues will be essential for a more accurate and comprehensive understanding of isoform-level events across species and disease stages.

3 Genome-wide 5mC differential methylation analysis via long-read DNA sequencing in the *App*$^{NL-G-F}$ Alzheimer's Disease mouse model

3.1 Introduction

Alzheimer's Disease (AD) is characterized by complex interactions among genetic, epigenetic, and environmental factors that influence disease progression. While genome-wide association studies (GWAS) have identified risk variants associated with AD, the functional mechanisms through which these variants contribute to the disease remain unclear (Kunkle et al., 2019; Lambert et al., 2013). One key avenue of exploration is epigenetic regulation, where DNA methylation and histone acetylation significantly influence gene expression, splicing, or other regulatory processes critical to neurodegeneration in AD (De Jager et al., 2014; Lardenoije et al., 2019; Marzi et al., 2018).

DNA methylation involves the formation of 5-methylcytosine (5mC) through a covalent bond with the cytosine 5′ carbon site of the CpG dinucleotides, mediated by DNA methyltransferases (DNMTs). 5mC may disrupt the transcription factor binding at promoter recognition sites or recruit methyl-CpG-binding proteins that suppress transcription by altering chromatin structure (Du et al., 2015; Jones, 2012).

Short-read sequencing, particularly bisulfite sequencing remains the gold standard for 5mC detection due to its high throughput and cost-effectiveness. BSseq distinguishes methylated from unmethylated cytosines by converting unmethylated cytosines to uracils, allowing differentiation upon sequencing (Ni et al., 2023). However, this method has notable limitations, including the potential for biases introduced during the bisulfite conversion process, which can lead to inaccurate methylation estimates (Tran et al., 2016).

Unlike traditional short-read sequencing, long-read sequencing directly detects DNA methylation at single-nucleotide resolution without bisulfite conversion, preserving the DNA in its native state and bypassing PCR, which preserves methylation and other base modifications (Logsdon et al., 2020; Wang et al., 2021). Long-read sequencing

also achieves comprehensive coverage of complex genomic regions, such as repetitive sequences and structural variants, which are often implicated in AD (Ebbert et al., 2019). Generating longer reads that span entire genes or regulatory regions enables the comprehensive identification of methylation patterns across extensive genomic areas, offering a more precise and detailed understanding of specific methylation types, such as 5mC. This capability is crucial for elucidating how these methylation modifications impact gene function, including potential changes in gene expression and splicing profiles (Liu et al., 2021)

In this chapter, we examine 5mC genome-wide epigenetic modifications through long-read DNA sequencing, focusing on methylation patterns in an AD mouse model versus controls to profile the responses to amyloid, simulating early AD stages. This study aims to identify differentially methylated regions (DMRs) that contribute to disease progression and investigate pathways potentially affected by methylation patterns, using both mouse and human AD brain RNA-seq datasets.

3.1.1 Aims and objectives

Hypothesis: Genome-wide DNA methylation patterns in the *App*$^{NL-G-F}$ AD mouse model contribute to altered gene regulation and are associated with specific cell types and pathways relevant to AD pathology. Identifying these methylation patterns will provide insights into cellular responses to the amyloid pathology, reflecting the early-stage AD pathology. This chapter focuses on the identifying DMRs in response to amyloid accumulation and annotating these regions to understand the broader impact of genome-wide methylation changes.

To achieve these objectives, the following analyses were performed:

1. Identification and characterisation of DMRs: To investigate genome-wide methylation differences between the *App*$^{NL-G-F}$ AD mouse model and WT controls, long-read DNA sequencing was performed using the same samples and experimental design as in Chapter 2.
2. Functional annotation and pathway analysis of DMRs: To assess the biological significance of the identified DMRs, GO and pathway enrichment analyses were performed on promoter regions and across all DMRs marking the coding regions of genes, enabling a focused investigation of regulatory regions (promoters) and

capturing broader biological processes influenced by methylation changes in response to amyloid pathology.

3. Cell-type enrichment analysis of DMR genes: To identify cell types most affected by differential methylation, cell-type enrichment analysis was conducted to explore cell-type-specific methylation patterns that might contribute to AD pathology.

3.2 Methods

3.2.1 Sample preparation

DNA was extracted from the same homogenized mouse cortex samples used for RNA extraction, specifically from *App*$^{NL\text{-}G\text{-}F}$ (n=6*)* and WT (n=6) mice. Genomic DNA (gDNA) was purified using the NEB Monarch gDNA purification kit (T3010). Sequencing libraries were prepared with the Ligation Sequencing gDNA Kit and native barcoding (SQK-LSK109 with EXP-NBD104 and EXP-NBD114). For each sample, 2 µg of DNA was sheared to approximately 20 kb fragments using a Megaruptor and then pooled for sequencing. Six pools, each containing a mix of one *App*$^{NL\text{-}G\text{-}F}$ sample and one WT sample, were loaded onto PromethION flow cells using R9.4.1 chemistry for sequencing.

3.2.2 Base calling and methylation detection

Base calling was performed using Guppy version 6.5.7 (high-accuracy model), generating 450 bp reads with integrated 5mC base modification detection. Methylation calls (modBAM) were generated during base calling using Guppy, a tool from Oxford Nanopore Technologies for base calling, also detecting base modifications such as 5mC.

3.2.3 Mapping

The modBAM files, containing 5mC base-modification calls, were aligned to the reference genome to map methylation sites. Base called reads were aligned with Minimap2 using the mouse reference genome (mm10) due to available tools for downstream analysis, and BAM files were sorted using samtools. The aligned BAM files were processed using the *modkit* tool to convert them into bedMethyl format. These bedMethyl files were then incorporated into the analysis pipeline for downstream genome-wide differential methylation analysis.

3.2.4 Data Import and Initial Preprocessing

The first step in the data analysis involved reading methylation data files in tabular format from individual pools, stratified by sample type. bedMethyl files were converted into TSV files to be used in the Dispersion Shrinkage for Sequencing (DSS) workflow (Feng et al., 2014; Park & Wu, 2016; Wu et al., 2013, 2015). The raw methylation data consisted of ~25M sites across the mouse genome.

3.2.5 Filtering and Quality Control

To include high-quality CpG sites across the methylation regions, filtering was applied to select those with sufficient coverage and variability. Specifically, the filtering criteria was the threshold a minimum of 10 reads per site in all samples (covMin) and a methylation standard deviation (SD) of at least 5% (methSD). These criteria were applied using the *filterRegions()* function from the *comethyl* package in R (Mordaunt et al., 2022). The thresholds for covMin and methSD were determined to capture regions with adequate variability and coverage were retained (Sigurpalsdottir et al., 2024).

The *filterRegions()* function was applied after optimizing these cutoffs using the *getRegionTotals()* and *plotRegionTotals()* functions, which summarise the read count and methylation distribution for each region. After filtering, the regions were re-examined using *plotRegionStats()* to visually compare the methylation profiles of the *App*$^{NL\text{-}G\text{-}F}$ and WT samples (n=6).

3.2.6 Differential methylation analysis

3.2.6.1 Model Design and Fitting

To perform differential methylation analysis, we used the DSS package, to robustly detect the differentially methylated loci and regions by count-based methylation data. The analysis was conducted on a filtered methylation object that included CpG sites meeting abovementioned pre-defined thresholds for coverage and methylation variability.

Model fitting was performed using the *DMLfit.multiFactor()* function in the *DSS* package, which supports the incorporation of multiple experimental factors into the model. The design matrix for the model included key

covariates: Genotype, Sex, Batch (samples pooled toegether for sequencing), and the interaction between Genotype and Sex.

The following model was applied to correct for batch and sex effects, while retaining the genotype effect:

$$Methylation \sim Genotype + Sex + Batch + Genotype{:}Sex\ Interaction$$

This function also performs Wald tests at each CpG site to assess whether the effect of the Genotype covariate significantly alters the methylation levels. Specifically, we tested whether methylation levels were significantly different between the *App*$^{NL\text{-}G\text{-}F}$ genotype and the reference (WT), focusing on the Genotype term in the model.

This analysis results in differentially methylated locis (DMLs); while informative, may not always reflect the broader biological implications of methylation changes. The methodologies for identifying DMRs from whole genome bisulfite sequencing, indicates that the aggregation of DMLs into DMRs can provide a more comprehensive understanding of methylation patterns across the genome (Peters et al., 2021a). This suggests that while DMLs are useful for pinpointing specific methylation changes, they may lack the contextual significance that DMRs provide when considering gene regulation on a larger scale.

3.2.6.2 DMR Identification

callDMR() function from the *DSS* package was used to identify DMRs from the CpG-level results obtained from the strain-specific differential methylation test. The function consolidates individual CpG sites with significant methylation changes into larger regions, helping to identify regional methylation patterns rather than relying on individual CpG changes, which can be more susceptible to noise.

The *areaStat* statistic was calculated for each DMR, which represents the summed test statistics (e.g. Wald statistics) for all CpG sites within the region. A higher *areaStat* indicates a stronger overall signal of differential methylation across the DMR, reflecting more substantial and widespread methylation changes in that genomic region.

3.2.7 Annotation of DMRs

The identified DMRs were converted into a *GRanges* object, which is a genomic ranges format that allows efficient representation and manipulation of genomic intervals. We loaded the *TxDb.Mmusculus.UCSC.mm10.knownGene* database, which contains known gene annotations for the mouse genome (mm10). This database is used in the annotation process to associate DMRs with specific genomic features such as promoters, exons, introns, and transcription start sites (TSS).

Using the *annotatePeak()* function from the *ChIPseeker* package (Wang et al., 2022; Yu et al., 2015), DMRs were annotated with genomic features to link them to nearby genes. In *annotatePeak()* function, the *tssRegion* argument was set to include regions 2 kb upstream and 500 bp downstream of the TSS, defining typical promoter regions. *annotatePeak()* then compares the coordinates of each DMR with these predefined genomic features. The function checks if each DMR overlaps with, or is within a set distance from features like promoters, exons, or intergenic regions. If a DMR is within the specified distance from a TSS, it will be annotated as related to that gene's promoter region. For DMRs that do not fall directly within promoter regions, *annotatePeak()* assigns the closest gene based on the distance to the nearest TSS. This association shows if gene might be regulated by methylation changes in that DMR, even if the DMR is located in an intron, exon, or intergenic region.

The areaStat parameter, derived from the DMR analysis, is a measure that quantifies the magnitude and direction of methylation changes across a region. It is used to assess whether the region is hypermethylated (more methylated in the experimental condition) or hypomethylated (less methylated in the experimental condition) relative to the control condition. DMRs were divided into these two groups, with hypermethylated and hypomethylated regions based on the areaStat parameter (areaStat>0 hypermethylated, and area<0 hypomethylated regions) of the DMR table and analyzed separately to explore their functional roles.

The primary focus was on promoter regions, which were used to subset for subsequent GO annotations, as these regions are key regulatory elements influencing gene expression. The *rGREAT* package was used to annotate genomic regions

overlapping DMRs (Gu & Hübschmann, 2023a). For the GO analysis, the simplifyEnrichment R package (Gu & Hübschmann, 2023b) was used to cluster enriched GO terms. Clustering was performed using the 'binary_cut' method, which simplifies and organizes GO terms into distinct clusters. Keyword enrichment analysis was then conducted for each GO cluster, and gene associations were retrieved for the top clusters, displaying annotations that included the distances of DMRs to the nearest TSS.

3.2.8 Cell-type enrichment

To determine the enrichment of specific cell types in the gene sets containing DMRs, we conducted a cell-type enrichment analysis using the Expression Weighted Cell-type Enrichment (EWCE) method (Skene & Grant, 2016). EWCE evaluates the over-representation of genes associated with specific cell types by comparing input genes to background gene expression data from reference cell types. The Mouse reference dataset was used for the cell-type enrichment analysis (Zeisel et al., 2015).

3.2.9 Methylation site visualisation

To visualize DMRs identified in the methylation dataset, the *DMR.plot* function from the *DMRcate* package was used, and a customised version, *DMRPlot2* function (Peters et al., 2015, 2021b, 2024). Additionally, the *bsseq* package (Hansen et al., 2012) was utilized to visualize DNA methylation levels at both single and multiple loci. The *plotRegion* function from *bsseq* was used to generate plots of methylation levels for specific genomic coordinates, incorporating smoothed estimates of methylation across the region. This function provided a granular view of methylation changes, to compare methylation levels across individual loci. For a broader view, *plotManyRegions* from the *bsseq* package was used to visualize methylation levels across multiple genomic regions simultaneously.

3.2.10 Data Availability

Raw data files will be made available online on GEO database.

The information about supplementary tables and their URLs are available in Appendix A.

3.2.11 Software and Algorithms

A snakemake pipeline created to pre-process modBAM files and generate the methylation tables (bedmethyl files) can be accessed through the GitHub repository: https://github.com/umranyaman/ONT_DNA_pipeline.

3.3 Results

3.3.1 Alignment statistics

The alignment statistics for the *App*$^{NL\text{-}G\text{-}F}$ and WT datasets reveal that there is no significant variation between the two groups, indicating downstream analyses were not affected by technical variability between mice and samples (Fig. 3.1). The comparison of alignment statistics between the *App*$^{NL\text{-}G\text{-}F}$ and WT datasets shows no statistically significant differences across summary key metrics, as determined by two-sample t-tests. The total read count for *App*$^{NL\text{-}G\text{-}F}$ (2.83M) and WT (3.29M) yielded a p-value of 0.273, indicating no significant difference. N50 values, which represents the length at which 50% of the total sequences that are contained in contigs or scaffolds equal to or longer than this length, are nearly identical between the two groups (*App*$^{NL\text{-}G\text{-}F}$: 13,994, WT: 14,035), with a p-value of 0.978. The mean coverage for *App*$^{NL\text{-}G\text{-}F}$ (10.82) is slightly lower than that of WT (12.02), but this difference is not statistically significant (p-value: 0.354). Similarly, the median read length for *App*$^{NL\text{-}G\text{-}F}$ (10,244.5) compared to WT (9,326.5) produced a p-value of 0.286, indicating no significant difference. Although WT has a higher total yield (33.07 Gb, versus 29.75 Gb for *App*$^{NL\text{-}G\text{-}F}$), the p-value of 0.354 confirms that this difference is not statistically significant.

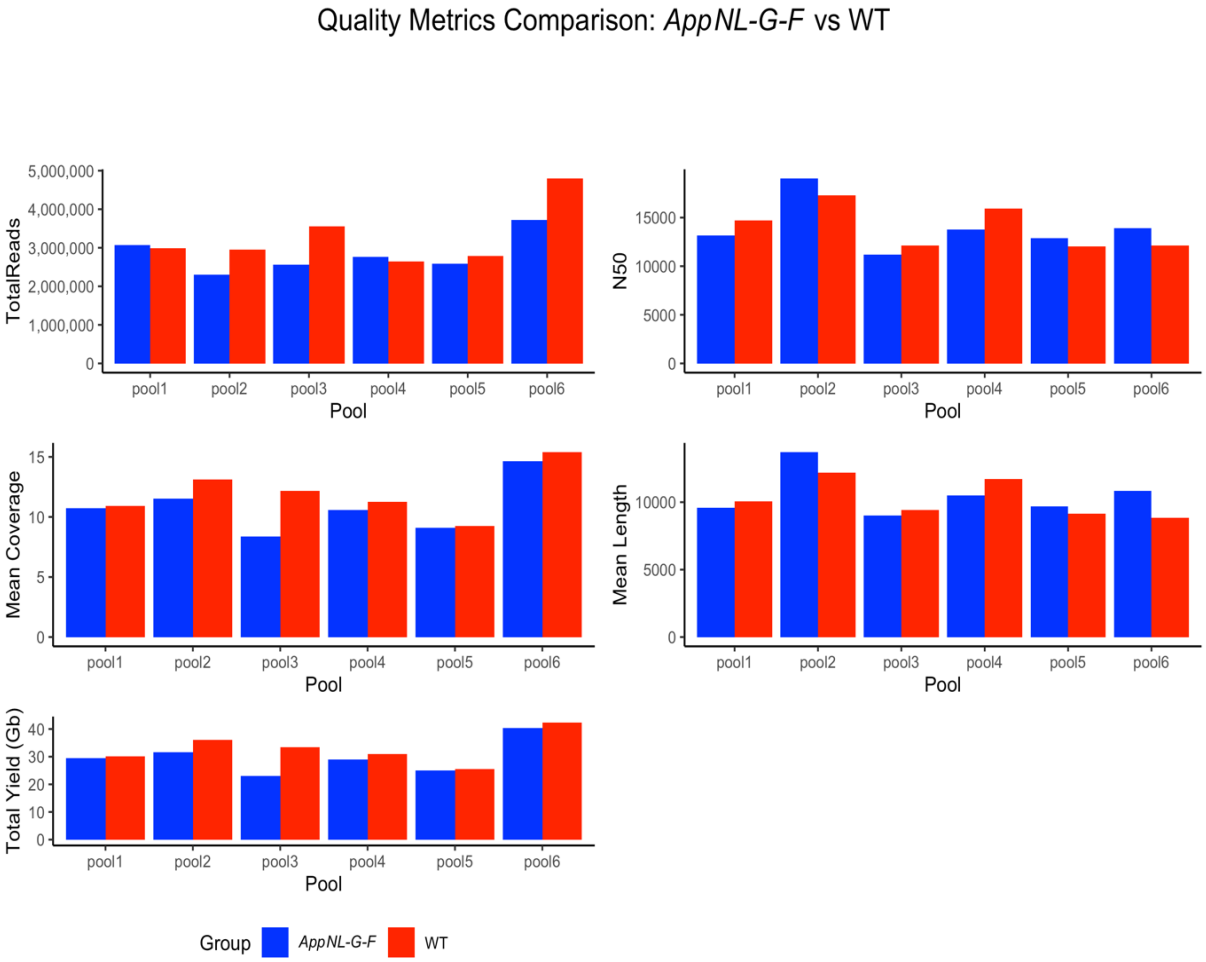

Figure 3-1 Comparison of five key sequencing metrics: Total reads, N50, mean coverage, mean length, and total yield (Gb) between the two groups: *App*$^{NL-G-F}$ (red bars) and WT (blue bars).

The samples were distributed across six pools (pool1 to pool6, where each pool contained one *App*$^{N-G-F}$ and one WT mouse sample). Total reads represent the total number of reads generated, N50 indicates the read length such that 50% of total bases are in reads of this length or longer, mean coverage shows the average depth of coverage across the genome, mean length represents the average read length, and total yield (Gb) represents the overall amount of data generated. Bars for *App*$^{NL-G-F}$ and WT are shown side by side for direct comparison in each pool (n=6). A t-test shows no significant differences between genotypes for any of the metrics.

3.3.2 Overall methylation metrics for *App*$^{NL-G-F}$ and WT mouse replicates

When comparing the 5mC modification between *App*$^{NL-G-F}$ and WT mouse samples, *App*$^{NL-G-F}$ showed a slightly higher pass fraction (fraction of counts passing QC) of

70.05% at a threshold of 0.7552, compared to 69.85% in WT samples at a threshold of 0.7526 (Table 3.1). In contrast, for the canonical (unmodified) cytosine (C), WT samples had a pass fraction of 30.15%, while *App*$^{NL-G-F}$ samples had a comparable pass fraction of 29.95%. This slight difference suggests a marginally higher prevalence of the 5mC modification in *App*$^{NL-G-F}$ samples. However, t-test for both 5mC and C modifications indicate no significant differences between the genotypes. For 5mC, the p-values associated with the number of counts passing QC per mouse (0.938), the fraction of counts passing QC (0.640), total counts per mouse (0.938), and the fraction of total counts passing QC (0.634) are all above the standard significance threshold of 0.05, suggesting no statistical difference. Similarly, for the C modification, p-values for these metrics (0.987, 0.640, 0.982, and 0.634, respectively) indicate no significant differences between *App*$^{NL-G-F}$ and WT samples.

Table 3-1 Summary of base modification and threshold analysis in *App*$^{NL-G-F}$ and WT samples.

group	*base*	*mod*	*threshold*	*pass_count*	*pass_frac*	*all_count*	*all_frac*
WT	C	5mC	0.752	10160054	0.698	11024943	0.682
WT	C	C	0.752	4374657	0.301	5110780	0.317
App^{NL-G-F}	C	5mC	0.755	10234370	0.700	11105926	0.684
App^{NL-G-F}	C	C	0.755	4381324	0.299	5121588	0.315

group: Indicates the sample genotype group (*App*$^{NL-G-F}$ or WT mice).

base: Specifies the type of base analyzed, either cytosine (C) or 5-methylcytosine (5mC).

mod: Lists the modification type, either 5mC or unmethylated C.

threshold: Provides the threshold value applied to determine base modification presence.

pass_count: Shows the number of bases that passed the threshold.

pass_frac: Presents the fraction of the total bases passing the threshold.

all_count: Represents the total count of all measured bases.

all_frac: Shows the fraction of all measured bases.

3.3.3 Differential methylated region analysis

The DMR analysis was performed using CpG sites within filtered regions, identified using *comethyl* package (Mordaunt et al., 2022)(Fig. 3.2), yielding 8,208,368 CpG sites in each sample from regions meeting the covMin and methSD criteria (minimum of 10 reads in all samples and a methylation standard deviation of at least 5%) as detailed in Section 3.2.5.

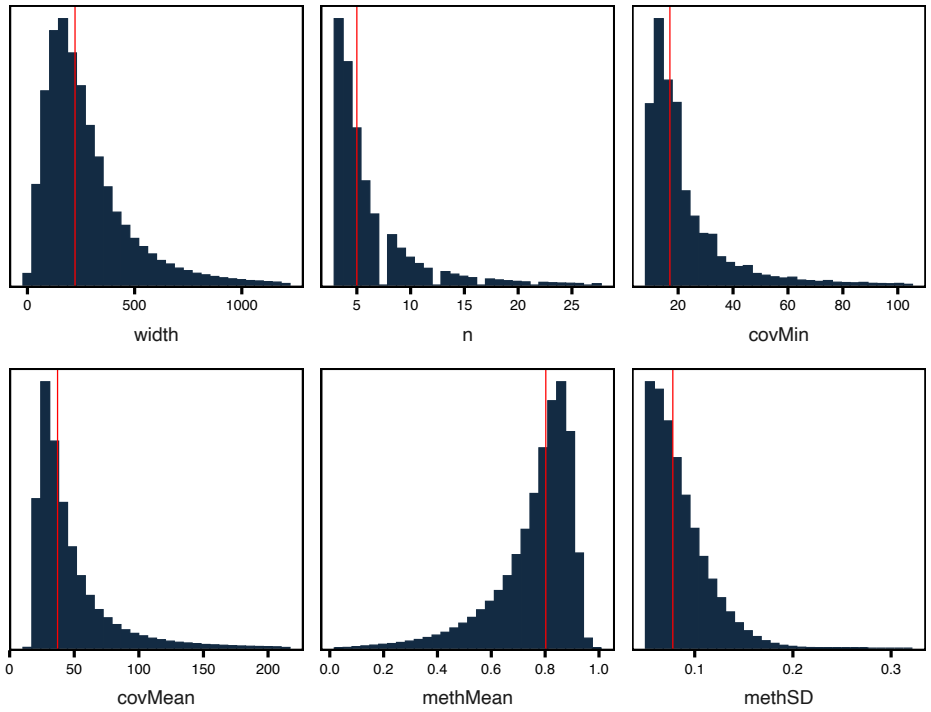

Figure 3-2 Histograms of region-level statistics before filtering for methylated regions

The distribution of several region-level methylation metrics for all samples (n=12). The width panel shows the distribution of region sizes in base pairs, with a red vertical line indicating the median width. The n panel represents the number of CpG sites within each region. The covMin panel illustrates the minimum coverage (number of reads) across samples for each region. The covMean panel reflects the average coverage across all CpGs in each region. The methMean panel depicts the average methylation level across CpGs within each region, with values clustering around 0.8, suggesting high methylation in most regions. The methSD panel indicates the standard deviation of methylation across CpGs within each region. Red vertical lines in each panel represent the median value for each metric.

From this dataset, 1,174 hypermethylated sites and 974 hypomethylated sites were detected with differential methylation analysis (Table 3.2, full table in Supplementary Table 3.1), spanning various genomic features, including promoters, exons, and intergenic regions. These DMRs are ranked based on their area statistics, which indicate the magnitude of methylation change between the *App*$^{NL\text{-}G\text{-}F}$ and WT groups (Fig. 3.3 and Fig. 3.4).

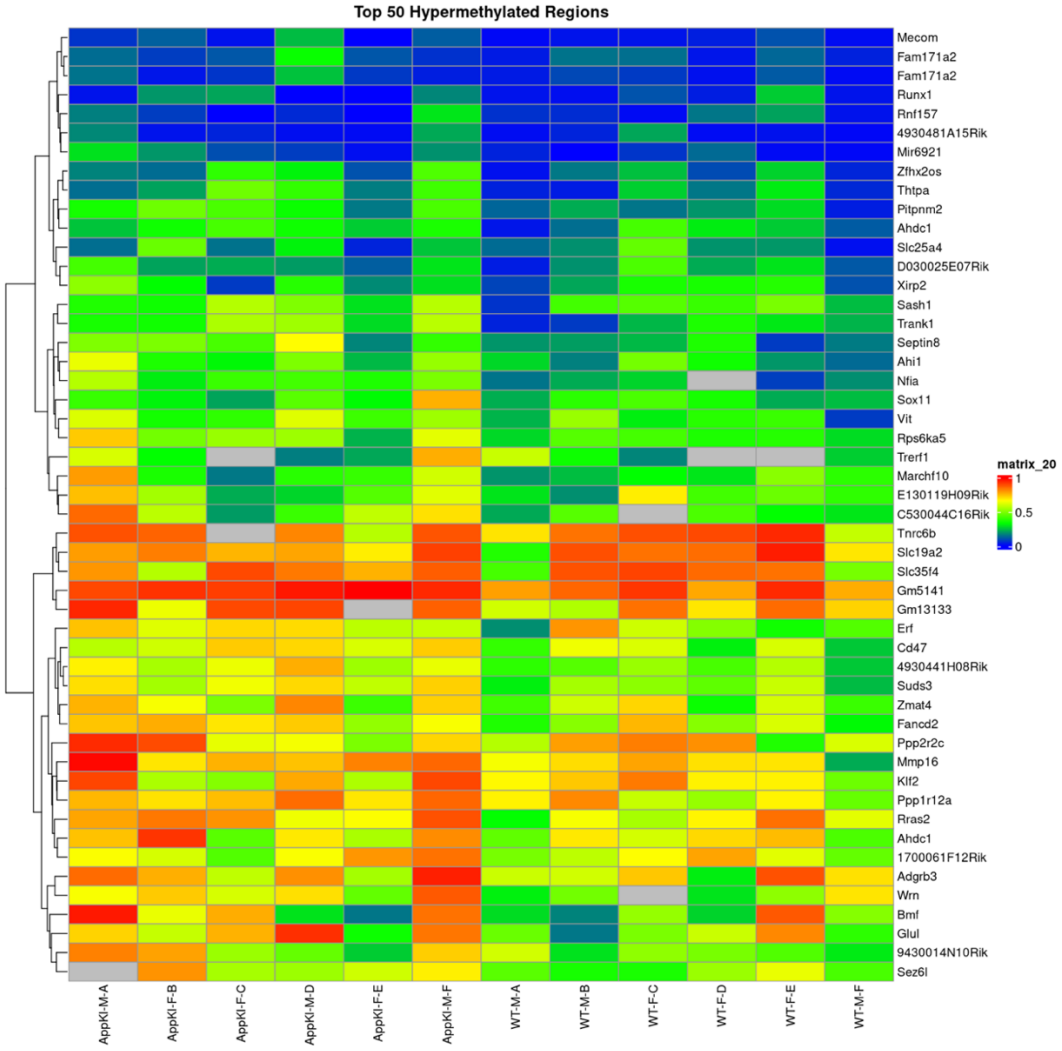

Figure 3-3 Heatmap of Top 50 Hypermethylated Regions in *App*$^{NL-G-F}$ mice Across Samples

The heatmap displays the top 50 hypermethylated differentially methylated regions (DMRs), with each row representing a DMR (gene symbols) and each column representing a sample. The colour scale represents beta methylation values which represents the percentage of CpG sites in that region are methylated; where blue indicates low methylation, green/yellow represents intermediate levels, and red indicates high methylation. Rows are hierarchically clustered to reveal distinct methylation patterns across samples. Gene symbols that appear more than once, indicate multiple distinct DMRs identified within or near the same gene, highlighting separate regions with significant methylation differences. Grey values in the heatmap denote missing or low-quality data for specific regions in certain samples, where beta methylation values could not be reliably measured. Sample columns (*App*$^{NL-G-F}$ or WT) are labelled with sex (M or F) and pool allocation (pool A to F).

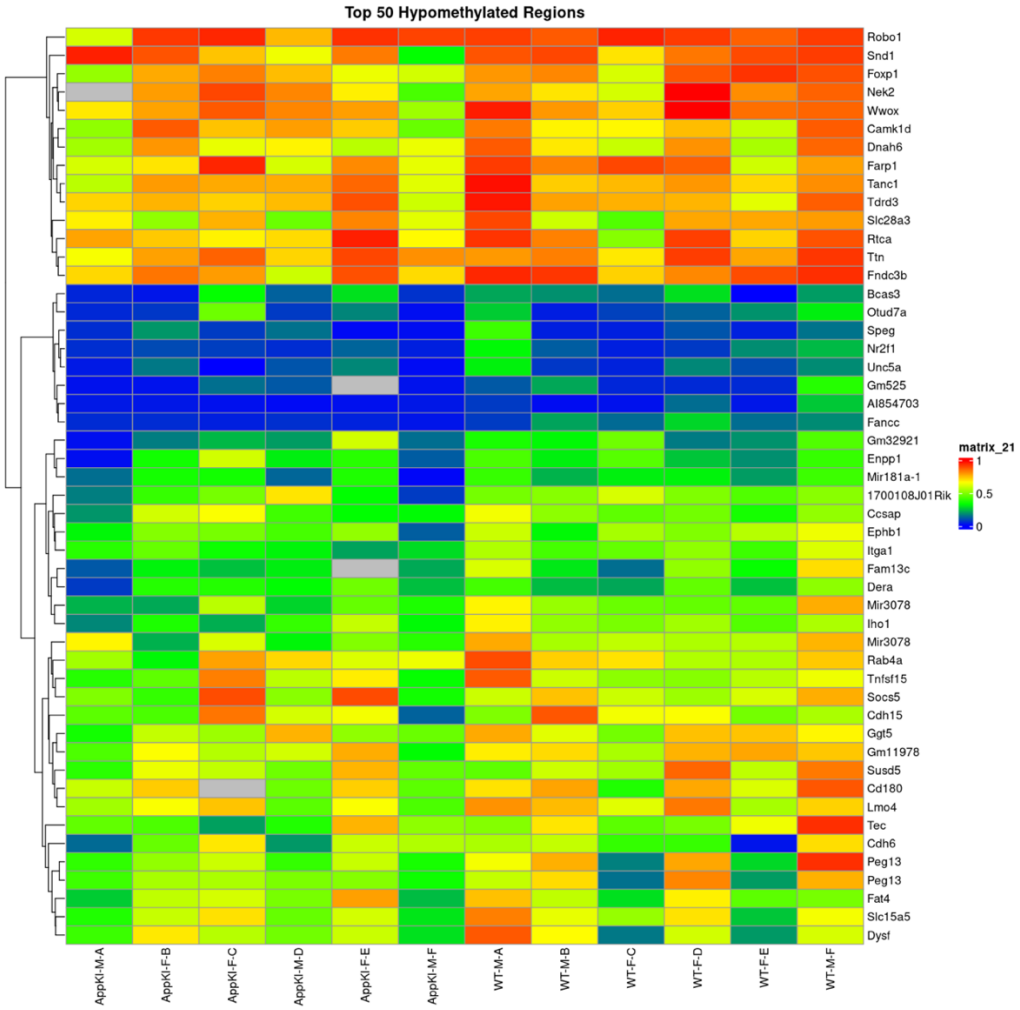

Figure 3-4 Heatmap of Top 50 Hypomethylated Regions in *App*$^{NL-G-F}$ mice Across Samples

This heatmap displays the top 50 hypomethylated differentially methylated regions (DMRs), with each row representing a DMR (gene symbols) and each column representing a sample. The colour scale indicates beta methylation values, where blue represents low methylation, green/yellow represents intermediate levels, and red indicates high methylation. Gene symbols appearing more than once, such as in cases of multiple DMRs near the same gene, highlight distinct regions with separate methylation patterns. Grey values in the heatmap denote missing or low-quality data for specific regions in certain samples, where beta methylation values could not be reliably measured. Sample columns (*App*$^{NL-G-F}$ or WT) are labelled with sex (M or F) and pool allocation (pool A to F).

3.3.3.1 Annotation of DMRs and Genomic Feature Associations

To explore the biological relevance of these DMRs, each was annotated based on its genomic context – whether located within promoters, exons, introns, or intergenic regions, and proximity to known genes. This annotation step mirrors the potential biological relevance of the DMRs, as regions located near or within genes are more likely to be affected by methylation differences observed in the region. DMRs were annotated relative to known genomic features such as TSS, promoters (defined as -2,000 to +500 bp relative to TSS), exons, introns, and intergenic regions (Yu et al., 2015). We assessed whether these DMRs corresponded to AD risk genes (Bellenguez et al., 2022), and DMRs detected in human AD brain (Altuna et al., 2019) regulatory elements, by comparing methylation patterns between $App^{NL\text{-}G\text{-}F}$ and WT mice, and the response to amyloid pathology.

Specifically, DMRs were annotated to genomic features using *ChIPseeker* and the *TxDb.Mmusculus.UCSC.mm10.knownGene* annotation database. Although the majority of DMRs were hypothesized to reside within promoter regions due to the TSS window applied, many were annotated to distal intergenic and intronic regions, particularly those located farther from the TSS. This observation aligns with previous reports that enhancer regions and distant regulatory elements, which are often intergenic, play critical roles such as regulation of gene expression, particularly in the context of complex diseases (Agrawal et al., 2019; Andersson et al., 2014).

Table 3-2 Summary of genomic annotation of the top 10 differentially methylated regions (DMRs) identified in the methylation analysis.

chr	*start*	*end*	*length*	*nCG*	*areaStat*	*annotation*	*distanceToTSS*	*symbol*	*gene name*
chr15	72809480	72811193	1714	83	-254.496	Promoter (<=1kb)	0	*Peg13*	*paternally expressed 13*
chr6	48629266	48629730	465	56	-116.461	Exon (2 of 4)	1331	*AI854703*	*expressed sequence AI854703*
chr14	55052648	55054356	1709	43	112.997	Promoter (<=1kb)	0	*Zfhx2os*	*zinc finger homeobox 2*
chr10	20953895	20954702	808	31	92.767	Exon (1 of 1)	1348	*Ahi1*	*Abelson helper integration site 1*
chr6	138006604	138007360	757	28	-72.570	Intron (4 of 5)	29543	*Slc15a5*	*solute carrier family 15, member 5*
chr2	118554971	118555340	370	31	68.242	Distal Intergenic	-5284	*Bmf*	*BCL2 modifying factor*
chr1	75388161	75388368	208	27	-57.232	Exon (4 of 41)	2551	*Speg*	*SPEG complex locus*
chr19	5406526	5407047	522	25	51.951	Promoter (<=1kb)	0	*4930481A15Rik*	*RIKEN cDNA 4930481A15 gene*
chr3	128520088	128520657	570	19	49.973	Distal Intergenic	-288482	*D030025E07Rik*	*RIKEN cDNA D030025E07 gene*
chr9	114073632	114074068	437	21	-49.555	Intron (3 of 4)	16492	*Susd5*	*sushi domain containing 5*

chr: Chromosome where the DMR is located.
start and end: Genomic coordinates marking the start and end of each DMR.
length: Length of the DMR in base pairs.
Number of CpG sites within each DMR.
Area statistic indicating the magnitude of methylation change for the DMR.
annotation: Genomic location of the DMR, such as promoter, exon, intron, or intergenic region.
distanceToTSS: Distance in base pairs from the DMR to the nearest transcription start site (TSS).
Symbol: Gene symbol for the nearest gene associated with the DMR.
Gene name: Full name of the nearest gene, highlighting the potential biological relevance of the DMR.

The analysis revealed that methylation changes in *App*$^{NL-G-F}$ model are most concentrated within intronic and distal intergenic regions, which comprise the highest proportions of both hypomethylated and hypermethylated sites (Fig 3.5). Specifically, intronic regions exhibit substantial methylation changes (Table 3.2, Supp. Table 3.1.), which may play a role in regulating splicing (Petibon et al., 2016) or modulating gene expression (Xue et al., 2011). Distal intergenic regions, often associated with enhancer activity, also display significant methylation alterations, indicating that long-range regulatory elements may be impacted in this disease context. Exonic and promoter regions contain fewer but still meaningful methylation changes, with hypermethylation in exonic regions potentially contributing to isoform-specific regulation and promoter hypomethylation suggesting increased gene activation. Minimal methylation changes were observed in the 3' and 5' UTRs and downstream regions, indicating that these areas are less affected by differential methylation in this analysis.

The feature distribution of DMRs further shows that most regions are annotated as gene body or intergenic regions (Fig. 3.5). Intronic regions, which can harbour regulatory elements such as enhancers and silencers, suggest that observed methylation changes in these regions may influence the expression levels of nearby genes. A concentration of binding sites within the "0-1 kb" range from TSSs supports a broader regulatory framework, implying that methylation's impact on gene expression may not be solely dependent on DMRs' absolute position but may also involve higher-order chromatin structures and DNA looping that bring together distant regulatory elements and TSSs.

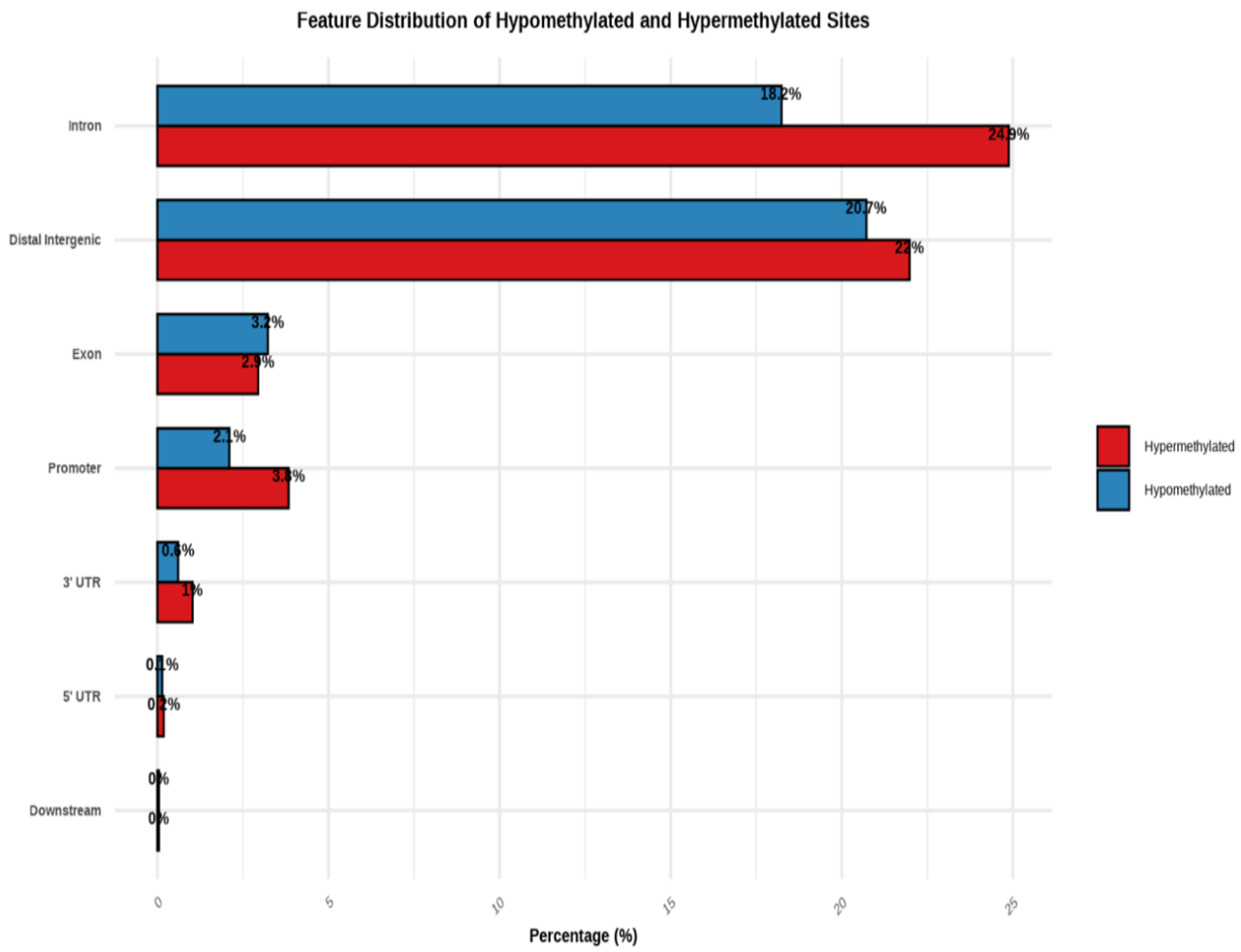

Figure 3-5 Feature distribution of hypomethylated and hypermethylated sites
The bar plot shows the genomic context of hypomethylated (blue) and hypermethylated (red) sites as a percentage of total sites analysed. Features include intronic, distal intergenic, exonic, promoter, 3' UTR, 5' UTR, and downstream regions. Intronic and distal intergenic regions account for the highest percentage of methylation changes, followed by exonic and promoter regions. Minimal methylation changes are observed in the UTR and downstream regions. Annotations of genomic regions obtained via mm10 database using *annotatePeak()* function from *ChiPseeker* package.

When considering known protein binding sites, a majority of hypermethylated and hypomethylated binding sites cluster within the "0-1 kb" range from the TSS, with 48.3% of hypermethylated and 47.9% of hypomethylated sites in this category (Fig. 3.6). Considering the genomic context of these methylated sites as in Figure 3.5., distance from TSS were small in not only promoter areas but other regions such as

gene body and distal intergenic region. This expected proximity suggests a direct influence on transcription factor binding, potentially modulating gene expression. Hypermethylated sites also show a significant proportion of binding loci within the "10-100 kb" range (24.9%), suggesting that long-range regulatory elements may be impacted by hypermethylation. Conversely, hypomethylated regions display a notable percentage of binding sites further from the TSS, with 16.5% located in the ">100 kb" range, potentially affecting distal regulatory elements such as enhancers or insulators.

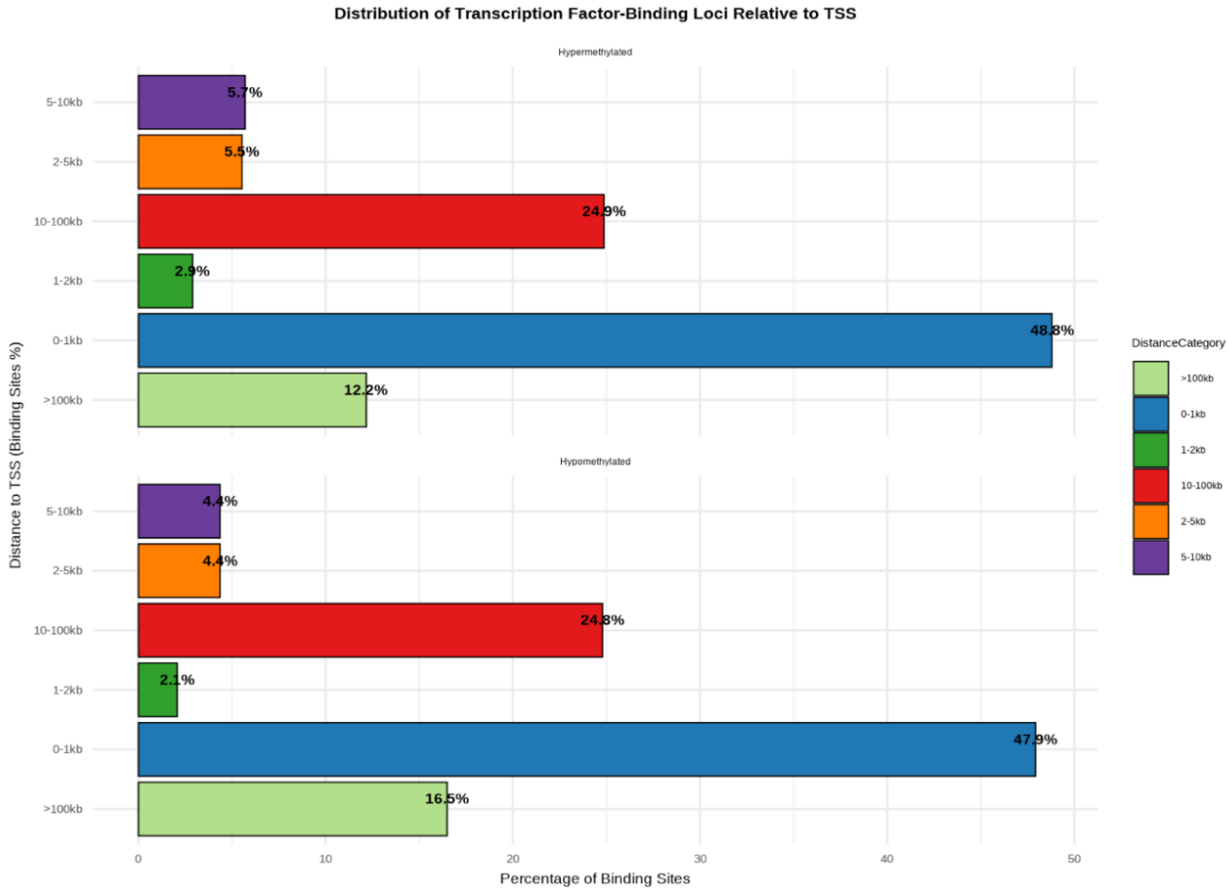

Figure 3-6 Distribution of Transcription Factor-Binding Loci Relative to TSS

Distribution of transcription factor-binding loci relative to the transcription start site (TSS) for hypermethylated (top) and hypomethylated (bottom) regions annotated via mm10 genome using *annotatepeak()* function from *ChiPseeker* package. Binding sites are grouped by their distance from the TSS into six categories: "0-1kb", "1-2kb", "2-5kb", "5-10kb", "10-100kb", and ">100kb". Each bar represents the percentage of binding sites in each distance category, with the percentage values displayed at the end of the bars. The colour-coding corresponds to the different distance categories. Hypermethylated sites show the highest binding activity within "0-1kb" regions

(48.3%), while hypomethylated regions also exhibit a significant percentage (47.9%) within the same range. Additionally, a considerable proportion of hypermethylated sites are located in the "10-100kb" range (24.9%), whereas hypomethylated regions have a notable number in the ">100kb" range (16.5%).

3.3.3.2 Gene ontology annotation of promoters

GO analysis of differentially methylated promoters in the *App*$^{NL\text{-}G\text{-}F}$ model reveals distinct biological processes influenced by hypermethylation and hypomethylation, with clustering illustrating functional similarities among GO terms. For hypermethylated promoters, GO analysis identified 42 clustered terms (Fig. 3.7), encompassing biological processes such as splicing, cellular maintenance, communication, metabolic processes, and positive regulation (Supplementary Table 3.2). The presence of splicing-related GO terms suggests that hypermethylation in these promoter regions could impact alternative splicing and RNA-binding proteins, potentially disrupting transcript processing. The identified clusters suggest potential repression of genes involved in cellular maintenance, communication, and metabolic processes, which may contribute to impaired cellular function and possibly play a role in AD progression. Hypomethylated promoters are associated with 35 distinct GO terms, which primarily cluster into categories related to dephosphorylation pathways (FDR = 0.007). Immune response-related terms also rank second, although they did not reach statistical significance (Fig. 3.8; Supplementary Table 3.4). These findings imply that hypomethylation in these promoter regions may enhance processes associated with immune activity and cell signaling, potentially contributing to an adaptive or compensatory immune response to amyloid pathology in AD.

Several key genes within enriched pathways were noted for their potential relevance to AD. For example, Semaphorin 4D *(Sema4d*), enriched in the phosphorylation pathway, exhibited differential splicing and isoform switching in response to amyloid (Supplementary Table 2.9 and 2.10). *Sema4d* is known to mediate cell-cell interactions and is implicated in neuroinflammatory responses, potentially affecting glial cell behaviour and neuron-glia interactions. Similarly, *Protein Phosphatase 2 Catalytic Subunit Beta (Ppp2cb)* and *Protein Phosphatase 3 Catalytic Subunit Gamma (Ppp3cc)* showed isoform switching, with roles in signaling pathways relevant to synaptic plasticity and neuronal survival. Their intersection with *Sema4d* may reflect an interplay between phosphatase activity and neuroinflammatory signaling,

potentially modulating neuronal responses to amyloid build-up. Other genes within the dephosphorylation pathway, such as Dual specificity protein phosphatase 2 (*Dusp2),* known for its dephosphorylation of MAPK family members, may mitigate hyperactivation of Mitogen-Activated Protein Kinase (MAPK) pathways, potentially offering protection against amyloid-induced stress. Additionally, *Protein Tyrosine Phosphatase, Non-Receptor Type 22 (Ptpn22)* has increased expression in these conditions and plays a role in immune cell signaling, indicating a pathway by which the immune response may be modulated during amyloid accumulation.

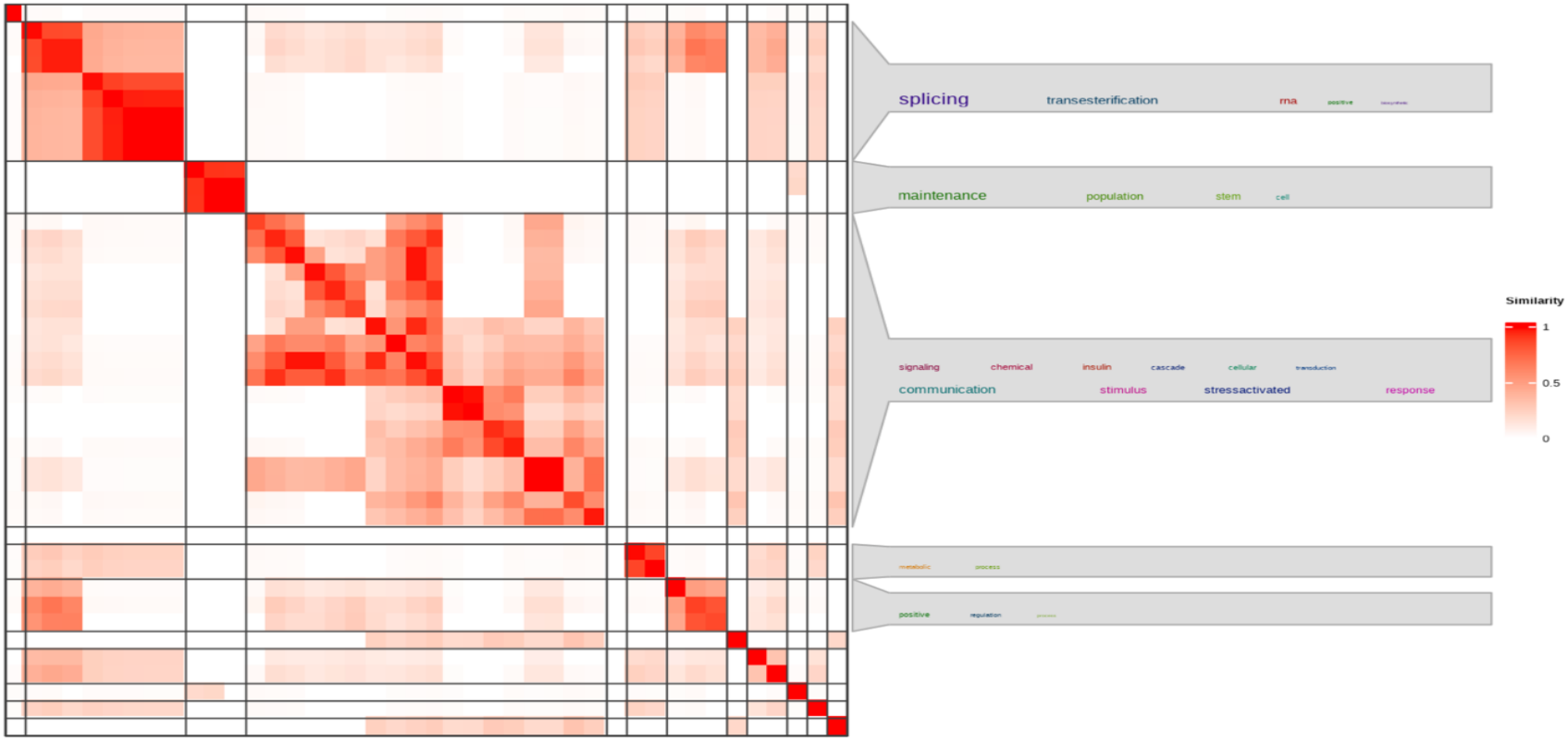

Figure 3-7 Gene ontology annotation of hypermethylated promoters

This heatmap shows clustering of 42 GO terms linked to hypermethylated promoters, identified using rGREAT and refined with simplifyEnrichment. Terms are grouped by functional similarity, with red shading indicating similarity strength, and clusters labeled as "splicing," "maintenance," "communication," "metabolic processes," and "positive regulation." Larger text indicates more significant enrichment (p-value < 0.05). Enrichment was assessed using a hypergeometric test with Benjamini-Hochberg correction, highlighting key pathways associated with hypermethylation.

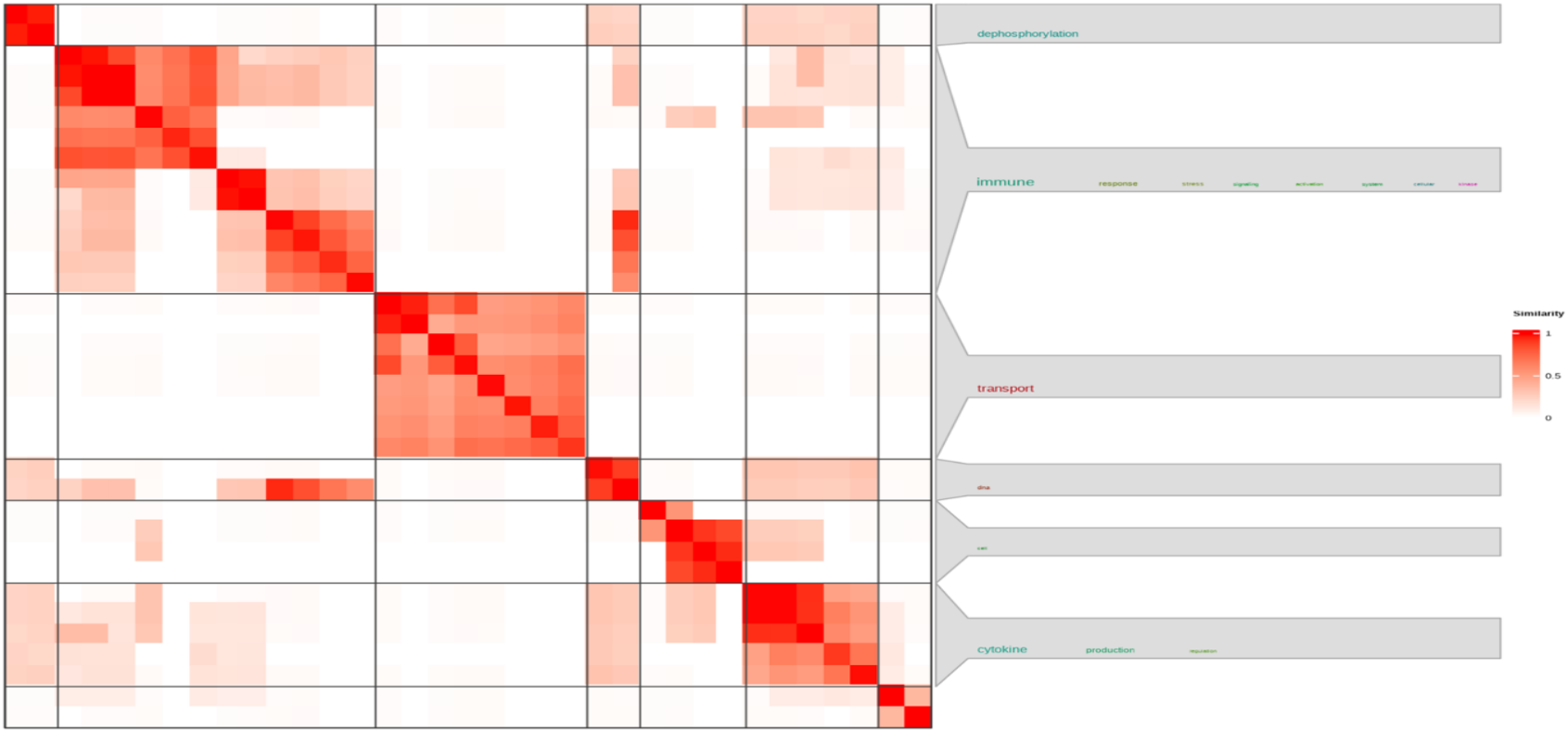

Figure 3-8 Gene ontology annotation of hypomethylated promoters

This heatmap shows clustering of 35 GO terms linked to hypomethylated promoters, identified using rGREAT and refined with simplifyEnrichment. GO terms are grouped by biological similarity, with red shading indicating higher similarity and white representing lower or no similarity. Functional categories like "dephosphorylation" and "immune response" are highlighted, with hierarchical clustering revealing groupings of hypomethylation that may influence gene expression. Larger text indicates more significant enrichment (pp-value < 0.05). Enrichment was assessed using a hypergeometric test with Benjamini-Hochberg correction.

3.3.3.3 Gene ontology annotation of all DMR sites

Amongst the top hypermethylated and hypomethylated DMRs, many regions were within the gene body and the intergenic regions. Limiting our analysis to promoter regions alone can lead to a significant oversight. In gene bodies, differential methylation has been suggested to correlate with gene expression levels, for instance, hypermethylation within gene bodies is often associated with increased transcriptional activity, while hypomethylation with reduced expression (Zhou et al., 2021). Intragenic methylated CpG islands have been shown to be tissue-specific and significantly associated with transcriptional activity (Almamun et al., 2015). The regulatory roles of methylation in the 3' and 5' untranstated regions (UTRs) could potentially modulate gene expression by influencing mRNA stability and translation.

Many crucial regulatory elements, such as enhancers and silencers, are located outside of the gene. These elements can greatly influence gene expression and play vital roles in the complex regulatory networks that operate within the genome. For that purpose, all the DMRs, regardless of position to the TSS, were annotated with the GO database, but separately for hypomethylation and hypermethylation .

The gene sets associated with both hypermethylated and hypomethylated sites showed enrichment in similar pathways (Fig. 3.7 and 3.8), primarily related to metabolic processes and cell development (Supplementary Table 3.3 and 3.5). In the hypomethylated region, the “cellular development” enrichment term had the largest cluster, meaning the majority of the genes regulated via hypomethylated regions shared the developmental/differentiation relevant annotations as observed in hypermethylation annotations. Immune cell migration and trafficking was specific to hypomethylated genes, including genes such as *Colony Stimulating Factor 3 Receptor* (*Csf3r)* (areaStat= -8.261). *CSF3R* encodes the receptor for colony-stimulating factor 3, a cytokine that controls the production, differentiation, and function of granulocytes, and identified AD blood biomarker (Seligmann et al., 2023).

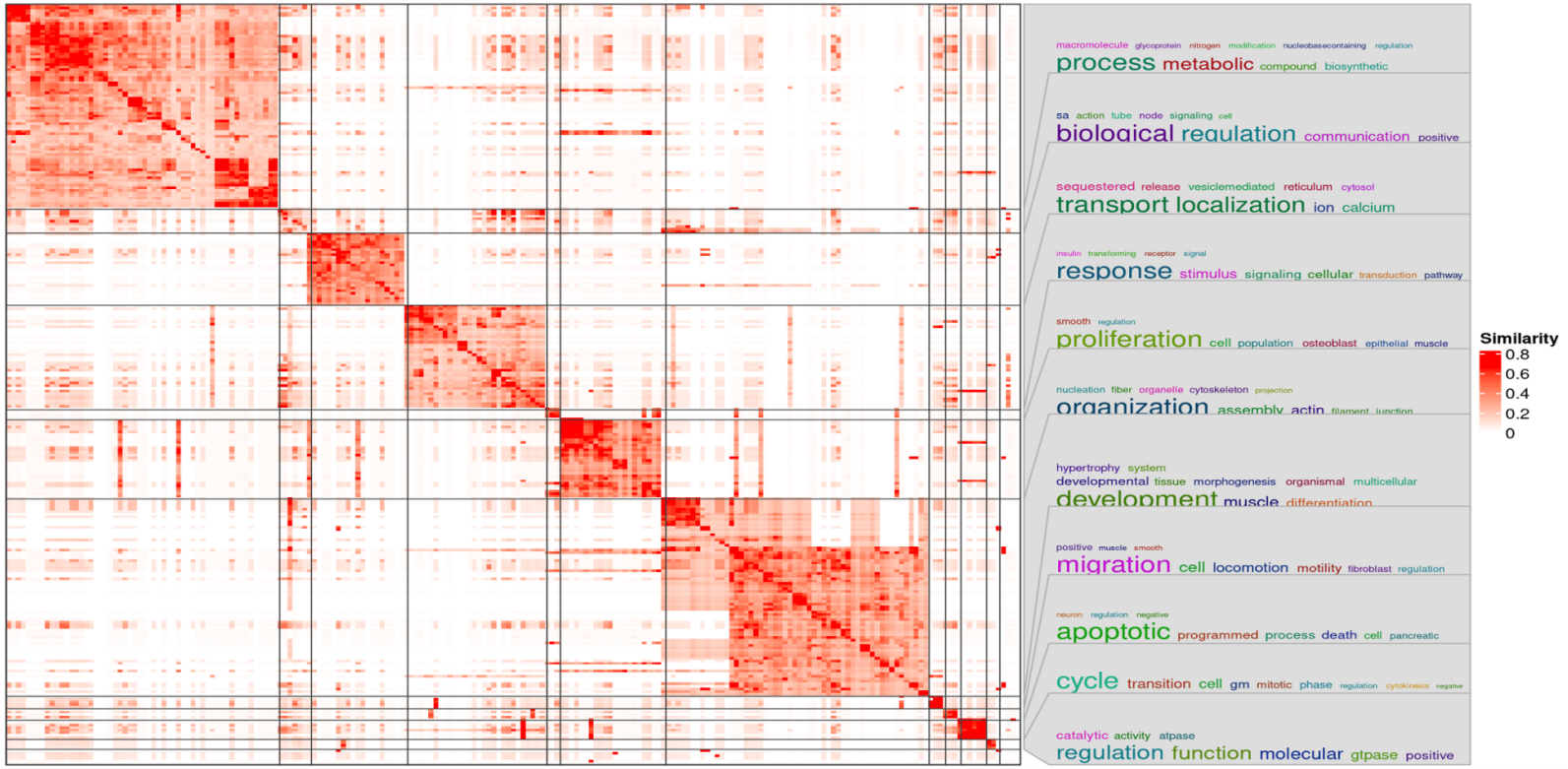

Figure 3-9 Gene ontology annotation of hypermethylated DMRs

This heatmap shows clustering of 990 GO terms associated with hypermethylated regions in gene bodies and intergenic regions, identified using rGREAT and refined with simplifyEnrichment. GO terms are grouped by biological similarity, with red shading indicating higher similarity and white indicating lower similarity. Functional categories, such as "metabolic pathways," "protein localization," and "neuronal and microglial response," are highlighted on the right. Text size reflects enrichment significance, with larger text indicating greater significance (pp-value < 0.05). Enrichment was assessed using a hypergeometric test with Benjamini-Hochberg correction.

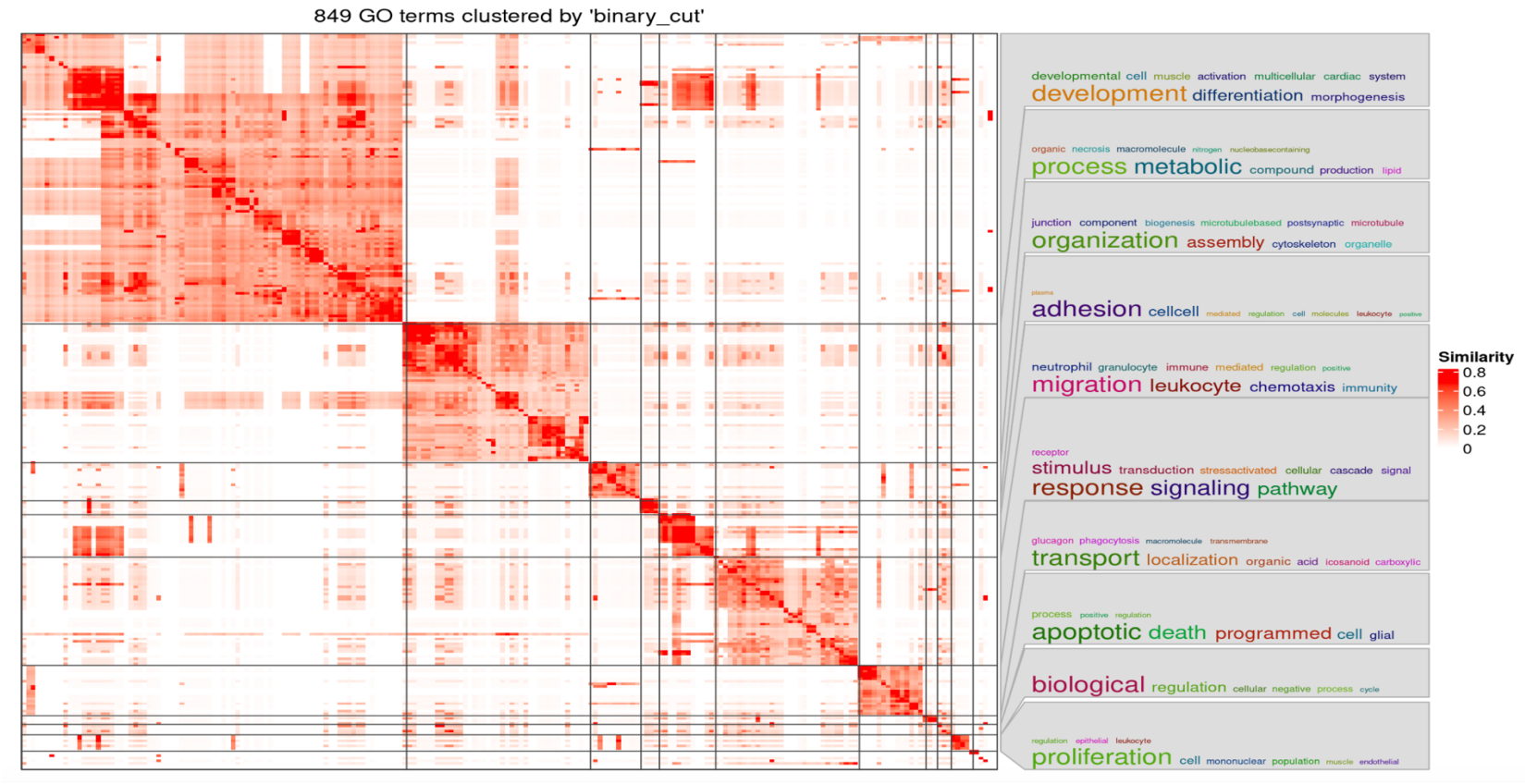

Figure 3-10 Gene ontology annotation of hypomethylated DMRs

This heatmap displays clustering of 849 GO terms associated with hypomethylated regions, covering DMRs across promoters, gene bodies, and intergenic regions. These regions are linked to genes regulating processes such as development, cell migration, cellular response to the environment, and protein localization. Text size reflects enrichment significance, with larger text indicating greater significance (pp-value < 0.05, adjusted for FDR). Enrichment analysis was performed using a hypergeometric test with Benjamini-Hochberg correction, based on 849 significant GO terms derived from hypomethylated regions.

3.3.4 Cell type enrichment of differentially methylated genes with a mouse reference dataset

To investigate the cell type specificity of genes potentially regulated by hypermethylation or hypomethylation, we first focused on promoter regions, given their well-established role as critical regulators of gene expression through methylation dynamics (Jones, 2012). Promoter methylation is known to influence transcriptional activity by modulating transcription factor binding and chromatin modifier recruitment, where hypermethylation is typically associated with transcriptional repression and hypomethylation with activation (Bird, 2002). These properties made promoter regions an ideal starting point for this analysis.

Using EWCE, we analyzed differentially hypermethylated and hypomethylated promoters (-2000 to +500 bp relative to the TSS) separately to determine cell type-specific enrichment. As illustrated in Figure 3.11, the enrichment test was conducted across 48 cell types (Zeisel et al., 2015) Although there were no significant associations observed for promoters overall, there was a tendency of hypomethylated genes to be expressed specifically in interneurons (Int), vascular endothelial cells (Vend), perivascular microglial cells (Pvm) and microglial cells (Mgl) subtypes. While on the gene expression level, microglial enrichment was observed in Chapter 2, isoform-level alterations pointed out more diverse enrichment of the cell types. A variety of neuronal cell types express the genes marked with hypermethylation and hypomethylation but overall did not reach statistical significance for enrichment when only promoter regions were analysed.

When all DMRs (including non-promoter regions) were incorporated into the analysis, a stronger enrichment for neuronal cell types was detected, as shown in Figure 3.12. A total of 664 genes were retained for hypermethylation, while 542 genes were identified for hypomethylation after filtering (Fig. 3.12). This broader inclusion revealed significant enrichment ($q < 0.05$) in three cell types for both hypermethylated and hypomethylated genes. Both hypermethylated and hypomethylated genes were enriched in pyramidal neurons in the hippocampal CA1 region, pyramidal neurons in the somatosensory Layer 2/3 region, and interneurons in the hippocampal CA1 region. These results indicate that DNA methylation changes, particularly those outside

promoter regions, may play a role in regulating isoform-level alterations, as suggested by the overlap with findings from Chapter 2.

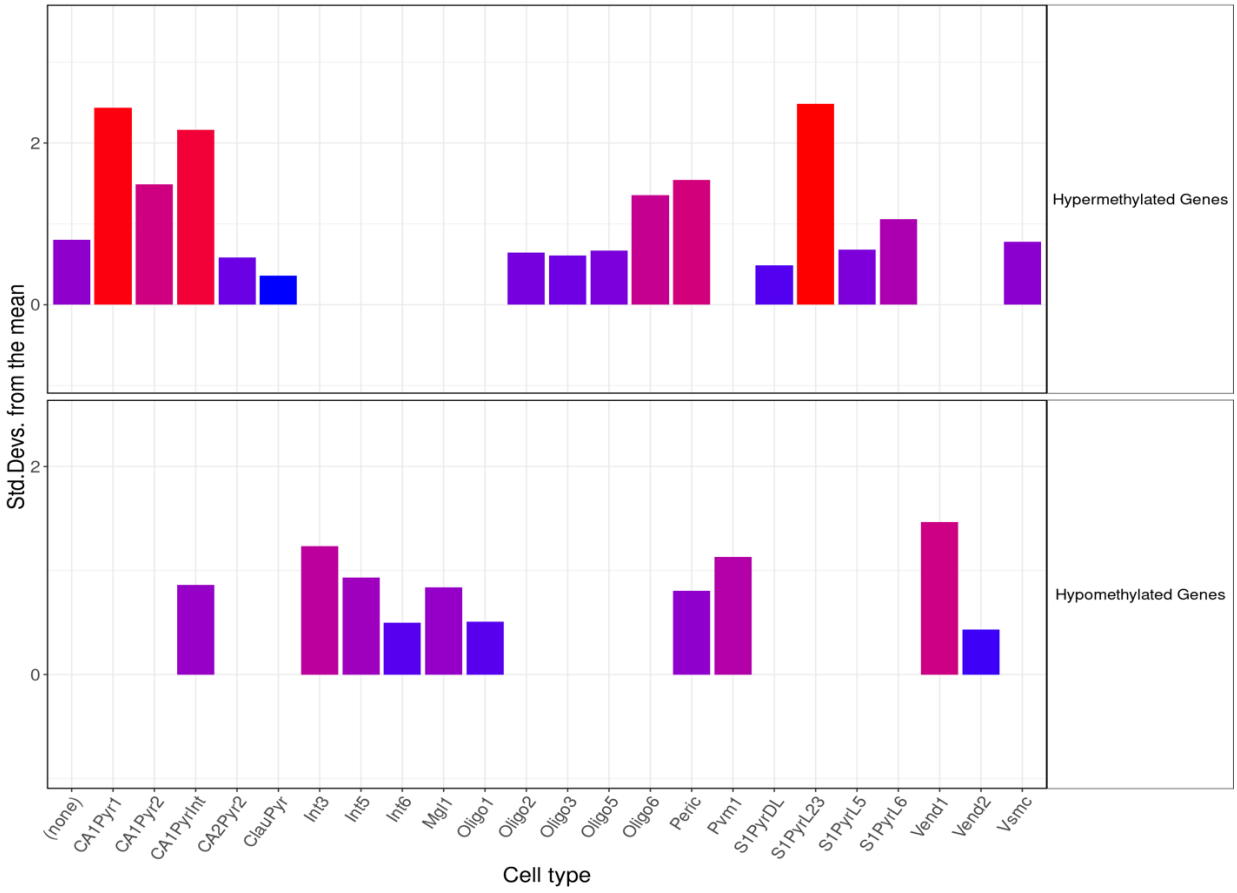

Figure 3-11 Cell type enrichment of expressed genes marked by differentially hypermethylated and hypomethylated promoters

The top panel illustrates the enrichment of expressed genes marked by hypermethylation (only promoters) across cell types, while the bottom panel depicts the enrichment of expressed genes marked by hypomethylation (promoters). The y-axis represents the number of standard deviations from the mean expression for each gene within a given cell type. Genes not classified into one of the defined major cell types are grouped under “None.” Enrichment analysis was performed using the bootstrap enrichment test from the EWCE package, with Benjamini-Hochberg (BH) adjustment applied to control the False Discovery Rate (FDR) and calculate q-values (n=6). No significant enrichment was observed for either hypermethylated or hypomethylated promoter regions. The colour scale represents the degree of deviation, with intensity corresponding to the magnitude of enrichment in standard

deviations. Statistical significance was determined based on the adjusted p-values (q-values).

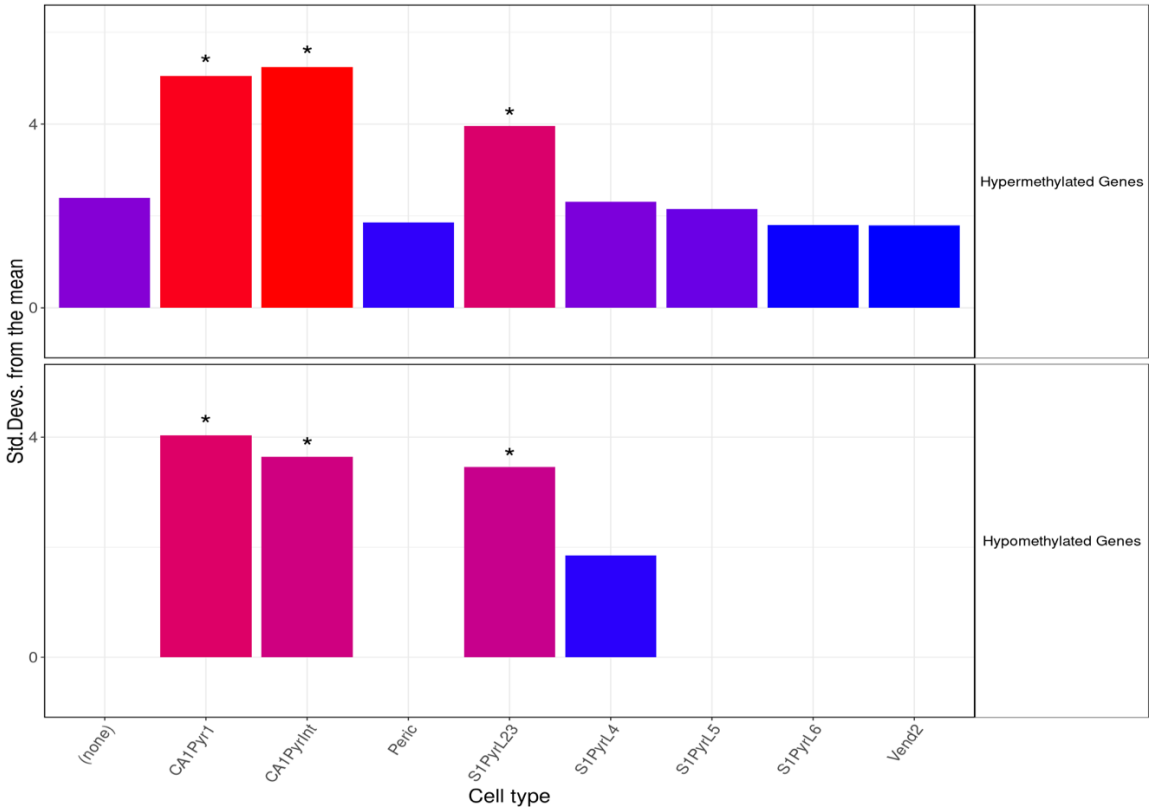

Figure 3-12 Cell type enrichment of expressed genes marked by differentially hypermethylated and hypomethylated regions outside of promoters

The top panel represents expressed genes marked by hypermethylation, and the bottom panel shows those marked by hypomethylation outside of promoters. The y-axis indicates the number of standard deviations from the mean expression for each gene per cell type. Genes marked with asterisks (*) demonstrate statistically significant deviations from the mean, highlighting the most differentially methylated genes in that cell type. Enrichment analysis was performed using the bootstrap enrichment test from the EWCE package, with Benjamini-Hochberg (BH) adjustment applied to control the False Discovery Rate (FDR) and calculate q-values (n=6). No significant enrichment was observed for either hypermethylated or hypomethylated promoter regions. The colour scale represents the degree of deviation, with intensity corresponding to the magnitude of enrichment in standard deviations. Statistical significance was determined based on the adjusted p-values (q-values).

3.4 Discussion

3.4.1 Long-read DNA sequencing quality control metrics and alignment statistic between genotypes

A limited number of studies have used long-read DNA sequencing to capture genome-wide methylation signatures. Particularly, for mouse models or for AD or neurodegenerative disease contexts, so the number of resources for cross-referencing are quite limited. Thus, this study presents a novel and comprehensive methylation profile of the *App*$^{NL\text{-}G\text{-}F}$ AD mouse model and WT controls. The dataset could be also used as a reference for detecting repetitive elements or large structural variants in mouse studies.

The reliability of the data assessed by quality control metrics for both the *App*$^{NL\text{-}G\text{-}F}$ and WT samples (Fig. 3.1), compare very favourably with published benchmarks for long-read methylation studies. The total read counts of approximately 2.83 million for *App*$^{NL\text{-}G\text{-}F}$ and 3.29 million for WT are in line with previous studies using long-read sequencing technologies, which generally report similar read counts as sufficient for comprehensive CpG methylation analysis (Yuen et al., 2021). The total yield of 29.75 Gb for *App*$^{NL\text{-}G\text{-}F}$ and 33.07 Gb for WT is well within the range expected for high-coverage, genome-wide methylation studies, further supporting the depth and reliability of the dataset (Beyter et al., 2021) Additionally, the mean sequencing accuracy of 93.15% for *App*$^{NL\text{-}G\text{-}F}$ and 93.08% for WT is above the 90% threshold considered necessary for confident methylation calling, and the data is of high fidelity (Simpson et al., 2017). The N50 values (~14 kb) and median read lengths (~10 kb) meet the benchmarks set by studies such as those by Simpson et al. (2017), which emphasize the importance of longer read lengths for covering critical regulatory regions, including promoters and enhancers (Simpson et al., 2017).

A recent benchmarking study by Sigurpalsdottir et al. (2024) evaluated the accuracy of two long-read sequencing platforms—Nanopore and SMRT sequencing by Pacific Biosciences—against oxidative bisulfite sequencing (oxBS), a gold-standard method for DNA methylation profiling. The study found a strong correlation between sequencing coverage and the accuracy of methylation detection (Sigurpalsdottir et al., 2024). Based on these findings, we included methylated regions with more than 10x coverage in our analysis.

Promoters, typically located within a few hundred to a couple of thousand base pairs around transcription start sites (TSS), play a critical role in regulating gene expression. Sigurpalsdottir et al. (2024) demonstrated that long-read sequencing reliably replicates biological patterns of methylation, such as the characteristic depletion near TSS. Their study highlights the importance of methylation changes in promoter regions and their potential impact on transcriptional regulation.

In summary, the number of reads, N50 values, median read lengths, and coverage all meet or exceed the standards necessary for conducting a comprehensive methylation analysis. The high base calling accuracy and sufficient sequencing yield further support the validity of our results. Our study provides a novel resource and a reference, given there are few prior studies that have used long-read sequencing to capture genome-wide methylation patterns in AD mouse models, which provides novel insights into the epigenetic mechanisms of the response of the brain to amyloid pathology (Liu et al., 2021; Logsdon et al., 2020).

3.4.2 Global methylation pattern shows high percentage of variation near transcription start sites

The analysis of transcription factor-binding loci within promoters revealed both hypermethylation and hypomethylation in response to amyloid at the highest percentage within 0-1 kb from the TSS. This proximity to the TSS strongly suggests that methylation changes in these regions could directly influence gene transcription by modulating transcription factor binding. Hypermethylation near the TSS is commonly associated with transcriptional repression, as methylation can prevent the binding of transcription factors. In contrast, hypomethylation in these regions can result in increased transcriptional activity. Proximity of methylation changes to TSS was consistent in both hypermethylated and hypomethylated regions, suggesting that AD-related methylation changes could either suppress or activate the transcription of genes directly implicated in disease pathology. Furthermore, the distribution of hypermethylated sites in the 10-100 kb range suggests that some methylation changes may be acting on distant regulatory elements, such as enhancers, which can modulate gene expression and DNA looping over long genomic distances. Hypomethylated sites, on the other hand, showed a notable enrichment in the >100 kb range, indicating the involvement of more distal elements. This reinforces the

importance of considering non-coding regions when studying the epigenetic landscape of AD. The GO analysis of hypermethylated and hypomethylated promoters further supports the functional relevance of the observed methylation changes in neuronal adaption of different neuronal populations.

3.4.3 Gene ontology annotation of DMRs

The GO analysis of differentially methylated promoters in the *App*$^{NL\text{-}G\text{-}F}$ model provided new insights into the regulatory roles that hypermethylation and hypomethylation may play in AD pathology. Promoters, essential for initiating transcription, are particularly sensitive to methylation changes that can alter transcription factor binding and gene expression (Lubliner et al., 2015; Xu et al., 2015). Hypermethylation in promoter regions, as observed in the *App*$^{NL\text{-}G\text{-}F}$ model, may repress gene expression by restricting transcription factor access (Bubnova et al., 2023). Specifically, hypermethylated promoters in this model are enriched for processes related to RNA splicing, cellular maintenance, metabolic pathways, and stress-activated signaling, suggesting that methylation of these regions might silence genes vital for neuronal homeostasis. RNA splicing is crucial for generating protein diversity, and alterations in this process could lead to abnormal protein isoforms, contributing to the cellular dysfunction. Notably, 49 out of 100 differentially spliced genes (Supplementary Table 2.10, and Supplementary Table 3.1) overlap with the DMRs in the same model in response to amyloid pathology, although only a few DMRs directly overlap with spliced regions. This observation suggests that hypermethylation may control factors mediating splicing control of genes, although the DNA methylation of promoters themselves may not directly regulate splicing. DNA methylation changes altering expression of splice factors may potentially compounding regulatory feedback loops that drive disease progression.

Hypermethylation in the promoters of stress-response genes involved in the response to amyloid are a series of serine/threonine (MAPK) kinase cascade, such as glutathione s-transferase pi 2 *(Gstp2),* SMAD family member 3 *(Smad3),* mitogen-activated protein kinase 6 *(Map2k6),* malignant fibrous histiocytoma amplified sequence 1 *(Mfhas1),* and SAM and SH3 domain containing 1 *(Sash1),* suggesting a mechanism by which the cell's ability to manage oxidative and inflammatory stress may be impaired. The MAPK cascade is a critical signaling pathway activated in

response to cellular stress, regulating processes like inflammation, apoptosis, and repair. This indicates that in response to amyloid accumulation, hypermethylation-mediated suppression of these stress-activated pathways, which increases oxidative stress and chronic inflammation, and could modulate the resilience of cells to external stimuli, which may exacerbate neuronal damage. It was previously shown that inhibiting the MAPK signaling pathway can alleviate memory decline and hippocampal damage in AD by reducing the expression of *APP* and *SP1* (Y. Du et al., 2019). Thus, this suggests a potential protective early mechanism via hypermethylation in $App^{NL\text{-}G\text{-}F}$.

Genes of hypomethylated promoters were enriched in GO terms related to immune response, cellular transport, and cytokine production, aligning with the inflammatory and immune changes that characterize AD pathology. Hypomethylation in immune-related promoters may lead to an upregulation of pro-inflammatory genes, promoting the chronic neuroinflammation seen in AD. For instance, immune-modulating genes such as *Ptpn22* may be upregulated due to hypomethylation, amplifying immune signaling and potentially accelerating disease pathology.

Hypomethylation was also enriched in transport-related genes such as *Cacna1e* and *Syt1*, suggesting possible upregulation of pathways involved in synaptic vesicle transport and neurotransmitter release. These processes are essential for maintaining synaptic function and plasticity, and disruptions in these transport mechanisms can impair synaptic communication and adaptability, contributing to the cognitive deficits associated with AD. Interestingly, while gene-level upregulation was not observed, isoform-level analyses (Chapter 2) reveal changes in specific isoforms associated with pre- and post-synaptic functions, alongside immune response pathways. These finding suggests a nuanced interaction between immune cell upregulation and specific synaptic isoform alterations, as reflected by methylation patterns. Such a pattern implies that while methylation changes indicate an immune response, they may also highlight specific isoform shifts relevant to synaptic processes. Despite limitations associated with GO terms covering only the biology studied to date by the community, these GO terms begin to provide new insights into how DNA methylation marks contribute to the progression of AD.

Overall, these findings suggest that differential methylation in promoter regions impacts a range of pathways relevant to AD, from immune responses to neuronal stress signaling, potentially contributing to amyloid adaptation and disease progression.

3.4.4 Cell type enrichment of differentially methylated promoters and all DMRs in AD

The enrichment analysis of genes marked by differentially methylated genes revealed distinct cell type associations for both hypermethylation and hypomethylation, emphasizing the regulatory potential of methylation changes across genomic contexts. Promoter regions, traditionally recognized as critical regulators of gene expression (Bird, 2002; P. A. Jones, 2012), provided an important focus for initial analyses. Methylation at promoters modulates transcription factor binding and chromatin structure, with hypermethylation typically silencing gene expression and hypomethylation often facilitating activation.

When analyzing promoter-specific methylation changes, significant enrichment patterns of marked genes were not observed for any cell types. However, when including the hypomethylated and hypermethylated DMRs from all regions, significant enrichment was observed in pyramidal neurons in the hippocampal CA1 region, pyramidal neurons in the somatosensory Layer 2/3 region, and interneurons in the hippocampal CA1 region. This finding suggests that methylation in intronic and intergenic regions, beyond promoters, also plays a key role in gene regulation in these cell types, as supported by studies linking such methylation events to alternative splicing and isoform-specific expression (Moore et al., 2013).

Genes marked by hypomethylated promoters, in particular, exhibited enrichment (ns) in microglial cells, aligning with findings from Chapter 2, where microglial activation was evident at both gene and isoform levels. These data point to a potential link between hypomethylation and microglial activation, a hallmark of neuroinflammation, observed in AD (Benitez et al., 2021; Keren-Shaul et al., 2017; Mathys et al., 2019; Olah et al., 2020; Parachikova et al., 2007; Sala Frigerio et al., 2019). The genes enriched in this microglial phenotype, ARM/DAM, were not entirely observed within the DMRs, indicating that the methylation is not the sole mechanism that drives the ARM/DAM phenotype. However, genes such as *Capg*, and *Dkk2* were found to be

enriched in the ARM/DAM response to amyloid pathology (Keren-Shaul et al., 2017; Sala Frigerio et al., 2019). *CAPG* was also found to be increased in proteomics in cerebro spinal fluid (CSF) from early onset AD individuals (Drummond et al., 2022). *CAPG* was also reported as consistently increased at protein level in AD proteomics studies, which signifies its role as a potential biomarker (Askenazi et al., 2023).

The inclusion of all DMRs, including those outside of classical promoters, revealed significant enrichment in pyramidal neurons and interneurons in hippocampal and somatosensory regions in both hypermethylated and hypomethylated regions, highlighting the importance of broader methylation events beyond promoter regions. These findings suggest that differential methylation may affect neuronal function by accompanied isoform usage or alternative splicing or by regulating these processes .

The combined evidence from promoter and genome-wide analyses highlights the complexity of methylation's role in cell type-specific gene regulation. Future studies integrating methylation with other epigenetic modifications and transcriptomic data will be essential to elucidate the functional consequences of these methylation changes, particularly in the context of neuronal and microglial interactions in neurodegenerative diseases. In longer term work it would be useful to disrupt methylated loci using CRISPR to investigate functional changes of DNA methylation on gene expression and splicing.

3.4.5 Potential role of DMRs in AD pathology

Differentially methylated regions (DMRs) identified in the analysis were associated with key genes involved in pathways essential for neuronal function and AD pathology. These pathways included neuronal signaling and plasticity, glutamate metabolism, oxidative stress, mitochondrial function, lipid metabolism, neuroinflammation, cytoskeletal integrity, and DNA repair (Fig. 3.9 and Fig. 3.10)

Epigenetic alterations in genes critical for neuronal function and signaling highlight potential mechanisms underlying AD pathology. For instance, *Forkhead Box P1 (Foxp1)*, a gene associated with neuronal signaling and plasticity, displayed significant and multiple hypermethylation changes in intergenic and intronic areas, as well as a hypomethylation in promoter area. *Foxp1* is essential for neurotrophic factor regulation and supports neuronal survival and synaptic plasticity (Kong et al., 2013). Its

dysregulation, driven by altered methylation, could impair neurotrophic support and disrupt the PI3K/Akt signaling pathway that is shown to be protective against AD progression (Peng et al., 2023). Reticulon *4 (Rtn4),* known for modulating BACE1 activity and amyloid production, exhibited differential hypomethylation in its promoter. Its role in amyloid plaque formation and synaptic dysfunction formation (Kulczyńska-Przybik et al., 2021; Masliah et al., 2010) suggests that epigenetic regulation may significantly influence amyloid-related pathology in AD. Other notable genes include zinc finger homeobox 2 *(Zfhx2os),* which exhibited hypermethylation in its promoter region, potentially silencing its expression. *Zfhx2os* is implicated in neuronal differentiation and function (Komine et al., 2012), and its repression could impair neurogenesis and synaptic plasticity.

Hypermethylation changes were also observed in exonic regions of abelson helper integration site 1 *(Ahi1),* and both hypermethylation and hypomethylation in solute carrier family 15 member 5 *(Slc15a5),* affecting its intronic regions. *Ahi1* is vital for ciliary trafficking and signaling, particularly in the hippocampus and cerebral cortex, brain regions notably affected in AD. Alterations in *Ahi1* methylation may disrupt ciliary function, leading to impaired neuronal signaling, which could contribute to cognitive impairments like memory loss commonly seen in AD (Hsiao et al., 2021). Hypomethylation changes in the promoter of genes related to glutamate metabolism and excitotoxicity, such as glutamate-ammonia ligase *(Glul)* and solute carrier family 25 member 12 *(Slc25a12),* also underline the potential role of excitotoxicity in AD. Excessive glutamate activity leads to excitotoxic neuronal damage, particularly in regions vulnerable to AD.

In addition to pathway-specific findings, cell-type enrichment analysis indicates changes in neuronal populations, highlighting genes such as *Cacna1e, Cux2,* Glutamate decarboxylase 1 *(Gad1), Nefm, Satb2,* and *Syt1* as central to neuronal function and behaviour. For example, *Cacna1e,* encoding the alpha-1E subunit of voltage-gated calcium channels (VGCCs), and essential for neurotransmission and synaptic plasticity, showed hypermethylation in an intronic region, potentially disrupting calcium homeostasis (Lauerer & Lerche, 2023). *Cux2* hypermethylation in the promoter region may further modulate neuronal signaling, possibly impacting synaptic strength and plasticity (Cubelos et al., 2010). *Gad1,* encodes glutamate

decarboxylase, a key enzyme responsible for synthesizing GABA, the primary inhibitory neurotransmitter in the brain. Hypermethylation in the 3' UTR of *Gad1* gene, could lead to loss of inhibitory input as observed previously in early AD (Guennewig et al., 2021; Ramos-Miguel et al., 2017). Similarly, *Syt1,* a key player in synaptic vesicle exocytosis, showed intronic hypermethylation, which could affect neurotransmitter release and synaptic communication (Chang et al., 2018). In contrast, hypomethylation in genes like RAR-related orphan receptor beta *(Rorb)* may indicate differential regulatory mechanisms affecting neuronal subtypes (Leng et al., 2021). Hypomethylation in these genes may correspond with increased gene expression, potentially influencing pathways related to neurogenesis and cellular differentiation.

The GO annotation of the promoter DMRs uncovered a significant enrichment of immune response pathways within hypomethylated regions. Hypomethylation, was notably associated with genes such as *Foxp1, Cd14, Rtn4,* and *Ipo5.* These genes are integral to immune response activation and suggest that chronic neuroinflammation in AD may not simply be a downstream consequence of neurodegeneration. Instead, hypomethylation appears to play an early role in initiating and sustaining the immune response.

In addition to the immune response pathways, a distinct enrichment for immune cell migration pathways was observed within hypomethylated regions, including hematopoietic stem cell migration ($q = 0.024$), which highlights the potential involvement of cellular migration pathways. While cell migration pathways were marked by both hypomethylated and hypermethylated regions, hypomethylated regions uniquely included genes such ectodomain phosphatase/phosphodiesterase-1 *(Enpp1),* B-cell lymphoma/leukemia 11B *(Bcl11b),* and exostosin glycosyltransferase *1 (Ext1).* These genes are known regulators of immune cell activity and movement, suggesting that hypomethylation may facilitate microglial migration around amyloid plaques or immune cell infiltration into the brain. This infiltration could contribute to the inflammatory environment characteristic of AD.

Interestingly, this regulation of immune cell migration pathways by hypomethylation may serve as an adaptive mechanism in early AD. Increased immune surveillance or cellular migration might initially be protective, enabling the brain to respond to

emerging damage. However, as the disease progresses, prolonged and heightened immune activation could foster chronic inflammation, contributing to neurodegeneration and worsening the disease's pathological landscape.

Further insights were provided by the enrichment of apoptotic pathways within hypomethylated regions, which included genes such as *Enpp1* again*,* peripheral myelin protein 22 *(Pmp22)* and a kinase anchoring Protein 12 *(Akap12)*. The upregulation of these genes suggests a potential link between immune activation and neuronal apoptosis, a hallmark of AD pathology. While apoptosis may serve as a protective mechanism in early stages of the disease, clearing damaged or dysfunctional cells, excessive activation of apoptotic pathways can lead to a detrimental reduction in neuronal populations. This loss of neuronal resilience can exacerbate cognitive decline, creating a vicious cycle of damage and cell death (Glass et al., 2010). Although the mouse model used here may does not show obvious neuron death, these pathways in early disease may prime neurons for death in late disease.

The inclusion of all DMRs revealed significant enrichment in neuronal cell types, particularly pyramidal neurons and interneurons in hippocampal and somatosensory regions. This suggests that methylation changes in intronic and intergenic areas, implicates broader regulatory elements in neuronal cell types. These non-coding DNA elements, often acting as enhancers, can influence transcriptional expression through enhancer-promoter interactions (Carullo & Day, 2019; Marzi et al., 2018). Some tissue-specific enhancers are located within intronic regions, suggesting that these non-coding regions are not merely passive but actively participate in the regulation of gene activity (Borsari et al., 2021). For instance, intronic hypermethylation in *Cacna1e* and *Syt1* may alter calcium signaling and synaptic vesicle release, thereby affecting neurotransmission and plasticity in these cells.

Several imprinted genes were observed in the DMRs list; genomic imprinting is an epigenetic mechanism that results in parent-of-origin-specific gene expression, meaning certain genes are expressed in a monoallelic manner based on whether they were inherited from the mother or father. In mammals, many imprinted genes are situated in the non-coding genome including paternally expressed gene 13 *(Peg13),* a long non-coding RNA implicated in AD due to its role in genomic imprinting and association with brain development. *PEG13* is located in a genomic region linked to

an AD-associated imprinting control region (Cevik et al., 2023). Its expression is regulated by brain-specific enhancers and is influenced by genomic imprinting, resulting in the selective expression of either the paternal or maternal allele in neural tissues (Claxton et al., 2022; Court et al., 2014). *Peg13* was identified as a top hypomethylated gene within the promoter region (≤1kb from the TSS) on chromosome 15.

Another key imprinted gene, ubiquitin-protein ligase E3A *(Ube3a),* is maternally expressed in neurons and plays a critical role in protein ubiquitination. *Ube3a* has been shown to be reduced with age in the AD Tg2576 mouse model, accompanied by a loss of dendritic spine density and behavioural deficits. In our analysis, *Ube3a* was found to be hypermethylated in a distal intergenic region, though this did not significantly impact gene expression at the bulk level. Additionally, hypermethylation was observed in the intronic region of scm-like with four mbt Domains 2 *(Sfmbt2)*, which is a maternally imprinted gene, meaning its expression is primarily derived from the paternal allele. It encodes a polycomb group (PcG) protein involved in chromatin remodeling and transcriptional repression, contributing to the regulation of developmental and cellular processes, and has been found to be differentially methylated in Braak Lewy body stage neuroinflammatory pathways, and increased in microglial cells (Oliveros et al., 2023). Finally, another imprinted gene, disks large-associated protein 2 *(Dlgap2)* was found to be differentially methylated in our study, where hypermethylation in an intronic region was observed. *DLGAP2,* a synaptic gene coding for DLGAP2 protein in the middle layer of the postsynaptic density complex (Li et al., 2017)*,* is highly expressed in the brain, and showed differential expression in brain regions associated with cognitive decline across diverse populations in AD (Ouellette et al., 2020). *Dlgap2* also has been identified as a potential key regulator of age-related cognitive decline in AD, with specific spine phenotypes linked to cognitive resilience against AD pathology (Ouellette et al., 2020). Since its protein form plays a critical role in synaptic plasticity by interacting with scaffold proteins in the postsynaptic density, disruption of these complexes can lead to synaptic stability and signaling dysfunction, emphasizing the importance of *DLGAP2* in preserving synaptic integrity and cognitive function. Thus, methylation changes in non-coding regions, may act, in part via imprinted genes to coordinate the cellular response to AD pathology.

3.4.6 Potential Implications for AD Pathology

The methylation landscape identified in this study provides key insights into the regulatory impact of DMRs on pathways and cellular processes essential for neuronal function and AD pathology. Both hypermethylation and hypomethylation were associated with distinct, yet interconnected, pathways, reflecting their potential roles in either suppressing or activating processes relevant to neurodegeneration.

Hypermethylation was observed in genes critical for neuroprotection, cellular maintenance, and stress responses. For example, hypermethylation in stress-response genes suggests a mechanism by which neuronal defences against oxidative stress and inflammation could be altered. In AD, where oxidative damage and inflammation are pervasive, this alteration in expression could indicate affected response of the cell's ability to disease-related stressors, which may cause neurons to be more vulnerable to degeneration. Although hypermethylation suggests repression, it's also worth considering that not all methylation changes necessarily lead to repression. Also, the change in the response to the stressor could be also a compensatory mechanism, to counteract other dysregulated pathways. This raises the possibility that certain methylation patterns could reflect adaptive responses rather than purely pathological processes.

Specific genes, such as *Foxp1,* exhibited hypermethylation that could interfere with their roles in transcriptional regulation and alternative splicing. *FOXP1* is also essential for neurotrophic signaling and synaptic plasticity (Khandelwal et al., 2024a), and its silencing has been implicated in promoting neuronal death or synaptic dysmorphogenesis (Louis Sam Titus et al., 2017). Similarly, hypermethylation in genes associated with cytoskeletal integrity and synaptic maintenance, such as *Nefm* and *Rtn4,* suggests potential impacts on axonal stability and synaptic communication.

In a study by Altuna *et al.* (2019), *RHOB* was identified as one of the hypermethylated genes in the human hippocampus, where it was significantly correlated with tau pathology (Pearson cor = 0.99, pvalue=0.013) (Altuna et al., 2019). In the $App^{NL\text{-}G\text{-}F}$ model, this hypermethylation occurs in a distal intergenic region approximately 13.7 kb from the *Rhob* TSS on chromosome 12 (chr12: 8,486,110–8,486,345) in response

to amyloid, suggesting the region serves as a regulatory element or enhancer rather than a promoter. The significance of *Rhob* hypermethylation in a distal intergenic region is two-fold. Firstly, such regions often harbour regulatory elements like enhancers that modulate gene expression by interacting with promoters via chromatin looping. Secondly, hypermethylation at these sites typically reduces transcription factor accessibility, potentially repressing *Rhob* expression. In AD, such repression may impair processes like cell-cell interaction and adhesion, both critical for maintaining synaptic integrity. As neurodegeneration progresses, this loss of function could exacerbate synaptic dysfunction.

Given the role of *Rhob* in cell signaling, structural integrity, and adhesion, its hypermethylation may reflect an epigenetic response to amyloid pathology. This modification may suppress its activity, disrupting cellular interactions and structural support crucial for countering amyloid-induced stress. Since *Rhob* supports synaptic communication and cognitive resilience, its dysregulation may contribute to the decline in synaptic integrity characteristic of AD.

Overall, the enrichment of hypermethylated genes across various pathways, including RNA splicing, and their presence in multiple neuronal cell types underscore the far-reaching impact of methylation on neuronal resilience. These epigenetic changes warrant further investigation to determine whether methylation directly contributes to neuronal dysfunction, thereby accelerating cognitive decline and disease progression.

Hypomethylation shared common pathways with the hypermethylation, however hypomethylation of promoters was exclusively associated with immune responses, cellular transport, and cytokine production. Immune-modulating genes, such as *Ptpn22*, showed hypomethylation that likely resulted in the upregulation of pro-inflammatory pathways. While this immune activation may initially serve as a protective response to protect against neuronal damage, sustained activation can lead to maladaptive chronic neuroinflammation.

Transport-related genes, including those involved in synaptic vesicle transport and neurotransmitter release, also exhibited hypomethylation, suggesting potential upregulation of synaptic activity. Genes such as *Cacna1e* and *Syt1,* which regulate calcium influx and synaptic vesicle exocytosis, respectively, may reflect compensatory

mechanisms aimed at preserving synaptic function in the face of disease-relevant stressors. However, persistent dysregulation in these pathways could result in impaired synaptic communication and plasticity, ultimately contributing to the cognitive deficits observed in AD.

Interestingly, while gene-level upregulation of transport pathways was not apparent at the bulk RNA level, isoform-level analyses revealed changes in synaptic and immune response pathways (as detailed in Chapter 2). These findings suggest that methylation changes might regulate these pathways more subtly through isoform-specific expression patterns or through discrete cell populations, highlighting complex interactions between immune activation and synaptic adaptations in the AD brain.

3.4.7 Broader Implications

The differential methylation patterns observed across both promoter and non-promoter regions reveal an influence on neuronal and immune-related pathways in response to amyloid pathology, occurring during the earliest phase of AD. These findings emphasise that the role of epigenetic changes in AD may extend beyond simple gene expression regulation.

In the next chapter, we explore the integration of methylation data with transcriptomics to uncover the mechanisms through which these modifications may shape AD pathology. This chapter emphasizes the need for validating methylation sites and integrating these data with transcriptomic analyses to elucidate the precise pathways influenced by epigenetic changes in AD. Additional experimental validation approaches, will be pivotal for directly testing the functional consequences of methylation modifications. For instance, site-specific demethylation or hypermethylation using CRISPR-dCas9 systems fused with TET1 or DNMT3A may reveal how these changes impact transcriptional activity, splicing regulation, or chromatin architecture (Gallego-Bartolomé et al., 2018; Morita et al., 2020). Chromatin immunoprecipitation (ChIP) can assess transcription factor binding at methylated loci, while reporter assays can evaluate promoter functionality (M. Lambert et al., 2018; Lassar et al., 2019; P. J. Park, 2009). Functional studies in neuronal and glial cell models derived from induced pluripotent stem cells (iPSCs) can provide a cellular

context to assess the impact of methylation changes on AD-relevant phenotypes, such as synaptic physiology, oxidative stress response, and immune activation.

In summary, this study advances our understanding of the methylation epigenetic landscape in early AD by identifying potentially critical pathways and regulatory mechanisms that may serve as potential therapeutic targets. The interplay between hypermethylation and hypomethylation highlights a complex regulatory network modulating immune responses, neuronal resilience, and synaptic function in an amyloid mouse model, representing the earliest phase of AD.

Chapter 4

4 Investigating the correlation patterns between methylation and expression in *App*$^{NL-G-F}$ and WT mice, and comparison to human AD brain

4.1 Introduction

DNA methylation is the process by which methyl groups (-CH3) are added to the cytosine residues of the DNA molecule, typically in the context of CpG dinucleotides (cytosine-guanine dinucleotides). This epigenetic modification can have profound effects on gene expression by influencing chromatin structure and the accessibility of transcriptional machinery. Thus, the correlation between gene expression and DNA methylation in Alzheimer's disease (AD) may provide understanding into the epigenetic mechanisms that influence gene regulation and may contribute to disease progression.

Numerous studies have demonstrated that the hypermethylation of CpG-rich regions, often found in gene promoters, is generally associated with transcriptional silencing. Promoter hypermethylation can lead to the recruitment of methyl-CpG-binding proteins (MBPs), which in turn recruit histone deacetylases (HDACs) and other chromatin-remodelling complexes. This results in a more compact and inaccessible chromatin structure, effectively blocking the binding of transcriptional activators and the recruitment of the RNA polymerase complex, thereby repressing gene expression. Conversely, the hypomethylation of CpG-rich promoter regions is generally associated with transcriptional activation. In the absence of DNA methylation, the promoter region becomes more accessible to transcriptional activators, allowing for the recruitment of the RNA polymerase complex and the initiation of gene transcription (Deaton & Bird, 2011; P. A. Jones, 2012). The complex dynamics of DNA methylation is critical to understand in AD, where aberrant methylation patterns correlate with altered gene expression (De Jager et al., 2014; Lardenoije et al., 2019; Lunnon et al., 2014; Marzi

et al., 2018; Roubroeks et al., 2020; Shireby et al., 2022), potentially driving disease-related changes.

Building on insights from previous analyses of gene expression, isoform-level changes, and epigenetic modifications presented in Chapters 2 and 3, this chapter examines the relationship between DNA methylation and gene expression in response to amyloid pathology in matching samples. This chapter also includes an analysis of whether the methylation patterns observed in an AD mouse model align with gene expression levels in a human snRNA-seq AD dataset, comprised of 440,000 single-nuclei RNA-sequencing (snRNA-seq) profiles from human early AD, late AD, and age-matched controls from the ROSMAP project (Xiong et al., 2023). Evaluating the human orthologues of mouse-identified genes may help identify methylation biomarkers in the early stage of the disease reflecting broader disease mechanisms and to inform future therapeutic strategies.

4.1.1 Aims and objectives

Hypothesis**:** Hypermethylation is associated with decreased gene expression, while hypomethylation correlates with increased expression, with some conserved expression patterns between mouse models of AD and human early AD compared to controls.

This chapter aims to:

1. Investigate the correlation between DNA methylation and gene expression in the *App*$^{NL\text{-}G\text{-}F}$ mouse model of AD, focusing on promoter regions and all genomic regions overlapping with DMRs to assess how methylation patterns influence gene expression.
2. Identify genotype-specific methylation-expression relationships by comparing *App*$^{NL\text{-}G\text{-}F}$ and WT genotypes, calculating differential correlations to uncover potential epigenetic regulatory differences. Examine whether shared genes with methylation-expression correlations exhibit isoform alterations, including isoform switching, alternative splicing, or transcript usage changes, integrating findings from Chapter 3.

3. Perform cell-type enrichment analysis using DMR-associated gene sets from AD mouse model to explore potential cell-specific roles of methylation patterns in human datasets.
4. Examine whether there are any genes from the mouse DNA-methylation and transcript correlation patterns that could be detected in the human AD brain, with a focus on cell-type-specific roles across disease stages using data from human snRNA-seq analyses.

4.2 Methods

4.2.1 Datasets

Long-read RNA- and DNA-sequencing data from Chapters 2 and 3 were used to compare expression and methylation patterns. Gene expression data, represented as log2-transformed normalized counts, were filtered to include approximately 6,100 shared genes of interest. Methylation data, derived from pre-processed long-read DNA sequencing, were mapped, quality-checked, and extracted for specific regions such as promoters, exons, and regulatory elements using the *bsseq* R package. Correlation analysis was conducted on both promoter regions and all regions overlapping with DMRs, focusing on shared genes to explore the relationship between methylation and gene expression across diverse functional elements.

4.2.2 Data Integration

For both promoter-specific and all-region analyses, promoters were defined as regions upstream of transcription start sites (TSS), typically spanning 1–2 kb, while all regions included promoters, gene bodies (both intronic and exonic regions), and distal intergenic regions. A GRanges object was created for each region based on genomic coordinates to facilitate methylation data extraction from the pre-processed dataset. Gene expression data from the filtered matrix were aligned with methylation data at the sample level, ensuring consistent sample IDs for reliable integration.

4.2.3 Pearson Correlation Analysis

The association between methylation levels at CpG sites and gene expression was assessed using Pearson correlation. For each CpG site in a defined region:

1. Methylation levels (percentages) were treated as independent variables (x).

2. Gene expression levels (log2-transformed normalized counts) were treated as dependent variables (y).
3. The Pearson correlation coefficient (r) was calculated as:

For each gene:

Genotype-Specific Correlation:

Pearson correlation coefficients were calculated separately for *App*$^{NL\text{-}G\text{-}F}$ and WT using the formula:

$$r = \frac{\sum(x_i - \bar{x})\,(y_i - \bar{y})}{\sqrt{\sum(x_i - \bar{x})^2 \sum(y_i - \bar{y})^2}}$$

where x_i and y_i are the methylation and expression values for sample i, and $\bar{x}$ and $\bar{y}$ are their means. A t-test was used to assess the statistical significance of the correlation coefficient.

For both promoter and all-region analyses, Pearson correlation coefficients were calculated separately for the *App*$^{NL\text{-}G\text{-}F}$ and WT genotypes:

$r_{AppNLGF}$: Correlation between methylation and expression within the AppNL-G-F genotype.

r_{WT}: Correlation between methylation and expression within the WT genotype.

Differential Correlation:

The difference in correlation ($\Delta r = r_{AppNLGF} - r_{WT}$) was calculated. A correlation test (t-test) was applied to the differences in methylation and expression values between the two genotypes.

4.2.4 snRNA-seq Analysis

Publicly available snRNA-seq data from Xiong et al. (2023), comprising ~414,000 cells from 92 individuals across different AD stages, were also used (Xiong et al., 2023). Low-quality cells were filtered based on gene count and mitochondrial expression levels, retaining cells with >750 to 8,000 gene counts, and <10% mitochondrial content (Luecken & Theis, 2019; Zheng et al., 2017).

Data were normalised to adjust for variations in sequencing depth and cell cycle effects. Normalization was performed using a log transformation with a scaling factor of 10,000 by default, which standardizes the data across cells and mitigates the impact of sequencing depth differences. Scaling was then applied to account for cell cycle effects by regressing out variations associated with cell cycle phases, using the cell cycle markers within *CellCycleScoring()* function. Principal Component Analysis (PCA) was applied to transform the gene expression data into principal components that capture the maximum variance in the dataset. Manifold Approximation and Projection (UMAP) performed for the low-dimensional representation of the data, and visualization and interpretation of cellular relationships. DoubletFinder was used to identify and remove potential doublets, with an estimated doublet rate of 4%.

4.2.5 Cell Type Enrichment Analysis

Expression Weighted Cell-type Enrichment (*EWCE*) was applied to determine the cell type enrichment within the gene sets of interest, which compares input gene lists against a background reference of brain cell types from both mouse and human datasets (Skene & Grant, 2016). Input gene lists were cross-referenced with orthologue conversion lists, retaining one-to-one orthologues for clear interpretation. The reference dataset included scRNA-seq data annotated for specific cell types in human AD (Xiong et al., 2023) and mouse brain (Zeisel et al., 2015). *EWCE*-based cell-type enrichment analysis was conducted using gene names derived from DNA methylation peak annotations.

While both *EWCE* and *rGREAT* were used to analyze DMR-associated gene sets, their differing methodologies result in expected differences. Unlike *rGREAT*, which extends associations to nearby regulatory elements, *EWCE* directly links genes to peaks, creating a more focused gene list. This targeted approach is especially advantageous for cell-type enrichment, as it can identify cell type-specific expression patterns with greater power. However, *EWCE*'s reliance on direct annotations excludes many of the indirect regulatory associations captured by *rGREAT*, resulting in a more refined view centred on gene products directly associated with genomic peaks. This distinction allows for a complementary analysis, where *EWCE* offers focussed insights into cell type-specific associations, while *rGREAT* captures a broader regulatory context surrounding each DMR.

The CellTypeDataset (CTD) summarizes gene expression for each cell type in the reference scRNA-seq dataset. Average expression levels were calculated for each gene within each cell type, alongside specificity scores that highlight genes predominantly expressed in specific cell types. Standardization was applied, aligning gene names to a consistent format and normalizing expression values across the dataset. These adjustments make the CTD suitable for downstream analyses.

EWCE's bootstrap_enrichment_test function to evaluate whether the input gene set was significantly enriched in specific cell types. This test compares the observed expression of the input gene list to randomly generated gene lists of the same size, sampled from the background gene set. The enrichment analysis was performed across multiple annotation levels in the ROSMAP dataset (Xiong et al., 2023) with level 1 (major cell types) and level 2 (cell subtypes) cell type annotations available. The bootstrapping process involved 10,000 repetitions as suggested (Skene & Grant, 2016).

4.2.6 Gene expression and methylation data integration and correlation

For each DMR mapped to genes/loci, the corresponding gene expression data were examined to determine the relationships between methylation and expression levels. Gene symbols from the DMR data were matched to those in the gene expression dataset, which was extracted from log-normalized counts from that used in differential expression analysis. Methylation data was processed using the *bsseq* package (Hansen et al., 2012) as described in Chapter 3.

The initial analysis focused on promoter regions due to their known regulatory impact on gene expression. To extend the analysis, methylation patterns in gene bodies and intergenic regions were also evaluated for associations with gene expression. Gene names were consistently referenced when comparing the isoform-level changes, such as isoform usage and switch.

Pearson correlation tests were conducted for both genotypes ($App^{NL\text{-}G\text{-}F}$ and WT) to assess the correlation between differential methylation and gene expression levels, and to uncover any genotype-specific patterns. Methylation data was summarized both as mean methylation percentages per sample genome-wide, and per-CpG basis only for some genes, resulting in two separate analyses:

1. Averaged methylation per region
2. Per-CpG site methylation

4.2.7 Code Availability

The R markdown file used for the snRNA-seq analysis and other relevant data processing scripts is available at: https://github.com/umranyaman/brain_AD.

4.2.8 Data Availability

The public snRNA-sequencing dataset used for analysis (Xiong et al., 2023) can be accessed from Kellis Lab' website at https://compbio.mit.edu/ad_epigenome/.

4.3 Results

4.3.1 Correlation between promoter methylation and gene expression

Promoter regions overlapping with DMRs were analyzed to explore their relationship with gene expression, recognizing that DMRs may span diverse functional elements. The study focused on around 6,100 genes shared between the methylated DNA loci and the filtered expression matrix used for differential gene expression analysis. Instead of genome-wide, correlation analysis was performed on set of shared genes. Promoter methylation levels of these genes that overlapped with gene expression data were examined to assess how methylation patterns in these regions correlate with gene expression.

Overall, there was a discernible trend suggesting a relationship between methylation and expression. However, these correlations were generally not statistically significant when analyzed within the *App*$^{NL\text{-}G\text{-}F}$ and WT groups individually (Figure 4.1). For example, the promoter region of the App gene displayed a noticeable correlation between methylation and expression, yet differences in methylation and expression levels were not statistically significant. This indicates a potential association that reflects trends rather than robust and consistent patterns.

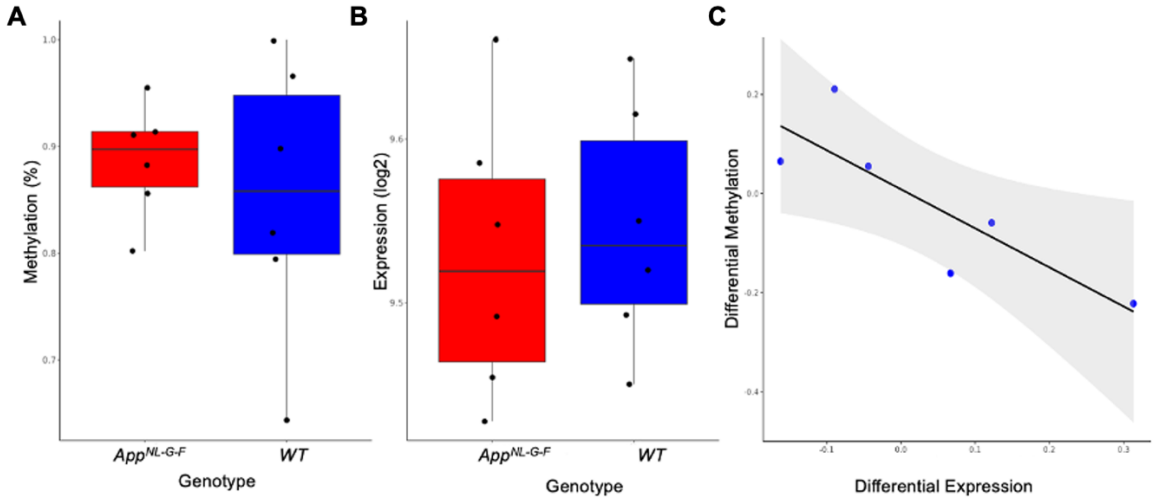

Figure 4-1 Differential methylation of the *App* promoter and expression correlation in *App*$^{NL-G-F}$ and WT mice in each pool

(A)Mean methylation levels of *App* promoter region in each samples per genotype (n=6). (B) Gene expression levels (log2) of *App* gene each samples per genotype (n=6). (C) The *App* promoter showed a significant correlation between differential expression and methylation values (Pearson correlation = -0.842, p = 0.035, t-test). Despite these strong correlations, no significant differential methylation was detected in the *App* promoter region overall (pane A), or differential expression of *App* between *App*$^{NL-G-F}$ and WT (p=0.7, Wald test, panel B).

The findings showed that some genes had correlations between methylation levels and gene expression in specific regions or at certain CpG sites (Table 4.3; full list in Supplementary Table 4.1). In several cases, strong positive correlations suggested potential biological importance, given the affected genes were linked to changes in expression or methylation associated with AD pathology.

4.3.2 Individual CpG sites may show stronger genotype-specific correlations with gene expression beyond DMRs

The promoter region of capping actin protein *(Capg)* gene, one of the amyloid - associated microglial (ARM/DAM) genes, and shown to have differential expression and isoform usage/switching in Chapter 2, was also identified within differentially methylated genes in Chapter 3 (areaStat: -8.757, Wald test, see Supplementary Table 3.1. and 4.1). Hypomethylation in the promoter region (within ±1 kb of the TSS, distance to TSS = 236 bp) of *Capg* displayed no overall correlation to expression (Figure 4.2). However, CpG site-specific analysis of promoter methylation against

expression changes highlighted a single CpG site correlated to expression (Table 4.1.).

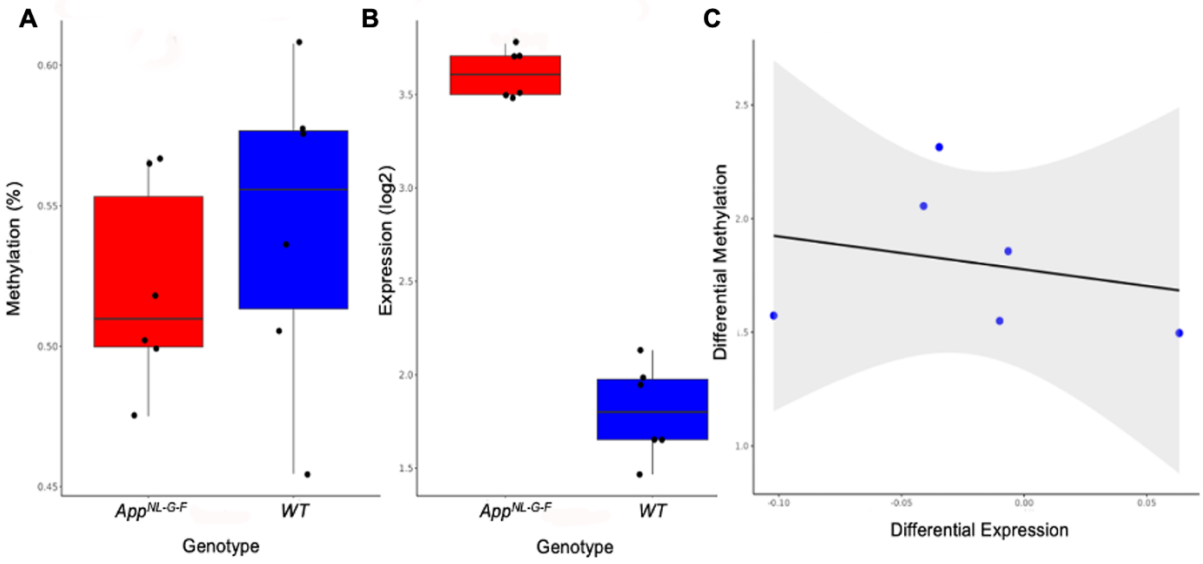

Figure 4-2 Methylation and expression comparisons for the *Capg* gene between *App*$^{NL-G-F}$ and WT genotypes.

Panel A shows a modest, however, significant change in the promoter methylation levels (areaStat= -8.757, Wald test, hypomethylation) between *App*$^{NL-G-F}$ and WT mice. Panel B shows significantly elevated expression in the *App*$^{NL-G-F}$ genotype compared to WT (p = 5e-07, Wald test). Panel C illustrates the no significant association (r = -0.239, p = 0.648) between methylation and expression changes per sample for *Capg*.

Although the overall DMR for *Capg* (Figure 4.2) showed limited correlation with gene expression, specific CpG sites revealed more significant relationships. For example, CpG Site 3 in the *App*$^{NL-G-F}$ group showed a positive correlation with gene expression (r = 0.814, p = 0.048; Table 4.1). This suggests that while genotype differences in promoter methylation and expression were not substantial, specific CpG sites (Figure 4.4 and Figure 4.5) might better predict expression levels within the same sample set.

Table 4-1 Summary of methylation and expression correlations at CpG sites of the *Capg* gene in $App^{NL\text{-}G\text{-}F}$ and WT genotypes, including differential correlations.

Site	*Pearson correlation $App^{NL\text{-}G\text{-}F}$*	*P val. $App^{NL\text{-}G\text{-}F}$*	*Pearson correlation WT*	*P val. WT*	*Pearson Corelation Diff* [2]	*P val. Diff*
Aggregated[1]	-0.225	0.667	-0.031	0.95	-0.239	0.648
CpG Site 1	0.348	0.498	0.375	0.462	0.574	0.233
CpG Site 2	0.564	0.242	-0.133	0.800	0.318	0.537
CpG Site 3	**0.814**	**0.048**	-0.139	0.791	0.382	0.454
CpG Site 4	-0.374	0.464	-0.150	0.775	-0.163	0.757

[1]The correlation between methylation and expression levels across all CpG sites combined (Pearson correlation=0.814, p.value=0.048, t-test)

[2]The correlation between methylation differences and expression differences in matching samples of $App^{NL\text{-}G\text{-}F}$ and WT groups, per pool (n=6) where each pool is one $App^{NL\text{-}G\text{-}F}$ and one WT mouse sequenced simultaneously.

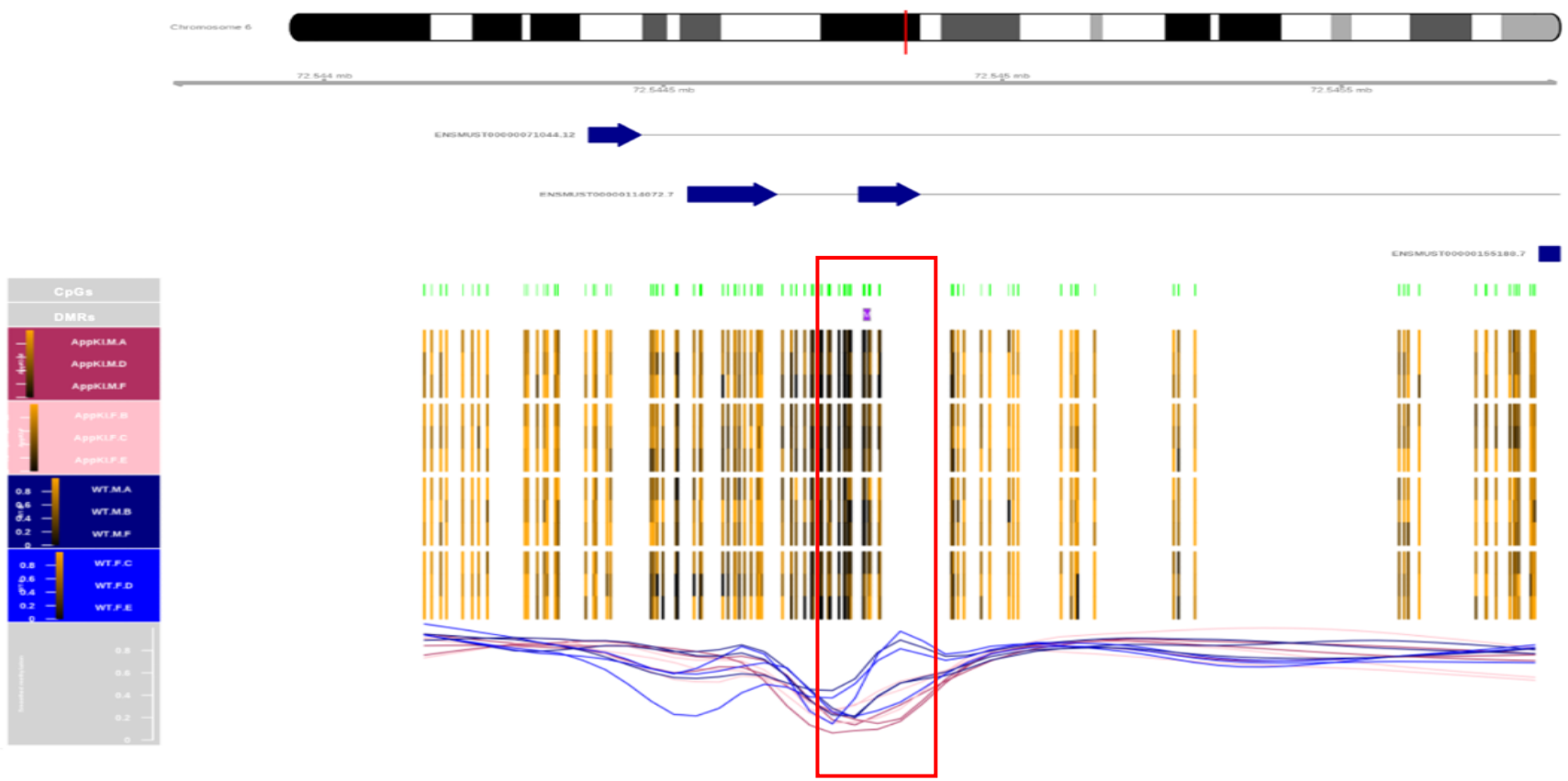

Figure 4-3 The differential hypomethylation level in the *Capg* promoter across *App*$^{NL\text{-}G\text{-}F}$ and WT groups.

Each row represents a different sample (pink and blue annotations, marked Ap or WT, then sex M/F, then the pool number A-F), with methylation levels indicated in the heatmap by colour intensity—lighter yellow shades represent higher methylation in WT (hypomethylation). The *Capg* promoter DMR, indicated by purple in the "DMRs" row, highlights areas of hypomethylation. The lines at the bottom of the figure illustrate the variation in methylation levels across samples, capturing the epigenetic differences between the genotype groups.

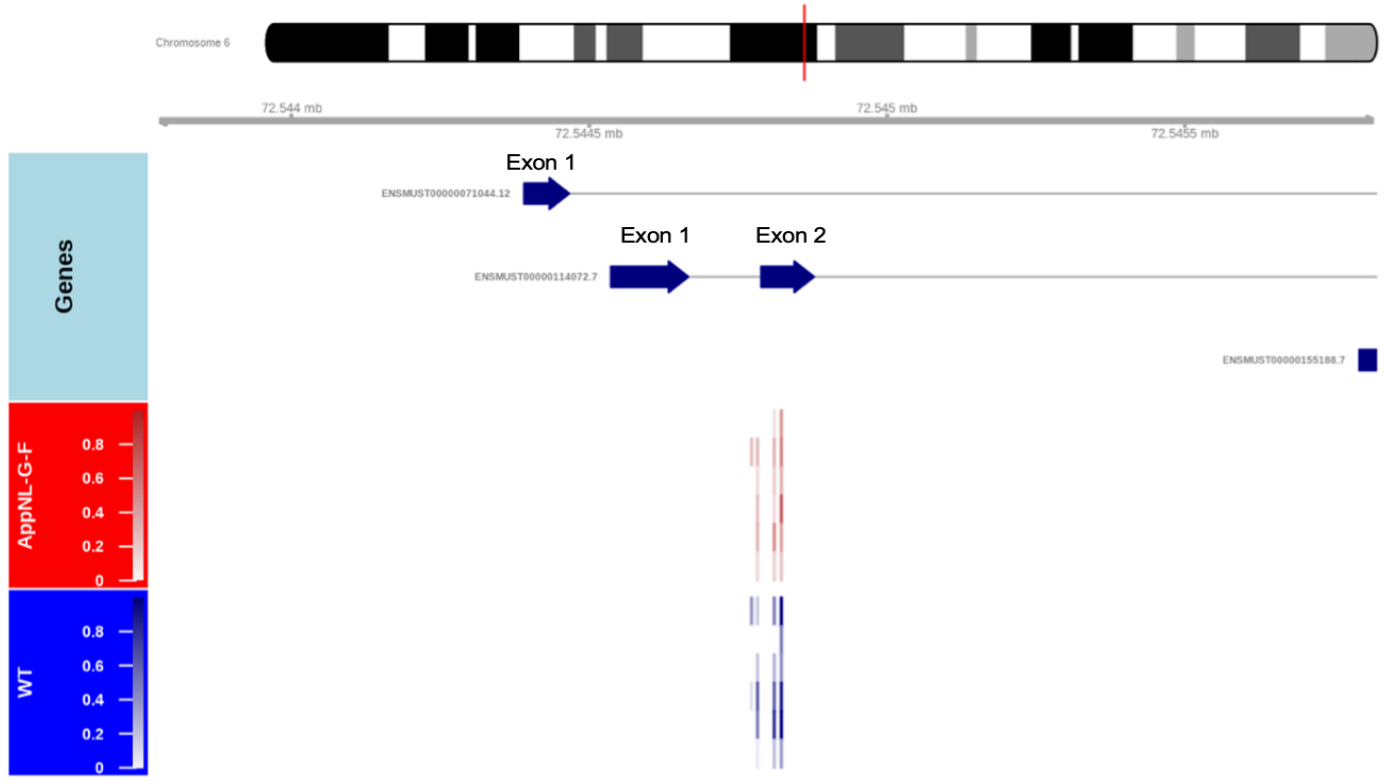

Figure 4-4 Hypomethylation in CpG sites in the *Capg* promoter region.

The figure shows DMR region for *Capg* promoter only (areaStat=-8.756, Wald test, hypomethylation). The top track shows *Capg* transcripts, with two major isoforms indicated by blue arrows. Below, heatmaps display methylation levels at four CpG sites for the $App^{NL\text{-}G\text{-}F}$ and WT groups. The $App^{NL\text{-}G\text{-}F}$ heatmap (red gradient) shows lower methylation per CpG site, with darker red indicating higher levels. The WT heatmap (blue gradient) shows higher WT methylation, with darker blue representing higher levels. Differences in methylation patterns between the groups are evident at multiple CpG sites.

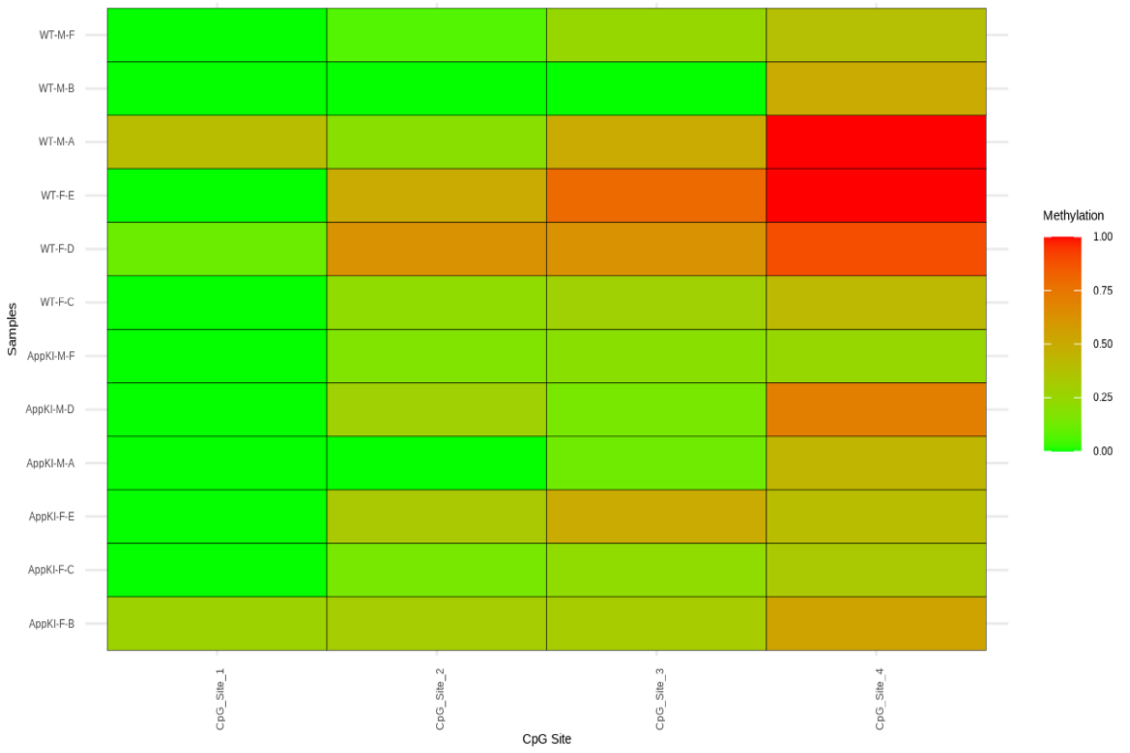

Figure 4-5 Heatmap showing methylation levels of four CpG sites within the *Capg* promoter region across samples.

Methylation levels range from low (green) to high (red), as indicated by the colour gradient on the right. Each row represents a biological sample from either the WT (wild-type) or *App*$^{NL\text{-}G\text{-}F}$ group, and each column corresponds to a CpG site (CpG_Site_1, CpG_Site_2, CpG_Site_3, CpG_Site_4). The heatmap visualizes differences in methylation patterns between genotypes, with notable hypermethylation observed in specific CpG sites in the *App*$^{NL\text{-}G\text{-}F}$ samples compared to WT. M/F, male/female; A-F, pool ID.

The colony-stimulating factor 1 *(Csf1)* shows a trend of a correlation between hypomethylation in a distal intergenic area and gene expression, where methylation at CpG site 1 tended to positively correlate with the increasing expression. Although the correlations are not statistically significant, the observed patterns may indicate that differential methylation of specific CpG sites may affect *Csf1* expression based on pathology and sample group.

Table 4-2 Summary of methylation and expression correlations at CpG sites of the *Csf1* gene in *App*$^{NL\text{-}G\text{-}F}$ and WT genotypes, including differential correlations.

Site	Pearson correlation App^{NL-G-F}	P val. App^{NL-G-F}	Pearson correlation WT	P val. WT	Pearson Corelation Diff[2]	P val. Diff
Aggregated[1]	-0.646	0.165	0.569	0.238	0.433	0.390
CpG Site 1	-0.431	0.393	0.622	0.186	**0.751**	**0.084**
CpG Site 2	-0.357	0.486	0.547	0.260	0.127	0.809
CpG Site 3	-0.207	0.693	0.499	0.313	0.694	0.126
CpG Site 4	-0.545	0.262	0.524	0.285	0.098	0.852
CpG Site 5	-0.660	0.153	0.448	0.372	-0.653	0.159

[1]The correlation between methylation and expression levels across all CpG sites combined.

[2]The correlation between methylation differences and expression differences in matching samples of App^{NL-G-F} and WT groups, per pool (n=6) where each pool is one App^{NL-G-F} and one WT mouse sequenced simultaneously.

Table 4-3 Correlation between promoter methylation and gene expression in genes of *App*$^{NL\text{-}G\text{-}F}$ and WT models

gene	*App*$^{NL\text{-}G\text{-}F}$ *correlation*	*App*$^{NL\text{-}G\text{-}F}$ *pvalue*	*WT correlation*	*WT pvalue*	*Diff. correlation*[1]	*Diff. pvalue*	*Hypo Increased Expression*[2]	*Hyper Decreased Expression*[3]	*Gene Exp*[4]
Irf9	-0.190	0.719	-0.349	0.497	-0.913	0.011	FALSE	FALSE	Yes
Ugt1a7c	-0.672	0.144	0.371	0.469	0.898	0.015	TRUE	FALSE	Yes
Gcnt1	-0.831	0.040	-0.524	0.286	-0.877	0.022	FALSE	FALSE	Yes
Fcer1g	0.693	0.127	0.458	0.361	0.858	0.029	TRUE	FALSE	Yes
Unc93b1	0.383	0.453	0.489	0.325	0.848	0.033	TRUE	FALSE	Yes
Tmem176b	-0.364	0.479	-0.405	0.425	-0.810	0.051	FALSE	FALSE	Yes
Cd180	-0.106	0.842	-0.251	0.632	-0.807	0.052	FALSE	FALSE	Yes
Gfap	0.067	0.899	-0.263	0.615	-0.806	0.053	TRUE	FALSE	Yes
Igf1	-0.736	0.096	-0.529	0.280	-0.803	0.054	TRUE	FALSE	Yes
Irf5	0.026	0.961	-0.747	0.088	-0.761	0.079	TRUE	FALSE	Yes
Cd300c2	0.035	0.947	0.330	0.523	0.754	0.083	TRUE	FALSE	Yes
Il4i1	-0.109	0.837	-0.112	0.832	-0.744	0.090	TRUE	FALSE	Yes

[1]The correlation between methylation and expression differences in matching samples of *App*$^{NL\text{-}G\text{-}F}$ and WT groups, per pool (n=6 pools)
[2]Indicates whether hypomethylation is associated with increased gene expression.
[3]Indicates whether hypermethylation is associated with decreased gene expression.
[4]"Yes" indicates observed increased gene expression in *App*$^{NL\text{-}G\text{-}F}$ (Figure 2.3, see Supplementary Table 2.1.)

4.3.3 DMRs in gene body and intergenic regions correlate with gene expression

To investigate the potential influence of DMRs located within gene bodies and intergenic regions on gene expression, correlations between differential methylation and gene expression were systematically explored. Pearson correlation coefficients were calculated for these annotated DMRs to evaluate the relationship between methylation and gene expression per pool (n=6 pools), with pools containing individual samples that were matched for both long-read DNA and RNA-sequencing (methylation called from DNA-sequencing, with one *App*$^{NL\text{-}G\text{-}F}$ and one WT mouse per pool). As outlined in the "Diff. correlation" column of Table 4.4, this column shows how the difference between *App*$^{NL\text{-}G\text{-}F}$ and WT samples within pools, correlates between one modality to the other across all the samples. It was of particular interest to determine whether methylation alterations outside promoter regions could significantly modulate transcriptional activity.

Table 4.4 details the correlation values, p-values, and the relationships between methylation status of gene body, intergenic regions and their correlation to the gene expression in both the *App*$^{NL\text{-}G\text{-}F}$ and WT mouse models. The results indicate that DMRs within non-promoter regions did indeed correlate with expression changes, suggesting that regulatory mechanisms may extend beyond promoter-associated methylation.

Hypomethylation in the intronic region of cadherin-13 *(Cdh13)* was linked to a significant negative differential correlation between expression and methylation. In the *App*$^{NL\text{-}G\text{-}F}$, *Cdh13* displayed a correlation of differential expression and methylation of -0.933 ($p < 0.01$) (Supplementary Table 4.2), suggesting that hypomethylation within the gene body could be associated with increased gene expression. *Cdh13* is notably abundant in cell types including excitatory, inhibitory neurons and oligodendrocyte progenitor cells (Mossink et al., 2022; Xiong et al., 2023). In human AD brain, *CDH13* is found to be increased in early AD and decreased as the disease progresses to the late stage in excitatory cell types (Figure 4.6) in single-cell ROS-MAP data, suggesting that elevated *CDH13* may initially enhance synaptic stability and plasticity, potentially serving a protective role. *CDH13* supports the formation and maintenance of excitatory synapses, crucial for cognitive functions such as learning and memory (Saunders et

al., 2022). This increased expression in early AD may counteract synaptic loss and help preserve cognitive function, as synaptic dysfunction is an early pathological feature of AD.

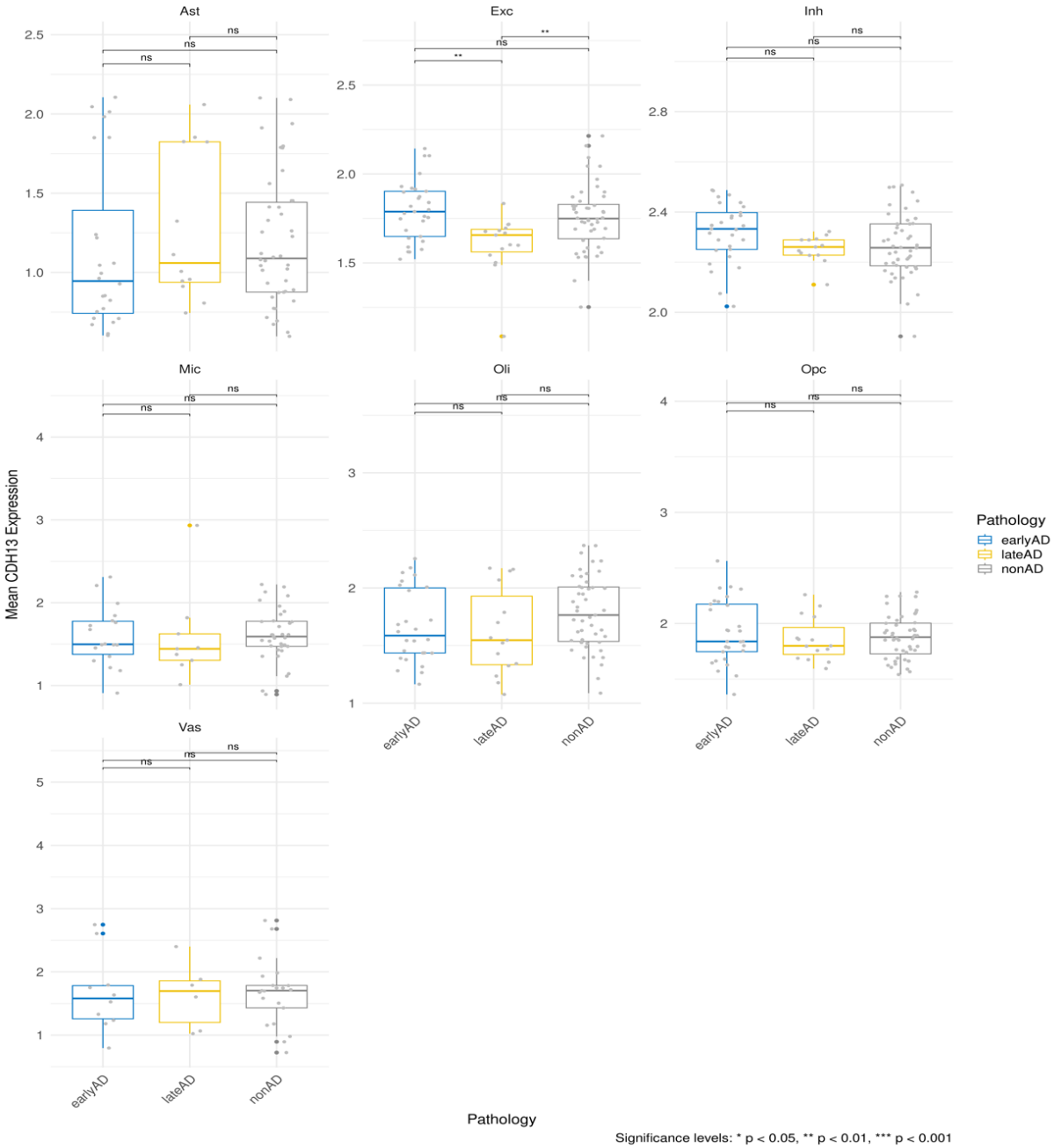

Figure 4-6 *CDH13* expression across human cell types and AD stages

Boxplots display the mean expression levels of *CDH13* across different cell types (Ast, Exc, Inh, Mic, Oli, Opc, Vas) for early AD (n=29 individuals), late AD (n=15), and non-AD stages (n=48), with individual data points representing individuals re-analyzed from the ROSMAP dataset (Xiong et al., 2023). The boxplots depict the interquartile

range (IQR; 25th–75th percentiles), with the central line representing the median expression level, while whiskers extend to the minimum and maximum values within 1.5×IQR. Statistically significant differences between pathology stages, assessed by pairwise Wilcoxon tests with FDR correction, are marked by asterisks (* $p<0.05$, ** $p<0.01$, *** $p<0.001$), while non-significant comparisons are labelled as "ns." *CDH13* expression increases in excitatory cells in the earlier stage, followed by decreased levels in the late stage of AD.

A high positive differential correlation was observed for protein tyrosine phosphatase, receptor type E *(Ptpre)*(0.979, $p < 0.001$), accompanied by a significant increase in the expression of the specific isoform *ENSMUST00000209979 (Ptpre-204)*, as detailed in Chapter 2, in response to amyloid pathology. This correlation, while reflecting overall gene expression, strongly suggests that hypermethylation within intronic region of the *Ptpre* gene may preferentially drive the expression of the *Ptpre-204* isoform, notable for being the only isoform to show increased expression in the $App^{NL\text{-}G\text{-}F}$ model. Interestingly, in the human AD dataset, *PTPRE* expression decreases as the disease progresses (Figure 4.7). Given that *PTPRE* has been implicated in the regulation of angiogenesis and vascular health — both crucial for maintaining blood-brain barrier (BBB) integrity (Cai et al., 2024) — pointing to altered microglia-endothelial cell interactions across AD stages. The observed increased expression of *Ptpre-204* in the $App^{NL\text{-}G\text{-}F}$ model may signify an adaptive or protective response during early AD, aiming to sustain BBB function and vascular stability, which may then breakdown with Tau pathology in humans (which is not seen in the $App^{NL\text{-}G\text{-}F}$ mice).

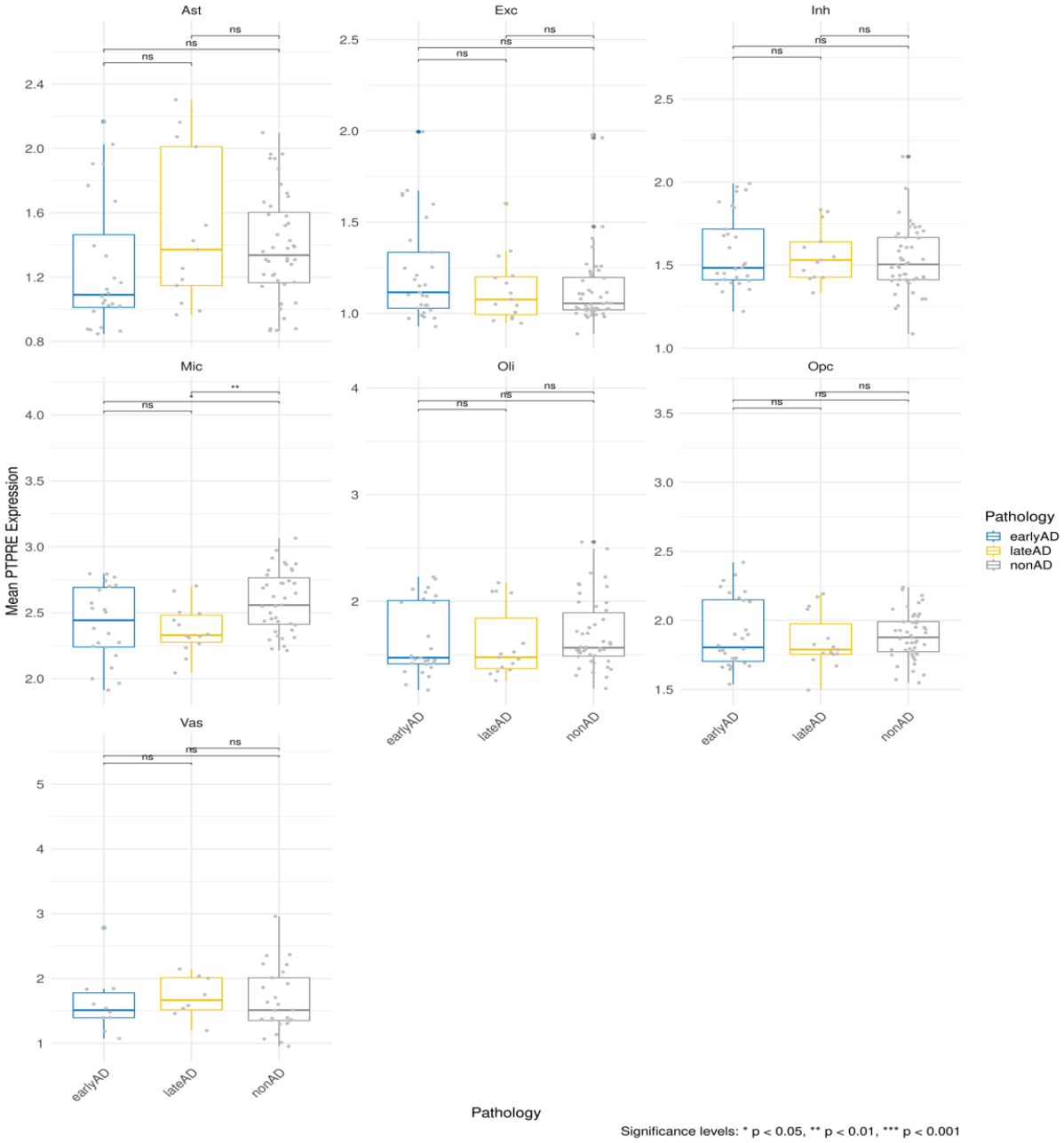

Figure 4-7 *PTPRE* expression across human cell types and AD stages

Boxplots display the mean expression levels of *PTPRE* across different cell types (Ast, Exc, Inh, Mic, Oli, Opc, Vas) for early AD (n=29 individuals), late AD (n=15), and non-AD stages (n=48), with individual data points representing individuals. The boxplots depict the interquartile range (IQR; 25th–75th percentiles), with the central line representing the median expression level, while whiskers extend to the minimum and maximum values within 1.5×IQR. Statistically significant differences between pathology stages, assessed by pairwise Wilcoxon tests with FDR correction, are marked by asterisks (* p<0.05, ** p<0.01, *** p<0.001), while non-significant comparisons are labelled as "ns." Statistically significant differences between

pathology stages, assessed by pairwise Wilcoxon tests with FDR correction, are marked by asterisks (* $p<0.05$,** $p<0.01$,*** $p<0.001$), while non-significant comparisons are labelled as "ns". *PTPRE* expression decreases in the microglia as the disease progresses.

These examples illustrate that in specific situations DMRs within gene bodies and intergenic regions can correlate with gene expression changes, reinforcing the hypothesis that methylation outside of promoter regions may also play an essential role in transcriptional regulation. This could influence isoform-specific responses and overall gene expression dynamics in the context of amyloid pathology.

Table 4-4 Summary of differentially methylated regions (DMRs) correlating with gene expression in *App*$^{NL-G-F}$ and WT mouse models.

gene	*App*$^{NL-G-F}$ *correlation*	*App*$^{NL-G-F}$ *pvalue*	*WT correlation*	*WT pvalue*	*Diff. correlation*[1]	*Diff. pvalue*	*Hypo Increased Expression*[2]	*Hyper Decreased Expression*[3]	*Gene Exp*[4]
Ptpre	0.222	0.672	0.448	0.373	0.979	0.001	FALSE	FALSE	No
Sik3	0.593	0.215	0.538	0.271	0.956	0.003	FALSE	FALSE	No
Stard5	-0.558	0.250	-0.262	0.616	-0.956	0.003	FALSE	FALSE	No
Cdh13	-0.800	0.056	-0.776	0.069	-0.933	0.007	TRUE	FALSE	No
Runx1	-0.051	0.924	-0.767	0.075	-0.929	0.007	FALSE	FALSE	Yes
Sipa1l3	0.670	0.146	0.399	0.434	0.891	0.017	FALSE	FALSE	No
Lmo7	0.772	0.072	0.341	0.508	0.884	0.020	FALSE	TRUE	No
Arap2	0.397	0.436	0.656	0.157	0.868	0.025	FALSE	FALSE	No
Pmp22	0.259	0.621	-0.369	0.472	0.816	0.048	FALSE	FALSE	No
Zfp804b	-0.166	0.753	0.166	0.790	0.879	0.050	FALSE	TRUE	No
Kit	0.575	0.233	0.672	0.144	0.778	0.068	FALSE	TRUE	No
Tec	0.109	0.837	0.837	0.037	0.760	0.080	TRUE	FALSE	Yes

[1]The correlation between methylation and expression differences in matching samples of *App*$^{NL-G-F}$ and WT groups, per pool (n=6 pools)
[2]Indicates whether hypomethylation is associated with increased gene expression.
[3]Indicates whether hypermethylation is associated with decreased gene expression.
[4]"Yes" indicates observed increased gene expression in *App*$^{NL-G-F}$ (Figure 2.3, see Supplementary Table 2.1.)

4.3.4 DMRs and AD risk genes

A gene set overlap of DMRs within known AD risk genes was conducted to explore potential epigenetic mechanisms influencing gene expression in AD pathology. Bellenguez et al. (2022) identified 89 genetic loci associated with AD, many of which are involved in processes such as lipid metabolism, immune response, and synaptic function (Bellenguez et al., 2022). Gene names associated with differentially DMRs were compared to AD risk genes identified through genome-wide association studies (GWAS), highlighting DMR sites within genes common to both datasets (Table 4.5.). The DMR analysis revealed hypomethylation or hypermethylation in many AD causal and risk genes, such as *App, Mapt,* epidermal growth factor receptor *(Egfr),* signal peptide peptidase-Like 2A *(Sppl2a),* ubiquitin associated and SH3 domain containing B *(Umad1),* and membrane metalloendopeptidase *(Mme)* across the individual CpG sites or the DMR in response to amyloid.

In *App,* both hypomethylation in distal intergenic region (areaStat=12.070) and hypermethylation in an intronic region (areaStat=-9.961) were observed. Hypermethylation in the *App* intronic region showed a weak non-significant trend of negative correlation with gene expression (Pearson correlation = -0.649, p = 0.162). A hypomethylated distal intergenic region exhibited a trend towards a negative correlation with *App* expression (Pearson correlation = -0.751, p = 0.085), suggesting that hypomethylation might contribute to an increase in *App* expression. A specific hypomethylated CpG site demonstrated the strongest correlation exclusively within the $App^{N\text{-}L\text{-}GF}$ group (Pearson correlation = -0.808, p = 0.051), highlighting the potential for non-promoter regions to exert regulatory control over the gene activity.

Egfr showed a significant positive correlation at CpG site 4 in the $App^{NL\text{-}G\text{-}F}$ group (Pearson correlation = 0.835, p = 0.039), suggesting that hypermethylation may be linked to increased expression in this group. Conversely, CpG site 1 in the WT group exhibited a strong positive correlation (Pearson correlation = 0.848, p = 0.033), implying that methylation at specific CpG sites in the interaction with *APP* and Aß signalling may regulate *Egfr* expression differently between the groups.

In *Sppl2a*, hypomethylation in exonic regions showed a strong negative overall correlation with gene expression in the WT group (Pearson correlation = -0.917, p =

0.010), indicating a potential inverse relationship between methylation and expression. Although CpG site-specific correlations in the *App*$^{NL\text{-}G\text{-}F}$ group were weaker and not statistically significant, these results suggest methylation may regulate *Sppl2a* expression, particularly in the WT group.

For *Umad1*, hypomethylation in the intronic region was associated with a significant positive overall correlation with gene expression across all samples (Pearson correlation = 0.659, p = 0.020). Although individual CpG site correlations within the *App*$^{NL\text{-}G\text{-}F}$ and WT groups were moderate, they did not reach statistical significance, highlighting the need for further investigation.

Finally, *Mme* showed a weak trend of a positive correlation between hypomethylation in the distal intergenic region and gene expression, particularly in the WT group at CpG site 2 (Pearson correlation = -0.719, p = 0.107). Although these correlations were not statistically significant, they suggest further methylation testing at this region should be investigated in relation to gene expression.

Interestingly, *Mapt* exhibited hypermethylation in an intronic region in both the AD mouse model (areaStat = 10.589) and human AD (p = 3.16e-07), with the latter reported by Wang et al. (2023). In both cases, the hypermethylation was associated with amyloid pathology burden.

Overall, these observations underscore the complexity of methylation regulation in AD-related genes, where both hypo- and hypermethylation in various genomic regions showed context-specific correlations with altered gene expression. The methylation patterns in genes *App*, *Egfr*, *Sppl2a*, *Umad1*, and *Mme* suggest that methylation may exert context-dependent regulatory effects influenced by the response to amyloid (*App*$^{NL\text{-}G\text{-}F}$ vs WT), CpG site location, and gene function.

Table 4-5 Differentially methylated regions in genes common with AD risk genes identified by GWAS

chr	*start*	*end*	*length*[1]	*nCG*[1]	*areaStat*[2]	*annotation*	*Distance ToTSS*[3]	*symbol*	*gene name*
chr16	84965424	84965476	53	4	12.070	Intron	-2043	*App*	*Amyloid beta precursor protein*
chr16	85269999	85270065	67	5	-9.962	Distal Intergenic	-96233	*App*	*Amyloid beta precursor protein*
chr14	45455942	45456152	211	5	11.262	Distal Intergenic	21709	*Fermt2*	*Fermitin family member 2*
chr11	16736261	16736435	175	5	10.307	Distal Intergenic	-15768	*Egfr*	*Epidermal growth factor recepto r*
chr11	16655442	16655553	112	4	8.848	Distal Intergenic	-96650	*Egfr*	*Epidermal growth factor recepto r*
chr7	131000000	131000000	228	4	8.558	Intron	5819	*Plekha1*	*Pleckstrin homology domain co ntaining, family A (phosphoinosi tide binding specific) member 1*
chr8	121000000	121000000	250	6	-13.322	Intron	58885	*Fendrr*	*Foxf1 adjacent non-coding developmental regulator y RNA*
chr3	51512071	51512304	234	5	-11.645	Distal Intergenic	24343	*Setd7*	*SET domain containing (lysine methyltransferase) 7*
chr2	127000000	127000000	226	5	-11.059	Exon	4495	*Sppl2a*	*Signal peptide peptidase like 2A*
chr17	48689345	48689471	127	4	-11.053	Distal Intergenic	226941	*Unc5cl*	*Unc-5 family C-terminal like*
chr6	8315739	8315853	115	4	-9.780	Intron	45273	*Umad1*	*UMAP1-MVP12 associated (UMA) dom ain containing 1*
chr3	63123237	63123537	301	4	-9.588	Distal Intergenic	-118000	*Mme*	*Membrane metallo endopeptida se*

[1]Specifies the size of each region in base pairs (bp) and the number of CpG sites contained within it.

[2]Reflects the degree of methylation changes within each region.

[3]Distance to TSS - Provides the genomic context (e.g., intronic, intergenic) and distance to the closest transcription start site (TSS). Gene set comparison between isoform switch and usage between *App*$^{NL-G-F}$ AD mouse model and human AD brain in base pairs

4.3.5 DMRs across human AD datasets

To explore the conservation of DMRs identified in the mouse model further, comparisons were made with three human AD datasets. Each dataset highlights DNA methylation changes in brain regions linked to amyloid pathology. The first dataset was from the middle temporal gyrus (MTG), a region highly vulnerable to amyloid plaque accumulation (PIP) (Piras et al., 2023). The second dataset was obtained from parahippocampal gyrus (PHG), an area closely associated with neuritic amyloid plaques and early cognitive dysfunction (WAP) (Wang et al., 2023). Finally, the ROSMAP dataset analyzed methylation changes in the prefrontal cortex (PFC), a region that is critically affected in the later stages of AD and associated with the burden of neuritic amyloid plaques (De Jager et al., 2014).

Comparisons between mouse DMR genes and these human datasets revealed varying degrees of overlap. In the ROSMAP dataset, five shared genes were identified, including striated muscle preferentially expressed protein kinase *(SPEG),* a regulator of cell signaling, and forkhead box protein K1 *(FOXK1),* a transcription factor involved in chromatin remodeling and cellular differentiation.

The WAP dataset showed a broader overlap with 39 genes. These included mitotic arrest deficient 1-like 1 *(MAD1L1),* a key regulator of mitotic checkpoints and chromosome segregation, and glutamate ionotropic receptor kainate type subunit 2 *(GRIK2),* which is involved in excitatory synaptic transmission. Other notable genes were transmembrane protein 2 *(TSPAN2),* which mediates cell adhesion and signalling, and protein phosphatase 2 regulatory subunit B gamma *(PPP2R2C),* a regulator of cell cycle progression.

The PIP dataset had the largest overlap, with 67 genes shared between mouse DMRs and human results. Among these were rho guanine nucleotide exchange factor 3 (*ARHGEF3),* a key player in actin cytoskeleton regulation, *FOXP1,* which is essential for neurodevelopment and synaptic maintenance. Other conserved genes included calcium/calmodulin-dependent protein kinase 1D *(CAMK1D),* which modulates calcium signalling pathways, and *MAD1L1,* underscoring its repeated relevance across datasets.

Notably, genes such as *MAD1L1, ARHGEF3,* and *GRIK2* were consistently identified across multiple human AD DNA methylation datasets, suggesting their strong association with amyloid-related pathology. A Venn diagram below shows the shared and unique overlaps between mouse DMRs and the three human datasets.

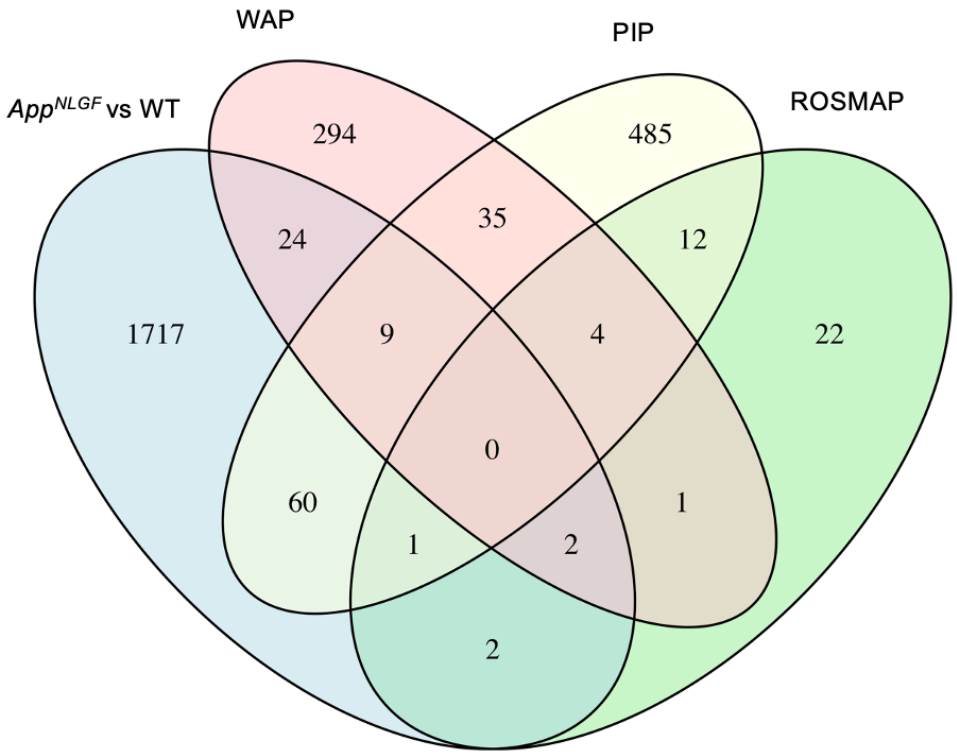

Figure 4-8 Conservation of DMRs between the *AppNL-G-F* mouse model and human AD datasets (WAP, PIP, ROSMAP)

The Venn diagram shows gene overlaps between the DMRs detected in $App^{NL\text{-}G\text{-}F}$ vs WT, and three human datasets: WAP (parahippocampal gyrus), PIP (middle temporal gyrus), and ROSMAP (prefrontal cortex). Numbers represent shared or unique genes, with 5 genes conserved across mouse and ROSMAP, 35 genes shared between mouse and WAP, and 70 genes shared between mouse and PIP datasets.

4.3.6 Cell-type specific DNA methylation dynamics in AD mouse model, and across human AD brain stages

In Chapter 3, cell type specificity of genes affected by hypermethylation and hypomethylation was examined using a mouse reference dataset. The analysis initially focused on promoter regions (-2000 to +500 bp relative to the TSS), given their central role in gene regulation. Although no significant associations were detected overall at the promoter level (Figure 3.11.), some of hypomethylated genes were expressed in

microglial cells in the human AD brain, consistent with the microglial activation described in Chapter 2. When the analysis was expanded to include all DMRs, both hypermethylated and hypomethylated genes were enriched in hippocampal CA1 pyramidal neurons, somatosensory Layer 2/3 pyramidal neurons, and CA1 interneurons (adjusted p-value<0.05), highlighting the contribution of methylation changes beyond promoters (Figure 3.12).

In Chapter 4, the focus shifted to human brain expression data, covering non-AD, early AD, and late AD stages. Using a human reference dataset captures the complexity of AD progression and addresses species-specific differences that limit the mouse model. This approach builds on the earlier findings and places them in a clinical context, providing a clearer view of how DNA methylation changes relate to cell type-specific processes at different stages of AD.

Although statistically insignificant, promoters of genes marked by hypomethylated regions exhibited a tendency for enrichment in early AD within endothelial cells (aEndo), excitatory neurons (Exc), and interneurons (Inh). In late AD, this tendency for enrichment extended to Exc, Inh, and microglial subtypes (Mic), whereas in non-AD samples, enrichment was observed in Inh and Mic cell types (Figure 4.9).

For promoters of genes marked by hypermethylated regions, early AD showed a tendency for enrichment in fibroblasts (Fib) and Inh, whereas late AD demonstrated enrichment in Exc and Inh. In non-AD samples, hypermethylated promoters tended to be enriched in excitatory neurons and interneurons (Figure 4.10).

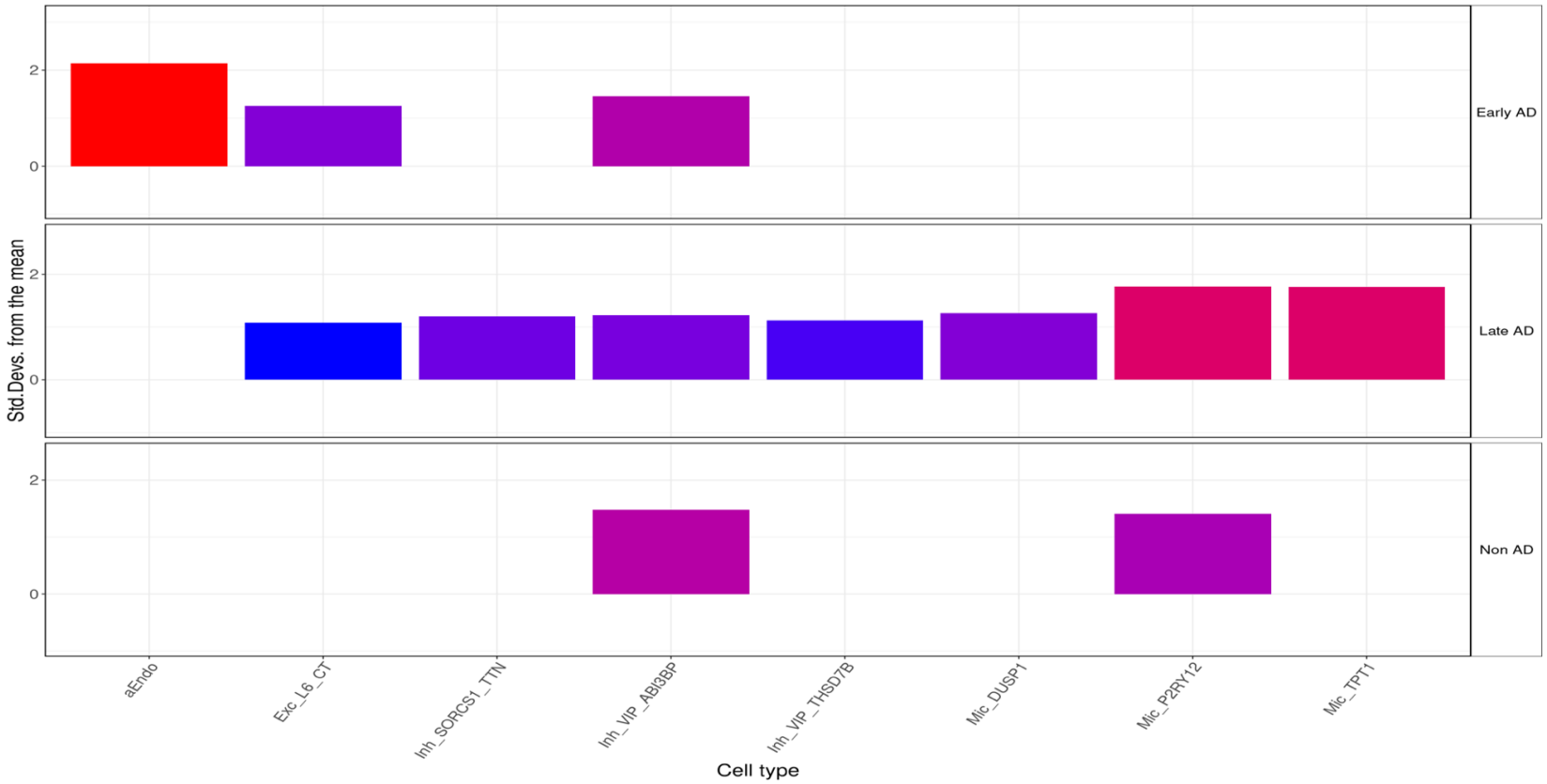

Figure 4-9 Cell type enrichment *test* of *genes with promoters marked by hypomethylated DMRs using the human single cell expression AD reference atlas, across early AD, late AD, and non-AD stages using EWCE.*

The top, middle, and bottom panels represent enrichment in early AD (n=29 individuals), late AD (n=15), and non-AD stages (n=48) respectively. The y-axis shows standard deviations from the mean expression per cell type, with the color scale indicating enrichment magnitude (red = higher, blue = lower). Enrichment analysis used Benjamini-Hochberg adjustment for FDR, with no significant enrichment observed for hypermethylated or hypomethylated promoter regions based on q-values.

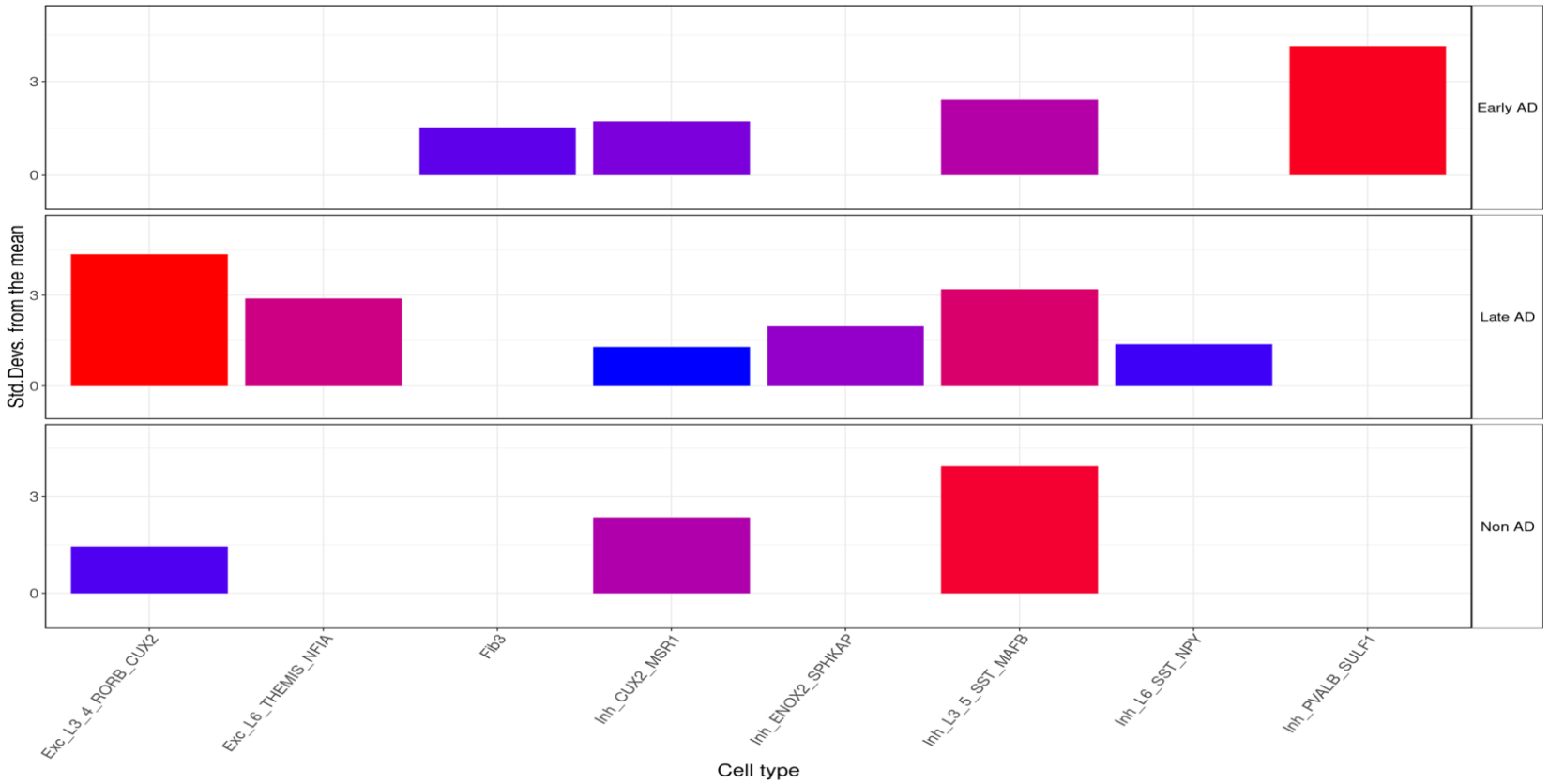

Figure 4-10 Cell type enrichment of hypermethylated promoter DMRs using the human AD reference atlas, across early AD, late AD, and non-AD stages using EWCE

The top, middle, and bottom panels represent hypermethylated promoter DMRs enrichment in early AD (n=29 individuals), late AD (n=15), and non-AD stages (n=48) respectively. The y-axis shows standard deviations from the mean expression per cell type, with the color scale indicating enrichment magnitude. While no significant enrichment is observed, deviations in excitatory and interneuron subtypes vary by disease stage. Enrichment analysis used Benjamini-Hochberg adjustment for FDR.

When incorporating DMRs from gene bodies and intergenic regions, cell type enrichment analysis revealed distinct patterns of hypomethylation across AD stages. Hypomethylated DMRs were significantly associated with subtypes of inhibitory neurons, and oligodendrocyte precursor cells (OPCs). In both early and late AD, these regions showed significant enrichment in inhibitory neuronal cell types, indicating a potential stage-specific and cell-type-specific methylation response to disease pathology. OPCs exhibited strong enrichment in both non-AD and late-AD stages (adjusted p-value < 0.05), suggesting a broader regulatory role extending beyond promoter regions to include gene bodies and intergenic regions (Figure 4.11).

The distribution of hypomethylated DMR-associated genes suggests these genes may serve distinct roles in different cell types and stages of disease. While no significant enrichment was observed in major cell types such as astrocytes, excitatory neurons, or interneurons, specific OPC subtypes, such as OPC_GPR4, showed consistent enrichment across stages, reflecting their potential involvement in disease-related cellular processes.

For hypermethylated DMRs, cell type enrichment analysis revealed distinct expression patterns across non-AD, early AD, and late AD stages. In early AD, hypermethylated regions were significantly enriched in an inhibitory neuronal subtype, reflecting a likely regulatory response to early pathological changes. By late AD, enrichment extended to multiple inhibitory subtypes and arterial endothelial cells (aEndo) (adjusted p-value < 0.05), suggesting hypermethylation may influence both neuronal and vascular functions in advanced disease stages (Figure 4.12).

Figures 4.11 and 4.12 illustrate these patterns, with standardized deviations highlighting the significance of methylation changes across specific cell types. Hypomethylation was prominent in OPCs, with non-significant trends observed in inhibitory cell types, while hypermethylation showed broader enrichment, including endothelial and inhibitory subtypes, particularly in early and late AD. These findings emphasize the role of stage-specific methylation changes in modulating cellular responses to AD pathology, with potential implications for both neuronal and vascular dysfunction during disease progression.

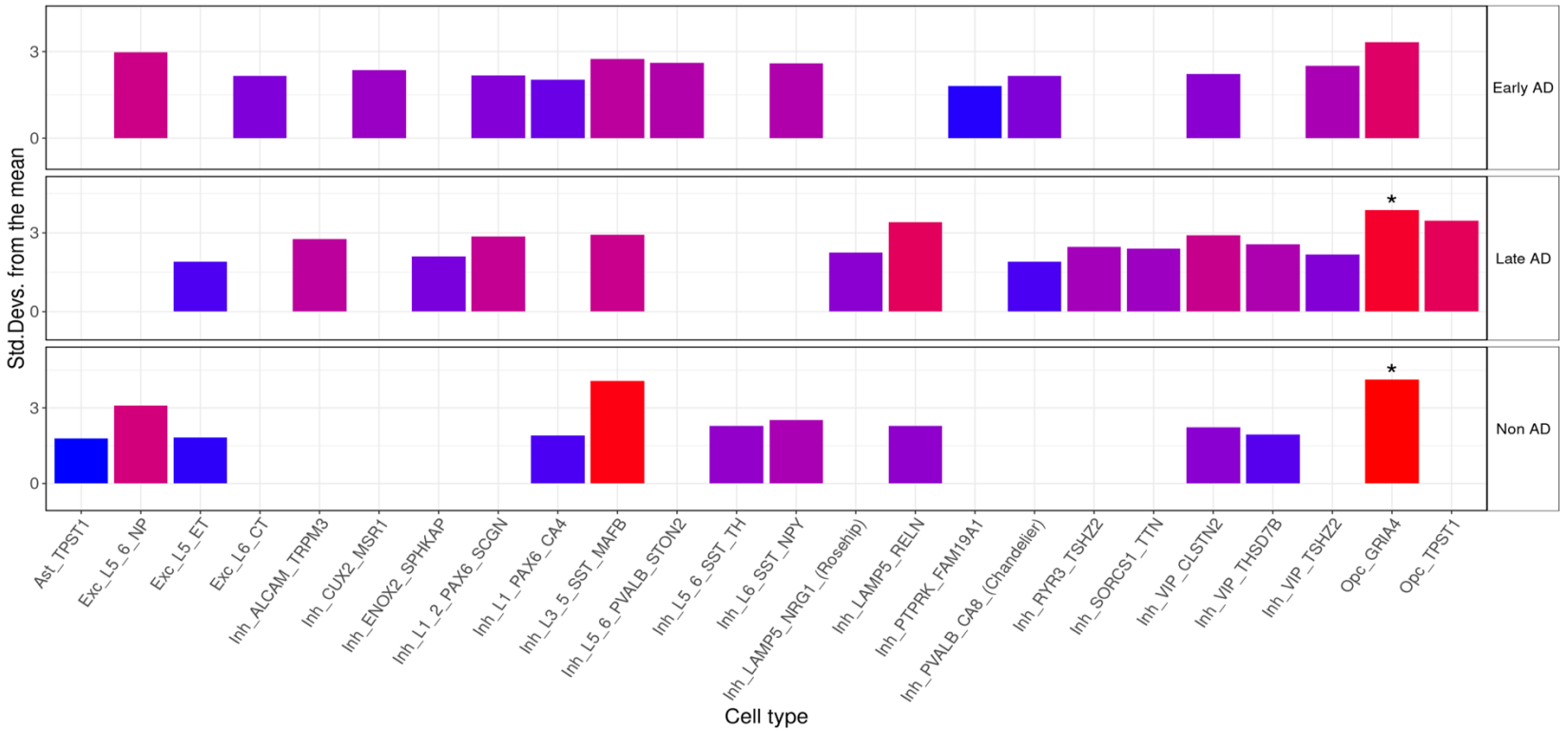

Figure 4-11 Cell type enrichment analysis of hypomethylated DMRs across non-AD, early AD, and late AD stages.

The cell-type enrichment bar plot shows the standardized deviation from the mean for cell type-specific enrichment of hypomethylated DMRs across early AD (n=29 individuals), late AD (n=15), and non-AD stages (n=48) respectively. Non-significant enrichment is observed in excitatory neurons (Exc), interneurons (Inh), astrocytes (Ast), and oligodendrocyte precursor cells (OPC). Significant enrichment is observed for OPC_GPR4 in non-AD, and late-AD, as indicated by the asterisks (*). These results highlight stage-specific patterns of hypomethylation and demonstrate that DMR gene lists may be enriched in distinct cell types at different stages of disease progression. Enrichment analysis used Benjamini-Hochberg adjustment for FDR.

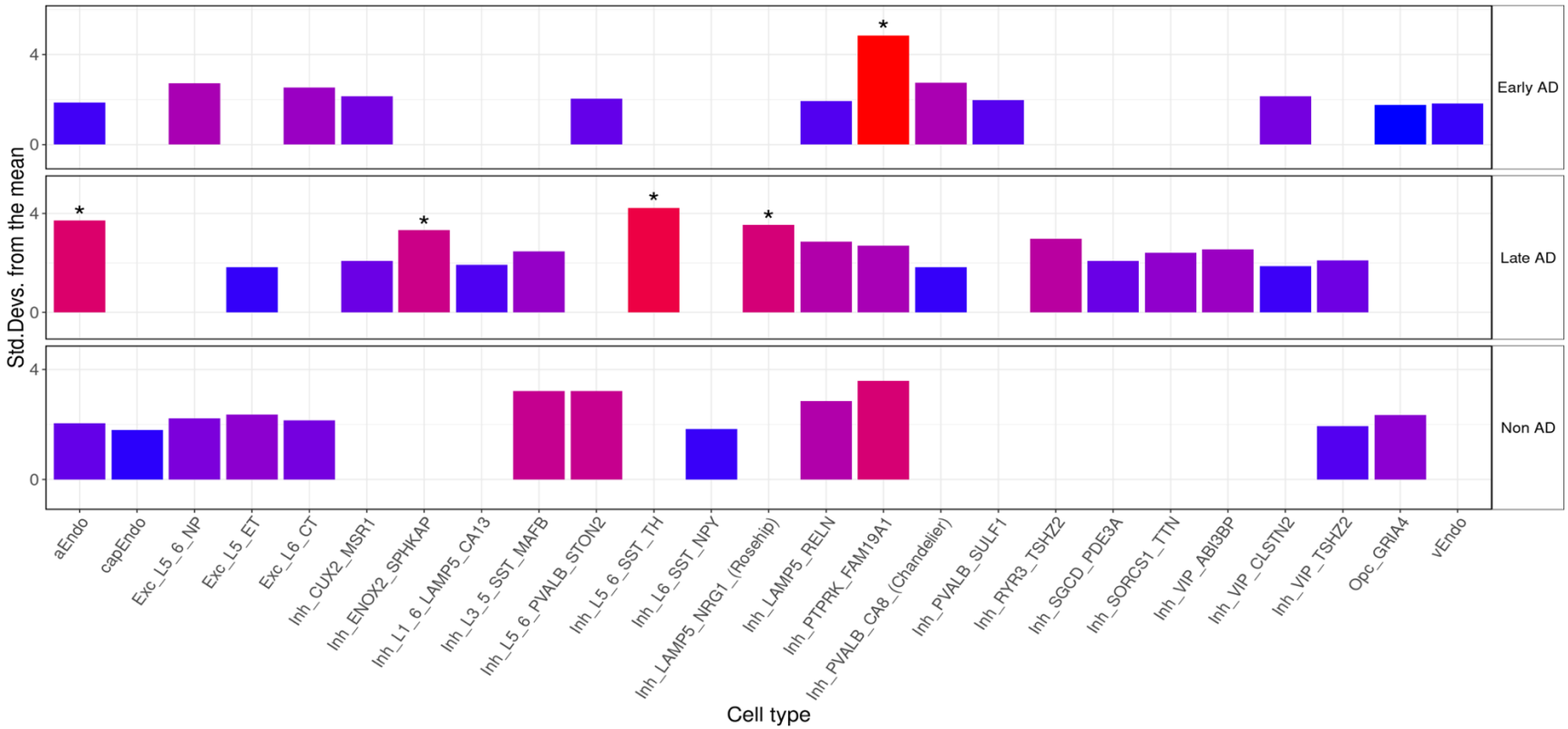

Figure 4-12 Cell type enrichment analysis of hypermethylated DMRs across early AD, late AD, and non-AD stages.

The cell-type enrichment bar plot shows the standardized deviation from the mean for cell type-specific enrichment of hypermethylated DMRs across early AD (n=29 individuals), late AD (n=15), and non-AD stages (n=48) respectively. Cell types are displayed on the x-axis, and the standardized deviation is plotted on the y-axis. Significant enrichment is observed in *aEndo* (arterial endothelial cells) and interneuron subtypes during early and late AD stages, as indicated by asterisks (*). These findings highlight enrichment trends of hypermethylation in specific cell types, particularly in endothelial cells and interneurons, across disease stages. Enrichment analysis used Benjamini-Hochberg adjustment for FDR.

4.4 Discussion

4.4.1 Correlation between methylation levels in promoter regions of genes with gene expression in matched samples

Several strong correlations or trends were observed between promoter methylation and gene expression. While significant inverse correlations were identified, differences in methylation or expression levels between *App*^{NLGF} vs WT were statistically insignificant. For example, hypomethylation in *Foxp1* (areaStat = -34.744, Wald test) showed a notable inverse correlation with its expression levels (Pearson correlation = -0.75, p = 0.080, t-test). Given the potential role of *Foxp1* in AD pathology, this correlation still warrants further investigation; *Foxp1* is a transcription factor involved in immune regulation neuronal function. Irie et al. (2023) demonstrated that *FOXP1* is expressed in NeuroD1-induced neuronal cells derived from microglia/macrophages, suggesting a role in modulating the inflammatory environment following brain injury (Irie et al., 2023). Khandelwal et al. (2024) reported that *FOXP1* can induce phenotypic rescue in striatal neurons, indicating its therapeutic potential in restoring normal synaptic function in neurodegenerative conditions (Khandelwal et al., 2024b). This suggests, as observed in the correlation analysis, increased *FOXP1* levels could counteract synaptic deficits observed in AD, potentially improving cognitive function and neuronal resilience.

Overall, while specific genes did exhibit statistically significant differential expression overall, there were an observable trends of increased gene expression in the *App*$^{NL\text{-}G\text{-}F}$ model compared to WT. These findings indicate that while expression levels for such genes remain relatively stable under current conditions, their sensitivity to methylation changes suggests that future shifts, possible through environment or experimental manipulations, could drive more pronounced changes, highlighting its regulatory potential in response to amyloid pathology.

This observation aligns with previous studies showing that even modest changes in promoter methylation can exert significant regulatory effects on gene expression (Jones, 2012). Methylation changes may not always result in immediate or detectable expression under controlled experimental conditions but could still have long-term impacts on gene regulation. Methylation may serve as a prelude to more substantial

transcriptional changes or as part of a complex regulatory mechanism that integrates multiple epigenetic and environmental inputs (Zhang et al., 2013), and it is important to bear in mind that the environments of the mice used in this study were carefully controlled, and so the true environmental impact is hard to assess. Thus, genes with significant correlations between methylation and expression could potentially be interesting, as they may represent key players in disease progression or response to environmental stimuli, particularly in humans.

4.4.2 CpG or regional correlation between methylation and expression

Several genes exhibited DMRs that correlated with the gene expression; however, in specific cases, individual CpG sites were more indicative of the expression changes rather than the overall methylation of a wider region. Microglial genes including *Capg* and *Csf1*, both of which are involved in critical biological processes such as immune response and neuroinflammation, showed CpG specific correlations. In the *Capg* gene, hypomethylation in the promoter was associated with upregulation of gene expression. The Pearson correlation coefficient for the aggregated methylation across the *Capg* promoter showed a moderately negative correlation (-0.3875), indicating that, at a regional level, methylation differences do not strongly correlate with gene expression in a straightforward, linear fashion.

Some CpG sites may exhibit methylation changes but have no impact on expression, whereas others might be tightly linked to transcriptional regulation. Identifying these functionally relevant CpG sites requires investigating them individually. Both DMR analysis and single CpG site analysis provide valuable, complementary insights. DMR analysis is effective for identifying broader regions of differential methylation and may reduce noise by aggregating information across sites. However, analysing each CpG site individually with the Pearson correlation is essential for uncovering fine-scale, functional relationships between methylation and gene expression. This higher-resolution view can highlight specific CpG sites with significant regulatory roles, and to help understand the heterogeneity within DMRs.

Examining individual CpG sites is critical, especially those near regulatory elements such as promoters or transcription factor binding sites, as these locations can have a significant influence on gene expression. Regional analyses, while useful for capturing

broader patterns, may obscure the functional importance of specific CpG sites. However, examining CpG sites comes with challenges, particularly in ensuring robust quality control. Variability in sequencing accuracy and coverage at individual CpG sites can introduce noise; thus, it is essential to implement stringent QC measures to filter for high-confidence sites. Focusing on these sites within differentially methylated regions or filtered sites, can provide more precise understanding of the potential regulatory roles these CpG sites play in modulating gene expression.

4.4.3 Stage-specific DNA methylation patterns reveal conserved cell type involvement in AD pathology

Cell type enrichment analysis revealed that DMRs were predominantly associated with neuronal genes. In Chapter 3, a mouse reference atlas was used to annotate hypermethylated and hypomethylated regions, providing insights into methylation patterns linked to amyloid pathology. Chapter 4 extended this approach by incorporating stage- and cell-type-specific annotations from human AD brains. This allowed for a comparative perspective, highlighting potential conserved methylation mechanisms across human and mouse models while exploring their associations with cell-type-specific pathology severity.

One relevant study, Lang et al. (2022), analyzed DNA methylation changes associated with AD neuropathologic change (ADNC) across eight brain regions, including the dentate gyrus and cingulate gyrus, using the Illumina 850k EPIC BeadChip array. Their study focused on bulk tissue methylation, identifying differentially methylated promoters linked to AD severity markers, such as Aβ plaques and neurofibrillary tangles. They employed cell-type deconvolution based on single-cell RNA sequencing (scRNA-seq) reference data, observing numerous differentially methylated promoters in neurons, as well as in astrocytes, endothelial cells, and oligodendrocyte precursor cells (OPCs). These methylation changes were associated with Aβ burden, as assessed using the National Institute on Aging-Alzheimer's Association (NIA-AA) criteria.

Similar to their findings, our enrichment analysis of DMRs from the AD mouse model annotated using human AD brain data revealed that common cell types and genes were involved with the two studies. While Lang et al. focused on promoter methylation, our analysis also included non-promoter regions, which allowed us to capture

significant methylation changes in neuronal, OPC, and endothelial genes. For example, Lang *et al.* (2022) identified promoter methylation of *CSF1* in the cingulate gyrus linked to Aβ and tangle severity, while our mouse model revealed DMRs in the distal intergenic area of *Csf1* associated with amyloid pathology.

There are notable methodological differences between the two studies. Lang et al. used cell-type deconvolution to infer methylation patterns, while our approach employed cell type enrichment using a curated marker list from dorsolateral prefrontal cortex snRNA-seq data. Additionally, Lang *et al*. (2022) focused solely on promoter methylation, while we included methylation changes across promoters, gene bodies, and intergenic regions, providing a broader view of epigenetic regulation.

Despite these differences, both studies identified shared methylation patterns in key genes. For instance, we observed conserved methylation changes in *SEMA4D* and glutamate ionotropic receptor AMPA Type subunit 1 *(GRIA1)* in neurons and astrocytes, as well as phosphatidylinositol-5-phosphate 4-kinase type 2 gamma (*PIP4K2C)* in oligodendrocytes/OPCs. While these findings highlight conserved methylation mechanisms, the functional roles of these changes may differ across species and brain regions.

Importantly, our analysis revealed that when including the regions beyond promoters, OPCs are more prominent in hypomethylated regions (Figure 4.11.) and distinct inhibitory cell subtypes are involved as disease progresses (Figure 4.12.), suggesting broader regulatory implications of methylation changes.

It is important to acknowledge that methylation changes do not always correlate directly with gene expression alterations. The reference datasets used in both studies serve solely to estimate the cell type enrichment of DMRs, without direct functional validation. A key strength of the cell type enrichment analysis in this study lies in the use of an annotation dataset with higher resolution, encompassing a greater number of cells, finer cell subtypes, and disease-stage-specific information. In this study, a large snRNA-seq dataset derived from the dorsolateral prefrontal cortex of human AD brains across various disease stages were used. Also, EWCE minimizes the bias toward highly expressed genes as it measures the standard deviation of the mean of

genes, providing a more precise evaluation of cell-type-specific associations and enhancing the reliability of our findings.

These findings emphasize the need for further research to determine whether the DMR-associated genes in the AD mouse model align with gene expression changes observed in human AD brains with matching methylation data. Such analyses could shed light on the functional implications of hypomethylation in disease progression and reveal how these epigenetic changes affect specific cell types on a gene-by-gene basis.

4.4.4 DMR overlap between mouse and human AD datasets

A gene set overlap analysis between three AD methylation datasets from various brain regions (De Jager et al., 2014; Piras et al., 2023; Wang et al., 2023) and our mouse model revealed significant degree of epigenetic conservation. The identification of multiple shared DMRs between our dataset and previous AD studies (Altuna et al., 2019; Shireby et al., 2022) suggests that certain epigenetic changes are preserved across different experimental and clinical cohorts. *RHOB* was specifically associated with neuritic plaque density, while *SPEG* and *CACNA1E* were also linked to tau pathology. The detection of common DMRs in both human and mouse datasets underscores potential conserved mechanisms of epigenetic regulation between human and mice.The enrichment of shared DMRs in genes involved in neurodevelopment, synaptic regulation, and cytoskeletal organization suggests that epigenetic dysregulation in these pathways may contribute to neuronal dysfunction in AD. This supports the utility of mouse models in capturing key epigenetic alterations observed in human AD.

Despite these parallels, notable differences about methodological approaches exists between our findings and those reported in previous AD studies. The DMRs in our study were identified using long-read DNA sequencing, in contrast to the bisulfite conversion-based methods employed in these datasets. While our dataset encompassed the whole cortex, the overlap with region-specific methylation changes observed in late AD from these human datasets is noteworthy. This suggests that some cortical methylation changes captured in our model may reflect alterations occurring in specific brain regions during late-stage AD. However, the exact DMR type

and its functional effects may differ due to the methodological and biological distinctions between datasets. The overlap of DMRs between different human brain regions was limited across human AD datasets, although mouse DMRs had intersections of genes with each of datasets, reinforcing the idea of brain-region specificity of methylation changes, underscoring the importance of studying methylation dynamics within distinct regions in AD. Despite these differences, the conservation of key pathways between mouse and human datasets strengthens the utility of mouse models in exploring epigenetic mechanisms underlying AD.

The relationship between methylation changes and gene expression in amyloid pathology is exemplified by *Cdh13* and *Ptpre*. In *the App^{NL-G-F}* model, intronic hypomethylation of *Cdh13* correlated with increased expression, suggesting a role in synaptic stability and plasticity during early AD (Mossink et al., 2022). This aligns with observations from human AD datasets, where *Cdh13* expression rises in early stages but declines in late AD, reflecting progressive synaptic dysfunction. Similarly, intronic hypermethylation of *Ptpre* in the *App^{NL-G-F}* model was linked to increased expression of the isoform *Ptpre-204*, suggesting isoform-specific regulation. In human AD, overall *PTPRE* expression declines with disease progression, potentially indicating disrupted microglia-endothelial interactions and blood-brain barrier dysfunction. These findings highlight the regulatory role of non-promoter methylation in gene expression and isoform-specific responses, emphasizing its contribution to cellular adaptations in AD and its potential as a therapeutic target.

DMRs in AD risk genes shared between studies and species, can help to uncover how epigenetic changes might influence disease progression, in relation to the amyloid pathology. The study revealed both hypomethylation and hypermethylation in key genes, including *App, Egfr, Sppl2a, Umad1,* and *Mme,* with distinct patterns linked to expression. For example, hypomethylation in an intergenic region of *App* was associated with increased expression, suggesting a potential role in regulating amyloid-related processes. In *Egfr,* methylation at specific CpG sites showed group-specific correlations with expression, pointing to differences in regulatory mechanisms between amyloid-pathology and control groups. *Mapt* displayed hypermethylation in an intronic region in both the AD mouse model and human AD as reported by Wang et al. (2023), where it was associated with amyloid pathology. The hypermethylation

of *Mapt* in an intronic region is particularly intriguing in the amyloid model of AD, as *Mapt* encodes tau, the key protein in neurofibrillary tangle formation. While the *App*$^{NL\text{-}G\text{-}F}$ mouse model primarily reflects amyloid pathology without tau tangles, this hypermethylation may represent an early or indirect link between amyloid accumulation and tau-related regulatory mechanisms. Alternatively, it could be a protective response to amyloid, potentially modulating *Mapt* expression to limit tau availability for pathological aggregation. These possibilities highlight the role of epigenetic regulation in amyloid pathology and its implications for amyloid-tau interactions and disease progression.

Although the precise methylation sites differ between species, and the methodology may differ, these findings underscore the importance of cross-species comparisons in identifying biologically relevant methylation patterns. Such comparisons can provide insights into the potential role of methylation in AD pathology, and help to prioritise experiments, even when the exact genomic coordinates are not directly comparable.

4.4.5 Implications for disease staging and therapeutic interventions

Uncovering preferential transcript usage, alternative splicing, and novel transcripts in response to amyloid pathology reveals critical insights into AD diagnosis, staging, and treatment strategies. These molecular shifts not only reflect underlying disease mechanisms but also suggest new biomarkers capable of refining disease staging and tracking progression with greater precision.

Adding to this complexity, the role of DNA methylation provides a better understanding of the complexities of gene regulation. Aberrant methylation patterns, particularly in immune response and synaptic physiology genes, have been observed in AD (De Jager et al., 2014; Lang et al., 2022; Piras et al., 2023; E. Wang et al., 2023). Hypermethylation in promoter regions may silence protective genes or isoforms, while hypomethylation could drive the expression of pathogenic genes or isoforms. Methylation changes outside of promoter regions, such as in enhancers or insulator elements, can also have regulatory consequences (Marzi et al., 2018; Nott et al., 2019).

Incorporating methylation data into analyses of gene or isoform level changes may enhance our ability to stage AD and devise therapeutic interventions. Epigenetic

modifications, such as CpG methylation changes, may serve as early biomarkers that can distinguish between disease stages, offering a tool for earlier diagnosis. Although the methylation and expression/splicing patterns warrant further investigation and validation in different models and samples, and with complementary technologies, targeting methylation patterns therapeutically may restore proper splicing and expression profiles. Correcting these methylation-driven disruptions holds promise for reducing synaptic dysfunction and slowing neurodegeneration.

These findings point to a dual approach: on the one hand, they enable the development of biomarkers to refine disease staging, and on the other, they highlight transcript isoforms as therapeutic targets. For example, antisense oligonucleotides (ASOs) could be employed to shift splicing toward protective isoforms and away from harmful variants (Chery, 2016). Such precision-targeted therapies offer the potential not only to slow disease progression but also to enhance synaptic resilience, presenting a promising new frontier for AD treatment.

5 Conclusions

AD is strongly associated with genetic risk variants identified through genome-wide association studies (GWAS) (Bellenguez et al., 2022; Kunkle et al., 2019; J. C. Lambert et al., 2013). However, the mechanisms by which these risk variants influence gene function remain largely unknown, especially if the variants is within the non-coding genome. These non-coding variants are thought to play significant roles in regulating gene expression, yet their functional implications are not fully understood.

Recent efforts to integrate DNA methylation and gene expression data from GWAS analyses have sought to bridge this gap by identifying the functional role of these variants in disease process (Marzi et al., 2018; Nott et al., 2019). For example, SNPs in the enhancer region of the *BIN1* gene, which is associated with AD, have been shown to regulate gene expression specifically in microglia, but not in other cell types, highlighting the cell-type-specific nature of genetic regulation in AD. While these studies offer important insights into how genetic variants may influence disease, the functional roles and regulation of many risk genes, particularly in the early stages of AD, remain unclear. This thesis has addressed some of these gaps by investigating how amyloid pathology, a hallmark of early AD, influences gene expression, transcriptome regulation and DNA methylation in the $App^{NL\text{-}G\text{-}F}$ mouse model of AD. By focusing on early disease mechanisms, this research offers new perspectives on the molecular pathways that may underlie AD progression and provides further understanding of how genetic risk variants contribute to the disease.

Our study investigated the impact of amyloid pathology on gene expression and DNA methylation in the $App^{NL\text{-}G\text{-}F}$ mouse model of AD using genome-wide long-read RNA and DNA sequencing. We identified genes involved in immune response and synaptic function that exhibited differential alternative splicing and unique isoform patterns in response to amyloid pathology. DNA methylation analysis revealed both hyper- and hypomethylation in neuronal subtypes, with differentially methylated CpG sites found in promoter, intronic, and intergenic regions of AD-associated genes. These findings suggest that early amyloid pathology induces cellular adaptations that may enhance resilience in microglia and neurons. Since this model is protected from overt tau

tangles and neuronal death, it provides a unique opportunity to explore the early molecular mechanisms underlying AD progression.

Chapter 2 focused on transcriptomic changes linked to amyloid pathology, identifying genes with distinct isoform usage and alternative splicing patterns. These included some disease-associated microglia genes such as *Capg, Syngr1, Csf1,* and cathepsins (including *Ctsa, and Ctsd*), as well as interestingly synaptic genes such as *Clta* and novel transcripts of AD risk genes. Thus immune-related genes expressed in microglia alongside synaptic genes demonstrated pronounced splicing variations, highlighting the cooperative adaptive responses of these cell types.

Chapter 3 explored DNA methylation changes, identifying DMRs in genes related to immune and synaptic processes, including *Capg, Csf1, App,* and *Mapt*. This analysis revealed methylation differences in promoters, intergenic regions, and non-coding elements, with bulk-level correlations indicating links between methylation and expression. Genes with DMRs mainly showed neuronal cell type enrichment of expression.

Chapter 4 integrated transcriptomic and methylation data, emphasizing some shared patterns between the mouse model and human AD datasets. These comparisons uncovered conserved molecular mechanisms, reinforcing the translational value of findings from the $App^{NL\text{-}G\text{-}F}$ model. Cross-species comparisons with human data demonstrated that some of these changes are conserved, providing insights into their relevance for human AD pathology.

5.1 Limitations

While this thesis provides valuable insights into the molecular mechanisms of AD, several limitations must be acknowledged. The analysis was performed on bulk tissue, which limited the ability to resolve cell-type-specific changes. The $App^{NL\text{-}G\text{-}F}$ mouse model represents a single stage of AD and does not capture the effects of ageing, which may play a critical role in disease progression. The mice used were relatively young, allowing the capture of data relevant for earlier amyloid deposition, however limiting the ability to study age-related methylation and splicing changes. Verifying the functional roles of spliced isoforms was constrained by the availability of specific antibodies, and mass spectrometry sensitivity and approaches. Additionally, while

some mechanisms were conserved between mouse and human, species-specific differences may limit the generalizability of findings. The controlled environment in which the mice were studied may also have restricted the range of methylation changes observed. Furthermore, this study did not distinguish between 5-methylcytosine (5mC) and 5-hydroxymethylcytosine (5hmC), which may have implications for interpreting methylation changes, particularly in the context of neuronal function where 5hmC is more abundant. Additionally, our focus was limited to CpG methylation, excluding non-CG (CH) methylation, which has been increasingly recognized for its role in neuronal epigenetics and may contribute to disease processes.

Given the limitations of using a mouse model, which may not capture the full complexity of human AD, future studies should aim to address these gaps for greater translational impact, such as performing similar studies in a mouse model of tau pathology, and mice that show amyloid-dependent tau pathology. Future research could focus on determining the functional roles of differentially expressed isoforms and DMRs in AD pathology. Applying gene-editing tools, like CRISPR/Cas9, or overexpression systems may clarify how these molecular changes and preferential isoform usage affect disease processes, particularly in microglia and neurons (Gallo et al., 2024). Furthermore, characterizing the identified DMRs and cross-validating with human epigenome-wide association studies could uncover new pathways involved in AD, shedding light on their functional roles in disease progression.

Acknowledging these limitations, future studies should aim to bridge these gaps to enhance the translational relevance of the findings.

5.2 Future Directions

Future research needs to address the current gaps in our understanding of AD. One important direction is to investigate changes in methylation and its functional roles on gene and isoform level alterations including isoform switching and splicing, within individual cell types, such as microglia and neurons across disease stages using long-read sequencing. These analyses could uncover subtle, cell-type-specific mechanisms that potentially drive AD pathology.

To gain a fuller understanding of disease progression, future studies should also include mouse models that incorporate tau pathology or amyloid-tau interactions. This expansion will allow for a more comprehensive examination of the disease, particularly in later stages, where the interplay between amyloid and tau may be critical and poorly understood. Additionally, investigating environmental factors—such as physical activity and diet—could offer valuable information on modifiable risk factors, opening new pathways for prevention and therapeutic interventions.

In parallel, it is essential to validate the functional significance of the observed molecular changes. Techniques such as CRISPR/Cas9 could be employed to directly modify methylation sites or splicing events, providing a way to test their impact on disease processes. Targeted mass spectrometry could be used to confirm whether these changes affect particular disease-relevant protein isoforms, while antisense oligonucleotide (ASO) therapies may offer a solution for correcting splicing abnormalities. Finally, conducting meta-analyses that combine findings from mouse models and human datasets will strengthen the reliability and translational relevance of these insights, ensuring that the knowledge gained from these studies can be effectively applied to human Alzheimer's disease.

5.3 Final Words

Our study employed genome-wide long-read RNA and DNA sequencing to investigate the effects of amyloid pathology on gene expression and DNA methylation in the *App*$^{NL\text{-}G\text{-}F}$ mouse model of AD. The analysis of both transcriptomic and methylation changes revealed distinct genes exhibiting alternative splicing and unique isoform patterns, particularly in pathways related to immune response and synaptic function. These findings suggest that early amyloid pathology triggers cellular adaptations that may enhance resilience in microglia and neurons, especially since this model is protected from overt tau tangles and neuronal death.

Several key immune-related genes, especially those expressed in microglia, were found to exhibit distinct isoforms and splicing variations. The methylation data further enriched these findings, revealing differential patterns of methylation between AD-model and control mice, particularly in genes associated with neuroinflammation and immune responses. The cross-species comparison with human data highlighted that

some of these splicing and methylation changes are conserved across species, suggesting that the molecular mechanisms observed in the mouse model may be relevant to human AD.

Despite the valuable insights gained from this work, the use of a mouse model inherently limits the ability to fully capture the complexity of human AD. Future studies should aim to address these gaps by using mouse models that incorporate tau pathology or amyloid-tau interactions at different disease stages. Expanding research to explore the functional roles of differentially expressed isoforms and differentially DMRs could provide further clarity on their contributions to AD pathology. Gene-editing tools such as CRISPR/Cas9 or overexpression systems could offer deeper insights into how these molecular changes affect disease processes, particularly in microglia and neurons (Gallo et al., 2024). Additionally, characterizing the identified DMRs and validating them against human epigenome-wide association studies (EWAS) could uncover novel pathways involved in AD, enhancing our understanding of their roles in disease progression.

The integration of long-read sequencing technologies into AD research represents a significant step toward unraveling the genetic and epigenetic complexities of the disease. By focusing on the functional analysis of isoforms and DMRs, future research could pave the way for developing targeted therapeutic strategies that address the multifaceted, cell-type and stage-dependent nature of AD pathology, ultimately advancing our understanding of the disease and potential treatments.

References

Abràmoff, M. D., Magalhães, P. J., & Ram, S. J. (2004). Image processing with imageJ. In *Biophotonics International* (Vol. 11, Issue 7). https://doi.org/10.1201/9781420005615.ax4

Agrawal, P., Heimbruch, K. E., & Rao, S. (2019). Genome-wide maps of transcription regulatory elements and transcription enhancers in development and disease. *Comprehensive Physiology*, *9*(1). https://doi.org/10.1002/cphy.c180028

Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther, O., Brunak, S., von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal peptide predictions using deep neural networks. *Nature Biotechnology*, *37*(4). https://doi.org/10.1038/s41587-019-0036-z

Alsina, F. C., Hita, F. J., Fontanet, P. A., Irala, D., Hedman, H., Ledda, F., & Paratcha, G. (2016). Lrig1 is a cell-intrinsic modulator of hippocampal dendrite complexity and BDNF signaling . *EMBO Reports*, *17*(4). https://doi.org/10.15252/embr.201541218

Altuna, M., Urdánoz-Casado, A., Sánchez-Ruiz De Gordoa, J., Zelaya, M. V., Labarga, A., Lepesant, J. M. J., Roldán, M., Blanco-Luquin, I., Perdones, Á., Larumbe, R., Jericó, I., Echavarri, C., Méndez-López, I., Di Stefano, L., & Mendioroz, M. (2019). DNA methylation signature of human hippocampus in Alzheimer's disease is linked to neurogenesis. *Clinical Epigenetics*, *11*(1). https://doi.org/10.1186/s13148-019-0672-7

Anders, S., Reyes, A., & Huber, W. (2012). Detecting differential usage of exons from RNA-seq data. *Genome Research*, *22*(10). https://doi.org/10.1101/gr.133744.111

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., Ntini, E., Arner, E., Valen, E., Li, K., Schwarzfischer, L., Glatz, D., Raithel, J., Lilje, B., Rapin, N., … Sandelin, A. (2014). An atlas of active enhancers across human cell types and tissues. *Nature*, *507*(7493). https://doi.org/10.1038/nature12787

Askenazi, M., Kavanagh, T., Pires, G., Ueberheide, B., Wisniewski, T., & Drummond, E. (2023). Compilation of reported protein changes in the brain in Alzheimer's disease. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-40208-x

Bakulski, K. M., Dolinoy, D. C., Sartor, M. A., Paulson, H. L., Konen, J. R., Lieberman, A. P., Albin, R. L., Hu, H., & Rozek, L. S. (2012). Genome-wide DNA methylation differences between late-onset alzheimer's disease and cognitively normal controls in human frontal cortex. *Journal of Alzheimer's Disease*, *29*(3). https://doi.org/10.3233/JAD-2012-111223

Bankhead, P., Loughrey, M. B., Fernández, J. A., Dombrowski, Y., McArt, D. G., Dunne, P. D., McQuaid, S., Gray, R. T., Murray, L. J., Coleman, H. G., James, J. A., Salto-Tellez, M., & Hamilton, P. W. (2017). QuPath: Open source software for digital pathology image analysis. *Scientific Reports*, *7*(1). https://doi.org/10.1038/s41598-017-17204-5

Bateman, R. J., Xiong, C., Benzinger, T. L. S., Fagan, A. M., Goate, A., Fox, N. C., Marcus, D. S., Cairns, N. J., Xie, X., Blazey, T. M., Holtzman, D. M., Santacruz, A., Buckles, V., Oliver, A., Moulder, K., Aisen, P. S., Ghetti, B., Klunk, W. E., McDade, E., … Morris, J. C. (2012). Clinical and Biomarker Changes in

Dominantly Inherited Alzheimer's Disease. *New England Journal of Medicine*, *367*(9). https://doi.org/10.1056/nejmoa1202753
Bellenguez, C., Küçükali, F., Jansen, I. E., Kleineidam, L., Moreno-Grau, S., Amin, N., Naj, A. C., Campos-Martin, R., Grenier-Boley, B., Andrade, V., Holmans, P. A., Boland, A., Damotte, V., van der Lee, S. J., Costa, M. R., Kuulasmaa, T., Yang, Q., de Rojas, I., Bis, J. C., … Lambert, J. C. (2022). New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nature Genetics*, *54*(4). https://doi.org/10.1038/s41588-022-01024-z
Benitez, D. P., Jiang, S., Wood, J., Wang, R., Hall, C. M., Peerboom, C., Wong, N., Stringer, K. M., Vitanova, K. S., Smith, V. C., Joshi, D., Saito, T., Saido, T. C., Hardy, J., Hanrieder, J., De Strooper, B., Salih, D. A., Tripathi, T., Edwards, F. A., & Cummings, D. M. (2021). Knock-in models related to Alzheimer's disease: synaptic transmission, plaques and the role of microglia. *Molecular Neurodegeneration*, *16*(1). https://doi.org/10.1186/s13024-021-00457-0
Beyter, D., Ingimundardottir, H., Oddsson, A., Eggertsson, H. P., Bjornsson, E., Jonsson, H., Atlason, B. A., Kristmundsdottir, S., Mehringer, S., Hardarson, M. T., Gudjonsson, S. A., Magnusdottir, D. N., Jonasdottir, A., Jonasdottir, A., Kristjansson, R. P., Sverrisson, S. T., Holley, G., Palsson, G., Stefansson, O. A., … Stefansson, K. (2021). Long-read sequencing of 3,622 Icelanders provides insight into the role of structural variants in human diseases and other traits. *Nature Genetics*, *53*(6). https://doi.org/10.1038/s41588-021-00865-4
Bird, A. (2002). DNA methylation patterns and epigenetic memory. In *Genes and Development* (Vol. 16, Issue 1). https://doi.org/10.1101/gad.947102
Borsari, B., Villegas-Mirón, P., Pérez-Lluch, S., Turpin, I., Laayouni, H., Segarra-Casas, A., Bertranpetit, J., Guigó, R., & Acosta, S. (2021). Enhancers with tissue-specific activity are enriched in intronic regions. *Genome Research*, *31*(8). https://doi.org/10.1101/gr.270371.120
Botía, J. A., Vandrovcova, J., Forabosco, P., Guelfi, S., D'Sa, K., Hardy, J., Lewis, C. M., Ryten, M., Weale, M. E., Ramasamy, A., Trabzuni, D., Smith, C., & Walker, R. (2017). An additional k-means clustering step improves the biological features of WGCNA gene co-expression networks. *BMC Systems Biology*, *11*(1). https://doi.org/10.1186/s12918-017-0420-6
Bubnova, A. N., Yakovleva, I. V., Korotkov, E. V., & Kamionskaya, A. M. (2023). In Silico Verification of Predicted Potential Promoter Sequences in the Rice (Oryza sativa) Genome. *Plants*, *12*(20). https://doi.org/10.3390/plants12203573
Cahoy, J. D., Emery, B., Kaushal, A., Foo, L. C., Zamanian, J. L., Christopherson, K. S., Xing, Y., Lubischer, J. L., Krieg, P. A., Krupenko, S. A., Thompson, W. J., & Barres, B. A. (2008). A transcriptome database for astrocytes, neurons, and oligodendrocytes: A new resource for understanding brain development and function. *Journal of Neuroscience*, *28*(1). https://doi.org/10.1523/JNEUROSCI.4178-07.2008
Cai, C., Gu, C., He, S., Meng, C., Lai, D., Zhang, J., & Qiu, Q. (2024). TET2-mediated ECM1 hypomethylation promotes the neovascularization in active proliferative diabetic retinopathy. *Clinical Epigenetics*, *16*(1). https://doi.org/10.1186/s13148-023-01619-1
Carro, E., Trejo, J. L., Gomez-Isla, T., LeRoith, D., & Torres-Aleman, I. (2002). Serum insulin-like growth factor I regulates brain amyloid-β levels. *Nature Medicine*, *8*(12). https://doi.org/10.1038/nm793
Carullo, N. V. N., & Day, J. J. (2019). Genomic enhancers in brain health and disease. In *Genes* (Vol. 10, Issue 1). https://doi.org/10.3390/genes10010043

Castellano, J. M., Kim, J., Stewart, F. R., Jiang, H., DeMattos, R. B., Patterson, B. W., Fagan, A. M., Morris, J. C., Mawuenyega, K. G., Cruchaga, C., Goate, A. M., Bales, K. R., Paul, S. M., Bateman, R. J., & Holtzman, D. M. (2011). Human apoE isoforms differentially regulate brain amyloid-β peptide clearance. *Science Translational Medicine*, *3*(89). https://doi.org/10.1126/scitranslmed.3002156

Cevik, S. E., Skaar, D. A., Jima, D., Liu, A., Whitson, H., Jirtle, R. L., Hoyo, C., & Planchart, A. (2023). DNA methylation of candidate imprint control regions associated with Alzheimer's disease in Non-Hispanic Blacks and Non-Hispanic Whites. *Alzheimer's & Dementia*, *19*(S12). https://doi.org/10.1002/alz.071944

Chang, S., Trimbuch, T., & Rosenmund, C. (2018). Synaptotagmin-1 drives synchronous Ca 2+ -triggered fusion by C 2 B-domain-mediated synaptic-vesicle-membrane attachment. *Nature Neuroscience*, *21*(1). https://doi.org/10.1038/s41593-017-0037-5

Chen PC, Han X, Shaw TI, Fu Y, Sun H, Niu M, & Wang Z. (2022). Alzheimer's disease-associated U1 snRNP splicing dysfunction causes neuronal hyperexcitability and cognitive impairment. *Nature Aging*, *2*(10), 923–940.

Chen, X., Firulyova, M., Manis, M., Herz, J., Smirnov, I., Aladyeva, E., Wang, C., Bao, X., Finn, M. B., Hu, H., Shchukina, I., Kim, M. W., Yuede, C. M., Kipnis, J., Artyomov, M. N., Ulrich, J. D., & Holtzman, D. M. (2023). Microglia-mediated T cell infiltration drives neurodegeneration in tauopathy. *Nature*, *615*(7953). https://doi.org/10.1038/s41586-023-05788-0

Chery, J. (2016). RNA therapeutics: RNAi and antisense mechanisms and clinical applications. *Postdoc Journal*, *4*(7). https://doi.org/10.14304/surya.jpr.v4n7.5

Claxton, M., Pulix, M., Seah, M. K. Y., Bernardo, R., Zhou, P., Aljuraysi, S., Liloglou, T., Arnaud, P., Kelsey, G., Messerschmidt, D. M., & Plagge, A. (2022). Variable allelic expression of imprinted genes at the Peg13, Trappc9, Ago2 cluster in single neural cells. *Frontiers in Cell and Developmental Biology*, *10*. https://doi.org/10.3389/fcell.2022.1022422

Cogan, G., Daida, K., Billingsley, K. J., Tesson, C., Forlani, S., Jornea, L., Arnaud, L., Tissier, L., LeGuern, E., Singleton, A. B., Ferrien, M., Bernard, H. G., Lesage, S., Blauwendraat, C., & Brice, A. (2024). Long-Read Sequencing Unravels the Complexity of Structural Variants in in Two Individuals with Early-Onset Parkinson's Disease. *Movement Disorders*, *39*(9), 1647–1648. https://doi.org/https://doi.org/10.1002/mds.29914

Corder, E. H., Saunders, A. M., Strittmatter, W. J., Schmechel, D. E., Gaskell, P. C., Small, G. W., Roses, A. D., Haines, J. L., & Pericak-Vance, M. A. (1993). Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science*, *261*(5123). https://doi.org/10.1126/science.8346443

Court, F., Camprubi, C., Garcia, C. V., Guillaumet-Adkins, A., Sparago, A., Seruggia, D., Sandoval, J., Esteller, M., Martin-Trujillo, A., Riccio, A., Montoliu, L., & Monk, D. (2014). The PEG13-DMR and brain-specific enhancers dictate imprinted expression within the 8q24 intellectual disability risk locus. *Epigenetics and Chromatin*, *7*(1). https://doi.org/10.1186/1756-8935-7-5

Cruchaga, C., Karch, C. M., Jin, S. C., Benitez, B. A., Cai, Y., Guerreiro, R., Harari, O., Norton, J., Budde, J., Bertelsen, S., Jeng, A. T., Cooper, B., Skorupa, T., Carrell, D., Levitch, D., Hsu, S., Choi, J., Ryten, M., Hardy, J., … Goate, A. M. (2014). Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. *Nature*, *505*(7484). https://doi.org/10.1038/nature12825

Cubelos, B., Sebastián-Serrano, A., Beccari, L., Calcagnotto, M. E., Cisneros, E., Kim, S., Dopazo, A., Alvarez-Dolado, M., Redondo, J. M., Bovolenta, P., Walsh,

C. A., & Nieto, M. (2010). Cux1 and Cux2 regulate dendritic branching, spine morphology, and synapses of the upper layer neurons of the cortex. *Neuron*, *66*(4). https://doi.org/10.1016/j.neuron.2010.04.038

Darnell, R. B. (2013). RNA protein interaction in neurons. In *Annual Review of Neuroscience* (Vol. 36). https://doi.org/10.1146/annurev-neuro-062912-114322

De Jager, P. L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L. C., Yu, L., Eaton, M. L., Keenan, B. T., Ernst, J., McCabe, C., Tang, A., Raj, T., Replogle, J., Brodeur, W., Gabriel, S., Chai, H. S., Younkin, C., Younkin, S. G., Zou, F., … Bennett, D. A. (2014). Alzheimer's disease: Early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. *Nature Neuroscience*, *17*(9). https://doi.org/10.1038/nn.3786

De La Fuente, L., Arzalluz-Luque, Á., Tardáguila, M., Del Risco, H., Martí, C., Tarazona, S., Salguero, P., Scott, R., Lerma, A., Alastrue-Agudo, A., Bonilla, P., Newman, J. R. B., Kosugi, S., McIntyre, L. M., Moreno-Manzano, V., & Conesa, A. (2020). TappAS: A comprehensive computational framework for the analysis of the functional impact of differential splicing. *Genome Biology*, *21*(1). https://doi.org/10.1186/s13059-020-02028-w

De Roeck, A., Van den Bossche, T., van der Zee, J., Verheijen, J., De Coster, W., Van Dongen, J., Dillen, L., Baradaran-Heravi, Y., Heeman, B., Sanchez-Valle, R., Lladó, A., Nacmias, B., Sorbi, S., Gelpi, E., Grau-Rivera, O., Gómez-Tortosa, E., Pastor, P., Ortega-Cubero, S., Pastor, M. A., … Blesa, R. (2017). Deleterious ABCA7 mutations and transcript rescue mechanisms in early onset Alzheimer's disease. *Acta Neuropathologica*, *134*(3). https://doi.org/10.1007/s00401-017-1714-x

De Strooper, B., & Karran, E. (2016a). The Cellular Phase of Alzheimer's Disease. *Cell*, *164*(4), 603–615. https://doi.org/10.1016/j.cell.2015.12.056

De Strooper, B., & Karran, E. (2016b). The Cellular Phase of Alzheimer's Disease. In *Cell* (Vol. 164, Issue 4). https://doi.org/10.1016/j.cell.2015.12.056

Deaton, A. M., & Bird, A. (2011). CpG islands and the regulation of transcription. *Genes and Development*, *25*(10). https://doi.org/10.1101/gad.2037511

Drummond, E., Kavanagh, T., Pires, G., Marta-Ariza, M., Kanshin, E., Nayak, S., Faustin, A., Berdah, V., Ueberheide, B., & Wisniewski, T. (2022). The amyloid plaque proteome in early onset Alzheimer's disease and Down syndrome. *Acta Neuropathologica Communications*, *10*(1). https://doi.org/10.1186/s40478-022-01356-1

Du, Q., Luu, P. L., Stirzaker, C., & Clark, S. J. (2015). Methyl-CpG-binding domain proteins: Readers of the epigenome. In *Epigenomics* (Vol. 7, Issue 6). https://doi.org/10.2217/epi.15.39

Du, Y., Du, Y., Zhang, Y., Huang, Z., Fu, M., Li, J., Pang, Y., Lei, P., Wang, Y. T., Song, W., He, G., & Dong, Z. (2019). MKP-1 reduces aβ generation and alleviates cognitive impairments in Alzheimer's disease models. *Signal Transduction and Targeted Therapy*, *4*(1). https://doi.org/10.1038/s41392-019-0091-4

Dujardin, S., Commins, C., Lathuiliere, A., Beerepoot, P., Fernandes, A. R., Kamath, T. V., De Los Santos, M. B., Klickstein, N., Corjuc, D. L., Corjuc, B. T., Dooley, P. M., Viode, A., Oakley, D. H., Moore, B. D., Mullin, K., Jean-Gilles, D., Clark, R., Atchison, K., Moore, R., … Hyman, B. T. (2020). Tau molecular diversity contributes to clinical heterogeneity in Alzheimer's disease. *Nature Medicine*, *26*(8). https://doi.org/10.1038/s41591-020-0938-9

Ebbert, M. T. W., Jensen, T. D., Jansen-West, K., Sens, J. P., Reddy, J. S., Ridge, P. G., Kauwe, J. S. K., Belzil, V., Pregent, L., Carrasquillo, M. M., Keene, D., Larson, E., Crane, P., Asmann, Y. W., Ertekin-Taner, N., Younkin, S. G., Ross, O. A., Rademakers, R., Petrucelli, L., & Fryer, J. D. (2019). Systematic analysis of dark and camouflaged genes reveals disease-relevant genes hiding in plain sight. *Genome Biology*, *20*(1). https://doi.org/10.1186/s13059-019-1707-2

Edwalds-Gilbert, G., Veraldi, K. L., & Milcarek, C. (1997). Alternative poly(A) site selection in complex transcription units: Means to an end? *Nucleic Acids Research*, *25*(13). https://doi.org/10.1093/nar/25.13.2547

Edwards, G. A., Gamez, N., Escobedo, G., Calderon, O., & Moreno-Gonzalez, I. (2019). Modifiable risk factors for Alzheimer's disease. In *Frontiers in Aging Neuroscience* (Vol. 11, Issue JUN). https://doi.org/10.3389/fnagi.2019.00146

Efthymiou, A. G., & Goate, A. M. (2017). Late onset Alzheimer's disease genetics implicates microglial pathways in disease risk. In *Molecular Neurodegeneration* (Vol. 12, Issue 1). https://doi.org/10.1186/s13024-017-0184-x

Fan, C., Chen, K., Zhou, J., Wong, P. pui, He, D., Huang, Y., Wang, X., Ling, T., Yang, Y., & Zhao, H. (2021). Systematic analysis to identify transcriptome-wide dysregulation of Alzheimer's disease in genes and isoforms. *Human Genetics*, *140*(4). https://doi.org/10.1007/s00439-020-02230-7

Felsky, D., Roostaei, T., Nho, K., Risacher, S. L., Bradshaw, E. M., Petyuk, V., Schneider, J. A., Saykin, A., Bennett, D. A., & De Jager, P. L. (2019). Neuropathological correlates and genetic architecture of microglial activation in elderly human brain. *Nature Communications*, *10*(1). https://doi.org/10.1038/s41467-018-08279-3

Feng, H., Conneely, K. N., & Wu, H. (2014). A Bayesian hierarchical model to detect differentially methylated loci from single nucleotide resolution sequencing data. *Nucleic Acids Research*, *42*(8). https://doi.org/10.1093/nar/gku154

Frankish, A., Mudge, J. M., Thomas, M., & Harrow, J. (2012). The importance of identifying alternative splicing in vertebrate genome annotation. *Database*, *2012*. https://doi.org/10.1093/database/bas014

Fransquet, P. D., Lacaze, P., Saffery, R., Phung, J., Parker, E., Shah, R., Murray, A., Woods, R. L., & Ryan, J. (2020). Blood DNA methylation signatures to detect dementia prior to overt clinical symptoms. *Alzheimer's and Dementia: Diagnosis, Assessment and Disease Monitoring*, *12*(1). https://doi.org/10.1002/dad2.12056

Friedman, B. A., Srinivasan, K., Ayalon, G., Meilandt, W. J., Lin, H., Huntley, M. A., Cao, Y., Lee, S. H., Haddick, P. C. G., Ngu, H., Modrusan, Z., Larson, J. L., Kaminker, J. S., van der Brug, M. P., & Hansen, D. V. (2018). Diverse Brain Myeloid Expression Profiles Reveal Distinct Microglial Activation States and Aspects of Alzheimer's Disease Not Evident in Mouse Models. *Cell Reports*, *22*(3). https://doi.org/10.1016/j.celrep.2017.12.066

Gallego-Bartolomé, J., Gardiner, J., Liu, W., Papikian, A., Ghoshal, B., Kuo, H. Y., Zhao, J. M. C., Segal, D. J., & Jacobsen, S. E. (2018). Targeted DNA demethylation of the arabidopsis genome using the human TET1 catalytic domain. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(9). https://doi.org/10.1073/pnas.1716945115

Gallo, C. M., Kistler, S. A., Natrakul, A., Labadorf, A. T., Beffert, U., & Ho, A. (2024). APOER2 splicing repertoire in Alzheimer's disease: Insights from long-read RNA sequencing. *PLOS Genetics*, *20*(7), e1011348-. https://doi.org/10.1371/journal.pgen.1011348

Garcia-Esparcia, P., Sideris-Lampretsas, G., Hernandez-Ortega, K., Grau-Rivera, O., Sklaviadis, T., Gelpi, E., & Ferrer, I. (2017). Altered mechanisms of protein synthesis in frontal cortex in Alzheimer disease and a mouse model. *American Journal of Neurodegenerative Diseases*, *6*(2).

García-Ruiz, S., Zhang, D., Gustavsson, E. K., Rocamora-Perez, G., Grant-Peters, M., Fairbrother-Browne, A., Reynolds, R. H., Brenton, J. W., Gil-Martínez, A. L., Chen, Z., Rio, D. C., Botia, J. A., Guelfi, S., Collado-Torres, L., & Ryten, M. (2023). Splicing accuracy varies across human introns, tissues and age. *BioRxiv*, 2023.03.29.534370. https://doi.org/10.1101/2023.03.29.534370

Gasparoni, G., Bultmann, S., Lutsik, P., Kraus, T. F. J., Sordon, S., Vlcek, J., Dietinger, V., Steinmaurer, M., Haider, M., Mulholland, C. B., Arzberger, T., Roeber, S., Riemenschneider, M., Kretzschmar, H. A., Giese, A., Leonhardt, H., & Walter, J. (2018). DNA methylation analysis on purified neurons and glia dissects age and Alzheimer's disease-specific changes in the human cortex. *Epigenetics and Chromatin*, *11*(1). https://doi.org/10.1186/s13072-018-0211-3

Ghosh, A., Mizuno, K., Tiwari, S. S., Proitsi, P., Gomez Perez-Nievas, B., Glennon, E., Martinez-Nunez, R. T., & Giese, K. P. (2020). Alzheimer's disease-related dysregulation of mRNA translation causes key pathological features with ageing. *Translational Psychiatry*, *10*(1). https://doi.org/10.1038/s41398-020-00882-7

Glass, C. K., Saijo, K., Winner, B., Marchetto, M. C., & Gage, F. H. (2010). Mechanisms Underlying Inflammation in Neurodegeneration. In *Cell* (Vol. 140, Issue 6). https://doi.org/10.1016/j.cell.2010.02.016

Grubman, A., Chew, G., Ouyang, J. F., Sun, G., Choo, X. Y., McLean, C., Simmons, R. K., Buckberry, S., Vargas-Landin, D. B., Poppe, D., Pflueger, J., Lister, R., Rackham, O. J. L., Petretto, E., & Polo, J. M. (2019). A single-cell atlas of entorhinal cortex from individuals with Alzheimer's disease reveals cell-type-specific gene expression regulation. *Nature Neuroscience*, *22*(12). https://doi.org/10.1038/s41593-019-0539-4

Gruijs da Silva, L. A., Simonetti, F., Hutten, S., Riemenschneider, H., Sternburg, E. L., Pietrek, L. M., Gebel, J., Dötsch, V., Edbauer, D., Hummer, G., Stelzl, L. S., & Dormann, D. (2022). Disease-linked TDP-43 hyperphosphorylation suppresses TDP-43 condensation and aggregation. *The EMBO Journal*, *41*(8). https://doi.org/10.15252/embj.2021108443

Gu, Z., & Hübschmann, D. (2023a). rGREAT: an R/bioconductor package for functional enrichment on genomic regions. *Bioinformatics*, *39*(1). https://doi.org/10.1093/bioinformatics/btac745

Gu, Z., & Hübschmann, D. (2023b). simplifyEnrichment: A Bioconductor Package for Clustering and Visualizing Functional Enrichment Results. *Genomics, Proteomics and Bioinformatics*, *21*(1). https://doi.org/10.1016/j.gpb.2022.04.008

Guennewig, B., Lim, J., Marshall, L., McCorkindale, A. N., Paasila, P. J., Patrick, E., Kril, J. J., Halliday, G. M., Cooper, A. A., & Sutherland, G. T. (2021). Defining early changes in Alzheimer's disease from RNA sequencing of brain regions differentially affected by pathology. *Scientific Reports*, *11*(1). https://doi.org/10.1038/s41598-021-83872-z

Hansen, K. D., Langmead, B., & Irizarry, R. A. (2012). BSmooth: from whole genome bisulfite sequencing reads to differentially methylated regions. *Genome Biology*, *13*(10). https://doi.org/10.1186/gb-2012-13-10-R83

Hansen, D. V., Hanson, J. E., & Sheng, M. (2018). Microglia in Alzheimer's disease. In *Journal of Cell Biology* (Vol. 217, Issue 2). https://doi.org/10.1083/jcb.201709069

Harris, M. A., Clark, J., Ireland, A., Lomax, J., Ashburner, M., Foulger, R., Eilbeck, K., Lewis, S., Marshall, B., Mungall, C., Richter, J., Rubin, G. M., Blake, J. A., Bult, C., Dolan, M., Drabkin, H., Eppig, J. T., Hill, D. P., Ni, L., … White, R. (2004). The Gene Oncology (GO) database and informatics resource. *Nucleic Acids Research*, *32*(DATABASE ISS.). https://doi.org/10.1093/nar/gkh036

Hartl, D., May, P., Gu, W., Mayhaus, M., Pichler, S., Spaniol, C., Glaab, E., Bobbili, D. R., Antony, P., Koegelsberger, S., Kurz, A., Grimmer, T., Morgan, K., Vardarajan, B. N., Reitz, C., Hardy, J., Bras, J., Guerreiro, R., Balling, R., … Singleton, A. (2020). A rare loss-of-function variant of ADAM17 is associated with late-onset familial Alzheimer disease. *Molecular Psychiatry*, *25*(3). https://doi.org/10.1038/s41380-018-0091-8

He, Z., Guo, J. L., McBride, J. D., Narasimhan, S., Kim, H., Changolkar, L., Zhang, B., Gathagan, R. J., Yue, C., Dengler, C., Stieber, A., Nitla, M., Coulter, D. A., Abel, T., Brunden, K. R., Trojanowski, J. Q., & Lee, V. M. Y. (2018). Amyloid-β plaques enhance Alzheimer's brain tau-seeded pathologies by facilitating neuritic plaque tau aggregation. *Nature Medicine*, *24*(1). https://doi.org/10.1038/nm.4443

Holtzman, D. M., Bales, K. R., Tenkova, T., Fagan, A. M., Parsadanian, M., Sartorius, L. J., Mackey, B., Olney, J., McKeel, D., Wozniak, D., & Paul, S. M. (2000). Apolipoprotein E isoform-dependent amyloid deposition and neuritic degeneration in a mouse model of Alzheimer's disease. *Proceedings of the National Academy of Sciences of the United States of America*, *97*(6). https://doi.org/10.1073/pnas.050004797

Hong, S., Beja-Glasser, V. F., Nfonoyim, B. M., Frouin, A., Li, S., Ramakrishnan, S., Merry, K. M., Shi, Q., Rosenthal, A., Barres, B. A., Lemere, C. A., Selkoe, D. J., & Stevens, B. (2016). Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science*, *352*(6286). https://doi.org/10.1126/science.aad8373

Hsiao, Y. C., Muñoz-Estrada, J., Tuz, K., & Ferland, R. J. (2021). The transition zone protein AHI1 regulates neuronal ciliary trafficking of MCHR1 and its downstream signaling pathway. *Journal of Neuroscience*, *41*(17). https://doi.org/10.1523/JNEUROSCI.2993-20.2021

Huang, K. L., Marcora, E., Pimenova, A. A., Di Narzo, A. F., Kapoor, M., Jin, S. C., Harari, O., Bertelsen, S., Fairfax, B. P., Czajkowski, J., Chouraki, V., Grenier-Boley, B., Bellenguez, C., Deming, Y., McKenzie, A., Raj, T., Renton, A. E., Budde, J., Smith, A., … Goate, A. M. (2017). A common haplotype lowers PU.1 expression in myeloid cells and delays onset of Alzheimer's disease. *Nature Neuroscience*, *20*(8). https://doi.org/10.1038/nn.4587

Irie, T., Matsuda, T., Hayashi, Y., Matsuda-Ito, K., Kamiya, A., Masuda, T., Prinz, M., Isobe, N., Kira, J. I., & Nakashima, K. (2023). Direct neuronal conversion of microglia/macrophages reinstates neurological function after stroke. *Proceedings of the National Academy of Sciences of the United States of America*, *120*(42). https://doi.org/10.1073/pnas.2307972120

Iwata, A., Nagata, K., Hatsuta, H., Takuma, H., Bundo, M., Iwamoto, K., Tamaoka, A., Murayama, S., Saido, T., & Tsuji, S. (2014). Altered CpG methylation in sporadic alzheimer's disease is associated with APP and MAPT dysregulation. *Human Molecular Genetics*, *23*(3). https://doi.org/10.1093/hmg/ddt451

Jaffrey, S. R., & Wilkinson, M. F. (2018). Nonsense-mediated RNA decay in the brain: emerging modulator of neural development and disease. In *Nature*

Reviews Neuroscience (Vol. 19, Issue 12). https://doi.org/10.1038/s41583-018-0079-z

Jones, P. A. (2012). Functions of DNA methylation: Islands, start sites, gene bodies and beyond. In *Nature Reviews Genetics* (Vol. 13, Issue 7). https://doi.org/10.1038/nrg3230

Kang, S., Gim, J., Lee, J., Gunasekaran, T. I., Choi, K. Y., Lee, J. J., Seo, E. H., Ko, P. W., Chung, J. Y., Choi, S. M., Lee, Y. M., Jeong, J. H., Park, K. W., Song, M. K., Lee, H. W., Kim, K. W., Choi, S. H., Lee, D. Y., Kim, S. Y., … Lee, K. H. (2021). Potential novel genes for late-onset alzheimer's disease in east-asian descent identified by APOE-Stratified genome-wide association study. *Journal of Alzheimer's Disease*, *82*(4). https://doi.org/10.3233/JAD-210145

Kang, Y. J., Yang, D. C., Kong, L., Hou, M., Meng, Y. Q., Wei, L., & Gao, G. (2017). CPC2: A fast and accurate coding potential calculator based on sequence intrinsic features. *Nucleic Acids Research*, *45*(W1). https://doi.org/10.1093/nar/gkx428

Karch, C. M., & Goate, A. M. (2015). Alzheimer's disease risk genes and mechanisms of disease pathogenesis. In *Biological Psychiatry* (Vol. 77, Issue 1). https://doi.org/10.1016/j.biopsych.2014.05.006

Keren-Shaul, H., Spinrad, A., Weiner, A., Matcovitch-Natan, O., Dvir-Szternfeld, R., Ulland, T. K., David, E., Baruch, K., Lara-Astaiso, D., Toth, B., Itzkovitz, S., Colonna, M., Schwartz, M., & Amit, I. (2017). A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell*, *169*(7). https://doi.org/10.1016/j.cell.2017.05.018

Khandelwal, N., Kulkarni, A., Ahmed, N. I., Harper, M., Konopka, G., & Gibson, J. R. (2024a). FOXP1 regulates the development of excitatory synaptic inputs onto striatal neurons and induces phenotypic reversal with reinstatement. *Science Advances*, *10*(18), eadm7039. https://doi.org/10.1126/sciadv.adm7039

Khandelwal, N., Kulkarni, A., Ahmed, N. I., Harper, M., Konopka, G., & Gibson, J. R. (2024b). FOXP1 regulates the development of excitatory synaptic inputs onto striatal neurons and induces phenotypic reversal with reinstatement. *Science Advances*, *10*(18), eadm7039. https://doi.org/10.1126/sciadv.adm7039

Kim-Hellmuth, S., Aguet, F., Oliva, M., Muñoz-Aguirre, M., Kasela, S., Wucher, V., Castel, S. E., Hamel, A. R., Viñuela, A., Roberts, A. L., Mangul, S., Wen, X., Wang, G., Barbeira, A. N., Garrido-Martín, D., Nadel, B. B., Zou, Y., Bonazzola, R., Quan, J., … Volpi, S. (2020). Cell type–specific genetic regulation of gene expression across human tissues. *Science*, *369*(6509). https://doi.org/10.1126/SCIENCE.AAZ8528

King, D., Holt, K., Toombs, J., Xin, H. E., Dando, O., Okely, J. A., Tzioras, M., Rose, J., Gunn, C., Correia, A., Montero, C., McAlister, H., Tulloch, J., Lamont, D., Taylor, A. M., Harris, S. E., Redmond, P., Cox, S. R., Henstridge, C. M., … Spires-Jones, T. L. (2023). Synaptic resilience is associated with maintained cognition during ageing. *Alzheimer's and Dementia*, *19*(6). https://doi.org/10.1002/alz.12894

Klein, H. U., Bennett, D. A., & De Jager, P. L. (2016). The epigenome in Alzheimer's disease: current state and approaches for a new path to gene discovery and understanding disease mechanism. In *Acta Neuropathologica* (Vol. 132, Issue 4). https://doi.org/10.1007/s00401-016-1612-7

Kobayashi, N., Shinagawa, S., Niimura, H., Kida, H., Nagata, T., Tagai, K., Shimada, K., Oka, N., Shikimoto, R., Noda, Y., Nakajima, S., Mimura, M., Shigeta, M., & Kondo, K. (2020). Increased blood COASY DNA methylation levels a potential

biomarker for early pathology of Alzheimer's disease. *Scientific Reports*, *10*(1). https://doi.org/10.1038/s41598-020-69248-9
Komine, Y., Takao, K., Miyakawa, T., & Yamamori, T. (2012). Behavioral Abnormalities Observed in Zfhx2-Deficient Mice. *PLoS ONE*, *7*(12). https://doi.org/10.1371/journal.pone.0053114
Kong, J., Ren, G., Jia, N., Wang, Y., Zhang, H., Zhang, W., Chen, B., & Cao, Y. (2013). Effects of nicorandil in neuroprotective activation of pi3k/akt pathways in a cellular model of alzheimer's disease. *European Neurology*, *70*(3–4). https://doi.org/10.1159/000351247
Köster, J., & Rahmann, S. (2012). Snakemake-a scalable bioinformatics workflow engine. *Bioinformatics*, *28*(19). https://doi.org/10.1093/bioinformatics/bts480
Kulczyńska-Przybik, A., Dulewicz, M., Słowik, A., Borawska, R., Kułakowska, A., Kochanowicz, J., & Mroczko, B. (2021). The clinical significance of cerebrospinal fluid reticulon 4 (Rtn4) levels in the differential diagnosis of neurodegenerative diseases. *Journal of Clinical Medicine*, *10*(22). https://doi.org/10.3390/jcm10225281
Kunkle, B. W., Grenier-Boley, B., Sims, R., Bis, J. C., Damotte, V., Naj, A. C., Boland, A., Vronskaya, M., van der Lee, S. J., Amlie-Wolf, A., Bellenguez, C., Frizatti, A., Chouraki, V., Martin, E. R., Sleegers, K., Badarinarayan, N., Jakobsdottir, J., Hamilton-Nelson, K. L., Moreno-Grau, S., … Pericak-Vance, M. A. (2019). Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nature Genetics*, *51*(3). https://doi.org/10.1038/s41588-019-0358-2
Kwan, T., Benovoy, D., Dias, C., Gurd, S., Provencher, C., Beaulieu, P., Hudson, T. J., Sladek, R., & Majewski, J. (2008). Genome-wide analysis of transcript isoform variation in humans. *Nature Genetics*, *40*(2). https://doi.org/10.1038/ng.2007.57
Lambert, J. C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., Jun, G., DeStefano, A. L., Bis, J. C., Beecham, G. W., Grenier-Boley, B., Russo, G., Thornton-Wells, T. A., Jones, N., Smith, A. V., Chouraki, V., Thomas, C., Ikram, M. A., Zelenika, D., … Seshadri, S. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature Genetics*, *45*(12). https://doi.org/10.1038/ng.2802
Lambert, M., Jambon, S., Depauw, S., & David-Cordonnier, M. H. (2018). Targeting transcription factors for cancer treatment. In *Molecules* (Vol. 23, Issue 6). https://doi.org/10.3390/molecules23061479
Lardenoije, R., Roubroeks, J. A. Y., Pishva, E., Leber, M., Wagner, H., Iatrou, A., Smith, A. R., Smith, R. G., Eijssen, L. M. T., Kleineidam, L., Kawalia, A., Hoffmann, P., Luck, T., Riedel-Heller, S., Jessen, F., Maier, W., Wagner, M., Hurlemann, R., Kenis, G., … Van Den Hove, D. L. A. (2019). Alzheimer's disease-associated (hydroxy)methylomic changes in the brain and blood. *Clinical Epigenetics*, *11*(1). https://doi.org/10.1186/s13148-019-0755-5
Lassar, A. B., Martin, P. L., & Roeder, R. G. (2019). Transcription of class iii genes: Formation of preinitiation complexes. In *Biotechnology and Biological Frontiers*. https://doi.org/10.4324/9780429050329-27
Lauerer, R. J., & Lerche, H. (2023). Voltage-gated calcium channels in genetic epilepsies. In *Journal of Neurochemistry*. https://doi.org/10.1111/jnc.15983
Lee, S. H., Meilandt, W. J., Xie, L., Gandham, V. D., Ngu, H., Barck, K. H., Rezzonico, M. G., Imperio, J., Lalehzadeh, G., Huntley, M. A., Stark, K. L., Foreman, O., Carano, R. A. D., Friedman, B. A., Sheng, M., Easton, A., Bohlen,

C. J., & Hansen, D. V. (2021). Trem2 restrains the enhancement of tau accumulation and neurodegeneration by β-amyloid pathology. *Neuron*, *109*(8). https://doi.org/10.1016/j.neuron.2021.02.010

Leng, K., Li, E., Eser, R., Piergies, A., Sit, R., Tan, M., Neff, N., Li, S. H., Rodriguez, R. D., Suemoto, C. K., Leite, R. E. P., Ehrenberg, A. J., Pasqualucci, C. A., Seeley, W. W., Spina, S., Heinsen, H., Grinberg, L. T., & Kampmann, M. (2021). Molecular characterization of selectively vulnerable neurons in Alzheimer's disease. *Nature Neuroscience*, *24*(2). https://doi.org/10.1038/s41593-020-00764-7

Leung, S. K., Jeffries, A. R., Castanho, I., Bamford, R. A., Moore, K., Dempster, E. L., Brown, J. T., Ahmed, Z., O'Neill, P., Hannon, E., & Mill, J. (2023). Long-read transcript sequencing identifies differential isoform expression in the entorhinal cortex in a transgenic model of tau pathology. *BioRxiv*, 2023.09.20.558220. https://doi.org/10.1101/2023.09.20.558220

Leung, S. K., Jeffries, A. R., Castanho, I., Jordan, B. T., Moore, K., Davies, J. P., Dempster, E. L., Bray, N. J., O'Neill, P., Tseng, E., Ahmed, Z., Collier, D. A., Jeffery, E. D., Prabhakar, S., Schalkwyk, L., Jops, C., Gandal, M. J., Sheynkman, G. M., Hannon, E., & Mill, J. (2021). Full-length transcript sequencing of human and mouse cerebral cortex identifies widespread isoform diversity and alternative splicing. *Cell Reports*, *37*(7). https://doi.org/10.1016/j.celrep.2021.110022

Leyns, C. E. G., Gratuze, M., Narasimhan, S., Jain, N., Koscal, L. J., Jiang, H., Manis, M., Colonna, M., Lee, V. M. Y., Ulrich, J. D., & Holtzman, D. M. (2019). TREM2 function impedes tau seeding in neuritic plaques. *Nature Neuroscience*, *22*(8). https://doi.org/10.1038/s41593-019-0433-0

Li, D., McIntosh, C. S., Mastaglia, F. L., Wilton, S. D., & Aung-Htut, M. T. (2021). Neurodegenerative diseases: a hotbed for splicing defects and the potential therapies. In *Translational Neurodegeneration* (Vol. 10, Issue 1). https://doi.org/10.1186/s40035-021-00240-7

Li, H. (2018). Minimap2: Pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18). https://doi.org/10.1093/bioinformatics/bty191

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G., & Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, *25*(16). https://doi.org/10.1093/bioinformatics/btp352

Li, J., Zhang, W., Yang, H., Howrigan, D. P., Wilkinson, B., Souaiaia, T., Evgrafov, O. V., Genovese, G., Clementel, V. A., Tudor, J. C., Abel, T., Knowles, J. A., Neale, B. M., Wang, K., Sun, F., & Coba, M. P. (2017). Spatiotemporal profile of postsynaptic interactomes integrates components of complex brain disorders. *Nature Neuroscience*, *20*(8). https://doi.org/10.1038/nn.4594

Liang, Y., Ryan, N. S., Schott, J. M., & Fox, N. C. (2013). Imaging the onset and progression of Alzheimer's disease: Implications for prevention trials. In *Journal of Alzheimer's Disease* (Vol. 33, Issue SUPPL. 1). https://doi.org/10.3233/JAD-2012-129010

Lin, K. T., & Krainer, A. R. (2019). PSI-Sigma: A comprehensive splicing-detection method for short-read and long-read RNA-seq analysis. *Bioinformatics*, *35*(23). https://doi.org/10.1093/bioinformatics/btz438

Liu, W., Taso, O., Wang, R., Bayram, S., Graham, A. C., Garcia-Reitboeck, P., Mallach, A., Andrews, W. D., Piers, T. M., Botia, J. A., Pocock, J. M., Cummings, D. M., Hardy, J., Edwards, F. A., & Salih, D. A. (2020). Trem2 promotes anti-inflammatory responses in microglia and is suppressed under

pro-inflammatory conditions. *Human Molecular Genetics*, *29*(19). https://doi.org/10.1093/hmg/ddaa209
Liu, X., Jiao, B., & Shen, L. (2018). The Epigenetics of Alzheimer's Disease: Factors and Therapeutic Implications. In *Frontiers in Genetics* (Vol. 9). https://doi.org/10.3389/fgene.2018.00579
Liu, Y., Rosikiewicz, W., Pan, Z., Jillette, N., Wang, P., Taghbalout, A., Foox, J., Mason, C., Carroll, M., Cheng, A., & Li, S. (2021). DNA methylation-calling tools for Oxford Nanopore sequencing: a survey and human epigenome-wide evaluation. *Genome Biology*, *22*(1). https://doi.org/10.1186/s13059-021-02510-z
Liu, Y., Yu, J. T., Wang, H. F., Han, P. R., Tan, C. C., Wang, C., Meng, X. F., Risacher, S. L., Saykin, A. J., & Tan, L. (2015). APOE genotype and neuroimaging markers of Alzheimer's disease: Systematic review and meta-analysis. *Journal of Neurology, Neurosurgery and Psychiatry*, *86*(2). https://doi.org/10.1136/jnnp-2014-307719
Logsdon, G. A., Vollger, M. R., & Eichler, E. E. (2020). Long-read human genome sequencing and its applications. In *Nature Reviews Genetics* (Vol. 21, Issue 10). https://doi.org/10.1038/s41576-020-0236-x
Louis Sam Titus, A. S. C., Yusuff, T., Cassar, M., Thomas, E., Kretzschmar, D., & D'Mello, S. R. (2017). Reduced Expression of Foxp1 as a Contributing Factor in Huntington's Disease. *The Journal of Neuroscience*, *37*(27), 6575. https://doi.org/10.1523/JNEUROSCI.3612-16.2017
Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biology*, *15*(12). https://doi.org/10.1186/s13059-014-0550-8
Love, M. I., Soneson, C., & Patro, R. (2018). Swimming downstream: Statistical analysis of differential transcript usage following salmon quantification [version 1; peer review: 3 approved with reservations]. *F1000Research*, *7*. https://doi.org/10.12688/F1000RESEARCH.15398.1
Lubliner, S., Regev, I., Lotan-Pompan, M., Edelheit, S., Weinberger, A., & Segal, E. (2015). Core promoter sequence in yeast is a major determinant of expression level. *Genome Research*, *25*(7). https://doi.org/10.1101/gr.188193.114
Luecken, M. D., & Theis, F. J. (2019). Current best practices in single-cell RNA-seq analysis: a tutorial. *Molecular Systems Biology*, *15*(6). https://doi.org/10.15252/msb.20188746
Lunnon, K., Smith, R., Hannon, E., De Jager, P. L., Srivastava, G., Volta, M., Troakes, C., Al-Sarraj, S., Burrage, J., Macdonald, R., Condliffe, D., Harries, L. W., Katsel, P., Haroutunian, V., Kaminsky, Z., Joachim, C., Powell, J., Lovestone, S., Bennett, D. A., … Mill, J. (2014). Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. *Nature Neuroscience*, *17*(9). https://doi.org/10.1038/nn.3782
Lüth, T., Klein, C., Schaake, S., Tse, R., Pereira, S., Lass, J., Sinkkonen, L., Grünewald, A., & Trinh, J. (2021). Analysis of mitochondrial genome methylation using Nanopore single-molecule sequencing. *BioRxiv*.
Madrid, A., Hogan, K. J., Papale, L. A., Clark, L. R., Asthana, S., Johnson, S. C., & Alisch, R. S. (2018). DNA Hypomethylation in Blood Links B3GALT4 and ZADH2 to Alzheimer's Disease. *Journal of Alzheimer's Disease*, *66*(3). https://doi.org/10.3233/JAD-180592
Marques-Coelho, D., Iohan, L. da C. C., Melo de Farias, A. R., Flaig, A., Letournel, F., Martin-Négrier, M. L., Chapon, F., Faisant, M., Godfraind, C., Maurage, C. A., Deramecourt, V., Duchesne, M., Meyronnet, D., Streichenberger, N., de

Paula, A. M., Rigau, V., Vandenbos-Burel, F., Duyckaerts, C., Seilhean, D., … Costa, M. R. (2021). Differential transcript usage unravels gene expression alterations in Alzheimer's disease human brains. *Npj Aging and Mechanisms of Disease*, *7*(1). https://doi.org/10.1038/s41514-020-00052-5

Marzi, S. J., Leung, S. K., Ribarska, T., Hannon, E., Smith, A. R., Pishva, E., Poschmann, J., Moore, K., Troakes, C., Al-Sarraj, S., Beck, S., Newman, S., Lunnon, K., Schalkwyk, L. C., & Mill, J. (2018). A histone acetylome-wide association study of Alzheimer's disease identifies disease-associated H3K27ac differences in the entorhinal cortex. *Nature Neuroscience*, *21*(11). https://doi.org/10.1038/s41593-018-0253-7

Masliah, E., Xie, F., Dayan, S., Rockenstein, E., Mante, M., Adame, A., Patrick, C. M., Chan, A. F., & Zheng, B. (2010). Genetic deletion of Nogo/Rtn4 ameliorates behavioral and neuropathological outcomes in amyloid precursor protein transgenic mice. *Neuroscience*, *169*(1). https://doi.org/10.1016/j.neuroscience.2010.04.045

Matarin, M., Salih, D. A., Yasvoina, M., Cummings, D. M., Guelfi, S., Liu, W., NahabooSolim, M. A., Moens, T. G., Paublete, R. M., Ali, S. S., Perona, M., Desai, R., Smith, K. J., Latcham, J., Fulleylove, M., Richardson, J. C., Hardy, J., & Edwards, F. A. (2015). A Genome-wide gene-expression analysis and database in transgenic mice during development of amyloid or tau pathology. *Cell Reports*, *10*(4). https://doi.org/10.1016/j.celrep.2014.12.041

Mathys, H., Davila-Velderrain, J., Peng, Z., Gao, F., Mohammadi, S., Young, J. Z., Menon, M., He, L., Abdurrob, F., Jiang, X., Martorell, A. J., Ransohoff, R. M., Hafler, B. P., Bennett, D. A., Kellis, M., & Tsai, L. H. (2019). Single-cell transcriptomic analysis of Alzheimer's disease. *Nature*, *570*(7761). https://doi.org/10.1038/s41586-019-1195-2

McNair, K., Spike, R., Guilding, C., Prendergast, G. C., Stone, T. W., Cobb, S. R., & Morris, B. J. (2010). A role for RhoB in synaptic plasticity and the regulation of neuronal morphology. *Journal of Neuroscience*, *30*(9). https://doi.org/10.1523/JNEUROSCI.5386-09.2010

Mészáros, B., Erdös, G., & Dosztányi, Z. (2018). IUPred2A: Context-dependent prediction of protein disorder as a function of redox state and protein binding. *Nucleic Acids Research*, *46*(W1). https://doi.org/10.1093/nar/gky384

Miller, J. A., Cai, C., Langfelder, P., Geschwind, D. H., Kurian, S. M., Salomon, D. R., & Horvath, S. (2011). Strategies for aggregating gene expression data: The collapseRows R function. *BMC Bioinformatics*, *12*. https://doi.org/10.1186/1471-2105-12-322

Mills, J. D., Nalpathamkalam, T., Jacobs, H. I. L., Janitz, C., Merico, D., Hu, P., & Janitz, M. (2013). RNA-Seq analysis of the parietal cortex in Alzheimer's disease reveals alternatively spliced isoforms related to lipid metabolism. *Neuroscience Letters*, *536*(1). https://doi.org/10.1016/j.neulet.2012.12.042

Mistry, J., Chuguransky, S., Williams, L., Qureshi, M., Salazar, G. A., Sonnhammer, E. L. L., Tosatto, S. C. E., Paladin, L., Raj, S., Richardson, L. J., Finn, R. D., & Bateman, A. (2021). Pfam: The protein families database in 2021. *Nucleic Acids Research*, *49*(D1). https://doi.org/10.1093/nar/gkaa913

Mitsumori, R., Sakaguchi, K., Shigemizu, D., Mori, T., Akiyama, S., Ozaki, K., Niida, S., & Shimoda, N. (2020). Lower DNA methylation levels in CpG island shores of CR1, CLU, and PICALM in the blood of Japanese Alzheimer's disease patients. *PLoS ONE*, *15*(9 September). https://doi.org/10.1371/journal.pone.0239196

Moore, L. D., Le, T., & Fan, G. (2013). DNA methylation and its basic function. In *Neuropsychopharmacology* (Vol. 38, Issue 1). https://doi.org/10.1038/npp.2012.112

Mordaunt, C. E., Mouat, J. S., Schmidt, R. J., & Lasalle, J. M. (2022). Comethyl: A network-based methylome approach to investigate the multivariate nature of health and disease. *Briefings in Bioinformatics*, *23*(2). https://doi.org/10.1093/bib/bbab554

Morita, S., Horii, T., Kimura, M., & Hatada, I. (2020). Synergistic upregulation of target genes by TET1 and VP64 in the dCas9–suntag platform. *International Journal of Molecular Sciences*, *21*(5). https://doi.org/10.3390/ijms21051574

Mossink, B., van Rhijn, J. R., Wang, S., Linda, K., Vitale, M. R., Zöller, J. E. M., van Hugte, E. J. H., Bak, J., Verboven, A. H. A., Selten, M., Negwer, M., Latour, B. L., van der Werf, I., Keller, J. M., Klein Gunnewiek, T. M., Schoenmaker, C., Oudakker, A., Anania, A., Jansen, S., … Nadif Kasri, N. (2022). Cadherin-13 is a critical regulator of GABAergic modulation in human stem-cell-derived neuronal networks. *Molecular Psychiatry*, *27*(1). https://doi.org/10.1038/s41380-021-01117-x

Moutinho, M., Coronel, I., Tsai, A. P., Di Prisco, G. V., Pennington, T., Atwood, B. K., Puntambekar, S. S., Smith, D. C., Martinez, P., Han, S., Lee, Y., Lasagna-Reeves, C. A., Lamb, B. T., Bissel, S. J., Nho, K., & Landreth, G. E. (2023). TREM2 splice isoforms generate soluble TREM2 species that disrupt long-term potentiation. *Genome Medicine*, *15*(1). https://doi.org/10.1186/s13073-023-01160-z

Mudge, J. M., Frankish, A., Fernandez-Banet, J., Alioto, T., Derrien, T., Howald, C., Reymond, A., Guigó, R., Hubbard, T., & Harrow, J. (2011). The origins, evolution, and functional potential of alternative splicing in vertebrates. *Molecular Biology and Evolution*, *28*(10). https://doi.org/10.1093/molbev/msr127

Ni, P., Nie, F., Zhong, Z., Xu, J., Huang, N., Zhang, J., Zhao, H., Zou, Y., Huang, Y., Li, J., Xiao, C. Le, Luo, F., & Wang, J. (2023). DNA 5-methylcytosine detection and methylation phasing using PacBio circular consensus sequencing. *Nature Communications*, *14*(1). https://doi.org/10.1038/s41467-023-39784-9

Nichols, E., Steinmetz, J. D., Vollset, S. E., Fukutaki, K., Chalek, J., Abd-Allah, F., Abdoli, A., Abualhasan, A., Abu-Gharbieh, E., Akram, T. T., Al Hamad, H., Alahdab, F., Alanezi, F. M., Alipour, V., Almustanyir, S., Amu, H., Ansari, I., Arabloo, J., Ashraf, T., … Vos, T. (2022). Estimation of the global prevalence of dementia in 2019 and forecasted prevalence in 2050: an analysis for the Global Burden of Disease Study 2019. *The Lancet Public Health*, *7*(2). https://doi.org/10.1016/S2468-2667(21)00249-8

Nott, A., Holtman, I. R., Coufal, N. G., Schlachetzki, J. C. M., Yu, M., Hu, R., Han, C. Z., Pena, M., Xiao, J., Wu, Y., Keuelen, Z., Pasillas, M. P., O'Connor, C., Schafer, S. T., Shen, Z., Rissman, R. A., Brewer, J. B., Gosselin, D., Gonda, D. D., … Glass, C. K. (2019). Cell type-specific enhancer-promoter connectivity maps in the human brain and disease risk association. *BioRxiv*, *0793*(November).

Nott, A., Holtman, I. R., Coufal, N. G., Schlachetzki, J. C. M., Yu, M., Hu, R., Han, C. Z., Pena, M., Xiao, J., Wu, Y., Keulen, Z., Pasillas, M. P., O'Connor, C., Nickl, C. K., Schafer, S. T., Shen, Z., Rissman, R. A., Brewer, J. B., Gosselin, D., … Glass, C. K. (2019). Brain cell type–specific enhancer–promoter interactome maps and disease-risk association. *Science*, *366*(6469). https://doi.org/10.1126/science.aay0793

Oakley, H., Cole, S. L., Logan, S., Maus, E., Shao, P., Craft, J., Guillozet-Bongaarts, A., Ohno, M., Disterhoft, J., Van Eldik, L., Berry, R., & Vassar, R. (2006). Intraneuronal β-amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: Potential factors in amyloid plaque formation. *Journal of Neuroscience*, *26*(40). https://doi.org/10.1523/JNEUROSCI.1202-06.2006

Ohno, M. (2018). PERK as a hub of multiple pathogenic pathways leading to memory deficits and neurodegeneration in Alzheimer's disease. In *Brain Research Bulletin* (Vol. 141). https://doi.org/10.1016/j.brainresbull.2017.08.007

Olah, M., Menon, V., Habib, N., Taga, M. F., Ma, Y., Yung, C. J., Cimpean, M., Khairallah, A., Coronas-Samano, G., Sankowski, R., Grün, D., Kroshilina, A. A., Dionne, D., Sarkis, R. A., Cosgrove, G. R., Helgager, J., Golden, J. A., Pennell, P. B., Prinz, M., … De Jager, P. L. (2020). Single cell RNA sequencing of human microglia uncovers a subset associated with Alzheimer's disease. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-19737-2

Oliveros, G., Wallace, C. H., Chaudry, O., Liu, Q., Qiu, Y., Xie, L., Rockwell, P., Figueiredo-Pereira, M. E., & Serrano, P. A. (2023). Repurposing ibudilast to mitigate Alzheimer's disease by targeting inflammation. *Brain*, *146*(3). https://doi.org/10.1093/brain/awac136

Ouellette, A. R., Neuner, S. M., Dumitrescu, L., Anderson, L. C., Gatti, D. M., Mahoney, E. R., Bubier, J. A., Churchill, G., Peters, L., Huentelman, M. J., Herskowitz, J. H., Yang, H. S., Smith, A. N., Reitz, C., Kunkle, B. W., White, C. C., De Jager, P. L., Schneider, J. A., Bennett, D. A., … Kaczorowski, C. C. (2020). Cross-Species Analyses Identify Dlgap2 as a Regulator of Age-Related Cognitive Decline and Alzheimer's Dementia. *Cell Reports*, *32*(9). https://doi.org/10.1016/j.celrep.2020.108091

Ouyang, X., Wani, W. Y., Benavides, G. A., Redmann, M. J., Vo, H., van Groen, T., Darley-Usmar, V. M., & Zhang, J. (2023). Cathepsin D overexpression in the nervous system rescues lethality and Aβ42 accumulation of cathepsin D systemic knockout in vivo. *Acta Pharmaceutica Sinica B*, *13*(10). https://doi.org/10.1016/j.apsb.2023.07.015

Pankiewicz, J. E., Baquero-Buitrago, J., Sanchez, S., Lopez-Contreras, J., Kim, J., Sullivan, P. M., Holtzman, D. M., & Sadowski, M. J. (2017). APOE Genotype Differentially Modulates Effects of Anti-Aβ, Passive Immunization in APP Transgenic Mice. *Molecular Neurodegeneration*, *12*(1). https://doi.org/10.1186/s13024-017-0156-1

Parachikova, A., Agadjanyan, M. G., Cribbs, D. H., Blurton-Jones, M., Perreau, V., Rogers, J., Beach, T. G., & Cotman, C. W. (2007). Inflammatory changes parallel the early stages of Alzheimer disease. *Neurobiology of Aging*, *28*(12). https://doi.org/10.1016/j.neurobiolaging.2006.08.014

Park, P. J. (2009). ChIP-seq: Advantages and challenges of a maturing technology. In *Nature Reviews Genetics* (Vol. 10, Issue 10). https://doi.org/10.1038/nrg2641

Park, Y., & Wu, H. (2016). Differential methylation analysis for BS-seq data under general experimental design. *Bioinformatics*, *32*(10). https://doi.org/10.1093/bioinformatics/btw026

Patro, R., Duggal, G., Love, M. I., Irizarry, R. A., & Kingsford, C. (2017). Salmon provides fast and bias-aware quantification of transcript expression. *Nature Methods*, *14*(4). https://doi.org/10.1038/nmeth.4197

Peng, X., Guo, H., Zhang, X., Yang, Z., Ruganzu, J. B., Yang, Z., Wu, X., Bi, W., Ji, S., & Yang, W. (2023). TREM2 Inhibits Tau Hyperphosphorylation and Neuronal

Apoptosis via the PI3K/Akt/GSK-3β Signaling Pathway In vivo and In vitro. *Molecular Neurobiology*, *60*(5). https://doi.org/10.1007/s12035-023-03217-x
Pertea, G., & Pertea, M. (2020). GFF Utilities: GffRead and GffCompare. *F1000Research*, *9*. https://doi.org/10.12688/f1000research.23297.2
Pertea, M., Kim, D., Pertea, G. M., Leek, J. T., & Salzberg, S. L. (2016). Transcript-level expression analysis of RNA-seq experiments with HISAT, StringTie and Ballgown. *Nature Protocols*, *11*(9). https://doi.org/10.1038/nprot.2016.095
Pertea, M., Pertea, G. M., Antonescu, C. M., Chang, T. C., Mendell, J. T., & Salzberg, S. L. (2015). StringTie enables improved reconstruction of a transcriptome from RNA-seq reads. *Nature Biotechnology*, *33*(3). https://doi.org/10.1038/nbt.3122
Peters, T. J., Buckley, M. J., Chen, Y., Smyth, G. K., Goodnow, C. C., & Clark, S. J. (2021a). Calling differentially methylated regions from whole genome bisulphite sequencing with DMRcate. *Nucleic Acids Research*, *49*(19). https://doi.org/10.1093/nar/gkab637
Peters, T. J., Buckley, M. J., Chen, Y., Smyth, G. K., Goodnow, C. C., & Clark, S. J. (2021b). Calling differentially methylated regions from whole genome bisulphite sequencing with DMRcate. *Nucleic Acids Research*, *49*(19). https://doi.org/10.1093/nar/gkab637
Peters, T. J., Buckley, M. J., Statham, A. L., Pidsley, R., Samaras, K., V Lord, R., Clark, S. J., & Molloy, P. L. (2015). De novo identification of differentially methylated regions in the human genome. *Epigenetics and Chromatin*, *8*(1). https://doi.org/10.1186/1756-8935-8-6
Peters, T. J., Meyer, B., Ryan, L., Achinger-Kawecka, J., Song, J., Campbell, E. M., Qu, W., Nair, S., Loi-Luu, P., Stricker, P., Lim, E., Stirzaker, C., Clark, S. J., & Pidsley, R. (2024). Characterisation and reproducibility of the HumanMethylationEPIC v2.0 BeadChip for DNA methylation profiling. *BMC Genomics*, *25*(1). https://doi.org/10.1186/s12864-024-10027-5
Petibon, C., Parenteau, J., Catala, M., & Elela, S. A. (2016). Introns regulate the production of ribosomal proteins by modulating splicing of duplicated ribosomal protein genes. *Nucleic Acids Research*, *44*(8). https://doi.org/10.1093/nar/gkw140
Piras, I. S., Brokaw, D., Kong, Y., Weisenberger, D. J., Krate, J., Delvaux, E., Mahurkar, S., Blattler, A., Siegmund, K. D., Sue, L., Serrano, G. E., Beach, T. G., Laird, P. W., Huentelman, M. J., & Coleman, P. D. (2023). Integrated DNA Methylation/RNA Profiling in Middle Temporal Gyrus of Alzheimer's Disease. *Cellular and Molecular Neurobiology*, *43*(5). https://doi.org/10.1007/s10571-022-01307-3
Pivetta, E., Capuano, A., Vescovo, M., Scanziani, E., Cappelleri, A., Rampioni Vinciguerra, G. L., Vecchione, A., Doliana, R., Mongiat, M., & Spessotto, P. (2022). EMILIN-1 deficiency promotes chronic inflammatory disease through TGFβ signaling alteration and impairment of the gC1q/α4β1 integrin interaction. *Matrix Biology*, *111*. https://doi.org/10.1016/j.matbio.2022.06.005
Pivetta, E., Wassermann, B., Del Bel Belluz, L., Danussi, C., Modica, T. M. E., Maiorani, O., Bosisio, G., Boccardo, F., Canzonieri, V., Colombatti, A., & Spessotto, P. (2016). Local inhibition of elastase reduces EMILIN1 cleavage reactivating lymphatic vessel function in a mouse lymphoedema model. *Clinical Science*, *130*(14). https://doi.org/10.1042/CS20160064

Qazi, T. J., Quan, Z., Mir, A., & Qing, H. (2018). Epigenetics in Alzheimer's Disease: Perspective of DNA Methylation. In *Molecular Neurobiology* (Vol. 55, Issue 2). https://doi.org/10.1007/s12035-016-0357-6

Raj, B., & Blencowe, B. J. (2015). Alternative Splicing in the Mammalian Nervous System: Recent Insights into Mechanisms and Functional Roles. *Neuron*, *87*(1), 14–27. https://doi.org/10.1016/J.NEURON.2015.05.004

Raj, T., Li, Y. I., Wong, G., Humphrey, J., Wang, M., Ramdhani, S., Wang, Y. C., Ng, B., Gupta, I., Haroutunian, V., Schadt, E. E., Young-Pearse, T., Mostafavi, S., Zhang, B., Sklar, P., Bennett, D. A., & De Jager, P. L. (2018). Integrative transcriptome analyses of the aging brain implicate altered splicing in Alzheimer's disease susceptibility. *Nature Genetics*, *50*(11). https://doi.org/10.1038/s41588-018-0238-1

Ramos-Miguel, A., Sawada, K., Jones, A. A., Thornton, A. E., Barr, A. M., Leurgans, S. E., Schneider, J. A., Bennett, D. A., & Honer, W. G. (2017). Presynaptic proteins complexin-I and complexin-II differentially influence cognitive function in early and late stages of Alzheimer's disease. *Acta Neuropathologica*, *133*(3). https://doi.org/10.1007/s00401-016-1647-9

Reyes, A., & Huber, W. (2018). Alternative start and termination sites of transcription drive most transcript isoform differences across human tissues. *Nucleic Acids Research*, *46*(2). https://doi.org/10.1093/nar/gkx1165

Roubroeks, J. A. Y., Smith, A. R., Smith, R. G., Pishva, E., Ibrahim, Z., Sattlecker, M., Hannon, E. J., Kłoszewska, I., Mecocci, P., Soininen, H., Tsolaki, M., Vellas, B., Wahlund, L. O., Aarsland, D., Proitsi, P., Hodges, A., Lovestone, S., Newhouse, S. J., Dobson, R. J. B., … Lunnon, K. (2020). An epigenome-wide association study of Alzheimer's disease blood highlights robust DNA hypermethylation in the HOXB6 gene. *Neurobiology of Aging*, *95*. https://doi.org/10.1016/j.neurobiolaging.2020.06.023

Saito, T., Matsuba, Y., Mihira, N., Takano, J., Nilsson, P., Itohara, S., Iwata, N., & Saido, T. C. (2014). Single App knock-in mouse models of Alzheimer's disease. *Nature Neuroscience*, *17*(5). https://doi.org/10.1038/nn.3697

Sala Frigerio, C., Wolfs, L., Fattorelli, N., Thrupp, N., Voytyuk, I., Schmidt, I., Mancuso, R., Chen, W. T., Woodbury, M. E., Srivastava, G., Möller, T., Hudry, E., Das, S., Saido, T., Karran, E., Hyman, B., Perry, V. H., Fiers, M., & De Strooper, B. (2019). The Major Risk Factors for Alzheimer's Disease: Age, Sex, and Genes Modulate the Microglia Response to Aβ Plaques. *Cell Reports*, *27*(4). https://doi.org/10.1016/j.celrep.2019.03.099

Salih, D. A., Bayram, S., Guelfi, S., Reynolds, R. H., Shoai, M., Ryten, M., Brenton, J. W., Zhang, D., Matarin, M., Botia, J. A., Shah, R., Brookes, K. J., Guetta-Baranes, T., Morgan, K., Bellou, E., Cummings, D. M., Escott-Price, V., & Hardy, J. (2019). Genetic variability in response to amyloid beta deposition influences Alzheimer's disease risk. *Brain Communications*, *1*(1). https://doi.org/10.1093/braincomms/fcz022

Sanchez-Mut, J. V., Heyn, H., Vidal, E., Moran, S., Sayols, S., Delgado-Morales, R., Schultz, M. D., Ansoleaga, B., Garcia-Esparcia, P., Pons-Espinal, M., De Lagran, M. M., Dopazo, J., Rabano, A., Avila, J., Dierssen, M., Lott, I., Ferrer, I., Ecker, J. R., & Esteller, M. (2016). Human DNA methylomes of neurodegenerative diseases show common epigenomic patterns. *Translational Psychiatry*, *6*(1). https://doi.org/10.1038/tp.2015.214

Saunders, A., Huang, K. W., Vondrak, C., Hughes, C., Smolyar, K., Sen, H., Philson, A. C., Nemesh, J., Wysoker, A., Kashin, S., Sabatini, B. L., & McCarroll, S. A.

(2022). Ascertaining cells' synaptic connections and RNA expression simultaneously with barcoded rabies virus libraries. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-34334-1

Seligmann, B., Camiolo, S., Scully, K., Hernandez, M., Babic, M., Yeakley, J., Eastburn, D., McComb, J., & Sahagian, G. (2023). Diagnosis of Disease Using Whole Blood from a Fingerstick Spotted on Filter Paper: Gene Expression Signature Test for Alzheimer's and Parkinson's Diseases. *Blood*, *142*(Supplement 1). https://doi.org/10.1182/blood-2023-190237

Semick, S. A., Bharadwaj, R. A., Collado-Torres, L., Tao, R., Shin, J. H., Deep-Soboslay, A., Weiss, J. R., Weinberger, D. R., Hyde, T. M., Kleinman, J. E., Jaffe, A. E., & Mattay, V. S. (2019). Integrated DNA methylation and gene expression profiling across multiple brain regions implicate novel genes in Alzheimer's disease. *Acta Neuropathologica*, *137*(4). https://doi.org/10.1007/s00401-019-01966-5

Shankar, G. M., Leissring, M. A., Adame, A., Sun, X., Spooner, E., Masliah, E., Selkoe, D. J., Lemere, C. A., & Walsh, D. M. (2009). Biochemical and immunohistochemical analysis of an Alzheimer's disease mouse model reveals the presence of multiple cerebral Aβ assembly forms throughout life. *Neurobiology of Disease*, *36*(2). https://doi.org/10.1016/j.nbd.2009.07.021

Sheets, M. D., Fox, C. A., Hunt, T., Vande Woude, & Wickens, M. (1994). The 3'-untranslated regions of c-mos and cyclin mRNAs stimulate translation by regulating cytoplasmic polyadenylation. *Genes and Development*, *8*(8). https://doi.org/10.1101/gad.8.8.926

Shireby, G., Dempster, E. L., Policicchio, S., Smith, R. G., Pishva, E., Chioza, B., Davies, J. P., Burrage, J., Lunnon, K., Seiler Vellame, D., Love, S., Thomas, A., Brookes, K., Morgan, K., Francis, P., Hannon, E., & Mill, J. (2022). DNA methylation signatures of Alzheimer's disease neuropathology in the cortex are primarily driven by variation in non-neuronal cell-types. *Nature Communications*, *13*(1). https://doi.org/10.1038/s41467-022-33394-7

Sierksma, A., Lu, A., Mancuso, R., Fattorelli, N., Thrupp, N., Salta, E., Zoco, J., Blum, D., Buée, L., De Strooper, B., & Fiers, M. (2020). Novel Alzheimer risk genes determine the microglia response to amyloid-β but not to TAU pathology. *EMBO Molecular Medicine*, *12*(3). https://doi.org/10.15252/emmm.201910606

Sigurpalsdottir, B. D., Stefansson, O. A., Holley, G., Beyter, D., Zink, F., Hardarson, M., Sverrisson, S., Kristinsdottir, N., Magnusdottir, D. N., Magnusson, O., Gudbjartsson, D. F., Halldorsson, B. V., & Stefansson, K. (2024). A comparison of methods for detecting DNA methylation from long-read sequencing of human genomes. *Genome Biology*, *25*(1). https://doi.org/10.1186/s13059-024-03207-9

Simon Andrews. (2020). Babraham Bioinformatics - FastQC A Quality Control tool for High Throughput Sequence Data. In *Soil* (Vol. 5, Issue 1).

Simpson, J. T., Workman, R. E., Zuzarte, P. C., David, M., Dursi, L. J., & Timp, W. (2017). Detecting DNA cytosine methylation using nanopore sequencing. *Nature Methods*, *14*(4). https://doi.org/10.1038/nmeth.4184

Sims, R., Van Der Lee, S. J., Naj, A. C., Bellenguez, C., Badarinarayan, N., Jakobsdottir, J., Kunkle, B. W., Boland, A., Raybould, R., Bis, J. C., Martin, E. R., Grenier-Boley, B., Heilmann-Heimbach, S., Chouraki, V., Kuzma, A. B., Sleegers, K., Vronskaya, M., Ruiz, A., Graham, R. R., … Schellenberg, G. D. (2017). Rare coding variants in PLCG2, ABI3, and TREM2 implicate microglial-mediated innate immunity in Alzheimer's disease. *Nature Genetics*, *49*(9). https://doi.org/10.1038/ng.3916

Skene, N. G., & Grant, S. G. N. (2016). Identification of vulnerable cell types in major brain disorders using single cell transcriptomes and expression weighted cell type enrichment. *Frontiers in Neuroscience*, *10*(JAN). https://doi.org/10.3389/fnins.2016.00016

Smith, A. R., Smith, R. G., Macdonald, R., Marzi, S. J., Burrage, J., Troakes, C., Al-Sarraj, S., Mill, J., & Lunnon, K. (2021). The histone modification H3K4me3 is altered at the ANK1 locus in Alzheimer's disease brain. *Future Science OA*, *7*(4). https://doi.org/10.2144/FSOA-2020-0161

Smith, R. G., Pishva, E., Shireby, G., Smith, A. R., Roubroeks, J. A. Y., Hannon, E., Wheildon, G., Mastroeni, D., Gasparoni, G., Riemenschneider, M., Giese, A., Sharp, A. J., Schalkwyk, L., Haroutunian, V., Viechtbauer, W., van den Hove, D. L. A., Weedon, M., Brokaw, D., Francis, P. T., … Lunnon, K. (2021). A meta-analysis of epigenome-wide association studies in Alzheimer's disease highlights novel differentially methylated loci across cortex. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-23243-4

Snajder, R., Leger, A., Stegle, O., & Bonder, M. J. (2023). pycoMeth: a toolbox for differential methylation testing from Nanopore methylation calls. *Genome Biology*, *24*(1). https://doi.org/10.1186/s13059-023-02917-w

Suárez-Calvet, M., Kleinberger, G., Araque Caballero, M. Á., Brendel, M., Rominger, A., Alcolea, D., Fortea, J., Lleó, A., Blesa, R., Gispert, J. D., Sánchez-Valle, R., Antonell, A., Rami, L., Molinuevo, J. L., Brosseron, F., Traschütz, A., Heneka, M. T., Struyfs, H., Engelborghs, S., … Haass, C. (2016). sTREM 2 cerebrospinal fluid levels are a potential biomarker for microglia activity in early-stage Alzheimer's disease and associate with neuronal injury markers . *EMBO Molecular Medicine*, *8*(5). https://doi.org/10.15252/emmm.201506123

Suh, J., Choi, S. H., Romano, D. M., Gannon, M. A., Lesinski, A. N., Kim, D. Y., & Tanzi, R. E. (2013). ADAM10 Missense Mutations Potentiate β-Amyloid Accumulation by Impairing Prodomain Chaperone Function. *Neuron*, *80*(2). https://doi.org/10.1016/j.neuron.2013.08.035

Suire, C. N., Abdul-Hay, S. O., Sahara, T., Kang, D., Brizuela, M. K., Saftig, P., Dickson, D. W., Rosenberry, T. L., & Leissring, M. A. (2020). Cathepsin D regulates cerebral Aβ42/40 ratios via differential degradation of Aβ42 and Aβ40. *Alzheimer's Research and Therapy*, *12*(1). https://doi.org/10.1186/s13195-020-00649-8

Takata, A., Matsumoto, N., & Kato, T. (2017). Genome-wide identification of splicing QTLs in the human brain and their enrichment among schizophrenia-associated loci. *Nature Communications*, *8*. https://doi.org/10.1038/ncomms14519

Tardaguila, M., De La Fuente, L., Marti, C., Pereira, C., Pardo-Palacios, F. J., Del Risco, H., Ferrell, M., Mellado, M., Macchietto, M., Verheggen, K., Edelmann, M., Ezkurdia, I., Vazquez, J., Tress, M., Mortazavi, A., Martens, L., Rodriguez-Navarro, S., Moreno-Manzano, V., & Conesa, A. (2018). SQANTI: Extensive characterization of long-read transcript sequences for quality control in full-length transcriptome identification and quantification. *Genome Research*, *28*(3). https://doi.org/10.1101/gr.222976.117

Tiwari, S. S., Mizuno, K., Ghosh, A., Aziz, W., Troakes, C., Daoud, J., Golash, V., Noble, W., Hortobágyi, T., & Giese, K. P. (2016). Alzheimer-related decrease in CYFIP2 links amyloid production to tau hyperphosphorylation and memory loss. *Brain*, *139*(10). https://doi.org/10.1093/brain/aww205

Tollervey, J. R., Wang, Z., Hortobágyi, T., Witten, J. T., Zarnack, K., Kayikci, M., Clark, T. A., Schweitzer, A. C., Rot, G., Curk, T., Zupan, B., Rogelj, B., Shaw, C.

E., & Ule, J. (2011). Analysis of alternative splicing associated with aging and neurodegeneration in the human brain. *Genome Research*, *21*(10). https://doi.org/10.1101/gr.122226.111

Tran, H., Wu, X., Tithi, S., Sun, M. A., Xie, H., & Zhang, L. (2016). A Bayesian assignment method for ambiguous bisulfite short reads. *PLoS ONE*, *11*(3). https://doi.org/10.1371/journal.pone.0151826

Trinchero, M. F., Buttner, K. A., Sulkes Cuevas, J. N., Temprana, S. G., Fontanet, P. A., Monzón-Salinas, M. C., Ledda, F., Paratcha, G., & Schinder, A. F. (2017). High Plasticity of New Granule Cells in the Aging Hippocampus. *Cell Reports*, *21*(5). https://doi.org/10.1016/j.celrep.2017.09.064

Van den Berge, K., Soneson, C., Robinson, M. D., & Clement, L. (2017). stageR: A general stage-wise method for controlling the gene-level false discovery rate in differential expression and differential transcript usage. *Genome Biology*, *18*(1). https://doi.org/10.1186/s13059-017-1277-0

Vega, F. M., Thomas, M., Reymond, N., & Ridley, A. J. (2015). The Rho GTPase RhoB regulates cadherin expression and epithelial cell-cell interaction. *Cell Communication and Signaling*, *13*(1). https://doi.org/10.1186/s12964-015-0085-y

Vitting-Seerup, K., Porse, B. T., Sandelin, A., & Waage, J. (2014). SpliceR: An R package for classification of alternative splicing and prediction of coding potential from RNA-seq data. *BMC Bioinformatics*, *15*(1). https://doi.org/10.1186/1471-2105-15-81

Vitting-Seerup, K., & Sandelin, A. (2017). The landscape of isoform switches in human cancers. *Molecular Cancer Research*, *15*(9). https://doi.org/10.1158/1541-7786.MCR-16-0459

Vitting-Seerup, K., Sandelin, A., & Berger, B. (2019). IsoformSwitchAnalyzeR: Analysis of changes in genome-wide patterns of alternative splicing and its functional consequences. *Bioinformatics*, *35*(21). https://doi.org/10.1093/bioinformatics/btz247

Wang, E., Wang, M., Guo, L., Fullard, J. F., Micallef, C., Bendl, J., Song, W. min, Ming, C., Huang, Y., Li, Y., Yu, K., Peng, J., Bennett, D. A., De Jager, P. L., Roussos, P., Haroutunian, V., & Zhang, B. (2023). Genome-wide methylomic regulation of multiscale gene networks in Alzheimer's disease. *Alzheimer's and Dementia*, *19*(8). https://doi.org/10.1002/alz.12969

Wang, Q., Li, M., Wu, T., Zhan, L., Li, L., Chen, M., Xie, W., Xie, Z., Hu, E., Xu, S., & Yu, G. (2022). Exploring Epigenomic Datasets by ChIPseeker. *Current Protocols*, *2*(10). https://doi.org/10.1002/cpz1.585

WANG, Y., LIU, J., HUANG, B., XU, Y.-M., LI, J., HUANG, L.-F., LIN, J., ZHANG, J., MIN, Q.-H., YANG, W.-M., & WANG, X.-Z. (2015). Mechanism of alternative splicing and its regulation. *Biomedical Reports*, *3*(2). https://doi.org/10.3892/br.2014.407

Wang, Y., Zhao, Y., Bollas, A., Wang, Y., & Au, K. F. (2021). Nanopore sequencing technology, bioinformatics and applications. In *Nature Biotechnology* (Vol. 39, Issue 11). https://doi.org/10.1038/s41587-021-01108-x

Watson, C. T., Roussos, P., Garg, P., Ho, D. J., Azam, N., Katsel, P. L., Haroutunian, V., & Sharp, A. J. (2016). Genome-wide12 DNA methylation profiling in the superior temporal gyrus reveals epigenetic signatures associated with Alzheimer's disease. *Genome Medicine*, *8*(1). https://doi.org/10.1186/s13073-015-0258-8

Weigand, A. J., Thomas, K. R., Bangen, K. J., Eglit, G. M. L., Delano-Wood, L., Gilbert, P. E., Brickman, A. M., & Bondi, M. W. (2021). APOE interacts with tau PET to influence memory independently of amyloid PET in older adults without dementia. *Alzheimer's and Dementia*, *17*(1). https://doi.org/10.1002/alz.12173

Westra, H. J., Peters, M. J., Esko, T., Yaghootkar, H., Schurmann, C., Kettunen, J., Christiansen, M. W., Fairfax, B. P., Schramm, K., Powell, J. E., Zhernakova, A., Zhernakova, D. V., Veldink, J. H., Van Den Berg, L. H., Karjalainen, J., Withoff, S., Uitterlinden, A. G., Hofman, A., Rivadeneira, F., … Franke, L. (2013). Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nature Genetics*, *45*(10). https://doi.org/10.1038/ng.2756

Wood, J. I., Wong, E., Joghee, R., Balbaa, A., Vitanova, K. S., Stringer, K. M., Vanshoiack, A., Phelan, S. L. J., Launchbury, F., Desai, S., Tripathi, T., Hanrieder, J., Cummings, D. M., Hardy, J., & Edwards, F. A. (2022). Plaque contact and unimpaired Trem2 is required for the microglial response to amyloid pathology. *Cell Reports*, *41*(8). https://doi.org/10.1016/j.celrep.2022.111686

Wu, H., Wang, C., & Wu, Z. (2013). A new shrinkage estimator for dispersion improves differential expression detection in RNA-seq data. *Biostatistics*, *14*(2). https://doi.org/10.1093/biostatistics/kxs033

Wu, H., Xu, T., Feng, H., Chen, L., Li, B., Yao, B., Qin, Z., Jin, P., & Conneely, K. N. (2015). Detection of differentially methylated regions from whole-genome bisulfite sequencing data without replicates. *Nucleic Acids Research*, *43*(21). https://doi.org/10.1093/nar/gkv715

Wu, T., Hu, E., Xu, S., Chen, M., Guo, P., Dai, Z., Feng, T., Zhou, L., Tang, W., Zhan, L., Fu, X., Liu, S., Bo, X., & Yu, G. (2021). clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. *Innovation*, *2*(3). https://doi.org/10.1016/j.xinn.2021.100141

Wu, Y., Mumford, P., Noy, S., Cleverley, K., Mrzyglod, A., Luo, D., van Dalen, F., Verdoes, M., Fisher, E. M. C., & Wiseman, F. K. (2023). Cathepsin B abundance, activity and microglial localisation in Alzheimer's disease-Down syndrome and early onset Alzheimer's disease; the role of elevated cystatin B. *Acta Neuropathologica Communications*, *11*(1). https://doi.org/10.1186/s40478-023-01632-8

Xavier da Silva, T. N., Schulte, C., Alves, A. N., Maric, H. M., & Friedmann Angeli, J. P. (2023). Molecular characterization of AIFM2/FSP1 inhibition by iFSP1-like molecules. *Cell Death and Disease*, *14*(4). https://doi.org/10.1038/s41419-023-05787-z

Xiong, X., James, B. T., Boix, C. A., Park, Y. P., Galani, K., Victor, M. B., Sun, N., Hou, L., Ho, L. L., Mantero, J., Scannail, A. N., Dileep, V., Dong, W., Mathys, H., Bennett, D. A., Tsai, L. H., & Kellis, M. (2023). Epigenomic dissection of Alzheimer's disease pinpoints causal variants and reveals epigenome erosion. *Cell*, *186*(20). https://doi.org/10.1016/j.cell.2023.08.040

Xu, F., Sun, W., Li, P., Chen, J., Zhu, D., Sun, X., Wang, J., Feng, J., Song, K., & Duan, Y. (2015). Specificity protein 1 transcription factor regulates human ARTS promoter activity through multiple binding sites. *PLoS ONE*, *10*(3). https://doi.org/10.1371/journal.pone.0120072

Xue, Q., Zhou, Y. F., Zhu, S. N., & Bulun, S. E. (2011). Hypermethylation of the CpG island spanning from exon II to intron III is associated with steroidogenic factor 1 expression in stromal cells of endometriosis. *Reproductive Sciences*, *18*(11). https://doi.org/10.1177/1933719111404614

Yaman, U., Banks, G., Gustavsson, E. K., Mumford, P., Magusali, N., Taso, O. S., Macpherson, H., Carmona, S., Murray, M., Bains, R. S., Forrest, H., Stewart, M., Scott, C., Lipina, T. V, Cheng, Z., Tierney, A. L., Unwin, R. D., Botia, J. A., Frigerio, C. S., … Salih, D. A. (2024). Long-read transcriptomic identification of synaptic adaptation to amyloid pathology in the AppNL-G-F knock-in mouse model of the earliest phase of Alzheimer's disease. *BioRxiv*, 2024.08.10.607237. https://doi.org/10.1101/2024.08.10.607237

Yokoyama, A. S., Rutledge, J. C., & Medici, V. (2017). DNA methylation alterations in Alzheimer's disease. *Environmental Epigenetics*, *3*(2). https://doi.org/10.1093/eep/dvx008

Young, A. M. H., Kumasaka, N., Calvert, F., Hammond, T. R., Knights, A., Panousis, N., Park, J. S., Schwartzentruber, J., Liu, J., Kundu, K., Segel, M., Murphy, N. A., McMurran, C. E., Bulstrode, H., Correia, J., Budohoski, K. P., Joannides, A., Guilfoyle, M. R., Trivedi, R., … Gaffney, D. J. (2021). A map of transcriptional heterogeneity and regulatory variation in human microglia. *Nature Genetics*, *53*(6). https://doi.org/10.1038/s41588-021-00875-2

Yu, G., Wang, L. G., Han, Y., & He, Q. Y. (2012). ClusterProfiler: An R package for comparing biological themes among gene clusters. *OMICS A Journal of Integrative Biology*, *16*(5). https://doi.org/10.1089/omi.2011.0118

Yu, G., Wang, L. G., & He, Q. Y. (2015). ChIP seeker: An R/Bioconductor package for ChIP peak annotation, comparison and visualization. *Bioinformatics*, *31*(14). https://doi.org/10.1093/bioinformatics/btv145

Yuen, Z. W. S., Srivastava, A., Daniel, R., McNevin, D., Jack, C., & Eyras, E. (2021). Systematic benchmarking of tools for CpG methylation detection from nanopore sequencing. *Nature Communications*, *12*(1). https://doi.org/10.1038/s41467-021-23778-6

Zeisel, A., M̂oz-Manchado, A. B., Codeluppi, S., Lönnerberg, P., Manno, G. La, Juréus, A., Marques, S., Munguba, H., He, L., Betsholtz, C., Rolny, C., Castelo-Branco, G., Hjerling-Leffler, J., & Linnarsson, S. (2015). Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science*, *347*(6226). https://doi.org/10.1126/science.aaa1934

Zhang, B., Zhou, Y., Lin, N., Lowdon, R. F., Hong, C., Nagarajan, R. P., Cheng, J. B., Li, D., Stevens, M., Lee, H. J., Xing, X., Zhou, J., Sundaram, V., Elliott, G., Gu, J., Shi, T., Gascard, P., Sigaroudinia, M., Tlsty, T. D., … Wang, T. (2013). Functional DNA methylation differences between tissues, cell types, and across individuals discovered using the M&M algorithm. *Genome Research*, *23*(9). https://doi.org/10.1101/gr.156539.113

Zhang, L., Silva, T. C., Young, J. I., Gomez, L., Schmidt, M. A., Hamilton-Nelson, K. L., Kunkle, B. W., Chen, X., Martin, E. R., & Wang, L. (2020). Epigenome-wide meta-analysis of DNA methylation differences in prefrontal cortex implicates the immune processes in Alzheimer's disease. *Nature Communications*, *11*(1). https://doi.org/10.1038/s41467-020-19791-w

Zhang, Y., Yang, H. T., Kadash-Edmondson, K., Pan, Y., Pan, Z., Davidson, B. L., & Xing, Y. (2020). Regional Variation of Splicing QTLs in Human Brain. *American Journal of Human Genetics*, *107*(2). https://doi.org/10.1016/j.ajhg.2020.06.002

Zhang, Z. F., Pal, S., Bi, Y., Tchou, J., & Davuluri, R. V. (2013). Isoform level expression profiles provide better cancer signatures than gene level expression profiles. *Genome Medicine*, *5*(4). https://doi.org/10.1186/gm437

Zheng, G. X. Y., Terry, J. M., Belgrader, P., Ryvkin, P., Bent, Z. W., Wilson, R., Ziraldo, S. B., Wheeler, T. D., McDermott, G. P., Zhu, J., Gregory, M. T., Shuga,

J., Montesclaros, L., Underwood, J. G., Masquelier, D. A., Nishimura, S. Y., Schnall-Levin, M., Wyatt, P. W., Hindson, C. M., … Bielas, J. H. (2017). Massively parallel digital transcriptional profiling of single cells. *Nature Communications*, *8*. https://doi.org/10.1038/ncomms14049

Appendix A Supplementary Tables

Summary tables are available online in Yaman, U. (2024). Appendix A. [Zenodo]. https://doi.org/ 10.5281/zenodo.14502108

Chapter 2: Long-read transcriptomic identification of synaptic adaptation to amyloid pathology in the knock-in *App*$^{NL-G-F}$ mouse model of the earliest phase of Alzheimer's disease

Supplementary Table 2.1. Differentially Expressed Genes via DESeq2 (sex corrected) – Amyloid

Supplementary Table 2.2. WGCNA Gene Network

Supplementary Table 2.3. Differentially Expressed genes via DESeq2 (genotype corrected) – Sex

Supplementary Table 2.4. Differentially Expressed Isoforms via DESeq2 (sex corrected)- Amyloid

Supplementary Table 2.5. Mouse Genome Annotation Generated via Long Read RNA-seq

Supplementary Table 2.6. DEXSeq Isoform Usage Results

Supplementary Table 2.7. tappAS results

Supplementary Table 2.8. Isoform Switch Results via IsoformSwitchAnalyzer

Supplementary Table 2.9. DEXSeq Exon Usage Results

Supplementary Table 2.10. Isoform switch comparison with human AD results

Supplementary Table 2.11. PSI-Sigma Splicing Results

Chapter 3: Genome-wide 5mC differential methylation analysis via long-read DNA sequencing in the *App*$^{NL-G-F}$ Alzheimer's Disease mouse model.

Supplementary Table 3.1. Annotated DMRs

Supplementary Table 3.2. GO Annotations of Hypermethylated DMRs - Promoters only

Supplementary Table 3.3. GO Annotations of Hypermethylated DMRs

Supplementary Table 3.4. GO Annotations of Hypomethylated DMRs - Promoters only

Supplementary Table 3.5. GO Annotations of Hypomethylated DMRs

Chapter 4: Investigating the cell-type-specific roles of candidate genes through human Alzheimer's disease single-nucleus RNA sequencing (snRNA-seq) and identifying common genes across bulk RNA-seq datasets.

Supplementary Table 4.1. Significant correlation results of promoters - Integration based on all common promotors

Supplementary Table 4.2. Differentially methylated regions (DMRs) correlating with gene expression in *App*$^{NL\text{-}G\text{-}F}$ and WT mouse models