The *EPHA1* risk gene and its Alzheimer's Disease associated variants

A thesis submitted for the degree of Doctor of Philosophy

At
Cardiff University School of Medicine

Lauren Elizabeth Thorburn
2022

Thesis Summary

Alzheimer's disease (AD) is a progressive neurological disorder. Genetic studies have identified numerous risk or susceptibility genes/loci to be linked to the pathology of late onset AD. One such gene is the EphA1 receptor which contains numerous genetic variants within the introns of the gene or within its neighbouring *EPHA1-AS1* antisense gene. Genomic location and inheritance pattern of the variants make identifying the causal variant behind a loci's disease association difficult, requiring the overlay of both bioinformatic and functional data onto genetic data.

Two *EPHA1* variants rs11765305 and rs7810606 were found to have the ability to affect the binding affinity of nuclear proteins. Deletion of these variants resulted in increased expression of the *ZYX* gene, a loss of stemness morphology and a decrease in pluripotency genes was also noted. Bioinformatic analysis hints at the CEBPB, KLF family or EGR1 transcription factors being involved in the alterations to *ZYX* regulation observed. It is likely that these variants are behind the *EPHA1* loci's association with AD, through altered regulation of *ZYX* contributing to disease pathology such as inflammation and synaptic impairment.

A coding variant within the EphA1 receptor P460L has been linked to late onset AD. The exact role of the EphA1 receptor during AD and how alterations as a consequence of P460L affect disease progression is unknown. This thesis identified both reduced receptor membrane expression and reduced soluble receptor release as a result of the P460L variant. Transient serine residue dephosphorylation followed by tyrosine residue phosphorylation was noted upon ligand activation of the EphA1 receptor. However, subsequent increase in phosphotyrosine was not observed within the P460L mutant. This reduced receptor membrane expression and subsequent activity due to the P460L variant may impact any role the EphA1 receptor may play within the periphery or brain that may regulate AD.

Acknowledgements

I would like to express my greatest appreciation to my supervisor Professor Ann Ager, for not only the opportunity to conduct this fascinating project within your lab but also for the immense support, guidance and expertise throughout. I would like to thank my co-supervisors Professor Nick Allen and Professor Julie Williams. Thanks also to Dr Vera Knäuper for your invaluable support trying to suss out EphA1 receptor function.

A PhD thesis is never done in isolation, I would like to express the most heartfelt thanks to my fellow researchers Cerys Ballard, Katy Widnall and Dina Fathalla, this work would not be possible without you guys. Your never-ending love, pep talks and motivational speeches during our numerous tea/cake breaks were invaluable. I will miss you all terribly, thanks for making my time in Cardiff wonderful.

To my family, Helen, Kevin and Johnathan Hindhaugh, I will never be able to express my appreciation for everything you have done to enable me to complete such an achievement. It is only with your never-ending support, encouragement and belief in me that this has been possible. Johny your patience for proof reading is never ending! And John Eltis – Grandad, I owe the completion of this thesis to you. Thank you for all the support you gave me not only throughout this PhD but in all things, I am sorry you did not get to see it completed, but we did it!
To my husband, Josh, I owe so much to your unceasing love and belief that I can do hard things. Your support in all aspects of my life has allowed me to undertake such a project.

And finally, to our daughter Ellie. I dedicate this thesis to you; may your future be so bright and full of wonder.

Abbreviations

ΔRn: Delta normalised reporter value

Aβ: Amyloid beta

AD: Alzheimer's disease

ADAM: A Disintegrin and metalloproteinase domain-containing protein

ADGC: Alzheimer's disease genetics consortium

ADNI: Alzheimer's disease neuroimaging initiative

ADSP: Alzheimer's disease sequencing project

AF: Alexa Fluor

APOE: Apolipoprotein E

APP: Amyloid precursor protein

ATAC-Seq: Assay of transposase-accessible chromatin with sequencing

BACE1-AS1: β-site APP cleaving enzyme 1 antisense

BACE1: β-site APP cleaving enzyme 1

BBB: Blood brain barrier

BCA: Bicinchoninic acid

BSA: Bovine serum albumin

CBF: Cerebral blood flow

CCR2: C-C chemokine receptor type 2

CEBP: CCAAT enhancer-binding proteins

CERI/II: Cytoplasmic Extraction Reagent I and II

CHARGE: Genomic Epidemiology Consortium

CHD: Chromodomain helicase DNA binding

CNS: Central nervous system

CRISPR: Clustered regulatory interspersed short palindromic repeats

CSF: Cerebrospinal fluid

Ct: cycle threshold

CTCF: CCCTC-binding protein

CX3CR1: CXC3C chemokine receptor 1

DMSO: Dimethyl sulfoxide

E-cadherin: Epithelial cadherin

Abbreviations

EADI: European Alzheimer's Disease Initiative

EB: Embryoid body

ECM: Extracellular matrix

EGR1: Early Growth Response 1

ELISA: Enzyme-linked immunosorbent assay

EMSA: Electrophoresis mobility shift assay

ENCODE: Encyclopaedia of DNA Elements

EOAD: Early onset Alzheimer's disease

Eph: Erythropoietin-producing hepatocellular

EphA1-AS1: Ephrin type A1 receptor antisense

EphA1: Ephrin type A1 receptor

eQTL: Expression quantitative trait loci

FACS: Fluorescence-activated cell sorting

FAD: Familial Alzheimer's disease

FANTOM5: Functional Annotation of the Mammalian genome

FBS: Foetal bovine serum

FMO: Fluorescence minus one

FN: Fibronectin Type-III domain

GERAD/PERADES: Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium

GPI: Glycosylphosphatidylinositol

GWAS: Genome wide association study

HDR: Homologous-directed repair

HEK293: human embryonic kidney 293

HIPK2: Homeodomain Interacting Protein Kinase 2

hnRNP: Heterogenous nuclear ribonuclear protein

HOTTIP: HOXA transcript at distal tip

HBMECs: Human brain microvascular endothelial cells

ICAM-1: Intercellular Adhesion Molecule 1

IDT: Integrated DNA Technologies

IGAP: International Genomics of Alzheimer's Project

IL-1β: interleukin 1 beta

Abbreviations

INO80: inositol requiring 80

IPSC: Induced pluripotent stem cell

ISW1: Imitation switch

KEGG: Kyoto encyclopaedia of genes and genomes

KLF: Krüppel-like factor

LD: Linkage disequilibrium

LDCV: Large dense-core vesicles

lncHUB: Long non-coding DNA hub

lncRNA: Long non-coding RNA

LOAD: Late onset Alzheimer's disease

LRP1: Low density lipoprotein receptor-related protein 1

miRNA: Micro RNA

MLL: Mixed lineage leukaemia complex

MMP: Matrix metallopeptidase 9

NER: Nuclear Extraction Reagent

NET: Neutrophil extracellular traps

NFT: Neurofibrillary tangle

NHEJ: Non-homologous end joining

NK: Natural killer cells

NVU: Neurovascular unit

OCR: Open chromatin region

PAM: Protospacer adjacent motif

PBS-T: Phosphate-buffered saline containing 0.01% Tween 20

PCR: Polymerase chain reaction

PE: Phycoerythrin

PET: Positron emission tomography

PGC-ALZ: Psychiatric genetics consortium, Alzheimer's working group

piRNA: Piwi RNA

qRTPCR: Quantitative PCR

RT: Room temperature

RTKs: Receptor tyrosine kinase

SAM: sterile alpha motif

Abbreviations

SDS-PAGE: Sodium dodecyl-sulfate polyacrylamide gel electrophoresis

siRNA: Small interfering RNA

SMCS: structural maintenance of chromosomes

SNPS: Single nucleotide polymorphism

SP: specificity proteins

SWI/SNF: Switch/sucrose non-fermenting

TEER: trans endothelial electrical resistance

TF: Transcription factor

TGEN1: Translational genomics research institute

THRIL: The heterogenous nuclear ribonuclear protein L related immunoregulatory lncRNA

TMB: tetramethylbenzidine

TNF-α: Tumour Necrosis Factor alpha

tracrRNA: Trans-activating CRISPR RNA

VLDL: Very-low-density lipoprotein

Ybx1: Y-Box Binding Protein 1

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Table of Contents

Chapter 1: General Introduction

1.1 Background to Alzheimer's Disease

Alzheimer's disease (AD) is a progressive neurological disorder characterised by memory deficits that progress into cognitive impairment and dysfunction. Clinical diagnosis is often preceded by neuronal loss and build-up of neuropathic hallmarks (Lane et al. 2018; Rabinovici 2019).

Alzheimer's disease is the most common form of dementia accounting for 62% of all dementias within the UK (Qiu et al. 2009). In 2020, the world Alzheimer report estimated that there were 55 million people living with the disease worldwide. This prevalence is only set to increase in part due to an aging population, expected to reach around 139 million by 2050. At an annual global cost of $1.3 trillion, dementia is one of the largest socioeconomical burdens on health services across the world (International 2015).

Alzheimer's disease can be characterised into two different subgroups, early onset (EOAD) and late onset (LOAD). Aside from their age of onset, with early onset presenting at around 30-65 years of age and late onset developing after 65 years of age, these two forms of AD are largely indistinguishable from one another. EOAD is typically a familial form of AD and only accounts for 4-6 % of the total Alzheimer's disease cases. LOAD on the other hand is the most common form, accounting for approximately 95% of AD cases. This form is typically sporadic in nature (Mendez 2017).

1.2 Clinical symptoms and diagnosis criteria

An early AD diagnosis is imperative for appropriate AD pathology management. However, due to the heterogenous nature of symptom presentation clinical diagnosis is often difficult and delayed. AD is clinically diagnosed through a comprehensive patient history report

generated from interviews with the patient themselves and a knowledgeable informant along with cognitive examinations. These findings can be combined with neuroimaging or quantitation of cerebrospinal fluid (CSF) AD biomarkers.

A workgroup convened by the national institute of neurological and communicative disorders and stroke along with the Alzheimer's disease and related disorders association established a report of clinical criteria for the diagnosis of dementia and AD. The criteria for diagnosing all dementia centres around cognitive impairment, such as, impaired visuospatial and language ability, interference with the ability to function during usual activities or a decline in previous levels of functioning. For diagnosis of AD all the above criteria must be met along with a range of specific AD clinical symptoms such as, the insidious onset of symptoms over the course of months or years and a history of worsening cognition with executive dysfunction (McKhann et al. 2011).

Upon meeting all the clinical criteria, a diagnosis of AD maybe supported via the quantitation of CSF biomarkers such as deposits of Amyloid-Beta$_{42}$ (Aβ) peptides and hyperphosphorylated Tau and their comparison to normative standards. Imaging for biomarkers can also imply the presence of AD pathophysiology, for example a structural magnetic resonance imaging (MRI) can measure levels of cerebral atrophy or during a positron emission tomography (PET) scan decreases in fluorodeoxyglucose uptake can indicate synaptic dysfunction (Jack et al. 2010).

1.3 Alzheimer's Disease pathology

Alzheimer's disease is characterised by various neuropathological hallmarks such as extracellular amyloid plaques and intracellular neurofibrillary tangles (NFT). These lesions are a result of aggregated, misfolded Aβ and aggregated, hyperphosphorylated Tau protein respectively. In addition to these primary hallmarks, pathology such as vascular damage, neuroinflammation and neuronal cell loss is also evident (Lane et al. 2018).

Identification of these abnormalities within the brain is possible through post mortem studies using silver impregnation methods as seen in figure 1.1 or fluorescent dyes such a Thioflavin-S (Arnold et al. 1991; Braak and Braak 1991). Studies of this nature allowed a staging of progression of both amyloid plaques and NFT pathologies to be proposed, which helped the development of diagnostic criteria (Serrano-Pozo et al. 2011).
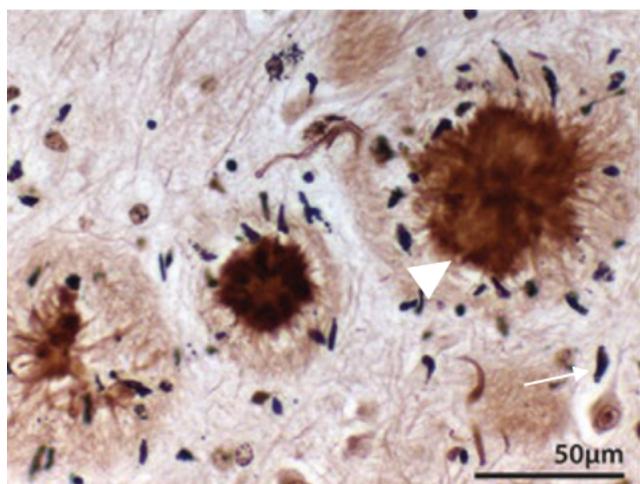
Figure 1.1 Neuropathology of Alzheimer's disease.
Silver staining of both extracellular amyloid plaques (white arrow head) and intracellular neurofibrillary tangles (white arrow) within post mortem sections of an Alzheimer's disease patient's cortex. Image taken from Serrano-Pozo et all, 2011.

1.3.1 Amyloid Plaques

Amyloid plaques primarily consist of insoluble deposits of Aβ which have aggregated between neuronal cells. These lesions are often associated with dystrophic neurites, neuronal loss and neuroinflammation.

The disturbed regulation of amyloid precursor protein (APP) processing is considered central to the amyloid pathogenesis of AD. APP is a single pass transmembrane protein mainly produced by neurons, which is then transported to axons and synapses. Its precise physiological function is not known, but studies have shown it to modulate synapses, neurite outgrowth and cell survival (O'Brien and Wong 2011).

APP has two processing pathways as described in figure 1.2, the amyloidogenic and non-amyloidogenic pathway. These pathways produce various products through sequential cleavage by proteases α (ADAM10), β (β-site APP cleaving enzyme 1 (BACE1)) and the γ-secretase complex (presenilins (PS), nicastrin, anterior pharynx defective 1 (APH-1), and presenilin enhancer 2 (PEN-2). The non-amyloidogenic pathway employs α-secretase to conduct the initial cleavage of APP, producing the products sAPPα, which is thought to have neuroprotective properties along with a C83 C-terminal fragment. This C83 fragment, which is retained within the membrane, is subsequently cleaved by γ-secretase to produce a small P3 fragment and the APP intracellular domain, which is a modulator of gene expression and actin cytoskeleton (Kojro and Fahrenholz 2005). In contrast to this, the amyloidogenic pathway is initiated by the protease β-secretase to produce the soluble sAPPβ product and a C99 fragment. As is the case with the non-amyloidogenic pathway this membrane bound C-terminal fragment (C99) is subsequently cleaved by γ-secretase, this time to release Aβ peptides of either 40 or 42 amino acids in length. This mainly consists of Aβ_{40} (90%) but Aβ_{42} is also produced (10%) (Qiu et al. 2015; Penke et al. 2017).
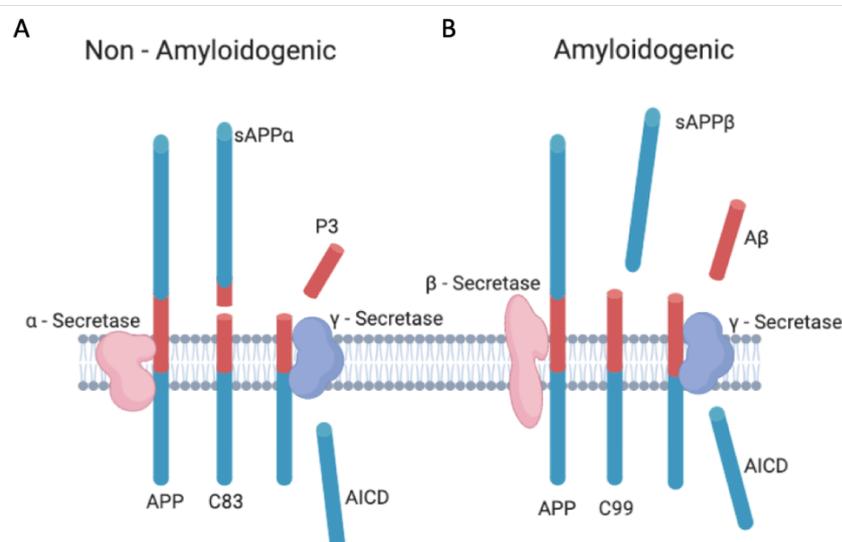
Figure 1.2 Processing pathways of the Amyloid Precursor Protein.
A. The non-amyloidogenic pathway employs sequential cleavage of APP by α and γ-secretases. The first cleavage event releases the sAPPα protein and a membrane bound C83 C-terminal fragment. This C-terminal fragment then undergoes a second cleavage event to release the amyloid precursor proteins intracellular domain (AICD) and a P3 fragment. **B.** The first cleavage event of the amyloidogenic pathway is initiated by β-secretase to release the sAPPβ protein and a C99 C-terminal fragment. The C99 fragment is then cleaved by γ-secretase to release the AICD along with amyloid-β peptides of both 40 and 42 amino acids in length. Figure created using BioRender.com.

Analysis of plaque ultrastructure via electron microscopy has revealed a dense core of filamentous Aβ peptides, which extends towards the periphery to interact with neuronal and glial cells (Serrano-Pozo et al. 2011). Amyloid plaques are mainly formed by aggregated Aβ due to its increased insolubility and tendency to form aggregates (Selkoe and Hardy 2016).

A staged progression of the amyloid plaque formation throughout the brain has been described by Braak and Braak *et al*. This 3-staged development initiates within the basal sections of the frontal and temporal lobes. During the second stage, amyloid aggregates spread within all isocortical regions. Finally, plaques appear throughout the thalamus, hypothalamus and cerebellum (Braak and Braak 1991). Accumulation of Aβ has also been identified within blood vessel walls in the form of cerebral amyloid angiopathy, however this tends to be the more soluble $A\beta_{40}$ species (Serrano-Pozo et al. 2011).

The discovery of the familial AD mutations within APP and Presenilin genes, which through altered protein processing result in increased Aβ production has provided the framework for developing the amyloid cascade hypothesis. This hypothesis states that imbalances in proteolytic cleavage and Aβ production trigger disease, resulting in processes such as NFT formation, synaptic dysfunction and neuronal cell death. The amyloid cascade hypothesis has been the most prevalent theory of AD pathology for many years. Many transgenic mice studies show Tau alterations and accumulation of NFTs occurring downstream of Aβ aggregation providing supportive evidence (Meziane et al. 1998; Lewis et al. 2001).

However, the amyloid cascade hypothesis is not without criticism. The main opposition comes from clinicopathological human and mouse data showing that amyloid plaque burden does not correlate well with cognitive impairment or neuronal cell death and therefore AD severity (Irizarry et al. 1997). Indeed, there have been cases in which individuals have high plaque load but are devoid of any AD symptoms (Giannakopoulos et al. 2003).

Therefore, it is suggested that Aβ is not acutely neurotoxic in itself but has the ability to disrupt neuronal processes and provoke neuroinflammation. Mounting evidence has turned attention to the smaller more soluble aggregations of the amyloid β peptide, Aβ oligomers. Many studies demonstrate that these soluble forms of Aβ correlate to a far greater extent with AD pathology such as synaptic dysfunction and neuronal loss (Jin et al. 2011). However, it is hypothesised that amyloid plaques may act as reservoirs releasing Aβ oligomers or indeed may be a potential protective mechanism via sequestering the cytotoxic Aβ species (Jin et al. 2011).

1.3.2 Amyloid Oligomers

In addition to Aβ peptides aggregating in large insoluble fibrils creating amyloid plaques, they can assemble as dynamic smaller, more soluble amyloid oligomers. Amyloid oligomers range from two to six monomers in size (Hayden and Teplow 2013; Chen et al. 2017).

Amyloid oligomers have been seen to reach levels of up to 70-fold higher within patients with AD compared to control individuals (Gong et al. 2003). Numerous studies isolating Aβ oligomers from the cortex of AD patients demonstrate their cytotoxic ability in both rodents and primary neuronal cultures as well as their ability to cause dysregulation the RNA binding proteins, heterogenous nuclear ribonucleoproteins (hnRNP) levels (Sackmann and Hallbeck 2020). Dysregulation of such multi-functional proteins contribute to disease pathology through their involvement in processes such as alternative splicing, transcriptional and translational regulation, cell cycle regulation, and axonal transport (Low et al. 2021). Such studies demonstrate that Aβ oligomers not only cause neuronal loss, but also disrupt the microtubule cytoskeleton, and induce Tau hyperphosphorylation (Shankar et al. 2008; Jin et al. 2011). Aβ oligomers have also been shown to have a dramatic effect on synapse structure and function causing a dose dependent decrease in synaptic number and inhibiting hippocampal long-term potential, resulting in disrupted signal transduction (Lambert et al. 1998).

This correlation of Aβ oligomers to early neurological dysfunction, occurring prior to neurodegeneration, explains impaired synaptic plasticity and subsequent memory dysfunction often seen during the early stages of disease before clinical presentation and diagnosis of AD (Lambert et al. 1998).

1.3.3 Neurofibrillary Tangles

Neurofibrillary tangles are another major hallmark of Alzheimer's disease. The major component of these intracellular aggregates is misfolded, hyperphosphorylated Tau protein. The Tau protein was one of the first microtubule associated proteins to be characterised. The main role of microtubules and their associated proteins within the neuron is to preserve the organisation and dynamics of the synapse (Arriagada et al. 1992; Lasagna-Reeves et al. 2012). The aberrant hyperphosphorylation of Tau seen under pathological conditions reduces the proteins' affinity to microtubules, resulting in cytoskeletal destabilisation.

The load and distribution of NFT are noted to correlate closely with the severity and clinical features of AD, such as synapse and neuronal loss. This is often another criticism of the amyloid hypothesis as there have been documented instances of AD with NFT pathology in the absence of Aβ plaques (Braak and Braak 1997).

NFTs have a standardised pattern of formation and progression through brain regions. The resulting degradation initiates within the entorhinal cortex, moving into the CA1 region of the hippocampus, before accumulating in the amygdala and thalamus and finally spreading to all the isocortical areas (Braak and Braak 1991). Along with this progressive topographical distribution, the development of NFTs occurs over three very distinct stages of morphological changes. Firstly, pre-NFTs are formed, where only punctate Tau staining can be identified within an otherwise normal neuron. The second stage is the formation of filamentous Tau, which aggregates within the cytoplasm, this is often referred to as mature fibrillar intraneuronal NTFs. This stage is also accompanied by pathogenic alterations within the cytoskeleton, distorting the appearance of the neuronal dendrites. Hyperphosphorylation results in migration of the Tau protein away from its primarily axonal

location towards the somato-dendritic compartment. The final stage is characterised by the presence of extra neuronal 'ghost' NFTs; these are formed on death of the neuron as a result of NFTs aggregation leaving behind only stainable cytoplasm (Braak et al. 1994).

These stages of NFT development appear to correlate with a distinct pattern and progression of Tau phosphorylation along with neuronal integrity (Augustinack et al. 2002). Early phosphorylation of Tau has been documented prior to filament formation and is thought to lead to conformational changes and pathological alterations within the cytoskeleton. For example, phosphorylation at residue Thr231 occurs within the early pre-tangle stage of NFT evolution, prior to Tau oligomerisation (Sengupta et al. 1998). This could support evidence of the 'prion-like' progression of hyperphosphorylated Tau aggregation throughout the brain (Clavaguera et al. 2009; Brundin et al. 2010).

1.3.4 Neuroinflammation

As is the case during peripheral inflammation, neuroinflammation is an inflammatory reaction within the central nervous system (CNS) that describes a set of immune responses initiated by variety of cues, such as infection, traumatic injury or toxic metabolites (Heneka et al. 2015). Glial cells such as microglia and astrocytes represent the main immune cells within the CNS. These cells act as the brain's first line of defence during injury through release of cytotoxic factors and phagocytosis of toxic molecules. They also play crucial roles in neurogenesis, neuronal plasticity and regeneration (Calsolaro and Edison 2016).

Some neuroinflammation is considered to have a beneficial and neuroprotective role. This has been shown through both mouse and cell culture experiments, indicating some neuroinflammation protects against AD progression via reducing Aβ plaque load through phagocytosis by activated glial cells (DiCarlo et al. 2001). However, this balance between neuroprotection and neurotoxicity is extremely complex. Excess neuroinflammation or its dysregulation can lead to tissue damage and disease pathology. As a result, chronic neuroinflammation is now considered to be another core feature of Alzheimer's Disease pathology (Kinney et al. 2018).

Chronic neuroinflammation and the subsequent increased production of pro-inflammatory cytokines by activated glial cells has been shown to have deleterious effects on brain health. Through their interactions with various AD pathologies they not only facilitate the creation of but exacerbate both Aβ and NFT lesions (Krstic et al. 2012; Lyman et al. 2014). Cytokines are cell signalling molecules secreted by immune cells, which act as chemical messengers to help regulate the immune response. These molecules can either be anti- or pro-inflammatory in nature. Pro-inflammatory cytokines such as Interleukin 1 Beta (IL-1β) and Tumour Necrosis Factor (TNF)$-\alpha$ play an integral role in the exacerbation of pathological conditions (Heneka et al. 2015). For example, overexpression of IL-1β exacerbates Tau hyperphosphorylation through its aberrant activation of the p38-MAPK kinase (Lee et al. 2010). Whereas, the TNF family of cytokines have been shown to be involved directly in neuronal loss through their involvement in the pro-apoptotic pathways (Micheau and Tschopp 2003). Therefore, both cytokines accelerate long term neurodegeneration and AD progression.

Low levels of chronic neuroinflammation and upregulation of pro-inflammatory cytokines correlates with age within both AD mouse models (Prolla 2002) and patient post mortem brain analysis (Bauer et al. 1991; Fillit et al. 1991; Fagiolo et al. 1993). As a result, neuroinflammation is considered a combined consequence of one or more AD pathologies and other risk factors such as genetics and age (Pizza et al. 2011).

1.3.4.1 Role of Microglia

Microglia are brain resident macrophages and considered to be the main immune cells within the CNS. They also play an important role in the maintenance of brain homeostasis and are involved in functions such as synaptic plasticity and remodelling as well as neuronal survival (Ji et al. 2013). It is becoming increasingly evident that an increased number of chronically activated microglia (microgliosis) play a major role in AD pathogenesis (Wang et al. 2015b).

Chapter 1: General Introduction

During homeostatic conditions 'resting' microglia adopt a highly ramified/branched morphology however, despite this name, they appear to be highly dynamic. Two-photon imaging in neocortex of mice show motile protrusions from these microglia to repeatedly and transiently sample their microenvironment, serving a house keeping function and enabling microglia to survey the entire brain (Nimmerjahn et al. 2005).

Microglia are derived from primitive macrophages within the yolk-sac which migrate during development to the CNS where they are maintained throughout life by local proliferation (Hickman et al. 2008). Microglia have been described to have alternative activation states characterised by specific ligands and marker expression, M1 and M2. Such activation states are driven by different stimuli similar to that seen in peripheral macrophages (Fig. 1.3). Within AD mouse models the pro-inflammatory M1 (classically activated) state, releases pro-inflammatory cytokines such as IL-1β, IL-12, IL-6 and TNFα. In contrast, the more anti-inflammatory M2 (alternatively activated) state is responsible for resolution and repair. This phenotype is consistent with deactivated macrophages and suppression of pro-inflammatory cytokines (Varnum and Ikezu 2012).
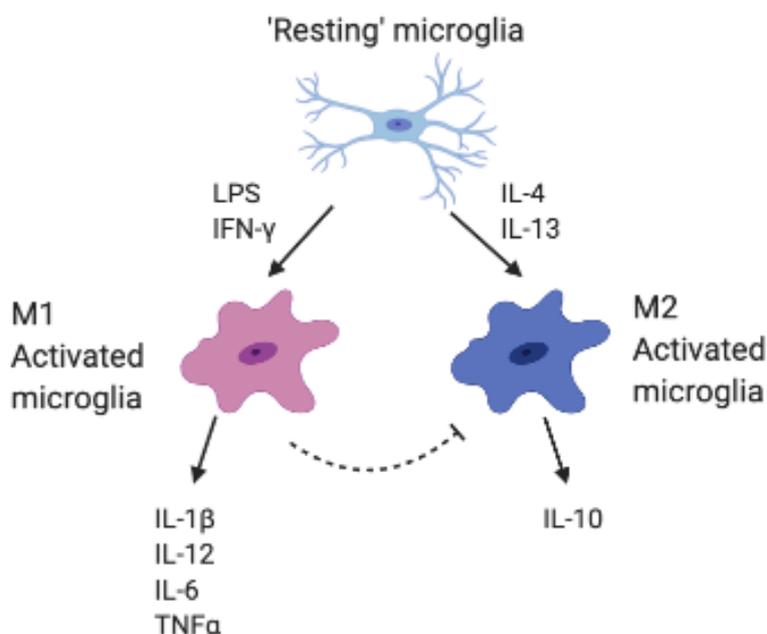
Figure 1.3 Microglial activation states.
Resting microglia exhibit a ramified morphology. Stimulation by Lipopolysaccharide (LPS)
and Interferon gamma (IFN-γ results in polarisation towards the M1 activation state and
release of pro-inflammatory cytokines such as IL-1β, IL-12, IL-6 and TNFα. In contrast IL-4
and IL-13 stimulate M2 activation, considered to be anti-inflammatory, releasing cytokines
such as IL-10. Figure created in Biorender.com, adapted from (Nakagawa and Chiba 2014).

However, this alternative activation state definition of microglia activation is now thought to
be over simplistic, with microglia in some studies demonstrating a mixture of both M1 and
M2 phenotypes (Zhou et al. 2017). Many studies showing M1/M2 activation states occur *in
vitro,* using purified stimuli and therefore fail to consider the tissue context of microglial
activation or their other homeostatic roles such as synaptic pruning and neuronal survival
(Ransohoff 2016).

Some level of neuroinflammation as a result of microglial activation, can be beneficial (Yang
et al. 2011). Upon activation, microglia undergo morphological changes to convert into a
more active and mobile ameboid shape, allowing their migration to sites of pathological
brain insult or injury (Varnum and Ikezu 2012). They have been noted in numerous studies
to co-localise to sites of Aβ plaques (Yokokura et al. 2011). The activated microglia can then
phagocytose Aβ through receptors such as CD36 scavenger receptors, reducing the Aβ
plaque load (Paresce et al. 1996).

Chronic microgliosis however, is considered an early pathological event during AD progression. It often correlates with synapse loss and cognitive decline (Edison et al. 2008; Okello et al. 2009), prior to NFT or Aβ plaque formation (Heneka et al. 2005; Yoshiyama et al. 2007) and potentially due to a result of brain vulnerabilities from other AD risk factors such as age. There is a characteristic shift from an M2 to a more M1 microglial activation state within aged brains. This results in an increased level of pro-inflammatory cytokines as well as decreased effectivity of the regulatory systems modulating pro and anti-inflammatory microglial phenotypes (Walker et al. 2009). This was observed to cause not only a decreased expression of Aβ clearance genes/receptors, resulting in increased Aβ plaque load within APP mouse models but also perpetuate the cycle of chronic age-related microgliosis contributing to disease pathology (Hickman et al. 2008; Stewart et al. 2010).

1.3.4.2 Peripheral contributions to neuroinflammation

There is now a large body of evidence that implicates systemic inflammation and blood brain barrier (BBB) dysfunction in the pathogenesis of AD, with brain infiltrating peripheral immune cells within aged and AD mouse models shown to contribute to disease progression (Lynch 2010; Ni Chasaide and Lynch 2020).

Chronic, systemic inflammation can have deleterious effects within the brain and serve to perpetuate the inflammation seen there. AD patients have been shown to display a more active adaptive immune system with increased levels of circulating peripheral cytokines (Goldeck et al. 2013). The relationship between the peripheral immune system and neuroinflammation was demonstrated in aged mice, where peripheral stimulation resulted in an amplified and prolonged neuroinflammatory response due to activation and priming of microglia (Godbout et al. 2005).

Numerous studies have shown increased presence of peripheral immune cells within the brains of AD patients (Stalder et al. 2005; Merlini et al. 2018). Peripheral immune cells migrate along gradients of cytokines and chemokines to infiltrate the brain. Indeed, Aβ$_{42}$ itself has chemoattractant abilities, increasing peripheral monocyte infiltration in BBB *in*

vitro models (Goldeck et al. 2013). Neutrophils also display preferential arrest in blood vessels containing Aβ and migrate towards amyloid plaques following extravasation (Baik et al. 2014).

Similar to brain resident glial cells, infiltrating peripheral immune cells display some Aβ phagocytosis ability, however, their clearance is often insufficient. Perivascular macrophages, for instance, have been linked to the severity of cerebral amyloid angiopathy, with increased levels reducing vascular amyloid deposits independent of Aβ clearance via microglia or astrocytes (Hawkes and McLaurin 2009). However, infiltrating immune cells have been implicated in perpetuating positive feedback loops, amplifying and sustaining chronic inflammation and tissue damage. Neutrophils for example, on infiltration into the CNS exert their neurotoxic effects through release of free oxygen radicals, proteolytic enzymes and neutrophil extracellular traps (NETs), perpetuating neuroinflammation and BBB damage (Manda-Handzlik and Demkow 2019). They have also been observed to physically block blood vessels within the brain due to increases in inflammatory adhesion receptors during AD pathology. The resulting reduced cerebral blood flow (CBF) is considered an early disease marker and helps to drive further inflammation and disease progression (Cruz Hernandez et al. 2019). Many studies have found that depletion of neutrophil numbers results in a protection from cognitive decline, reduced Aβ load and gliosis (Zenaro et al. 2015; Pietronigro et al. 2017). This build-up of infiltrating immune cells is thought to be a downstream event as a result of increased leukocyte adhesion molecules following microglia activation, Amyloid/Tau deposits and vascular remodelling (Togo et al. 2002; Rossi et al. 2011).

1.4 Blood Brain Barrier

The blood brain barrier is a highly specialised endothelial cell membrane of cerebral micro vessels. Its tightly regulated permeability allows the passage of essential nutrients and metabolites required for brain function, whilst providing a barrier to neurotransmitters and neurotoxic peripheral proteins such as albumin, pro-thrombin and plasminogen (Bradbury

1984). Together, these properties enable this highly complex system to provide a stable environment for neuronal function, homeostasis of brain proteins/metabolites along with regulation of cerebral blood flow (Muoio et al. 2014).

1.4.1 Structure of the blood brain barrier

Endothelial cells of the brain capillaries form the BBB through their tight junction contacts. Tight junctions seal the brain capillaries, limiting any flow of solutes into the paracellular space (Bradbury 1984). The BBB properties are supported and regulated through communication with various other cell types such as, pericytes, astrocytes and neurons as well as basement membrane proteins. All these components come together to structurally function as one neurovascular unit (NVU) (Fig 1.4) (Muoio et al. 2014).
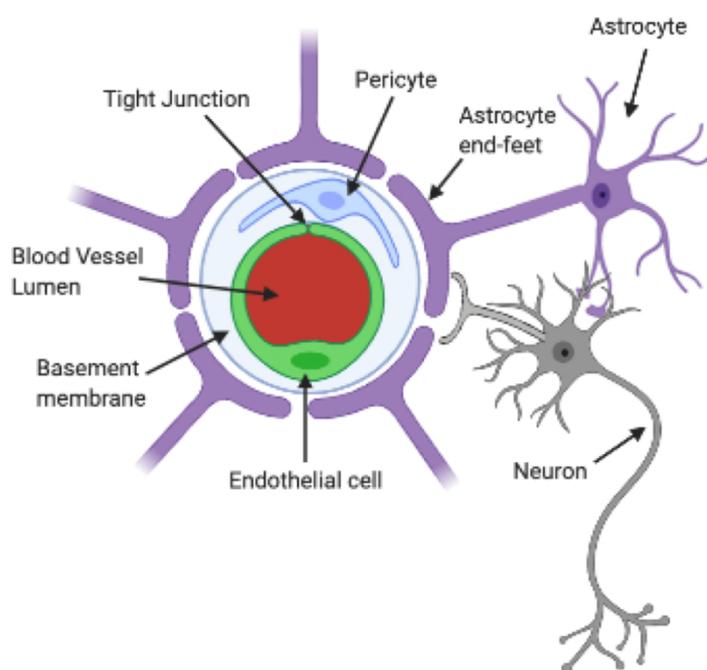
Figure 1.4: Structure of the Neurovascular Unit.
The Neurovascular Unit is comprised of brain capillary endothelial cells which form a barrier through tight junction contacts. BBB properties are maintained and regulated through contacts with other cells such as pericytes, astrocytes and neurons as well as basement membrane components. Schematic created in BioRender.com

1.4.2 Blood brain barrier dysfunction in Alzheimer's disease

AD is associated with neurovascular dysfunction and BBB breakdown (Nelson et al. 2016; Yamazaki and Kanekiyo 2017). Post mortem analysis of AD brains when compared to age-matched controls show neurovascular dysfunction, such as, a characteristic increase in pericyte and endothelial cell degradation, capillary leakages and disruption to the adhesion complexes responsible for maintaining BBB structure and a subsequent build-up of neurotoxic serum proteins (Thomas et al. 1996; Arvanitakis et al. 2016; Nelson et al. 2016). This occurs in an age-dependent manner with reductions in CBF and vascular dysfunction shown to be an early event during AD pathogenesis, occurring prior to and independently of Aβ and Tau pathology (Montagne et al. 2015; Nation et al. 2019). Indeed, an age-dependent breakdown of the BBB, especially within the hippocampus has been noted during normal aging (Montagne et al. 2015). Additionally, neurovascular uncoupling (failure of the CBF to react to neuronal activity) has also been noted to occur prior to cognitive decline (Ruitenberg et al. 2005; Bell et al. 2010; Kisler et al. 2017).

Neuroinflammation and chronic overexpression of pro-inflammatory cytokines play a key role in BBB dysfunction by altering the structure and function of cells within the NVU of the BBB. Gliosis as a result of activated astrocytes and microglia release cytokines, such as, Tumour Necrosis Factor alpha (TNFα), IL-6 and IL-1β, which have been shown to decrease the trans endothelial electrical resistance (TEER) of cerebral endothelial cells *in vitro* (de Vries et al. 1996; Ryu and McLarnon 2009). These cytokines also induce the release of matrix metalloproteinase (MMP) -9 and other inflammatory molecules from pericytes, resulting in further BBB breakdown (Takata et al. 2011).

Aβ and Aβ-induced oxidative stress also has the ability to cause BBB dysfunction. Aβ within the microenvironment of astrocytes can disrupt their structure and function, for example, decreasing the expression of the Aβ scavenger receptor lipoprotein related protein 1 (LRP1), therefore reducing the ability of cells within the NVU to endocytose and degrade Aβ (Donahue et al. 2006). Aβ also disrupts the tight junctions and adherens junctions within the NVU through causing reduced expression of the adhesion proteins, claudin-5, JAMs and ZO-

1 within brain endothelial cells, ultimately compromising the BBBs integrity (Carrano et al. 2012).

1.5 Genetic and non-genetic risk factors of Alzheimer's disease

1.5.1 Non-genetic risk factors

The complex aetiology of AD is thought to involve a combination of both environmental and genetic susceptibility. Therefore, in combination with genetic mutations and variations there is strong population evidence that LOAD risk is also influenced by modifiable risk factors such as, medical comorbidities, life style choices and environment (Xu et al. 2015).

A systematic review and meta-analysis of multiple population-based studies has highlighted the heterogeneity of such modifiable risk factors. These studies have indicated factors such as physical activity, education and socioeconomic status which may all provide protection against AD development, with factors such as high BMI during middle age having been identified as a risk factor. To this end, a healthy lifestyle with regular physical exercise and management of cardiovascular risk factors such as diabetes, hypertension and obesity may go some way to reduce the risk of cognitive decline (Baumgart et al. 2015).

Different sociodemographic factors also affect the prevalence of AD, such as age and sex. The proportion of people diagnosed with AD increases significantly with age, affecting approximately 5-8% of individuals aged 65, which increases to 25-50% of individuals over 85 years old (Duthey 2013). Sex also appears to have an impact on AD risk. In 2017, 16.3% of women within the UK died due to AD and other dementias compared to only 8.7% of men, making it the leading cause of female mortality (hUB 2018).

1.5.2 Genetic risk factors

The vast majority of AD cases are sporadic with no apparent family inheritance. However, a disease heritability of ~70% suggests some genetic influence. Late onset AD is thought to have numerous common, low-penetrance genetic variants that influence disease risk (Bellenguez et al. 2020), confirming the emerging consensus that common genetic variation plays a significant role in the aetiology of AD (Jansen et al. 2019).

1.5.3 Apolipoprotein E

The Apolipoprotein E4 (APOE) allele is the strongest genetic risk factor for sporadic AD. The apolipoprotein family is responsible for lipoprotein mediated lipid transport between organs. APOE, mainly produced by astrocytes and microglia is responsible for cholesterol and other lipid transportation to neurons as well as the clearance of Aβ (Mendez 2019; Rabinovici 2019). The ability and efficiency by which the APOE protein performs this function alters between alleles.

The APOE genes polymorphic alleles (E2, 3 and 4) are considered to be the primary genetic risk factor for AD (Bellenguez et al. 2020), resulting in a 3-15 fold increase in disease risk (Liu et al. 2013). The APOE4 allele has been associated with increased risk of developing the disease, with at least 40-70% of patients having at least one E4 allele, whilst the E2 allele has been associated with slight disease protection (Coon et al. 2007; De Luca et al. 2016). The population allele frequency of the E2 and E4 isoforms are 8.4% and 13.7% respectively (Rabinovici 2019).

The differences between these isoforms are the amino acid residue at positions 112 and 158. APOE2 has Cys residues and APOE4 has Arg residues at both the 112 and 158 positions. However, APOE3 has the residues Cys and Arg respectively. Alterations in these residues affects the stability of the lipoprotein binding N-terminal domain, producing an unstable region in the E4 isoform, ultimately effecting its Aβ processing ability (Zhong and Weisgraber 2009).

APOE is proposed to influence Aβ clearance through multiple mechanisms, such as cellular uptake and subsequent degradation, brain interstitial fluid to cerebrospinal fluid bulk flow or via transport across the BBB (Yamazaki et al. 2019). Clearance of Aβ across the BBB involves both the very low-density lipoprotein (VLDL) receptor and LRP1. However, on binding Aβ, APOE4 redirects clearance from LRP1 to the much less efficient VLDL receptor resulting in slower internalisation of the APOE-Aβ complex (Deane et al. 2008). As a result, the APOE4 allele is associated with increased aggregation and reduced clearance of Aβ, as well as increased severity of cerebral amyloid angiopathy (Schmechel et al. 1993; Kok et al. 2009). APOE4 status has also been associated with other hallmarks of AD such as, accelerated pericyte degeneration and loss of BBB integrity along with increased neuroinflammation (Halliday et al. 2016).

1.5.4 Common variants associated with Alzheimer's Disease

Despite APOE first being identified in 1993 (Saunders et al. 1993), it wasn't until 2009 that systematic high-throughput approaches, such as genome wide association studies (GWAS) started to identify common genetic variants such as single nucleotide polymorphisms (SNPs) to be associated with AD risk (Harold et al. 2009). Common risk loci typically have much smaller genomic effect sizes (odds ratio of <1.5), compared to that of the APOE allele (odds ratio of 3), therefore progress has been made to increase patient cohort sample sizes for genetic analysis. (Jansen et al. 2019).

The development of large consortia available from the UK Biobank has vastly increased the scale and statistical power of such studies. These include, the International Genomics of Alzheimer's Project (IGAP) containing multiple large GWAS data sets; Alzheimer Disease Genetics Consortium (ADGC), Cohorts for Heart and Aging Research in Genomic Epidemiology Consortium (CHARGE), The European Alzheimer's Disease Initiative (EADI), and Genetic and Environmental Risk in AD/Defining Genetic, Polygenic and Environmental Risk for Alzheimer's Disease Consortium (GERAD/PERADES), along with the use of GWAS by proxy data which allows the inclusion of individuals with immediate family members with have been diagnosed with AD. Recent studies have currently identified roughly 40 loci linked

with AD risk (Kamboh 2018). However, these loci are thought to only explain a small proportion of the heritability of AD, with further increases in sample size and therefore statistical power likely to yield additional variants.

The risk genes associated with these common genetic variants can be categorised based on their involvement in certain biological pathways; immune response and inflammation (these genes include, *CR1, MS4A, TREM2, CD33 and EPHA1)*, cholesterol and lipid metabolism (*ABCA7, APOE* and *CLU*) and endocytosis and synaptic function (*PICALM, CD2AP* and *BIN1*) (Misra et al. 2018). This indicates that genetic variants may interconnect and affect disease aetiology via converging on these common pathways (Kunkle et al. 2019a).

The vast majority of AD-associated variants identified through GWAS are located within non-coding regions of the genome. Only 2% of the 173 variants across 24 loci (excluding *APOE)*, identified using the IGAP GWAS data sets, are located within exons (Kunkle et al. 2019b). These non-coding variants are thought to contribute to disease pathology via affecting gene regulatory elements, altering their consensus sequences and subsequently influencing gene regulation.

Regulatory elements have been noted to act over considerable distances of up to 1Mb from the gene transcriptional start site. As a consequence, the target genes of these variants may not be the same gene within which the variants reside, but rather neighbouring genes (Amlie-Wolf et al. 2019). Insights taken from these genetic data aim to better characterise the pathophysiological pathways and gene expression alterations underpinning the biology of AD.

1.5.3 Polygenic Risk Score

AD is a multifactorial disease with multiple genetic and environmental components that combine to influence disease risk. A polygenic risk score (PGRS) aims to amalgamate all disease associated risk that is distributed throughout the genome into one predictive score (Escott-Price et al. 2015). High PGRS has been associated with poor memory and smaller

hippocampal volume in younger adults, with those individuals going on to develop greater cognitive decline and clinical progression. High PGRS has also been associated with AD-like levels of Aβ load (Mormino et al. 2016). Therefore, PGRS can act as a predictive tool for AD risk, enabling identification of individuals at high risk of AD, improving prognosis via early intervention.

Calculation of PGRS uses GWAS data sets but allows for the inclusion of variants that failed to reach the genome wide significance threshold ($p<5.10^{-8}$) (Lee et al. 2013). This allows them to increase the amount of disease heritability detected. Due to the small effect size of these common variants, collectively they are only able to explain part of AD heritability, leaving what has been termed 'missing heritability'. This describes the gap between AD estimated heritability and that which can be explained by known risk variants (Manolio et al. 2009). The high level of predictability of PGRSs in determining AD patients and unaffected individuals suggests that this unaccounted heritability is just not detected in GWAS due to its small effect size but has been captured (Escott-Price et al. 2017).

However, the genetic architecture and disease prediction still remains unclear. Zhang et al have noted greatest LOAD prediction accuracy when selecting only a small subset of variants, suggesting LOAD to be more oligogenic with possible <100 causal common variants as opposed to polygenic. They employed 22 of the highest scoring SNPs within their model which generated the genetic risk score with the highest level of disease prediction accuracy. The *EPHA1* locus SNP rs7791765, with a p-value of 7.1×10^{-14} was included in this list.

Their analysis potentially indicates only a small subset of all the known common AD-associated variants explain the majority of disease heritability and have the strongest influence on disease aetiology with other variants having a smaller more additive effect on disease progression (Zhang et al. 2020).

1.6 Gene regulation by the non-coding genome

Only around 1.5% of the whole genome is protein coding. The remaining non-coding genome was originally thought to consist of 'junk' DNA, however, projects such as the Encyclopaedia of DNA Elements project (ENCODE) has identified its crucial function in terms of gene regulatory ability (Consortium 2004). Further understanding the role of the non-coding genome and how its alterations may affect disease pathology is crucial to understanding the genetics behind AD progression.

1.6.1 Chromosome Structure

DNA is highly packaged into nucleoprotein complexes (nucleosomes), consisting of 147bp DNA wrapped around a histone octamer core. These cores are made up of either Histone 2A, 2B, 3 or 4 proteins linked together via small sections of linker DNA. Nucleosomes are further condensed via creating local interactions with neighbouring nucleosomes to form a 30nm chromatin fibre which are stabilised by the architectural histone protein H1 or H5 linker histones. The addition of further scaffold proteins such as the structural maintenance of chromosomes (SMCs)-type I and II proteins and fibre-fibre interactions allows super coiling of chromatin fibres into condensed heterochromatin and a higher degree of compaction that is observed in condensed chromosomes (Luger et al. 2012). Chromatin structure and packaging is illustrated in figure 1.5.
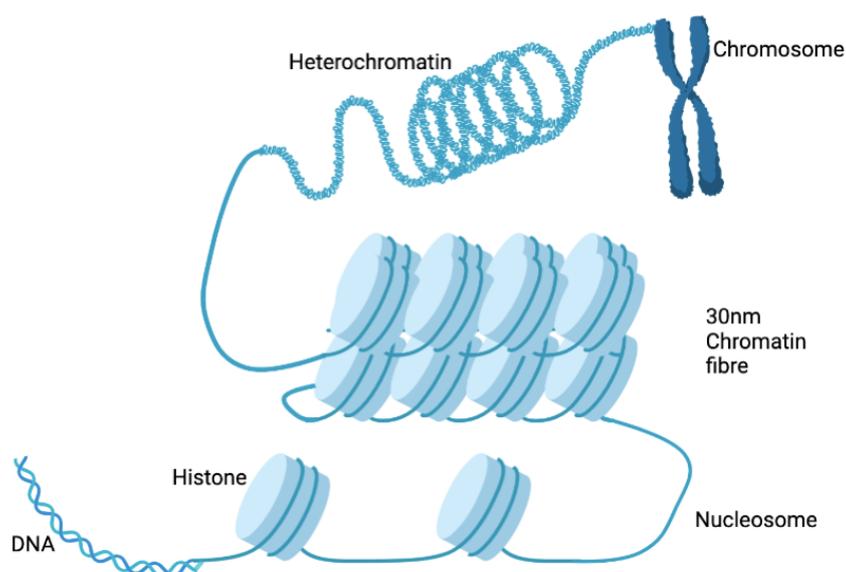
Figure 1.5: Illustration of chromatin structure.
Double stranded DNA is wrapped around Histone proteins to form nucleosomes which are joined by short sections of linker DNA. Interactions between neighbouring nucleosomes forms a 30nm chromatin fibre secondary structure. Chromatin fibres are then supercoiled to form the heterochromatin structure that is found within condensed chromosomes. Schematic created with BioRender.com

Several genomic structural elements, such as epigenetic histone modifications and 3D organisation of the chromatin structure, play a crucial role in allowing correct and efficient gene regulation. Tightly condensed heterochromatin is inaccessible to transcription factors (TFs) and transcriptional machinery resulting in repressed genes. In contrast a more open chromatin structure allows access of regulatory proteins to the underlying DNA and therefore, promotes gene expression. As a result dynamic modulation of chromatin packaging is important for gene regulation (Quina et al. 2006).

Post-transcriptional modification of Arginine or Lysine residues of the histone tails allows control over the extent of DNA packaging via affecting chromatin stability. Such modifications include phosphorylation, acetylation, methylation and ubiquitination. Certain Histone marks are associated with either silent or active genes, for example, acetylation of lysine residues 4 or 79 on H3 (H3K4ac and H3K79ac respectively) is associated with open chromatin and active genes. Whereas, trimethylation of lysine residues 27 or 9 of H3

(H3K27me3 and H3K9me3) are associated with silent genes and highly condensed heterochromatin (Bartova et al. 2008; Zhang et al. 2015).

Histone modifications are also associated with certain gene regulatory elements. For example, H3K4me3 is a highly conserved modification associated with promoters. This Histone modification facilitates gene expression by creating open chromatin regions with relaxed structure, allowing the recruitment of the RNA Polymerase II complex and various transcription factors. Whereas, H3K27ac and H3K4me1 are found at enhancer sites (Bannister and Kouzarides 2011). Therefore, specific histone modifications can be used to map the activation state and location of regulatory elements within the genome (Tessarz and Kouzarides 2014).

Chromatin remodelers also play a key role in the regulation of chromatin accessibility and therefore gene expression. Categorised into four families; SWI/SNF (switch/sucrose non-fermenting), INO80 (inositol requiring 80), ISWI (imitation switch) and CHD (chromodomain helicase DNA binding), remodelling proteins alter the confirmation of nucleosomes via the displacement and movement of histone octamers. They are recruited to specific regions of the genome through recognition of target DNA sequences or TFs (Clapier and Cairns 2009; Langst and Manelyte 2015). Chromatin remodelling proteins can also recognise and be recruited to histones themselves through the recognition of various modifications and other nucleosome features (Bannister and Kouzarides 2011).

1.6.2 Regulatory elements

The genome contains multiple regulatory elements such as such as enhancers, silencers and insulators which control gene expression by regulating chromatin organisation. Such elements are often located in non-coding regions and can act over large distances of up to 1Mb either up or downstream of their target gene's transcriptional start site (Perenthaler et al. 2019).

1.6.2.1 Transcription factors

Transcription factors either bind 6-12bp DNA sequences called TF binding motifs or bind to target DNA regions as part of a larger protein complex. TFs control gene expression in multiple different ways, such as, altering chromatin states via their recruitment of histone modification proteins, recruiting other accessory proteins required for gene expression such as co-repressors or co-activators or directly recruiting the RNA polymerase complex (Reiter et al. 2017). TFs are expressed in both tissue and cell type specific manners and alterations in their expression has been linked to diseases. For example, many polygenic disease-associated SNPs, identified through GWAS analysis have been shown to be enriched in regions of TF binding (Lambert et al. 2018). Therefore, disease-associated SNPs may impact TF function and subsequently gene expression via altering either the DNA-binding motif sequence specificity itself or if the variation lies outside the binding motif, disrupt or enhance the binding of additional proteins (Doane and Elemento 2017). This is seen with the SNP rs1421085 whose T to C allele alteration causes a disruption of a conserved repressor motif resulting in increased expression of the *IRX3* and *IRX5* genes during early adipocyte differentiation (Claussnitzer et al. 2015).

Therefore, investigations into how differing alleles of AD-associated SNPs may alter DNA-protein interactions between TF or other regulatory proteins is key to understanding the link between the casual SNPs at a given AD risk loci and identifying gene expression alterations involved in disease pathology. This is explored for the *EPHA1* risk locus within chapter 4 and 5.

1.6.2.2 Enhancers

Enhancers are cis-acting regulatory elements that facilitate gene expression. Located both up and downstream of either the target gene itself or a neighbouring gene, enhancers control gene expression via bringing together regulatory proteins such as trans-activators, recruiting other TFs or interacting with RNA polymerase itself (Heintzman et al. 2009; Deplancke et al. 2016; Doane and Elemento 2017).

Enhancers regulate gene expression often in a spatial- and cell type-specific manner, chromatin looping facilitates this by bringing them into close proximity to target promoters (Pennacchio et al. 2013).

Chromatin looping facilitates promoter and enhancer interactions allowing them to come together over spatial distances of up to 1Mb. Chromatin loops are 3D structures which separate regions of the genome that preferentially interact with each other, bringing for example, promoters and enhancers into closer proximity (Hnisz et al. 2016). Loop formation is facilitated by the architectural proteins cohesin and the zinc finger binding protein, CCCTC-binding protein (CTCF), this is illustrated in figure 1.6. CTCF proteins dimerise to set the boundary of chromatin and form a topologically associated domain (TAD) of up to 1Mb in size containing regions of the genome which preferentially interact with each other (Bell et al. 1999; Ong and Corces 2014).
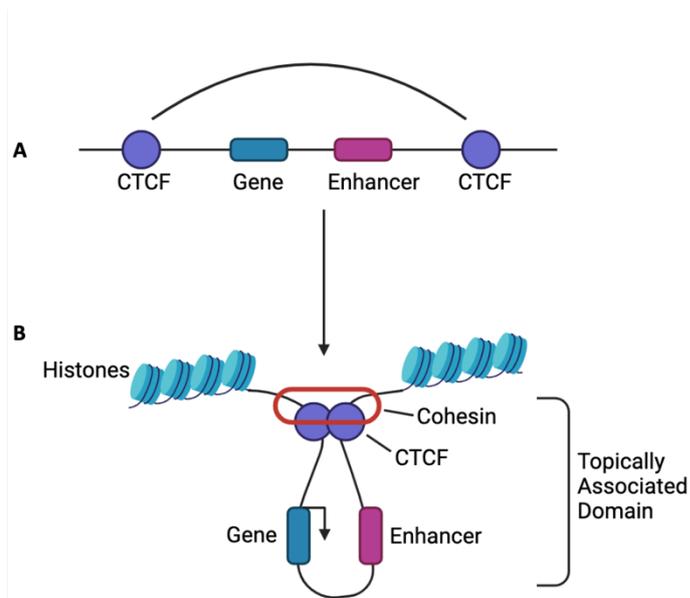
Figure 1.6 Gene regulation by chromatin looping.
A. Genome region containing gene and distant regulatory element such as an enhancer, flanked by two CTCF architectural proteins. **B.** CTCF proteins dimerise and along with another architectural protein, cohesion, create a chromatin loop or topically associated domain bringing the regulatory element and gene promoter into close proximity to drive gene expression. Illustration created in BioRender.com

1.6.2.3 Insulators and silencers

Additional regulatory elements include silencers and insulators. Silencers repress gene activity via binding repressor TFs and bringing them into proximity with their target promotor. Similarly to enhancers, silencers can function independently of distance and orientation to their target promoters (Ogbourne and Antalis 1998).

Insulators are approximately 0.5-3kbp regions of the genome that prevent gene transcription during neighbouring gene activity. This is done by preventing promoter-enhancer interactions and the spread of heterochromatin (Maston et al. 2006). For example, CTCF can act both as an activator or an insulator through either bringing together promoters and enhancers or binding their target insulator sequences to prevent transcription via blocking promoter-enhancer interactions (Kim et al. 2015).

1.6.3 Categories of non-coding DNA

The mammalian genome also undergoes extensive transcription of DNA regions which lack protein coding potential; around 50% of all mammalian transcripts are non-protein coding. Functional non-coding transcripts are categorised into two large families based on size (Morris and Mattick 2014). The small non-coding RNA family includes transcripts ranging from 18 to 200 nucleotides in length, these are split into three main categories; microRNA (miRNA), piwi-interacting RNA (piRNA) and small nuclear RNA. Transcripts over 200bp are categorised as long non-coding RNAs (lncRNA). LncRNAs are further sub-divided by genomic location, as illustrated in figure 1.7. For example, transcripts are located either within introns of coding genes (intronic), between coding regions (intergenic) or those transcribed from the antisense DNA strand which overlaps with a portion or the entirety of the sense coding gene (antisense) (Mercer et al. 2009; Derrien et al. 2012).
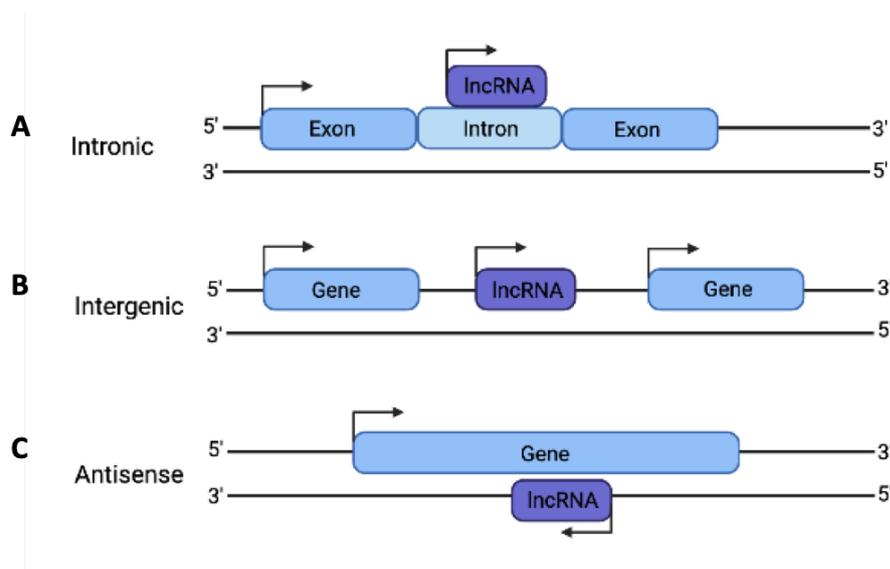
Figure 1.7: Categories of lncRNA.
LncRNAs are divided into categories based on their genomic location. **A.** Those which reside within gene introns (Intronic). **B.** In between two genes (Intergenic). **C.** On the antisense DNA strand overlapping a portion or the entirety of the coding gene. Schematic created using BioRender.com

Despite the development of technical approaches allowing the discovery and genome mapping of non-coding RNA transcripts, the exact functional role of lncRNAs remains elusive (Khalil et al. 2009). Expression pattern analysis has revealed that antisense lncRNAs are often co-expressed with their corresponding protein coding genes, which hints at co-regulatory ability (Guttman et al. 2009). As a result, hypotheses can be made regarding the function of specific lncRNAs based on that of the respective coding genes (Hung et al. 2011). Roughly 32% of all human lncRNAs are antisense to a coding gene, suggesting that gene regulation by antisense transcripts is a commonality (Derrien et al. 2012).

It is now widely recognised that non-coding RNA, mainly miRNA and lncRNA, possess regulatory ability, modulating cell function through altering gene expression often in a cell-specific manner (Morris and Mattick 2014).

1.6.4 Biology and function of long non-coding RNAs

Functional studies have implicated lncRNAs in multiple cellular processes, such as the cell cycle, pluripotency regulation and diseases, such as cancer and neurodegeneration

(Musunuru et al. 2010; Jendrzejewski et al. 2012). Recent studies have indicated that this regulation is orchestrated through a plethora of mechanisms both within the nucleus and the cytoplasm, such as regulation of chromatin remodelling, transcription, translation and mRNA stability as depicted in figure 1.8 (Rinn and Chang 2012). These proposed mechanisms fit within 3 main categories based on the interactions formed: RNA-RNA, RNA-DNA and RNA-protein (Spitale et al. 2011; Rinn and Chang 2012).

Within the nucleus lncRNAs can interact with various chromatin remodelling complexes, resulting in either transcriptional silencing or activation (Dreyfuss et al. 1993). Through acting as scaffolds for chromosome remodelling complexes, such as H3K9 or H3K4-specific methyltransferases, lncRNAs enable the addition of epigenetic markers at their target sites, resulting in the formation of open or closed chromatin regions (Fig 1.8A) (Nagano et al. 2008; Ghafouri-Fard et al. 2020).

Alternatively, lncRNAs can influence the transcription of genes as depicted in figure 1.8B. This allows them to act as either decoys to silence transcriptional activity via mimicking DNA binding domains, thus preventing regulatory proteins from binding target DNA (Nagano et al. 2008; Kino et al. 2010), or as chaperones to localise ribonuclear proteins to target DNA regions to drive expression (Li et al. 2014). Additionally, it can be the act of lncRNA transcription itself rather than its actions that may be responsible for its regulatory role. For example, transcription of lncRNAs may promote gene expression via the maintenance of open and active chromatin structure or interfere with other gene expression through their co-transcription causing polymerase collision (Long et al. 2017).

LncRNAs can also influence gene expression within the cytosol via altering the stability of mRNA transcripts or through translational regulation. The effect of lncRNAs on mRNAs can be seen through their simple binding and tagging of the transcript for decay, resulting in gene silencing or allowing gene expression by preventing mRNA degradation as seen in figure 1.8C (Zeng et al. 2019). Translational regulation by lncRNAs is very similar to the mechanisms employed for transcriptional regulation in which the lncRNA acts to recruit ribosomes or translational repressors, resulting in either gene expression or silencing (Fig. 1.8D) (Carrieri et al. 2012; Yoon et al. 2012).
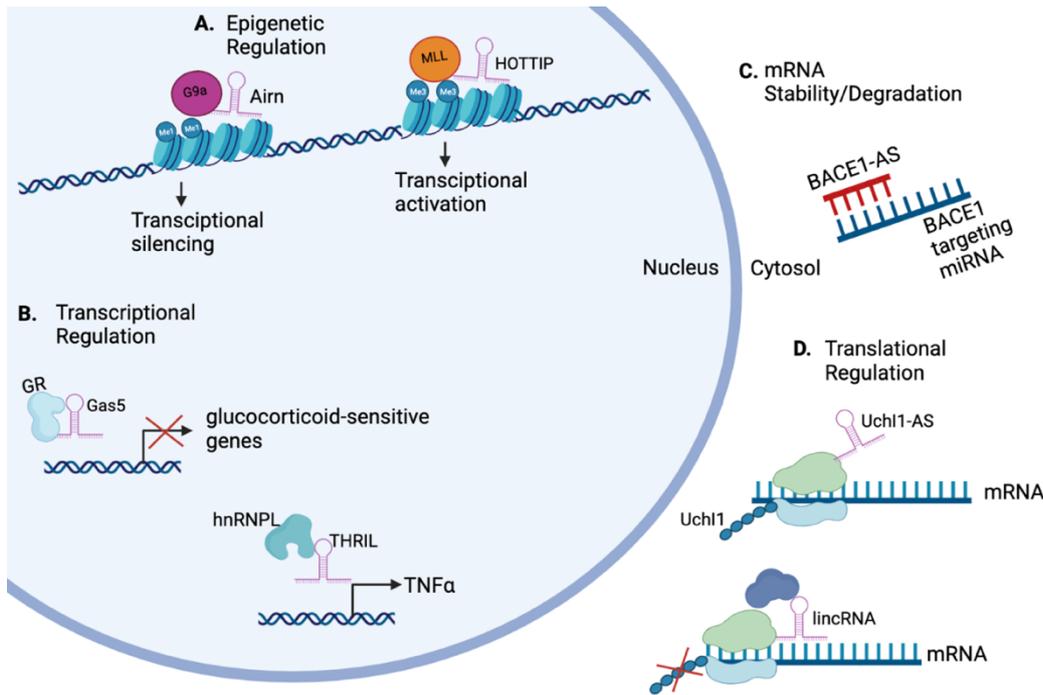
Figure 1.8: Gene expression regulation by lncRNAs.
A. Epigenetic regulation by lncRNAs. The lncRNA *Airn* can interact with the H3K9-methyltransferase G9a to bring about paternal silencing of *Igf2.* Gene activation can also be initiated this way via the interaction of lncRNAs such as HOXA transcript at distal tip (HOTTIP) with alternative chromosome remodelling complexes, such as the mixed lineage leukaemia complex (MLL), to allow methylation of lysine 4 on histone 3, resulting in activation of the HOX genes. **B.** Transcriptional regulation by lncRNAs. The lncRNA Gas5 binds glucocorticoid receptors preventing receptor activation via sequestering proteins away from their target DNA. This results in silencing of glucocorticoid receptor regulated genes. Gene expression alternatively can be driven via lncRNAs, such as the TNFα and the heterogenous nuclear ribonuclear protein L (hnRNPL) related immunoregulatory lncRNA (THRIL), interacting with ribonuclear proteins, such as hnRNPL, to allow transcription of target genes. **C.** The antisense to BACE1 (BACE1-AS1) prevents the degradation of BACE1 mRNA allowing gene expression. This is due to BACE1-AS1 sharing multiple miRNA-response elements with BACE1 enabling it to act as a decoy to sequester BACE1 targeting miRNAs preventing BACE1 mRNA degradation. **D.** Translational regulation by lncRNAs. Through their recruitment of ribosomes, lncRNAs can drive gene expression as seen with antisense UchI1 (UchI1-AS) and UchI1 protein expression or via the binding of translational repressors, this is seen during lincRNA-p21s silencing of its target mRNAs. Illustration created using BioRender.com.

1.6.5 Instances of lncRNAs within Alzheimer's Disease pathology

Due to their regulatory ability, it has been proposed that a skewed or altered balance of the gene coding sense RNA and the antisense lncRNA may lead to disease pathology (Li et al. 2010). Various antisense transcripts have been shown to influence multiple aspects of AD

pathology (Diniz and Teixeira 2011), for example, BACE1-AS, prevents the degradation of BACE1 mRNA, ultimately driving gene expression, as seen in figure 1.8C. This ultimately results in increased Aβ accumulation via an increase in the processing of APP through the amyloidogenic pathway (Kang et al. 2014; Zeng et al. 2019).

Such widespread implications of non-coding RNA in AD pathology has led to investigation into their therapeutic potential. One example of this is the creation of strategies which aim to interfere with actions of regulatory lncRNAs, blocking their alterations to gene expression which may drive disease pathology (Hung and Chang 2010). Another example is the administration of oligonucleotides to inhibit BACE1-AS, preventing its stabilisation of BACE1 mRNA. This results in reduced BACE1 expression attenuating BACE1-mediated APP cleavage, reducing the development of Aβ plaques (Liu et al. 2014).

1.7 Eph Receptors

1.7.1 Eph receptor structure

Amongst the earlier AD risk loci to be identified by GWAS were SNPs linked to the Ephrin type A1 receptor (EphA1) (Lambert et al. 2013). EphA1 is part of a large family of transmembrane erythropoietin-producing hepatocellular (Eph) tyrosine kinase receptors. Eph receptors are categorised into two sub-families based on their homology, the EphA sub-family (containing 9 receptors) and the EphB sub-family (containing 5 receptors). The main difference between the sub-families, detailed in figure 1.9B, lies within their ligand binding domains, resulting in Eph receptor binding specificity to ephrin ligands within their sub-type, although some promiscuity is found within this (Gale et al. 1996; Pasquale 2010). However, the main structure of both receptor subtypes is the same. The extracellular portion of the receptor is comprised of the N-terminal ephrin ligand-binding domain, a cysteine-rich region and two fibronectin type-III repeats. The intracellular C-terminal domain is comprised of a juxtamembrane region, a dual lobe kinase domain, a sterile α-motif (SAM) and a PDZ binding domain. Ephrin ligands are categorised based on their interaction with the cell membrane. The ephrin A (ephrinA1-5) sub-family are bound to the cell surface through a

glycosyl-phosphatidyl-inositol (GPI) anchor and ephrin B (ephrinB1-3) sub-family contain a transmembrane segment (Fig. 1.9A) (Zisch and Pasquale 1997; Taylor et al. 2017). EphA1 receptor signalling and receptor processing is analysed is greater detail within Chapter 3, Section 1.1.
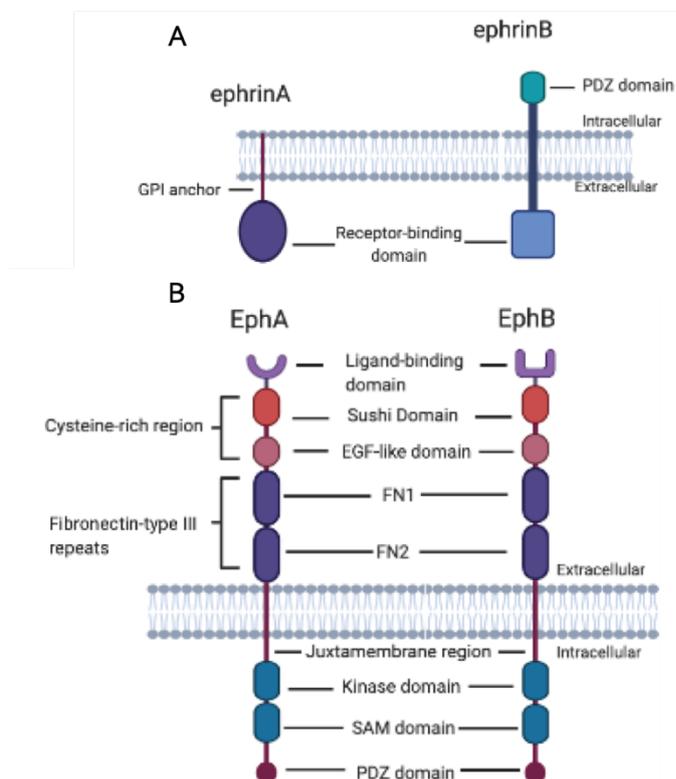
Figure 1.9: Eph and ephrin A/B structures.
A. Structure of both ephrinA and ephrinB ligands, detailing their glycosyl-phosphatidyl-inositol membrane anchor and transmembrane segment with PDZ domain respectively. **B.** Structures of both EphAs and EphBs, differing in their ligand binding domain. Each receptor consists of an extracellular N-terminal region containing either ephrinA or ephrinB ligand binding domain, a cysteine rich region (made up of a Sushi and EGF-like domain) and two fibronectin type-III repeats (FN1 and 2). Following an α-helix membrane spanning domain the intracellular region of Eph receptors is composed of a regulatory juxtamembrane domain, kinase and SAM domain. Figure designed using Biorender.com.

1.7.2 Eph receptors in Alzheimer's disease

Many hallmarks associated with AD such as inflammation, synaptic dysfunction and Tau hyperphosphorylation are thought to be regulated by kinases. Studies have indicated there

is an overall reduction in kinase activity within the early pre-clinical stages of AD (Rosenberger et al. 2016).

Emerging evidence suggests a role for Eph receptor signalling along with multiple additional effector kinases such as the phosphoinositide-3-kinase, src and Rho kinase families in AD aetiology (Brantley-Sieders et al. 2004; Knoll and Drescher 2004). For example, EphA4-ephrinA1 signalling has been implicated in contributing to both synaptic dysfunction within the hippocampus and BBB deterioration. This is caused via a decrease in tight junction proteins (ZO-1 and claudin-5) along with enhanced apoptotic induced cell death (Chen et al. 2018). EphA1 may also contribute to BBB dysfunction and perpetuate chronic neuroinflammation through its suggested role in the adhesion and transmigration of immune cells into the brain parenchyma in response to inflammation (Aasheim et al. 2005). Indeed, ephrinA1 expressed on endothelial cells can stimulate migration of CD4+ and CD8+ T-lymphocytes through activation of their EphA1 receptors (Hjorthaug and Aasheim 2007; Holen et al. 2010). This occurs to a greater extent in cell populations expressing L-selectin, an adhesion receptor involved in the transmigration of cells into high endothelial venules (Holen et al. 2010).

1.7.3 EPHA1 as an Alzheimer's Disease risk loci

EPHA1 was first identified as an AD risk locus in 2011 during a 3-stage association study conducted using the GERAD+ consortium containing 4 AD GWAS data sets; (GERAD, EADI, translational genomics research institute (TGEN1), and the Alzheimer's disease neuroimaging initiative (ADNI)) consisting of 6688 AD cases and 13,685 controls. The *EPHA1* SNP rs11767557 reached genome wide significance (p=6.0x10^{-10}) (Hollingworth et al. 2011; Naj et al. 2011).

Since this first identification, multiple replication studies have confirmed *EPHA1* as an AD risk locus. Multiple common variants within this locus reach genome-wide significance for AD association, the lead SNP however, fluctuates depending on GWAS data sets used. Jansen *et al* performed a large GWAS combining data sets from the IGAP, Alzheimer's

disease sequencing project (ADSP) and the Psychiatric genetics consortium (PGC-ALZ). This analysis produced variant rs7810606 (p=3.59x10^{11}) as the lead SNP within the *EPHA1* loci (Jansen et al. 2019). Another 2019 publication by Kunkle et al, using IGAP GWAS data containing the largest sample size of 21,982 cases and 41,944 cognitively normal controls, identified the most significant SNP at this locus to be rs10808026 (p=1.3x10^{-10}) (Kunkle et al. 2019a).

In addition to these common non-coding variants, a rare coding variation of EphA1, rs20217856 has been identified through targeted sequencing of 3 independent AD cohorts (176 patients from 124 Caribbean Hispanic families, 33 unaffected individuals from National Institute of Aging LOAD family study and 263 unrelated Canadian individuals of European dissent (210 patients and 53 controls)). This rare coding EphA1 variant segregated with instances of AD within an extended Caribbean Hispanic family, supporting its significant disease association (p=2.6x10^{-3}). However, this variant only reached nominal significance in Caucasian samples (p=3.07x10^{-2}) (Vardarajan et al. 2015). This rare EphA1 variation produces a nonsynonymous mutation at amino acid 460 located in the second fibronectin type III domain, causing a Proline to Leucine substitution (P460L). Crystal structures of EphA family members have shown a ~90° kink at the FN1-FN2 linkage region (Himanen et al. 2010). Therefore, it is hypothesised that this mutation may affect AD via altering receptor stability or trafficking to the membrane. Translating this genetic association into molecular pathway and gene expression alterations within specific cell types to ultimately understand the clinical significance of these AD-associated variants is extremely complex, requiring multiple functional and analytical annotations.

1.8 Hypothesis and overall aims

The overarching hypothesis of this thesis is that genetic alterations produced by the AD-associated GWAS variants contribute to disease pathology by impacting regulatory protein interactions with DNA.

Chapter 1: General Introduction

Using the *EPHA1* locus to establish proof-of-concept this thesis seeks to identify the common genetic variant/s that is/are responsible for the AD-association at the *EPHA1* loci discovered during GWAS. In addition, the variants' potential mode of action during AD pathology will be investigated via analysing their ability to regulate gene expression within cell types thought to be important during AD.

In parallel, the expression location and activation of the EphA1 receptor will be analysed. The rare P460L coding variant will be employed to draw comparisons to the wild type receptor to determine how this variant may alter receptor processing or activity to influence Alzheimer's pathology.

The overall aims of this thesis were:

- Investigate EphA1 receptor expression and ligand-dependent and -independent activation using HEK293 cell-based models

- Determine the impact of the variant P460L on EphA1 receptor biology

- Identify AD-associated variants at the *EPHA1* locus with the potential to alter gene expression

- Determine the impact of these variants on *EPHA1* expression and other genes within the locus using gene editing studies within induced pluripotent stem cell lines.

Chapter 2: Material and Methods

2.1 Cell lines

Multiple different cell systems were employed during this thesis to allow *EPHA1* receptor characterisation and variant investigation.

Human embryonic kidney 293 (HEK293) and the Flp-In expression system were used for the generation of isogenic cell lines containing both the wild type and P460L EphA1 receptor variant. Cell line generation and specific methods are detailed in Chapter 3.

The human monocytic cell line (THP1), human neuroblastoma cell line (SH-SY5Y) and the human microglia clone 3 (HMC3) cell line were employed as control cell lines for the analysis of *EPHA1* variant ability to bind nuclear proteins. The maintenance of these specific cell lines is outlined in Chapter 4.

Genetic editing to allow deletion of non-coding DNA regions for variant function analysis was conducted within the control human induced pluripotent stem cell (iPSC) line KOLF2-C1. IPSC specific maintenance, differentiation and CRISPR-Cas9 genome editing is detailed in Chapter 5.

2.2 Cell lysis

One well of a 6-well plate (Nunc) was lysed for western blot analysis once cell cultures had reached 80% confluency. This was conducted directly within the culture vessel, following any required cell treatment, via the addition of 200 μL 1X Cell Lysis Buffer (Cell Signalling Systems). After incubation at room temperature (RT) for 5 min the lysed cells were scraped and collected into a 1.5 mL Eppendorf tube. Cell lysate was then sonicated briefly and centrifuged at 300 g for 10 min at 4 °C. Supernatant was transferred to a fresh 1.5 mL Eppendorf tube.

Protein levels in cell lysates were measured using a Pierce bicinchoninic acid (BCA) Protein Assay Kit (ThermoFisher) as per manufacturer's protocol. Following protein quantification, a 1:1 ratio of Laemmli 2X Concentrate (Sigma-Aldrich) was added to each cell lysate and boiled for 5 min at 95 °C. Samples were then stored at -80 °C until required.

2.3 Sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS–PAGE)

The volume of cell lysate was altered to ensure 20 µg of protein was loaded for each condition. Samples were loaded alongside the Precision Plus Protein Standard (Bio-Rad) and were resolved on NuPAGE Novex 4-12% Bis-Tris gels using 1X MES SDS running Buffer (ThermoFisher Scientific) in an Invitrogen™ Mini Gel Tank and Blot Module Set at 100 V for 1 h or until the bromophenol blue dye of the Laemmli buffer had run to the bottom of the gel.

2.4 Western Blotting

The gel was floated off its plastic casing within a bath of 1X NuPAGE™ Transfer Buffer (ThermoFisher Scientific). A pre-soaked nitrocellulose membrane (ThermoFisher Scientific) was applied to one side of the gel and a sandwich created with blotting paper and sponges on each side. The transfer cassette was then inserted into the electrophoresis unit containing fresh 1X NuPAGE™ Transfer Buffer and 10 V was applied for 30 min.

Following transfer, the nitrocellulose membrane was blocked for 1 h with 5% milk in phosphate-buffered saline containing 0.01% Tween 20 (Sigma) (PBS-T) on a rocker at RT. After removing the milk and briefly washing with PBS-T, the membrane was transferred to a 50 mL universal tube (membrane curled inwards) and primary antibody (table 2.1) diluted in 5 mL PBS-T was added before being placed on a roller for 1 hr at RT. The universal tube and membrane were then moved to incubate at 4 °C overnight on a roller.

After another brief wash in PBS-T, the membrane was incubated in the dark with secondary antibody (table 2.1) diluted in 1% Milk in PBS-T for 1 h with rocking. The membrane was then washed for 30 min in PBS-T at RT with rocking.

Antibodies were then detected using a two-colour Odyssey infrared Imaging System (LI-COR).

Table 2.1: Western Blot antibodies.

Antibody	Species/Clonality	Dilution	Source	Catalogue number
V5 -Tag	Mouse Monoclonal IgG$_{2a}$	1/1000	Invitrogen	R960-25
N-terminal EphA1	Mouse Monoclonal IgG$_{2A}$	1/400	R&D Systems	MAB638
Anti-Phosphotyrosine	Mouse Monoclonal IgG$_{2bk}$	1/1000	Sigma-Aldrich	05-321
Anti-Phosphoserine	Mouse Monoclonal IgG$_{2bk}$	1/1000	ThermoFisher Scientific	MA1-91800
Anti-Nanog	Mouse Monoclonal IgG$_1$	1/500	ThermoFisher Scientific	MA1-017
Anti-Zyxin	Rabbit monoclonal IgG	1/1000	Abcam	ab109316
Secondary Antibody, Alexa Fluor Plus 800	Goat anti-Mouse IgG	1/15000	ThermoFisher Scientific	A32730
Anti-GAPDH	Mouse Monoclonal IgG$_{1k}$	1/1000	ThermoFisher Scientific	398600

2.5 Quantitation of western blot images

Western blots were quantified using the gel analysis software on ImageJ. A schematic of this quantitation process is detailed in figure 2.1.

On opening the western blot image within ImageJ, it was converted to a black and white 8-bit mode (Fig. 2.1A). The rectangle tool was used to draw a box around the bands within

each lane using the select first and next lane command under the gel analysis feature of the ImageJ software (Fig. 2.1B). Plotting the lanes provides a profile plot of peaks representing each gel band (Fig. 2.1C). As seen in figure 2.1D, the peaks do not extend to the bottom of the graph due to background noise. Therefore, it is necessary to close off the peaks using the line tool. The closed peaks were selected using the wand tool and the label peak feature was then used to express the size of each band peak as a percentage in relation to all peaks (Fig. 2.1E). This quantitation was carried out for both the experimental sample blots as well as the loading control blot and the percentage obtained used to calculate the relative density of each peak. This was done by dividing each sample peak percentage by that of the experimental control peak, therefore producing a value of fold change for of each sample and loading control relative to the experimental control.

Following from this the relative densities of the loading controls were used to scale those of the samples by dividing the sample relative density of each lane by their respective loading control relative density. This allows the production of an adjusted relative density for each sample considering slight alterations in protein loading within each lane. On plotting this final relative density within the Graphpad Prism V9 software, a T-test was preformed to determine if alterations in gene expression between two experimental conditions of interest was statistically significant.
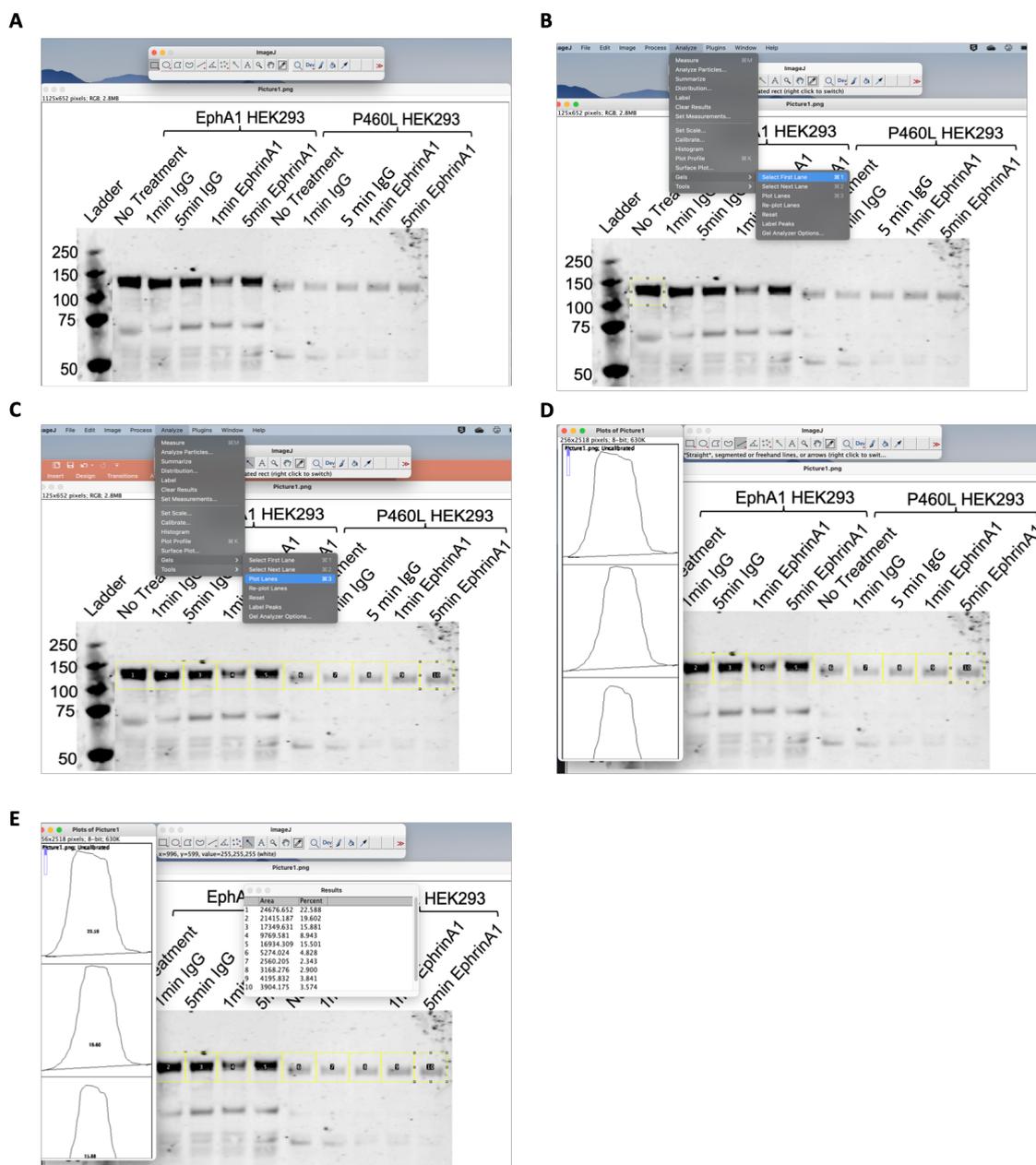
Figure 2.1: Schematic of western blot quantitation methodology.
A. A representative western blot image converted to black and white 8-bit. **B.** The band of interest within each lane were selected using the ImageJ square tool and the 'select first lane' and 'select next lane' feature of the gel analysis plugin. **C.** Once selected these lanes were plotted on a profile plot. **D.** Profile plot of each band within each selected lane. The line tool was used to close off each band peak. **E.** Each peak was then selection using the wand tool. This generated a percentage representing the band size in relation to all bands. This was then used to obtain a relative density of each band.

2.5 Stripping and re-staining of Western Blots

Antibodies were stripped from the blotting membrane using Restore™ Western Blot Stripping Buffer (ThermoFischer Scientific) following manufacturer's instructions.

Following initial imaging, blots were incubated in 20 mL Restore™ Western Blot Stripping Buffer for 30 min with gentle rocking. After incubation blotting membranes were washed briefly in 1X PBS and re-blocked in 5% milk in PBS-T for 1 hr at RT. Blotting membranes were re-stained with desired antibodies as per methods section 2.3.

2.6 qRTPCR analysis of gene expression

RNA was extracted from cells using an RNeasy mini kit (QIAGEN) as per manufacturer's instructions and quantified using a nanodrop spectrophotometer.

cDNA was synthesised using the QuantiTech Reverse Transcription Kit (QIAGEN). Samples were kept on ice when not incubating. A mixture of 2 μL of gDNA Wipeout buffer (7X), 1 μg of RNA and RNase-free water to a final volume of 14 μL was incubated for 2 min at 42 °C. Following addition of 4 μL Quantiscript Reverse Transcription buffer (5X), 1 μL Reverse transcription Primer Mix and 1 μL Quantiscript Reverse Transcriptase, samples were incubated at 42 °C for a further 30 min. Samples were subject to an additional incubation at 95 °C for 3 min in order to inactivate the reverse transcriptase. Samples were stored at -20 °C until required.

A quantitative real time PCR (qRTPCR) master mix was created containing 0.5 μL of both forward and reverse gene primers (final concentration of 200nM), 8 μL water and 10 μL SYBR green (ThermoFisher Scientific) per reaction. Primers where purchased from ThermoFisher Scientific, sequences are detailed in table 2.2. The qRTPCR master mix (19 μL) was added to desired wells of a 96-well plate followed by 1 μL cDNA. Samples were run on a QuantStudio 7 Flex RT-PCR machine using the parameters detailed in table 2.3.

Table 2.2: Primer Sequences used during q-PCR.

Primer	Sequence (5'-3')
EPHA1 Forward	CACGTGCTGAACCAGGATGAAGAACGGTAC
EPHA1 Reverse	CACTGTCATAGGAGAAGGAGAGTTTGGGGA
EPHA1-AS1 Forward	AAACTAAAGAAAAAGAAGGCAGCAACTGCT
EPHA1-AS1 Reverse	ACCCCTGAGAGTGATATTATTCTACAGAAAC
ZYX Forward	GGGGTCACCAAGGGGAGCTG
ZYX Reverse	CATCTGCTCGGGACAGGGTG
Nanog Forward	CTCCAACATCCTGAACCTCAGC
Nanog Reverse	CGTCACACCATTGCTATTCTTCG
E-Cadherin Forward	GCCTCCTGAAAAGAGAGTGGAG
E-Cadherin Reverse	TGGCAGTGTCTCTCCAAATCCG

Table 2.3: qRTPCR cycle parameters.

Step	Time	Temperature (°C)
PCR initial heat activation	15 min	95
3-step cycling*: Denaturation	 15 sec	 94
Annealing	30 sec	60
Extension	30 sec	72

* 40 cycles preformed

2.7 Quantitation of qRTPCR data

Amplification results were plotted as a delta normalised reporter value (ΔRn) against cycle number/cycle threshold (ct). Rn is the fluorescence signal from SYBR green normalised to the signal of the reference dye. The ΔRn therefore, is the experimental Rn minus the baseline Rn generated by the equipment. This enables the magnitude of the specific signal to be calculated.

Amplification curves were analysed to allow comparison of the ct values of samples relative to the housekeeping genes. EphA1 HEK293 cDNA was used as a positive control and GAPDH as the housekeeping gene. An example of amplification curves generated by the GAPDH

housekeeping and *EPHA1* primers within the undifferentiated iPS cells and iPSC-derived monocytes are shown in figure 2.2. Low ct values of around 10-20 cycles as in the case of the GAPDH housekeeping gene imply high gene expression as seen in figure 2.2A. Ct values of 30 and over was considered indicative of no gene expression, this is seen for EPHA1 expression within the iPSC-derived cell lines (Fig. 2.2B).
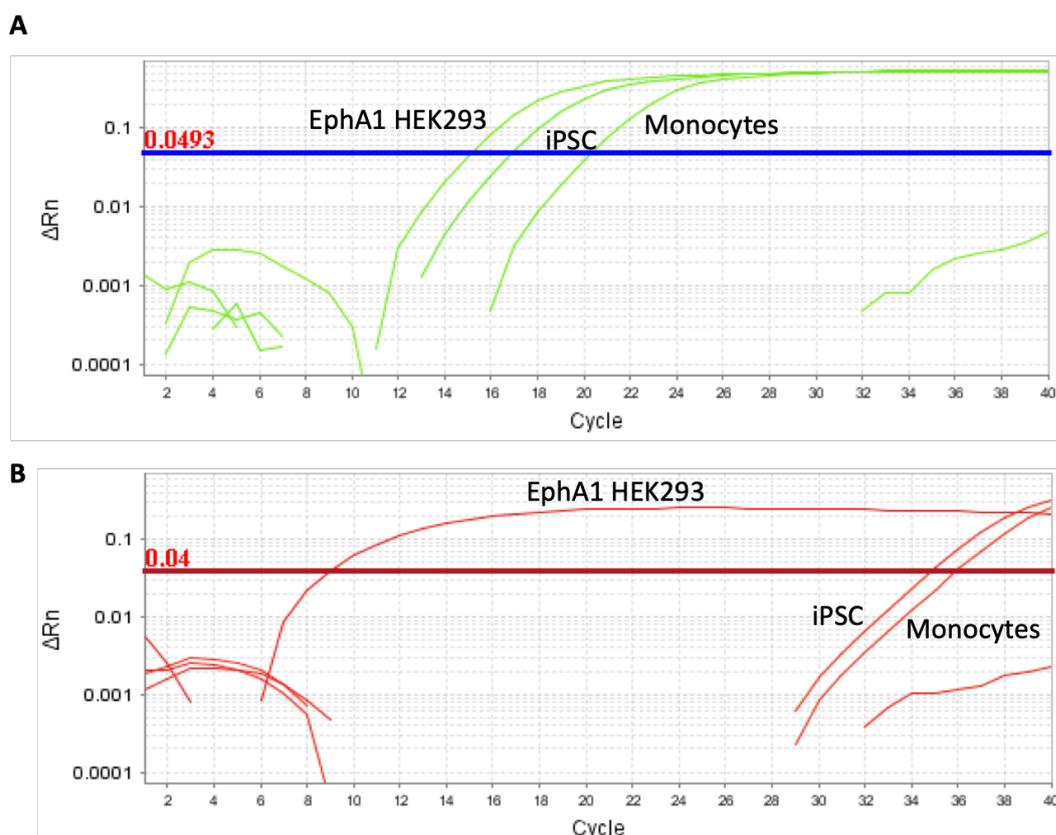
Figure 2.2: qRTPCR amplification curves.
A. Graphical representation of GAPDH qRTPCR amplification curves of EphA1 HEK293, undifferentiated iPS cells and iPSC-derived monocytes. **B.** Graphical representation of *EPHA1* qRTPCR amplification curves of EphA1 HEK293, undifferentiated iPS cells and iPSC-derived monocytes. Magnitude of amplification is represented by ΔRn and threshold is indicated by the horizontal line.

The double delta ct (2^ΔΔCt) formula: Expression fold change = 2^ - [(Experimental Ct-housekeeping Ct) - (Control Ct-housekeeping Ct)] was employed. Where 'experimental' represents the tested samples and 'control' represents the sample used as a reference, for example EphA1 HEK293 cells, CRISPR-Cas9 control induced pluripotent stem cell (iPSC) lines or undifferentiated iPSCs.

2.8 Statistical analysis

Data was visually determined to be normally distributed following examination of the Q-Q plot showing a roughly straight line.

To allow data analysis during this thesis two different statistical tests were performed dependent on the number of factors within the experiment. Where only two groups are present a Students t-test was employed to compare the two means. During experiments where multiple groups are present a 2-way ANOVA (analysis of variance) was performed. A Turkey's method analysis was also conducted as a post-hoc analysis following the ANOVA to allow the identification of specific differences between group means.

Chapter 3: Analysis of EphA1 receptor function and activation

3.1 Introduction

Despite EphA1 being the first Eph receptor to be identified and isolated, it still remains one of the least well characterised members of the family, with information about its function being inferred from its closest homolog, EphA2 (Hirai et al. 1987). The aim of this chapter is to better characterise the stability, turnover and activation of the EphA1 receptor within the cell membrane, under both homeostatic and ligand-activated conditions. In addition, alterations to such functions as a consequence of the AD-associated EphA1 receptor variant, rs20217856 (P460L), were investigated.

3.1.1 Eph receptor signalling and activation

A unique feature of Eph-ephrin interactions is that signalling can occur in a bidirectional manner. Forward signalling, as illustrated in figure 3.1B, is considered to be the canonical signalling pathway with activation resulting in downstream signalling cascades within the receptor-containing cell. Ephrin ligand binding initiates autophosphorylation of tyrosine residues within the juxtamembrane region of the receptor, producing a conformational change allowing further phosphorylation of the kinase domain (Darling and Lamb 2019). This forward signalling is thought to modulate cytoskeletal rearrangements allowing, for example, neurite outgrowth and cell motility (Yang et al. 2018). On the other hand, reverse signalling acts within the ephrin-expressing cell (Fig. 3.1B). The signalling cascade resulting from reverse signalling is not fully understood, especially in the case of the ephrinA ligand sub-type which lacks intracellular domains. It is thought however, that this signalling is orchestrated by other kinases, such as Fyn from the src family of kinases, which can phosphorylate tyrosine residues on the intracellular domain of the ephrinB (Pasquale 2010; Gucciardo et al. 2014).

Chapter 3: Analysis of EphA1 receptor function and activation

Often forward and reverse signalling cascades produce opposite cellular responses, allowing for tight regulation of processes within cells. For example, neuronal EphA4 and EphA5 forward signalling and glial ephrinA3 reverse signalling act together to regulate synaptic plasticity within the hippocampus (Das et al. 2016; Yang et al. 2018). Different responses resulting from Eph receptor activation can also be a consequence of whether the receptor is being activated through conventional trans signalling via neighbouring cell-cell contact, or through cis signalling, whereby an Eph receptor can be activated by an ephrin ligand expressed within the same cell membrane (Taylor et al. 2017).

Both receptor and ligand clustering are often required for efficient Eph signalling. Ligand binding results in the formation of Eph-ephrin heterodimers and tetramers, forming large receptor signalling arrays (Himanen et al. 2010; Janes et al. 2011). Receptor aggregation is thought to be mediated through the PDZ-binding motifs which act as scaffold proteins and regulate processes such as receptor clustering (Ye and Zhang 2013). The SAM domain is unique to the Eph family of receptor tyrosine kinases (RTKs) and has also increasingly been associated with mediating receptor clustering, with interactions within this domain shown to stabilise EphA3 dimers (Himanen et al. 2010; Singh et al. 2015).

Activation of Eph receptors can occur in a ligand-dependent or -independent manner, as illustrated in figure 3.1. Indeed, a number of active dimers in the absence of ligands are thought to exist (Artemenko et al. 2008). As with bi-directional signalling, these two activation methods also enable Eph receptor activation to result in opposing cellular outcomes, which allows tight regulation of cellular processes such as cell migration and repulsion (Miao and Wang 2012; Taylor et al. 2017). This is demonstrated by both the pro- and anti-oncogenic consequences of EphA2 activation, where a reciprocal regulatory loop is formed between EphA2 activation and the serine-threonine specific kinase Akt. Ligand-independent activation of EphA2 results in phosphorylation of the serine residue S897 by Akt, promoting cell migration and invasion (Fig. 3.1A). This site is subsequently dephosphorylated on ephrinA1 ligand activation of EphA2, thereby deactivating the Akt kinase and inhibiting cell migration (Fig. 3.1B) (Miao et al. 2009). Ligand-independent serine phosphorylation in this manner is conserved within the EphA1 receptor (Taylor et al. 2017)
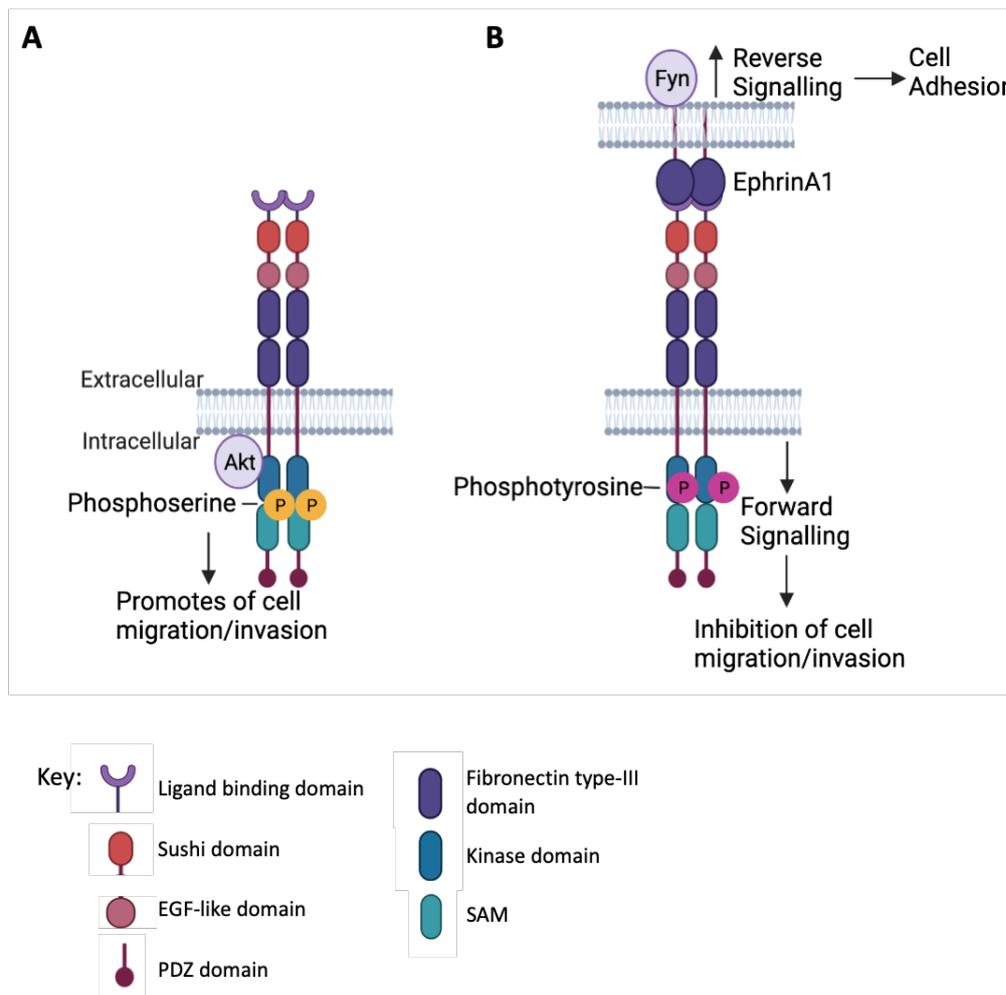
Figure 3.1: EphA1 and 2 receptors signalling and activation.
A. Ligand-independent activation of the EphA1/2 producing AKt-dependent phosphorylation of the serine residue S897, leading to cell invasion and migration. **B.** Ligand-dependent activation of the EphA1/2 receptor on ephrinA1 ligand binding, causing phosphorylation of tyrosine residues within the juxtamembrane and kinase domains. Signalling cascades within the receptor-containing cell is termed forward signalling. This classical form of activation often leads to cell migration and proliferation. Reverse signalling within the ephrinA1 ligand-expressing cell often leads to opposing cell behaviours, such as cell adhesion. Signalling in this manner is thought to be mediated by src family kinases, such as Fyn. Schematic created with BioRender.com

All these unique characteristics of Eph-ephrin signalling enables a highly varied cellular response to receptor activation, despite the high structural similarity between receptor sub-types. This results in multiple outcomes becoming possible based on the spatial and temporal expression patterns of both the receptor and ligand, as well as their cell and microenvironmental context (Darling and Lamb 2019). A perfect example of cell type-dependent alternative cytoskeletal rearrangements produced from EphA activation,

resulting in either the coordinated retraction of dendritic spines or the extension of astrocytic filopodia (Nestor et al. 2007).

Analysis of ligand binding interactions of EphA1 with different ephrinA-Fc fusion proteins has indicated that EphA1 binds ephrinA1 with the highest affinity with weaker binding to ephrinA3 and ephrinA4 (Coulthard et al. 2001). The ephrinA1 protein has also been shown to be the candidate ligand for EphA1 on T cells (Aasheim et al. 2005; Holen et al. 2010). To enable ligand signalling, ephrinA1 can either exist on the membrane of a neighbouring cell or in a monomeric soluble form following cleavage via multiple metalloproteases. Both monomeric soluble ephrinA1 and artificially clustered homodimers of ephrinA1-Fc have the same functional capacity in activating EphA2, producing ligand-dependent internalisation (Beauchamp et al. 2012). Therefore, within this chapter, ligand-dependent signalling of the EphA1 and P460L receptor were analysed using the artificially clustered ephrinA1-Fc ligand fusion protein.

3.1.2 EphA1 receptor processing by proteases

Proteases allow regulation of some receptors, such as RTKs, through promoting protein internalisation, degradation and turnover (Atapattu et al. 2014). Ligand activation of RTKs typically results in internalisation, after which the receptor can continue to signal until dephosphorylation and degradation (Goh and Sorkin 2013). There is wide acceptance that Eph receptors and their ephrin ligands are substrates for many different types of proteases such as MMPs, A Disintegrin And Metalloproteases (ADAMs) and γ - secretase (Mancia and Shapiro 2005; Lisabeth et al. 2013). During Eph signalling, both the Eph receptor itself and the Eph-ephrin complex can be internalised into the receptor-expressing cell through a process termed *trans*-endocytosis. This rac1-dependent internalisation of the Eph-ephrin complex effectively removes cell-cell contacts allowing cellular repulsion. On internalisation, Eph receptors are degraded via the Cbl ubiquitin ligase (Walker-Daniels et al. 2002).

Chapter 3: Analysis of EphA1 receptor function and activation

ADAMs are membrane-anchored metalloproteases belonging to the metzincin proteinase super-family. They mediate cleavage of a wide range of membrane proteins both constitutively and as a result of stimuli such as intracellular calcium levels and peptide growth factors (Hattori et al. 2000). ADAMs have been observed to interact with, and cause shedding of, ephrin ligands or Eph receptors to allow cell repulsion as a result of Eph-ephrin signalling. For example, as illustrated in figure 3.2A, ADAM10 interacts with the EphA4 receptor during the tight regulation of axonal guidance (Scilabra et al. 2018). ADAM10 has been reported to cleave ephrinA2 at locations where a neurite touches a fibroblast cell surface to allow cellular repulsion (Seals and Courtneidge 2003).

MMPs, similar to ADAMs, have a broad range of protein substrates and contribute to the homeostasis of many tissues during processes such as wound healing and angiogenesis. They exist as either secreted or membrane-anchored proteins. It is the membrane-bound subset that is mainly responsible for the shedding of cell surface and membrane-spanning proteins, such as Eph receptors and ephrin ligands (Loffek et al. 2011). Membrane type-1 MMP has been found to cleave EphA2 in *cis* within surface complexes on the same cell following receptor activation by the ephrinA1 ligand. This cleavage occurs within the fibronectin type-III domain 1 as depicted in figure 3.2B and results in RhoA activation and subsequent internalisation of the EphA2 receptor promoting cell detachment (Sugiyama et al. 2013).

γ-secretase is a membrane-embedded proteolytic complex that enables protein regulation through intramembrane proteolysis; at least half of the human genome's RTKs are regulated in this way (Merilahti and Elenius 2019). Regulatory cleavage by γ-secretase is generally a two-step process, firstly involving release of the target protein's ectodomain usually via an MMP or ADAM. This ectodomain release creates a substrate for the subsequent cleavage by the γ-secretase. Cleavage occurs at the protein's transmembrane region and creates a soluble intracellular domain that can interact with other proteins in a variety of cellular compartments (Fig. 3.2C). For example, this intracellular domain may interact with transcriptional regulators to influence processes such as cell proliferation or can facilitate receptor turnover and degradation. Although cleavage by γ-secretase has been noted to

occur constitutively, ligand activation is often required as seen in the case of ligand-dependent internalisation of EphB2 (Litterst et al. 2007). However, in a screen using the inhibitor DAPT, EphA1 was found not to be a substrate of γ-secretase regulated cleavage (Merilahti et al. 2017).
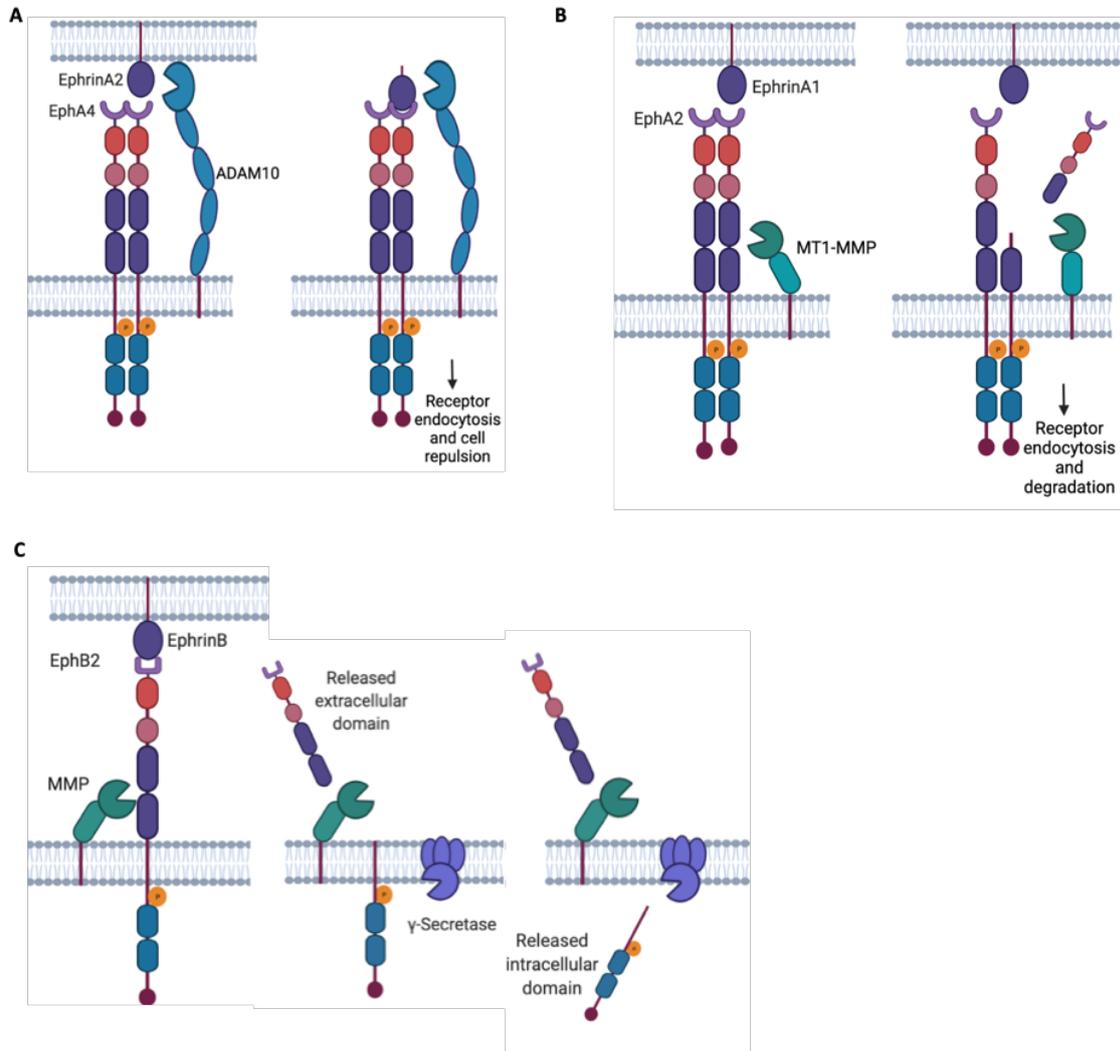
Figure 3.2: Proteolytic regulation of Eph receptors.
A. ADAM10 regulates ligand-activated EphA4 by cleaving membrane-bound ephrinA2 ligand allowing internalisation of the receptor-ligand complex and cell repulsion. **B.** Membrane type-1 MMP (MT1-MMP) regulation of EphA2 via receptor cleavage at the first fibronectin type-III domain, releasing the receptor's N-terminal ectodomain and causing receptor internalisation and degradation. **C.** The sequential, two-step process of γ-secretase regulation of EphB2. The first step is orchestrated by an MMP or ADAM, releasing the receptor's N-terminal domain. This is followed by γ-secretase cleavage of the intracellular domain. Schematic created using BioRender.com

3.1.3 Creation of Wild Type and P460L variant EphA1 isogenic cell lines

To allow effective comparison between different forms of the EphA1 receptor, isogenic Flp-In HEK293 cell lines (ThermoFisher Scientific) were generated by Dr Helen Owens (within the Ager/Knäuper Labs) which stably express functional, V5 and His-tagged EphA1 in either its full-length, wild type (EphA1 HEK293) or the P460L variant form (P460L HEK293). This is due to the Flp-In system allowing insertion of different cDNAs into the same genomic location, enabling protein expression to be controlled by the same promoter. A schematic of the receptor composition including the location of the P460L variant can be seen within figure 3.3. Receptor sequence with the location of the P460L variant are displayed in Appendix I.
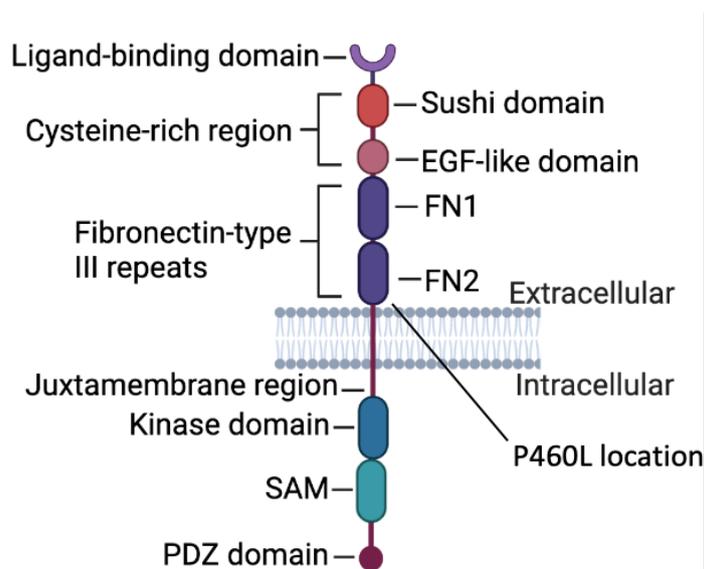
Figure 3.3: Structure of the EphA1 receptor.
The extracellular region of the receptor comprises of the ligand binding domain and the cysteine rich region containing the Sushi and EGF-like domains. This is followed by two Fibronectin-type III repeats (FN1 and FN2). The second of which contains the P460L receptor variation. The intracellular portion of the receptor contains the Juxtamembrane region, Kinase, SAM domain and PDZ domains. Schematic created using biorender.com

These transfected HEK293 cell lines were employed during receptor characterisation assays, for example, imaging flow cytometry analysis of the subcellular localisation of both homeostatic and activated receptor. Additionally, western blot analysis of phosphotyrosine and phosphoserine levels under both homeostatic and ligand induced receptor activation. Due to the nature of such isogenic cell lines any alterations in receptor expression or

activation levels could be attributed to effects produced as a result of the receptor variation and therefore provided comparison of both receptors.

The desired cell lines were created using the Flp-In recombination system which utilises different expression vectors in order to create isogenic cell lines. The initial pFRT/lacZeo vector creates a control cell line containing the flippase recognition target sites in addition to a lacZ-Zeocin^TM fusion gene (Fig. 3.4A) at a desired location within the genome.

The second vector, pcDNA5/FRT contains the *EPHA1* variant of interest (EphA1 or P460L EphA1) in addition to a V5 and His-tag (Fig 3.4B). Co-transfection into the above created control cell line with a pOG44 plasmid, allows the constitutive expression of the Flp recombinase under the control of a CMV promoter and drives homologous recombination between the two FRT sites. Thus the insertion of the pcDNA5/FRT vector and the *EPHA1* variant (Fig. 3.4C) (Broach et al. 1982). Cells which have successfully integrated the EphA1 variant can then be selected for based on hygromycin resistance.
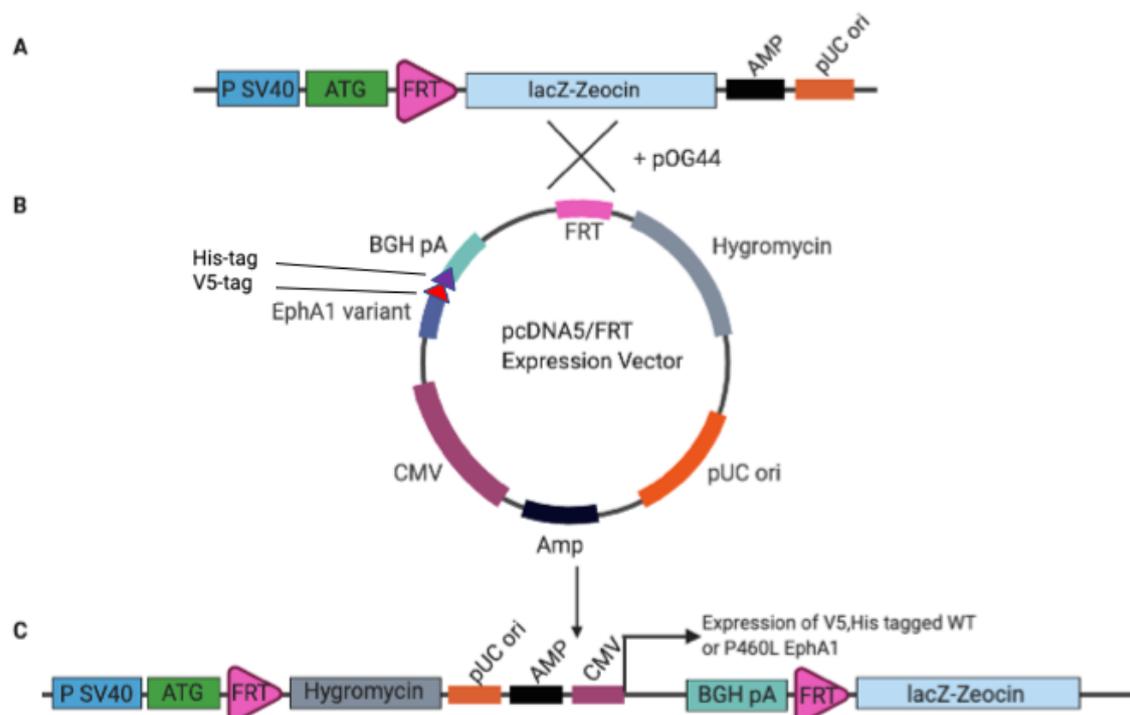
Figure 3.4: Generation of EphA1 isogenic HEK293 cell lines.
Illustration depicting vectors used in the Flp-In expression system to create isogenic HEK293 cell lines that stably express either wild type or P460L variant EphA1 receptor protein. **A.** The pFRT/lacZeo™ vector transfected into the HEK293 cell line. **B.** pcDNA5/FRT expression vector containing the EphA1 variant of interest with its V5 and His-tags and the hygromycin resistance gene. **C.** Outcome following insertion of the pcDNA/FRT expression vector into the genome of the Flp-In HEK293 cell line mediated by the co-transfection of the Flp recombinase expressing vector pOG44. Homologous recombination driven at the FRT insertion sites creates isogenic cell lines stably expressing the desired EphA1 variant.
 Schematic made using BioRender.com

3.1.4 The HEK293 cell line as a model

Since their generation, HEK293 cells have been used widely as a cellular model and expression tool for recombinant protein analysis. This is due to the presence of all the necessary machinery in terms of post-translational processing to produce mature and functional proteins (Thomas and Smart 2005; Chin et al. 2019). In conjunction, this robust cell line lacks endogenous EphA1 receptor (Fig. 3.5C) or ephrinA1 ligand expression (Fig. 3.5D), allowing the accurate analysis of the transfected EphA1 and P460L receptor under both homeostatic and controlled ligand activated conditions (Miao et al. 2001).

Chapter 3: Analysis of EphA1 receptor function and activation

3.1.5 Hypothesis

The EphA1 receptor variant P460L changes the sub-cellular location and activation of the EphA1 receptor. Such changes could result in downstream biological and functional consequences explaining this receptor variants link Alzheimer's disease pathology.

3.1.6 Overall Aim

The aim of this chapter is to compare expression levels, activity and N-terminal domain proteolytic processing of the EphA1 receptor with that of the P460L receptor variant. This will be conducted under both homeostatic and ligand-activated conditions.

3.1.7 Objectives:

1. Compare the cell surface expression and subcellular localisation of the EphA1 receptor within HEK293 cell lines transfected with either the EphA1 or P460L receptor variant.

2. Investigate the proteolytic processing of the EphA1 receptor and any alterations produced as a result of the P460L variant.

3. Compare ligand-dependent and -independent activation of the EphA1 receptor with that of the P460L receptor variant.

3.2 Materials and Methods

3.2.1 HEK293 Cell Culture and Maintenance

HEK293 cells were maintained in cell culture media as specified in table 3.1. The cells were grown in a humidified incubator at 37°C in the presence of 5% CO_2. Cells were fed once every two days with fresh culture media, pre-warmed to 37°C.

Once cell cultures had reached 70-80% confluency they were passaged. Spent media was removed and cell cultures washed with 1-5 mL PBS (Gibco) (depending on vessel size). Cells were removed from tissue culture vessel via incubation with 1-5 mL pre-warmed Trypsin/EDTA (0.05%/0.5 mM) (Gibco) (depending on vessel size), for 3 min or until cells detached upon mechanical agitation. An appropriate volume of pre-warmed media was added depending on vessel size in order to inactivate Trypsin/EDTA. Cells were centrifuged at 300 g for 5 min. The supernatant was removed and the cell pellet was resuspended in appropriate volume of pre-warmed fresh media before being dispensed into fresh culture vessels, at a dilution of 1:4-1:10.

3.2.2 Thawing and Cryopreservation of HEK293 cells

Cells were stored in liquid nitrogen for long term storage. Cells were removed from liquid nitrogen and thawed rapidly in a 37°C water bath until a small ice crystal remained. Pre-warmed fresh culture media was added (5 mL) in a dropwise fashion and the suspension was centrifuged at 300 g for 5 min. Cell pellet was re-suspended in appropriate volume of culture media for desired culture vessel size. A complete media change was performed 24 h post thaw and cell cultures maintained as per method 3.2.1.

Cells were cryopreserved from an 80% confluent T75 culture vessel following the above-mentioned passaging procedure with final re-suspension in 4 mL cryopreservation media (complete cell culture media as per table 3.1 containing 10% Dimethyl sulfoxide (DMSO),

Sigma). Cell suspension volumes of 1 mL were aliquoted into 4 cryovials. Cryovials were placed in a specialised cell-freezing container, which provides 1°C/min cooling rate required for successful cryopreservation of cells. Cells were held at -80°C for 24 h before being placed in liquid nitrogen for long term storage.

All cell culture plasticware was sourced from Greiner.

Table 3.1: Cell culture media composition.

Cell Line	Cell Culture Media
Parental HEK293	90% DMEM (Gibco), 10% foetal bovine serum* (FBS) (Sigma)
WT EphA1 HEK293	90% DMEM, 10% FBS 100µg/ml hygromycin B (ThermoFisher)
P460L HEK293	90% DMEM, 10% FBS 100 µg/mL hygromycin B

* All FBS is heat inactivated for 30 min at 56 °C.

3.2.3 HEK293 Cell Culture Treatment Procedure

Prior to treatment, cells were seeded at a density of 0.5×10^6 cells per well of a 6 well plate (Greiner). After 24 h, cells were treated with cell culture media (table 3.1) containing either 2 µg/mL ephrinA1-Fc (Merck), human IgG at 2 µg/mL (Merck) (this allowed for control of the EphA1-Fc chimera addition), 25 µM DAPT (Merck), 25 µM GM6001 (Merck) or equivalent volume of DMSO as vehicle control (Sigma). Cultures were incubated at 37°C with 5% CO_2 for a stipulated period of time. Treatment media was collected and aliquoted for EphA1 ectodomain quantification via an ELISA. Cells were then either lysed for activation analysis via western blot or stained for EphA1 protein expression and quantitated using flow cytometry.

3.2.4 Conventional Flow Cytometry

Conventional flow cytometry was performed on EphA1 and P460L HEK293 cell lines to determine cell surface EphA1 protein expression levels.

Chapter 3: Analysis of EphA1 receptor function and activation

Prior to flow cytometry analysis, cells were subject to a HEK293 extracellular staining protocol as follows: Cells were removed from their culture vessel via incubation with 1-5 mL Accutase (Sigma) for 3 min or until cells became detached on mechanical agitation. A Trypan Blue (Hyclone) cell count was performed on a Luna Dual Florescence cell counter using Luna cell counting slides. Following this, cells were centrifuged at 300 g for 5 min at 4°C. Cells were seeded at a density of $1x10^6$ cells per well in a round-bottom 96 well plate (Greiner). Each incubation was at 4°C for 30 min with centrifugation at 300 g for 5 min to allow supernatant removal, with 3 additional washes in Fluorescence-activated cell sorting (FACS) buffer (PBS with 10% FBS) between primary and secondary antibodies only.

Cells were incubated with 1 μg/mL LIVE/DEAD Fixable Aqua Dead Cell Stain (ThermoFisher Scientific) followed by incubation with unconjugated EphA1 N-terminal primary antibody (R&D Systems) before finally being incubated with Phycoerythrin (PE) labelled Goat anti-mouse secondary antibody (Biolegend) or alternatively a PE Isotype control. Antibodies used are detailed in table 3.2.

Table 3.2: Flow cytometry Antibodies

Antibody	Species/Clonality	Dilution	Source	Catalogue Number
N-terminal EphA1	Mouse Monoclonal IgG$_{2A}$	1/400	R&D Systems	MAB638
PE-anti mouse IgG	Goat Polyclonal	1/400	R&D Systems	Poly-4053
V5-tag Alexa Fluor 647	Mouse Monoclonal IgG$_{2A}$	1/1000	ThermoFisher Scientific	451098
PE Isotype control	Mouse IgG1κ	20 μL per test	BioLegend	555749

Cells were then resuspended in 200 μL of FACS buffer and loaded into FACS tubes for flow cytometry analysis on a FACSCanto II machine using DIVA software with data analysis preformed on FlowJo V10. The gating strategy employed to distinguish single, live HEK293 cells that are PE and therefore EphA1 positive for analysis is depicted in figure 3.5. UltraComp eBeads (ThermoFisher) were used for compensation during flow cytometry analysis following manufacturers protocols.
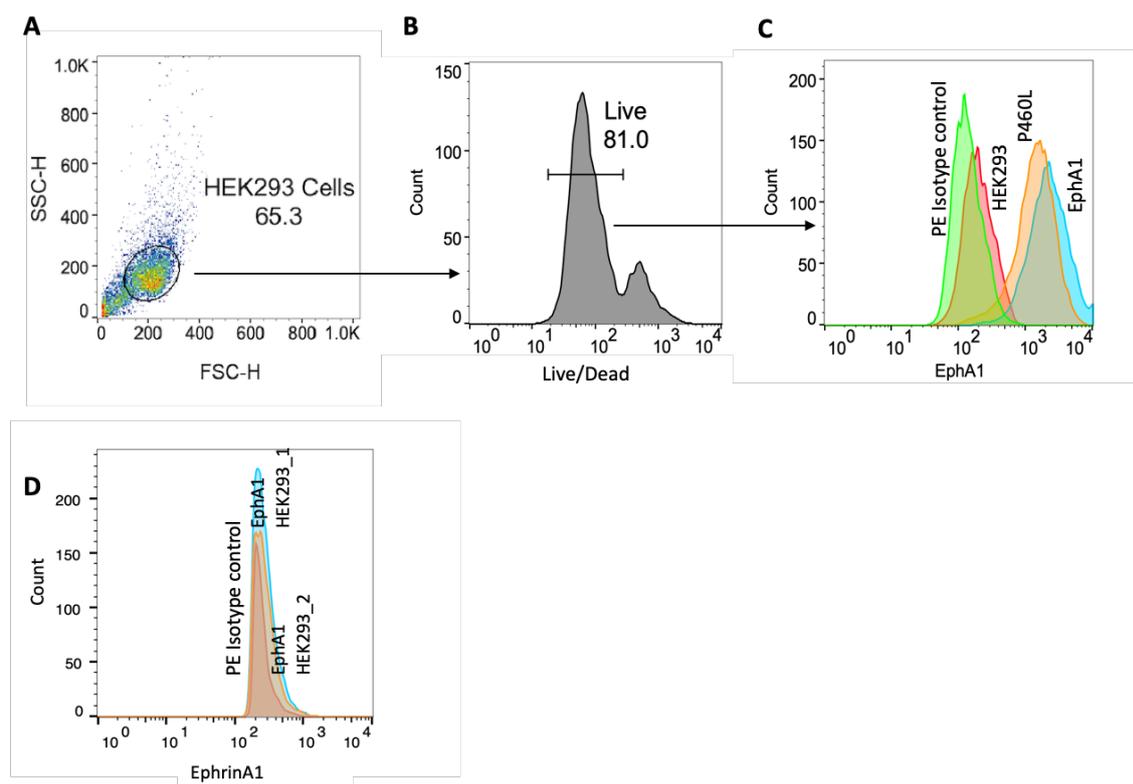
Figure 3.5: Flow cytometry gating strategy.
Representative plots depicting the gating strategy employed for the analysis of the EphA1 receptor expression on transfected HEK293 cell lines via flow cytometry. **A.** Scatter plot of forward and side scatter height to allow single HEK293 cell gating. **B.** Histogram of LIVE/DEAD fixable aqua dead cell stain, gating 81% live cells. **C.** Histogram depicting EphA1 expression levels. An isotype control (green) was used to identify EphA1 expression. EphA1 negative parental control HEK293 (red), EphA1 HEK293 (blue) and P460L HEK293 (orange).
D. Histogram depiction of ephrinA1 expression levels within two biological replicates of EphA1 HEK293 cells (blue and orange). Again, a PE isotype control was used to identify ephrinA1 expression (red).

3.2.5 Imaging Flow Cytometry

Cells were removed from culture vessels and seeded into round-bottom plates as per method section 3.2.4. Cells were incubated with 1 µg/mL LIVE/DEAD Fixable Near Infrared Dead Cell Stain (ThermoFisher Scientific) for 30 min at 4°C. Cells were fixed and permeabilized using the True-Nuclear Transcription Factor Buffer Set diluted as per manufacturer's instructions (BioLegend). Cells were fixed via incubation with 1X Fixation Buffer for 60 min at 4°C and permeabilized via resuspending cells in 1X Permeabilisation buffer and centrifuging cells at 300 g for 5 min a total of three times. Cells were stained with the Alexa Fluor (AF) 647 anti-V5-tag or EphA1 N-terminal antibody via incubation at 4°C for

30 min. Antibodies used detailed in table 3.2. Cells were then stained with NucBlue™ Live ReadyProbes™ Reagent (ThermoFisher) for 30 min at 4°C. Following staining, cells were resuspended in 30 µL of FACS buffer and placed into separate 1.5 mL Eppendorf tubes. Analysis was performed on an Amnis ImageStream system using IDEAS software.

Controls for ImageStream analysis were created through Florescence Minus One (FMO), single stains of each antibody following the above-mentioned method, with the exception of the LIVE/DEAD stain where cells were lysed prior to staining via boiling at 95°C for 5 min.

3.2.6 Creation of membrane and cytosol sub-cellular compartment masks for analysis of ImageStream data

Membrane and cytosolic sub-cellular compartment 'masks' were created in the imaging flow cytometry software programme IDEAS. This allowed for the accurate quantitation of EphA1 protein subcellular location from imaging flow cytometry data. A cartoon illustration of how these masks were determined is depicted in figure 3.6.
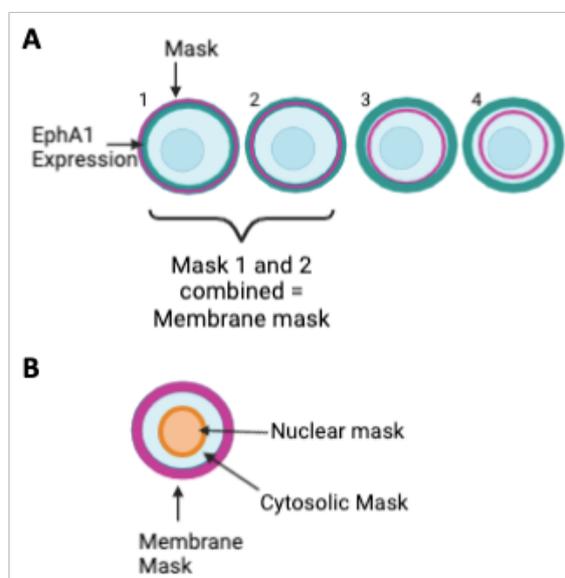
Figure 3.6: Cartoon illustration of membrane and cytosolic mask creation.
A. Concentric ring masks (pink) were created throughout the cell image. Masks which were noted to span the fluorescence intensity peak for EphA1 expression (teal) where combined to create the membrane mask (Masks 1 and 2 in this cartoon). **B.** EphA1 fluorescence intensity within the newly created membrane mask (pink) and a nuclear mask set around the NucBlue nuclear stain (orange) where subtracted from that of the whole cell to create a cytosolic mask. Schematic created using BioRender.com

Chapter 3: Analysis of EphA1 receptor function and activation

EphA1 HEK293 cells were stained with both the N-terminal EphA1 and V5-cytotail tag AF647 antibody (table 3.2) following the staining method detailed in section 3.2.5. These two antibodies were initially used in tandem to create the subcellular compartments and to determine if the V5-tag antibody alone can be used accurately to track both membrane and cytosolic EphA1 receptor levels.

Imaging flow cytometry follows a similar gating strategy as that applied in conventional flow cytometry detailed in section 3.2.4. As shown in figure 3.7A an unstained control and FMO stains where used to allow the gating of EphA1 expressing, single, live HEK293 cells that were within camera focus.

The adaptive erode feature within the IDEAS analysis software allowed the creation of masks covering a user-defined percentage of the cell image, as briefly outlined in figure 3.6. This feature was used to create concentric ring masks, each 1 pixel or 0.3 μm in size, throughout the cell image (Fig. 3.7B). The median fluorescence intensity values of both the N-terminal EphA1 and the V5-tag antibody were extracted and normalised to the sum intensity value for their respective fluorophore. This was plotted against mask number as seen in figure 3.7C. These data were used to plot the sum of two Lorentzian curves (Fig. 3.7D). The width of the Lorentzian curve, illustrated by the grey region on figure 3.7D corresponds to the membrane-associated fluorescence for each antibody. The masks within this grey region were merged to create a quantitatively defined membrane mask. This was then subtracted from an adaptive erode mask of the whole cell image to give a value of EphA1 receptor expression solely at the membrane.

Back applying this mask range to cell images (Fig. 3.7E) confirmed by co-localisation of the fluorescent signal that the mask ranges identified from the Lorentzian curve analysis correspond to membrane staining of both antibodies. Therefore, the V5-tag antibody was used in subsequent analysis to investigate EphA1 receptor subcellular localisation.

The cytosolic subcellular compartment was defined using an adaptive erode mask of the entire cell image minus a nuclear stain mask around the NucBlue nuclear staining and the

newly created membrane mask. This method of creating cellular compartment masks was created and optimised within the thesis of Dr Owen Moon (Moon 2020).

Figure 3.7: Creation of membrane and cytosolic sub-cellular masks for ImageStream data analysis.

A. A representative example of the imaging flow cytometry gating strategy employed for EphA1 expression analysis. An unstained control was used to allow accurate gating of HEK293 cells within camera focus. FMO stains allowed accurate gating of live cells and EphA1 expression. **B.** Representative cell images with V5-tag antibody (red) and N-terminal EphA1 antibody staining (yellow) overlaid with the concentric adaptive erode ring masks (blue) created throughout the cell image. **C.** Median fluorescence intensity for each mask converted to a percentage of total fluorescence intensity of the N-terminal EphA1 (yellow) and the V5-tag antibody (red). **D.** Two fit component Lorentzian curve analysis plotted with upper and lower centre values (grey region) representing membrane-associated fluorescence for each antibody. **E.** Back applied mask ranges from Lorentzian curve analysis to brightfield (ch01), V5-tag (ch11) and N-terminal EphA1 (ch03) show mask range identified corresponded to membrane staining of both antibodies. **F.** Co-localisation of N-terminal EphA1 (yellow) and V5-tag (red) antibody staining with purple nuclear stain.
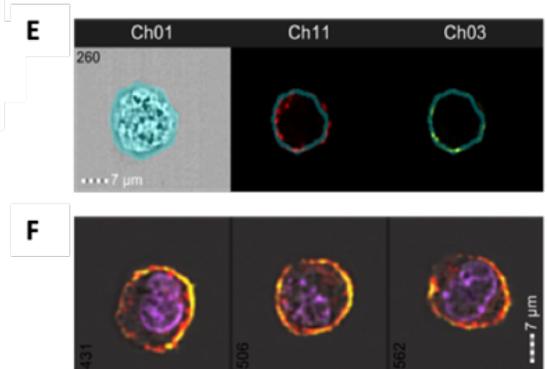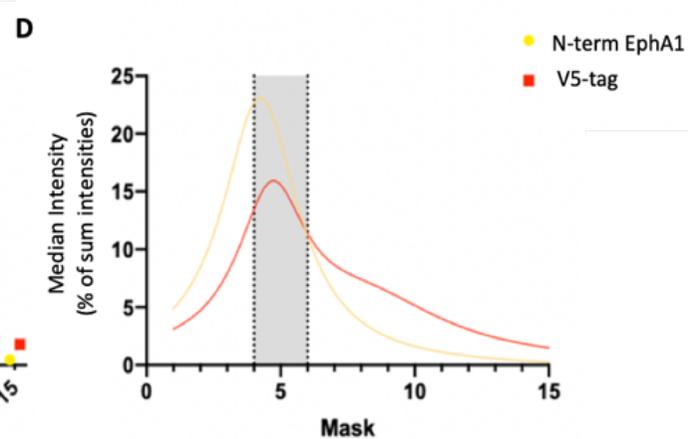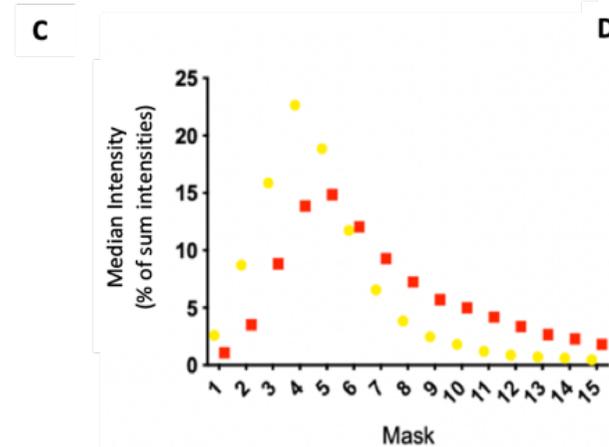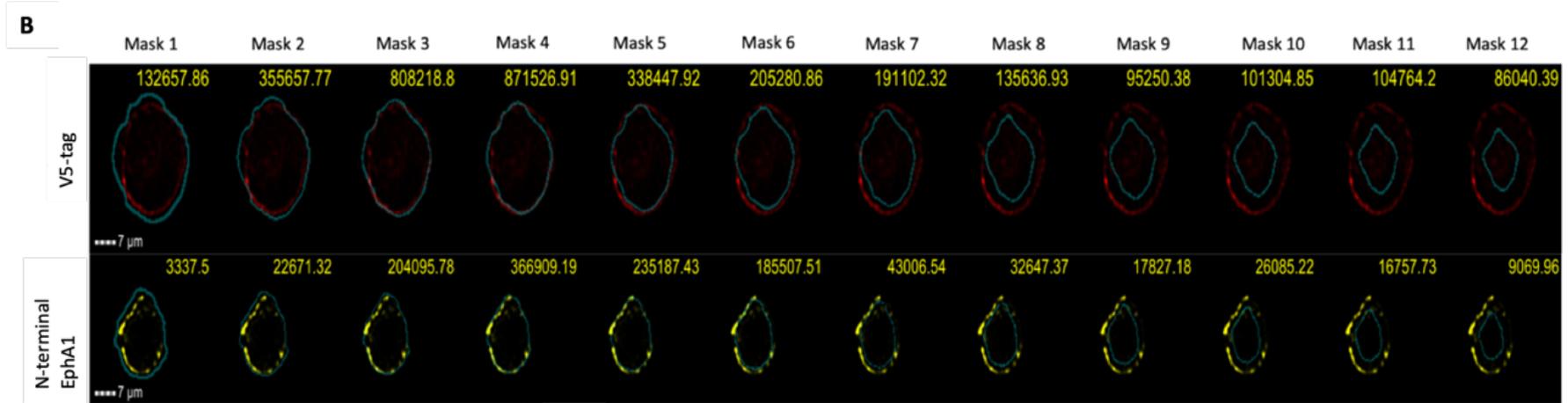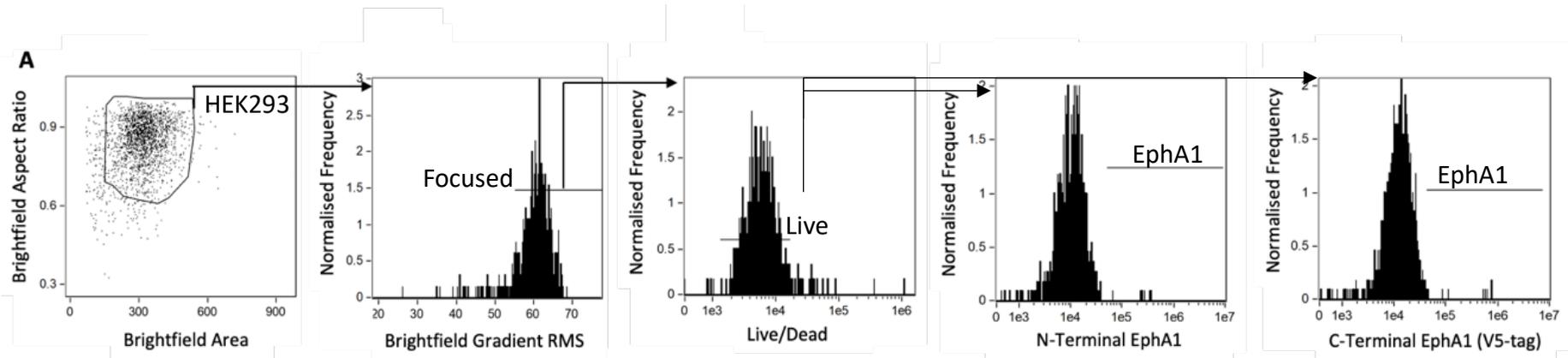
3.2.7 Enzyme-linked immunosorbent assay (ELISA)

The Human EphA1 ELISA kit (Millipore) was used to quantify the concentration of released soluble EphA1 receptor as per manufacturer's instructions. The minimal detection concentration for EphA1 was 0.085 ng/mL. A standard curve was produced for EphA1 using the human protein standard solution at a range of seven 2.5-fold serial dilutions (Fig. 3.8, circle). Standard or sample (100 μL) was pipetted into a Human EphA1 antibody-coated ELISA plate and incubated with gentle shaking at RT for 2.5 h. Solution was discarded and wells washed 4 times with 1X wash solution (provided in kit) followed by the addition of 100 μL of the supplied Biotinylated human detection antibody. The plate was covered and left to incubate at RT with gentle shaking for 1 h. After wash steps had been repeated 100 μL horseradish peroxidase-Streptavidin solution was added for 1 hr at RT with gentle shaking. Solution was discarded and wash steps repeated. ELISA colorimetric 3,3',5,5'-tetramethylbenzidine (TMB) one-step substrate reagent (100 μL) was added and the plate was incubated in the dark at RT for 30 min with gentle shaking. Finally, 50 μL stop solution was added and absorbance read at 450 nm immediately using a Clariostar microplate reader. Sample protein concentrations were derived from interpolating the x-values using the human EphA1 standard curve.

Normalisation of the level of soluble EphA1 receptor released into the cell media into the amount of receptor present at the cell membrane was quantitated using the total V5 mean fluorescent intensity and the membrane V5 mean fluorescent intensity to calculate a percentage of total receptor at the membrane for both the EphA1 and P460L HEK293 cell lines. The level of soluble receptor produced by the P460L HEK293 cells was then increased by this percentage difference to account for differences in receptor levels at the membrane between the two cell lines.

Prior to running experimental samples, the standard curve was analysed with the addition of 2 μg/mL ephrinA1-Fc ligand to confirm ligand-receptor interaction did not prevent binding of the ELISA kits anti-EphA1 detection antibody (Fig. 3.8, square).
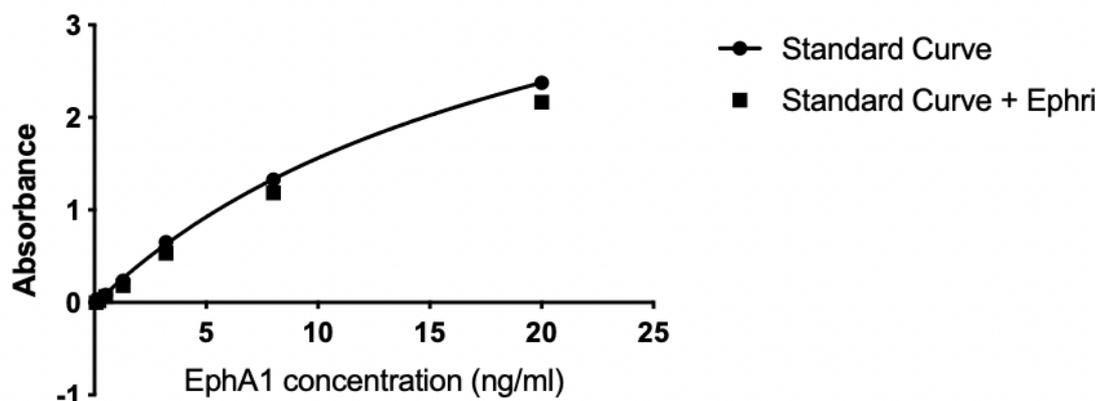
Figure 3.8: Human EphA1 ELISA Standard Curve
Absorbance read out of EphA1 standard curve produced using the human protein standard solution supplied in the human EphA1 ELISA kit at seven 2.5-fold dilutions (circle). The same standard curve plotted with the addition of 2 μg/mL ephrinA1-Fc ligand into each standard (square).

3.3 Results

3.3.1 Cell surface expression of the EphA1 receptor and its P460L variant under homeostatic conditions

The consequences of the P460L variant on the cell surface expression of the EphA1 receptor was assessed via flow cytometry (Fig. 3.9).

The HEK293 cell lines created expressing either the EphA1 receptor or the P460L receptor variant were stained using an anti-human EphA1 antibody following the protocol detailed in section 3.2.4 and gated for flow cytometry analysis as per figure 3.5 (Fig. 3.9Ai). Using the Parental HEK293 cell line (red) as a negative control, cell surface expression of EphA1 was seen in the EphA1 (blue) and P460L HEK293 cell lines (orange) (Fig. 3.9Aii). Upon quantitation, cell surface EphA1 expression in both cell lines was significantly greater than

the negative control (p<0.0001) with P460L HEK293 seen to express significantly less receptor compared to the EphA1 HEK293 cell line (p=0.0013) (Fig. 3.9B).
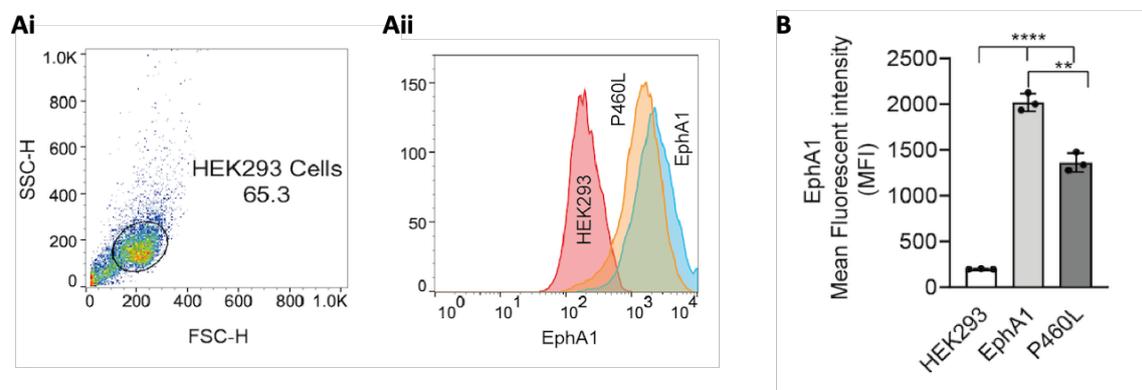
Figure 3.9: Cell surface expression of EphA1 and P460L receptor.
EphA1 and P460L HEK293 cells were stained for cell membrane EphA1 receptor expression and quantitated via flow cytometry. **Ai.** HEK293 cells were gated following protocol detailed in figure 3.5. **Aii.** Cell surface EphA1 expression within the negative control Parental HEK293 (red), EphA1 HEK293 (blue) and P460L HEK293 (orange) cell lines. **B.** Quantitation of cell surface EphA1 expression between negative control parental, EphA1 (p=0.0013) and P460L (p<0.0001) HEK293 cell lines. Error bars indicate mean ± SD of three independent data sets. Statistical analysis conducted, t-test. **, P<0.01 and ****, P<0.0001

3.3.2 Subcellular localisation of the EphA1 receptor

As a result of observing reduced cell surface expression of the P460L receptor on live cells, imaging flow cytometry was conducted to further analyse the subcellular distribution of the EphA1 receptor. Fixed and permeabilised cells were stained with an anti-V5-tag antibody to allow the detection and quantitation the subcellular protein expression of the EphA1 receptor as per method outlined in section 3.2.5. Cells were gated (Fig. 3.10A) and expression patterns designated as either within the membrane or cytosol based on analysis and software settings previously created (detailed in section 3.2.6).

Levels of total EphA1 (Fig. 3.10B), membrane EphA1 (Fig. 3.10C) and cytosolic EphA1 (Fig. 3.10D) were measured in both the EphA1 HEK293 and P460L HEK293 cell lines, using the Parental HEK293 cell line as a negative control. Representative images produced from imaging flow cytometry are shown in figure 3.10E. The negative control Parental HEK293 containing no EphA1 expression are shown on the left. The vast majority of EphA1 receptor

expression within the EphA1 HEK293 cells was seen to be located at the cell membrane (Fig. 3.10E, middle), this is in contrast to expression patterns seen within the P460L HEK293 cells, where the majority of expression was clustered within the cytosol (Fig. 3.10E, right). Despite no difference in total level of C-terminal EphA1 receptor expression being observed between the EphA1 and P460L EphA1 HEK293 cell lines (Fig. 3.10F), the receptor variant showed significantly altered subcellular localisation. A membrane to cytosol ratio of receptor expression was calculated to quantitate the distribution pattern (Fig. 3.10G). From this, it could be seen that the P460L variant HEK293 cells had significantly less membrane receptor expression when compared to the EphA1 HEK293 cell lines (p=0.0004).
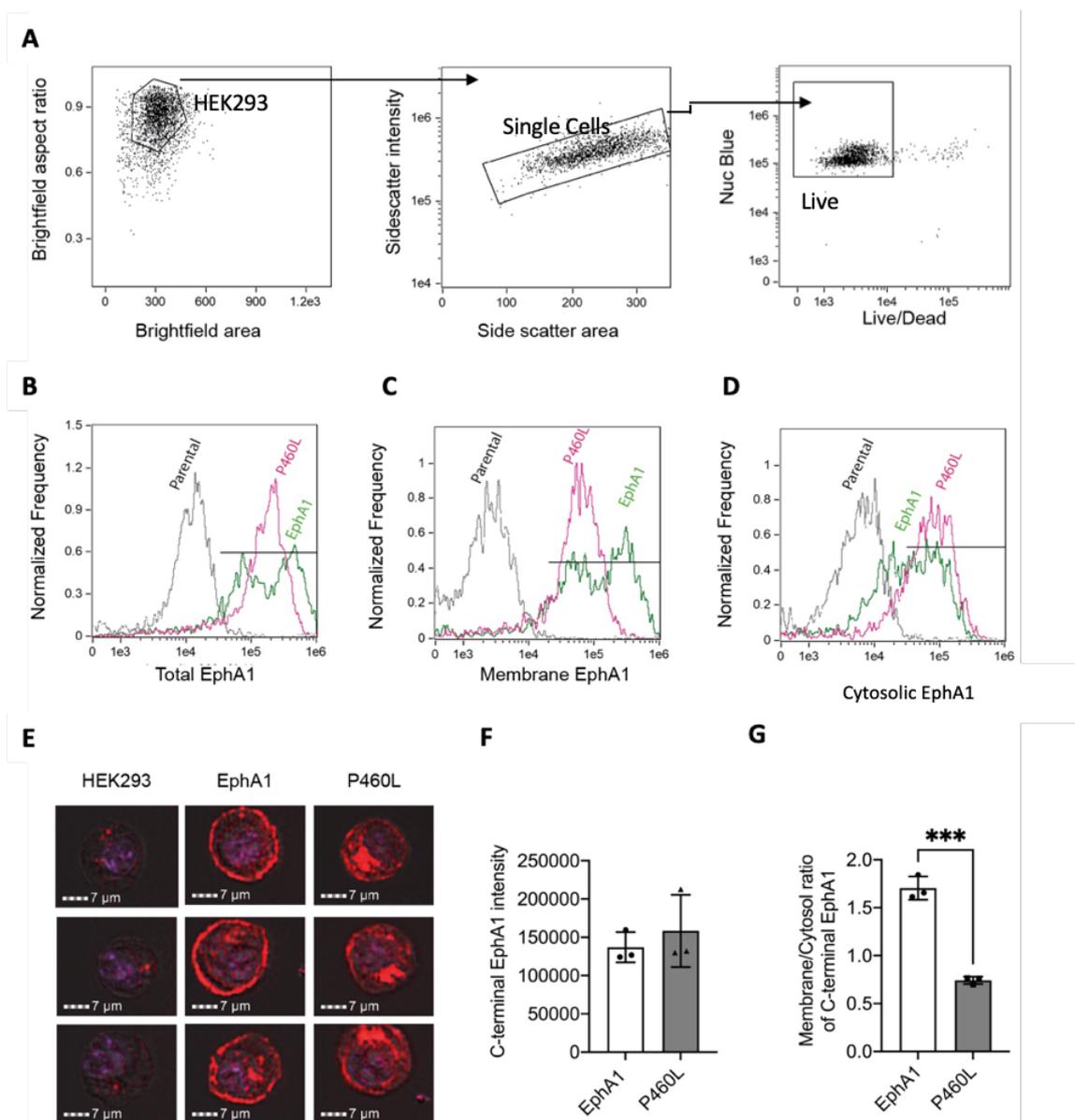
Figure 3.10: Subcellular localisation of the EphA1 receptor within EphA1 and P460L EphA1 HEK293 cells.

EphA1, P460L and the parental HEK293 cell lines were stained for EphA1 receptor expression using an AF647 V5-tag antibody and analysed via imaging flow cytometry. **A.** Gating strategy employed, as per figure 3.7, selects for live, single HEK293 cells containing a nuclear stain. **B.** Total level of C-terminal receptor expression within the Parental HEK293 cells (black), EphA1 HEK293 cells (green) and P460L HEK293 cells (pink). **C.** EphA1 receptor expression present within the membrane mask within the Parental HEK293 cells (black), EphA1 HEK293 cells (green) and P460L HEK293 cells (pink). **D.** EphA1 receptor expression present within the cytosolic mask within the Parental HEK293 cells (black), EphA1 HEK293 cells (green) and P460L HEK293 cells (pink). **E.** Representative images produced by the ImageStream showing the distribution of EphA1 receptor expression within the negative control Parental HEK293 cells (left), EphA1 (middle) and P460L HEK293 cells (right). **F.** Quantitation of total C-terminal levels of EphA1 receptor within EphA1 and P460L HEK293 cells. **G.** Quantitation of receptor expression distribution between the membrane and cytosolic subcellular compartment masks (p=0.0004). Error bars indicate mean ±SD of three independent data sets. Statistical analysis performed, t-test. ***, p<0.001.

3.3.3 Quantitation of EphA1 receptor N-terminal domain release and cleavage mechanism investigation

As discussed in section 3.1.2, Eph receptors can undergo proteolytic cleavage via a range of proteases such as MMPs or ADAMs and γ-secretase, shedding their N-terminus or ectodomain. To determine if the alterations in subcellular distribution of the P460L receptor were due to increased proteolytic cleavage and subsequent release of a soluble receptor in addition to receptor internalisation, the level of soluble EphA1 receptor released by both EphA1 and P460L HEK293 cell lines was investigated using a human EphA1 ELISA Kit.

As seen in figure 3.11A, under basal conditions the EphA1 receptor releases a soluble form into the cell culture medium. The P460L HEK293 cells produced significantly reduced levels of this soluble receptor product compared to levels seen from the EphA1 HEK293 cells (p=0.0006). On standardisation of levels of soluble EphA1 receptor released to the amount of receptor observed at the cell membrane, this reduction is still evident (p=0.0021) (Fig. 3.11B). This analysis was conducted via increasing the released receptor levels from the P460L HEK293 cell lines by the difference within EphA1 and P460L HEK293 membrane receptor expression observed via imaging flow cytometry. The level of soluble receptor present within the medium of both cell lines (Fig. 3.11C) along with total receptor levels (Fig. 3.11D) was not altered on receptor activation by the addition of ephrinA1-Fc ligand (2 μg/mL) for 2 hr.

To determine if the generation of soluble EphA1 is due to proteolysis by either an MMP/ADAM or as part of a multi-step cleavage mechanism involving γ-secretase (as detailed in Fig. 3.2), cells were incubated with either a broad-spectrum metalloprotease (MMP) inhibitor GM6001 (25 μM), a γ-secretase inhibitor DAPT (25 μM) or DMSO as a vehicle control for 2 hr. Again, levels of soluble EphA1 receptor within the cell culture medium, as well as receptor subcellular localisation, were analysed using a human EphA1 ELISA kit and imaging flow cytometry respectively. As seen in figure 3.11E, addition of either inhibitor showed no alteration in receptor subcellular localisation nor reduction in the level of soluble EphA1 released by EphA1 or P460L HEK293 cells (Fig. 3.11F). The EphA1 HEK293

cells, however, demonstrated receptor internalisation in response to ligand activation as illustrated in figure 3.11E (p=<0.0001).
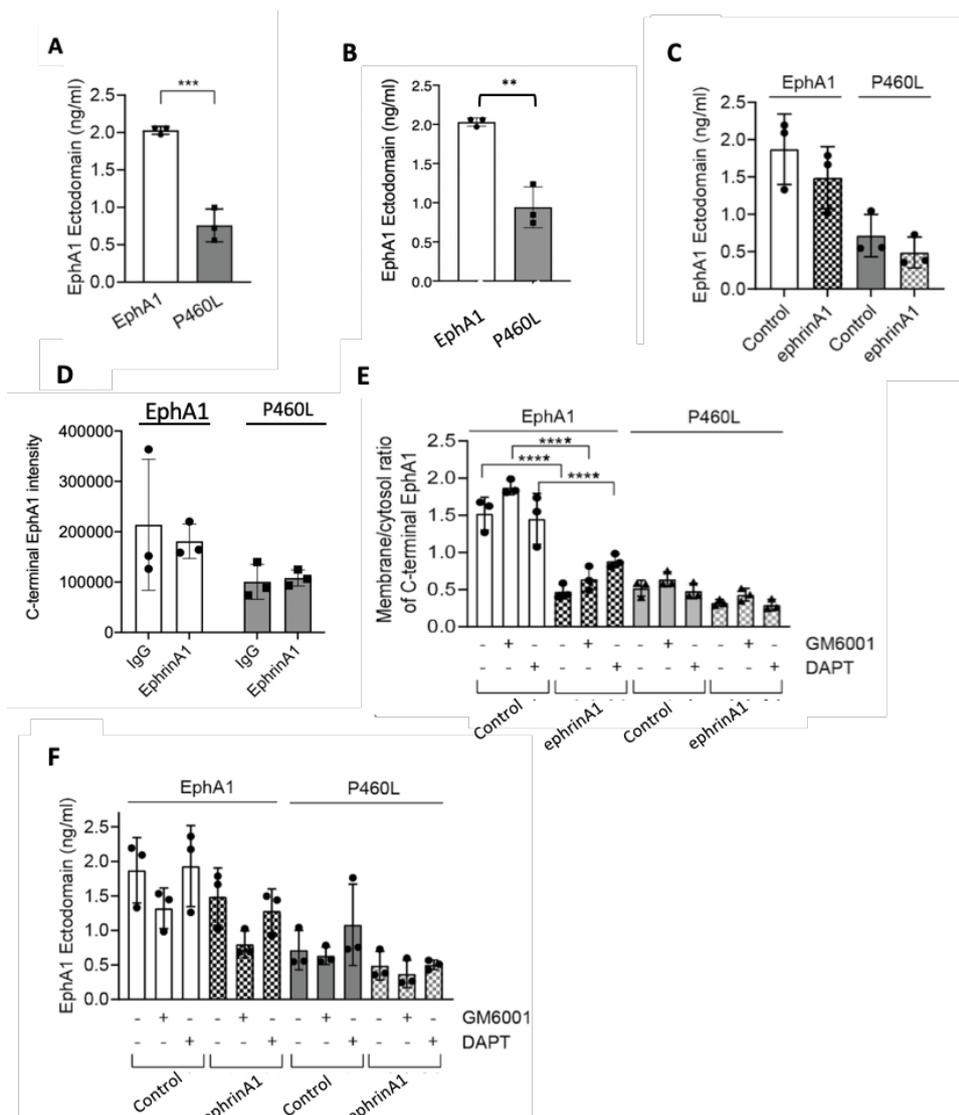
Figure 3.11: Quantitation of released soluble EphA1 via a human EphA1 ELISA and investigation into the proteolytic cleavage mechanism via imaging flow cytometry.

A. Quantitation of soluble EphA1 receptor released by EphA1 and P460L HEK293 cells under homeostatic conditions via a human EphA1 ELISA (p=0.0006). **B.** Amount of soluble EphA1 receptor released standardised to total receptor present at the cell membrane (p=0.0021). **C.** Quantitation of soluble EphA1 receptor released by EphA1 and P460L HEK293 cells after incubation with 2 μg/mL ephrinA1-Fc ligand or IgG control for 2 hr via a human EphA1 ELISA. **D.** Quantitation of total levels of EphA1 receptor within EphA1 and P460L HEK293 cells upon incubation with 2 μg/mL ephrinA1-Fc ligand of IgG control for 2 hr. **E.** Alterations in the subcellular localisation of EphA1 and P460L HEK293 cells quantitated via imaging flow cytometry following ligand activation in combination with inhibitor addition (GM6001 or DAPT (25 μM for 2 hr). Again, the addition of IgG was used as a control for ligand addition. **F.** Quantitation of soluble EphA1 receptor released by EphA1 and P460L HEK293 cells after incubation with 25 μM of either GM6001, DAPT or DMSO as a vehicle control for 2 hr, under both homeostatic and ligand-activated conditions (2 μg/mL ephrinA1-Fc or IgG control addition for 2 hr). Data from 3 independent data sets. Statistical analysis performed, 2-way ANOVA. Error bars indicate mean ±SD. **, $p<0.01$, ***, $p<0.001$. ****, $P<0.0001$.

3.3.4 Quantitation of ligand dependent EphA1 receptor activation via tyrosine phosphorylation analysis

It was hypothesised that the P460L variant of the EphA1 receptor may result in altered subcellular localisation and reduced cell membrane expression due to constitutive activity. Therefore, the activation levels of both the EphA1 and P460L receptor were analysed via quantitation of phosphotyrosine-EphA1 levels within both EphA1 and P460L HEK293 cell lysates.

EphA1 HEK293 and P460L HEK293 were stimulated with ephrinA1-Fc ligand (2 μg/mL) for either 1 or 5 min in order to activate the EphA1 receptor. Cell lysates were analysed via SDS-PAGE and western blotting, probing for total EphA1 receptor expression via V5-tag staining and tyrosine phosphorylation levels (Fig. 3.12). Levels of V5-tag expression as opposed to that of the total protein loading control, GAPDH was used as a control for EphA1 receptor protein levels. As despite levels of GAPDH being fairly consistent between all conditions due to the same amount of total protein being loaded, V5-tag expression within the P460L HEK293 cells was observed to be lower than that of the EphA1 HEK293 cells. This is potentially due to the P460L variant receptor being sequestered in a region of the cell/cytosol that is difficult to access via standard cell lysis methods, resulting in the observed decrease in P460L receptor expression. Therefore, V5-tag expression allowed for more accurate data normalisation to total levels of the EphA1 receptor present within the lysates. All antibodies used are detailed in Chapter 2, table 2.1. Human IgG treated cell lysates from each time point were used as experimental and normalisation controls, allowing quantitation of phosphotyrosine-EphA1 fold changes upon ephrinA1 stimulation as described in Chapter 2, section 2.4.

EphrinA1 ligand stimulation for 1 min produced a slight decrease in phosphotyrosine-EphA1 levels when compared to the 1 min IgG control lysates (p=0.0162). EphrinA1 ligand stimulation for 5min however, significantly increased phosphotyrosine levels within the EphA1 receptor when compared to the 5 min IgG control treatment control (p=0.0012) (Fig. 3.12A). This increase was not seen within the P460L receptor, where no significant

alterations in phosphotyrosine-EphA1 levels were noted (Fig. 3.12B). Representative western blot images showing total EphA1 protein and phosphotyrosine-EphA1 expression levels within both EphA1 and P460L HEK293 cell lysates are shown in figure 3.12C, black arrow head and figure 3.12D, blue arrow head respectively.
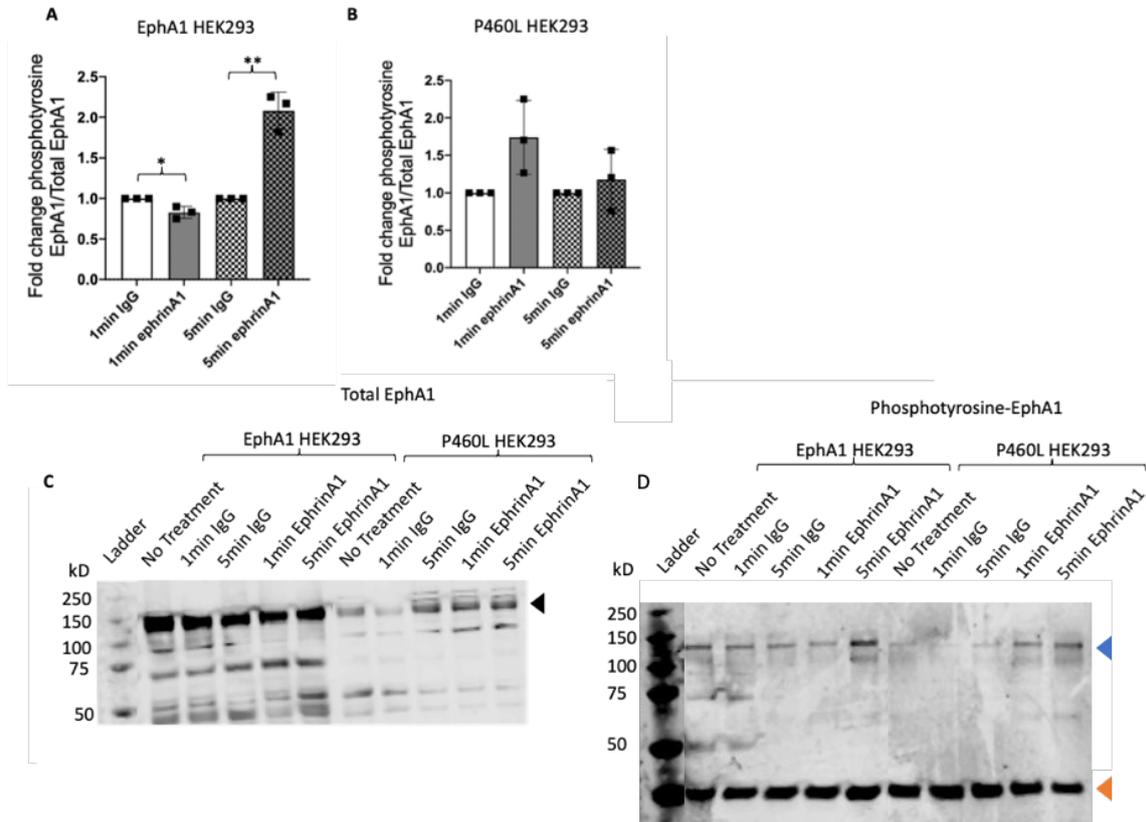
Figure 3.12: Analysis of ligand dependent EphA1 receptor activation via quantitation of phosphotyrosine-EphA1 levels.

EphA1 HEK293 and P460L HEK293 cells were treated with ephrinA1-Fc ligand (2 µg/mL) or human IgG control for 1 or 5 min. Cell lysates were then analysed by western blot, staining for either V5-tag or phosphotyrosine levels. **A.** Fold change in phosphotyrosine-EphA1 levels within EphA1 HEK293 cell lysates on 1 (p=0.0163) and 5 min ligand stimulation (p=0.0012). **B.** Fold change in phosphotyrosine-EphA1 levels within P460L HEK293 cell lysates on 1 and 5 min ligand stimulation. **C.** A representative western blot showing total EphA1 protein levels within both EphA1 and P460L HEK293 cell lysates using anti-V5-tag staining (black arrow head). **D.** A representative western blot showing phosphotyrosine-EphA1 within both EphA1 and P460L HEK293 cell lysates using anti-phosphotyrosine staining (blue arrow head) in addition to GAPDH loading control (orange arrow head).

Data from 3 independent data sets. Statistical analysis performed, t-test. Error bars indicate mean ±SD. *, p<0.05, **, P<0.01.

3.3.5 Quantitation of ligand independent EphA1 receptor activation via serine phosphorylation analysis

As discussed in 3.1.1, EphA1s closest homolog EphA2 is capable of ligand-independent activation through phosphorylation of a serine at residue 897 within the receptor's intracellular domain (Miao et al. 2009). Therefore, levels of phosphoserine-EphA1 during ligand-dependent and -independent receptor activated conditions were quantitated within both the EphA1 HEK293 and P460L HEK293 cell lysates.

The western blot membrane from the analysis of phosphotyrosine-EphA1 levels in section 3.3.4 was stripped and re-probed using a pan anti-phosphoserine antibody using methods described in methods section 2.5. Staining with a V5-tag antibody was again used to allow quantitation of total EphA1 receptor protein levels. All antibodies used are detailed in Chapter 2, table 2.2. Expression levels of phosphoserine-EphA1 from both cell line lysates were quantitated using the respective time points IgG cell lysates as normalisation controls, as detailed in Chapter 2, section 2.4 (Fig. 3.13).

EphA1 receptor showed decreased phosphoserine-EphA1 levels on addition of ephrinA1 ligand for 1 min in comparison with human IgG treated control cell lysates (p=0.0034) (Fig. 3.13A). No significant change was observed on 5 min ephrinA1 ligand stimulation. This decrease in phosphoserine-EphA1 after 1 min ephrinA1 ligand stimulation was also observed within the P460L receptor when compared to the IgG control cell lysates (p=0.0082) lysates (Fig. 3.13B). No significant change was also observed on 5 min ephrinA1 ligand stimulation.
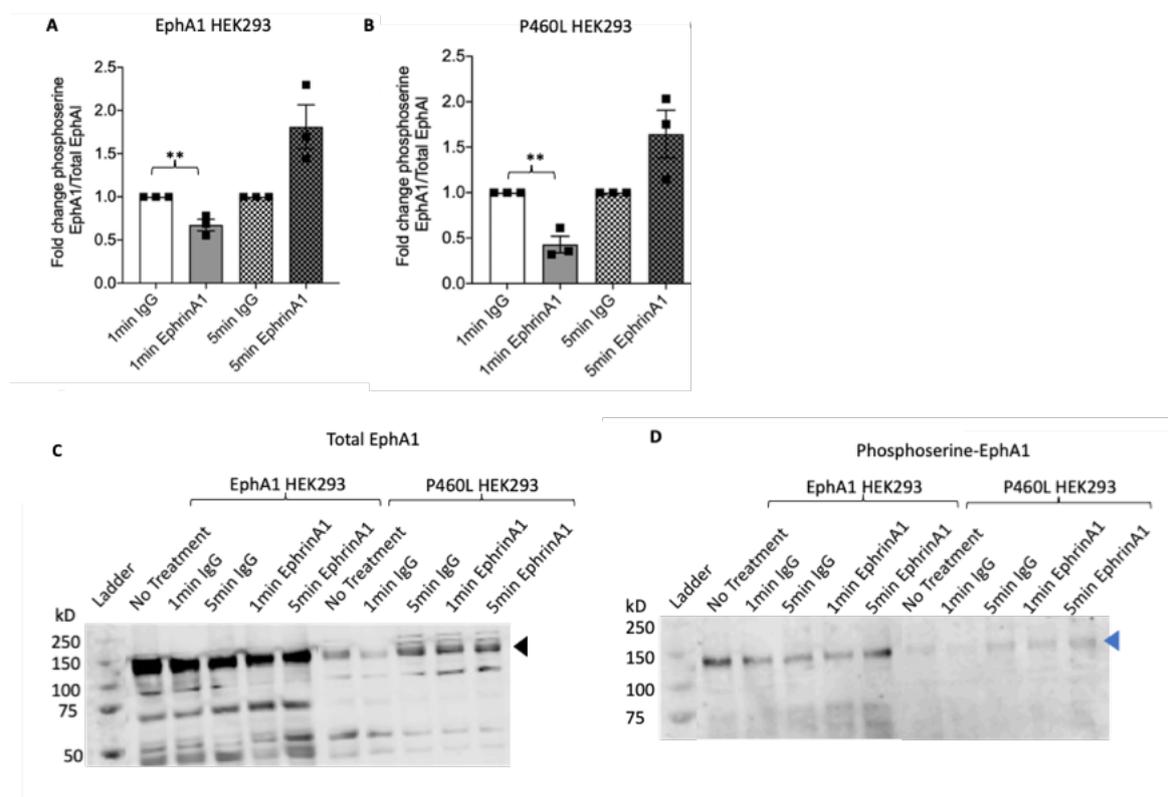
Figure 3.13: Analysis of ligand independent EphA1 receptor activation via quantitation of phosphoserine-EphA1 levels.

EphA1 HEK293 and P460L HEK293 cells were treated with ephrinA1-Fc ligand (2 µg/mL) or human IgG control for either 1 or 5 min. Cell lysates were analysed via western blot, staining for either V5-tag or phosphoserine levels. **A.** Fold change in phosphoserine-EphA1 levels within EphA1 HEK293 cell lysates on 1 (p=0.0034) and 5 min ligand stimulation. **B.** Fold change in phosphoserine-EphA1 levels within P460L HEK293 cell lysates on 1 (p=0.0082) and 5 min ligand stimulation. **C.** A representative western blot showing total EphA1 protein levels within both EphA1 and P460L HEK293 cell lysates using anti-V5-tag staining (black arrow head). **D.** A representative western blot showing phosphoserine-EphA1 within both EphA1 and P460L HEK293 cell lysates using anti-phosphoserine staining (blue arrow head).

Data from 3 independent data sets. Statistical analysis preformed, t-test. Error bars indicate mean ±SD. *, p<0.005, **, P<0.01.

3.4 Discussion

3.4.1 Cell surface expression and subcellular localisation of the EphA1 receptor

This chapter demonstrates that the Parental HEK293 cell line created provided an accurate negative control cell line for use alongside those containing the EphA1 receptor variants due to its lack of endogenous EphA1 expression (Fig. 3.5). Resulting in any differences produced

between the two EphA1 and P460L HEK2932 isogenic cell lines being a consequence of the specific EphA1 variant.

The HEK293 cell line containing the P460L receptor variant showed significantly reduced receptor expression at the cell surface following staining of live cells with the N-terminal EphA1 antibody (Fig. 3.9B). Further investigation into this reduced expression level was conducted via imaging flow cytometry, to analyse the subcellular distribution of the receptor (Fig. 3.10). Quantitation of C-terminal EphA1 receptor levels using an AF647 V5-tag antibody showed total levels of EphA1 and P460L receptor to be comparable (Fig. 3.10F). However, differences in the subcellular distribution of the receptor variants could be seen, with P460L clustering within the cytosol (Fig. 3.10E, right), showing significantly reduced expression at the cell membrane (Fig. 3.10G). In comparison, the majority of the EphA1 receptor expression was observed at the membrane (Fig. 3.10E, middle).

Imaging flow cytometry allowed the subcellular localisation of the EphA1 receptor to be analysed in a high through-put and quantitative manner. Other techniques however, such as conventual immunofluorescence would allow a greater degree of resolution. Despite not being as quantitative, immunofluorescence would enable co-localisation with cellular markers to better understand the exact receptor subcellular localisation.

3.4.2 Analysis of EphA1 receptor processing by proteases

Such alterations in the subcellular distribution is potentially a consequence of altered receptor processing or activation as a result of the P460L variant within the EphA1 receptor. For example, the P460L variant may increase constitutive or ligand-independent receptor activation leading to increased receptor internalisation. Alternatively, the P460L variant may be subject to increased protease processing as eluded to in section 3.1.2, also leading to increased receptor internalisation. The hypothesis of increased proteolytic cleavage was investigated by looking at levels of released soluble EphA1 receptor in the cell culture medium of both EphA1 HEK293 and P460L HEK293 cells (Fig. 3.11).

Chapter 3: Analysis of EphA1 receptor function and activation

Analysis of the human EphA1 ELISA data indicated both EphA1 and P460L HEK293 cells release soluble EphA1 receptor into the cell culture media. However, a significantly decreased concentration of soluble receptor was observed to be released by the P460L HEK293 cells (Fig. 3.11A). On standardisation to the levels of receptor at the membrane this reduced release was still observed (Fig. 3.11B). This confirms that the reduced soluble receptor release from the P460L HEK293 cells is not due to the reduced levels of receptor present at the membrane and is potentially a consequence of the P460L receptor variant.

Ligand activation by ephrinA1-Fc resulted in decreased membrane EphA1 receptor expression as seen in figure 3.11E. However, there was no alteration in either level of soluble receptor released (Fig. 3.11C) or total C-terminal receptor levels (Fig. 3.11D), indicating that on ligand activation by ephrinA1, EphA1 receptor undergoes internalisation that is not accompanied by degradation of the receptor C-terminal (Fig. 3.11B). The subcellular distribution of P460L receptor was not affected in this manner by ligand activation (Fig. 3.11E), indicating this receptor variant is not internalised upon ligand activation to similar levels seen in the EphA1 receptor.

The addition of either a broad spectrum MMP inhibitor (GM6001) or a γ-secretase inhibitor (DAPT) had no effect on levels of soluble EphA1 receptor released by either EphA1 or P460L HEK293 cells (Fig. 3.11F) or receptor variant subcellular distribution (Fig. 3.11E). Despite a range of MMPs such as MMP1-3, 8-10 and MMP13 as well as γ-secretase expression being found in HEK293 cells (Liu and Wu 2006; Placanica et al. 2010). This confirms that under both homeostatic and ligand activated conditions EphA1 cell surface levels are not regulated via protease processing by the actions of either MMPs or as part of a multi-step process involving γ-secretase. The lack of cleavage by γ-secretase is not surprising given the evidence from Merilahti *et al.* in section 3.1.2, showing that EphA1 is not a substrate for this particular protease (Merilahti et al. 2017). There is however, a slight trend of decreased soluble receptor release on the addition of GM6001 both under homeostatic and ligand activated conditions, which is accompanied by a general increase in membrane/cytosol receptor ratio. Therefore, potentially a repeated experiment altering inhibitor concentration or incubation time may allow small effects of the GM6001 inhibitor to be observed.

3.4.3 Quantitation of ligand dependent and independent EphA1 receptor activation

Alterations in receptor activation as a result of the P460L receptor variant were investigated via the quantitation of phosphotyrosine-EphA1 and phosphoserine-EphA1 levels within both EphA1 HEK293 and P460L HEK293 cell lysates on addition of ephrinA1-Fc ligand.

Ligand activation the EphA1 HEK293 cells for 5 min showed significantly increased levels of phosphotyrosine-EphA1 compared to the human IgG control lysates (Fig. 3.12A), indicating receptor activation induced via ephrinA1 stimulation. However, 1 min ligand stimulation produced a decrease in the phosphotyrosine-EphA1 levels. This may be due to this 1 min time point being too short to observe sufficient increase in tyrosine phosphorylation and receptor activation. Such ephrinA1 ligand-induced increase in activation was not observed within the P460L receptor when compared to their respective controls (Fig. 3.12B).

As outlined in section 3.1.1, EphA1's closest homologue EphA2 has been observed to undergo ligand-independent activation (Miao et al. 2009). The signalling ability of the EphA1 receptor in this manner was investigated via quantitation of phosphoserine-EphA1 levels within both EphA1 HEK293 and P460L HEK293 cell lysates.

Both EphA1 HEK293 (Fig. 3.13A) and P460L HEK293 cell lysates (Fig. 3.13B) displayed increased levels of phosphoserine-EphA1 upon 1 min ephrinA1 ligand stimulation with no significant change observed after 5 min ligand stimulation when compared to their respective IgG control cell lysates. The presence of phosphoserine-EphA1 implies ligand-independent activation of the EphA1 receptor, this is transient and dephosphorylated upon ligand stimulation to allow tyrosine phosphorylation. This correlates with observed data from the analysis of ligand independent activation of the EphA2 receptor, showing dephosphorylation of serine residues upon 10 min ephrinA1 ligand incubation (Miao et al. 2009).

These data suggest altered ligand-dependent receptor activation as a result of the P460L variant. Despite the expression levels of both EphA1 receptors being comparable within both the HEK293 cell lines EphA1 protein distribution was remarkable different. The more cytosolically expressed P460L receptor was not observed to elicit a response from ligand stimulation even upon normalisation to amount of receptor present at the membrane. This insinuates that this variant is capable of altering multiple aspects of the EphA1 receptors biology. When amalgamated together these alterations are likely to cause disruption to its downstream signalling pathways, influencing an individual's AD risk.

Possible limitations to analysis of receptor activation in this manner is the use of pan-phosphotyrosine and -phosphoserine antibodies. More accurate results may be produced through the use of antibodies against specific EphA1 residues such as Ser897 or Tyr781 due to the removal of background noise, potentially allowing smaller alterations within EphA1 activation to be observed.

Another limitation of receptor activation analysis in this manner is the differences in membrane receptor expression levels between EphA1 HEK293 and P460L HEK293 cells. Reduced membrane expression of the P460L receptor variant may skew the results to show reduced receptor activation within P460L HEK293 cells when compared to EphA1 HEK293 cells. This is due to less receptor being present on the membrane to become activated with a large body of un-activated receptor residing within the cytosol. Cell fractionation via centrifugation would allow the isolation and enrichment of a membrane fraction within the cell lysate, this would enable a more accurate analysis of receptor activation through only analysing membrane bound receptor capable of activation and not that located within the cytosol. Immunofluorescence may be a potential alternative method to allow receptor activation analysis. Through the use of specific phosphotyrosine/phosphoserine-EphA1 antibodies levels of receptor activation can be more accurately quantitated via only considering membrane bound receptor.

3.5 Conclusion

Taken together, these data elude to the details of receptor expression, subcellular localisation and activation of the EphA1 receptor in the absence and presence of ligand stimulation (Fig. 3.14A and 3.14C).

Soluble EphA1 receptor is released from the EphA1 HEK293 cell lines in the absence of ephrinA1 ligand (Fig. 3.8A). This soluble EphA1 is not generated by MMPs, but potentially by another protease not inhibited by GM6001, such as ADAM12 or a different category of protease such as serine proteases to release the extracellular domain of the receptor. Alternatively, this soluble product observed may be full length receptor released via vesicles.

Upon ligand activation, serine residues responsible for ligand-independent activation are transiently dephosphorylated and the tyrosine residues responsible for ligand-dependent activation phosphorylated (Fig. 3.13 and Fig 3.12 respectively), leading to EphA1 receptor internalisation (Fig. 3.11C). Serine dephosphorylation and tyrosine phosphorylation in response to ligand activation has also been noted within the EphA2 receptor leading to receptor internalisation and degradation (Miao et al. 2009). However, it still remains unclear whether this internalised EphA1 receptor is just the V5-tagged C-terminal tail, the full-length receptor or the entire EphA1-ephrinA1 signalling complex. This could potentially be investigated through imaging flow cytometry using an antibody against ephrinA1.

The impact of the P460L variant on EphA1 receptor expression, localisation and downstream signalling can also be inferred (Fig. 3.14B and D). These data indicate that the P460L variant results in altered receptor subcellular distribution with reduced levels at the membrane, including reduced release of soluble receptor (Fig. 3.10G and Fig. 3.11A respectively). Although transient serine dephosphorylation was observed on ligand activation was also noted within this receptor variant (Fig. 3.13C) no significant activation nor receptor internalisation was observed upon ligand stimulation in contrast to the EphA1 receptor (Fig. 3.12C and 3.11E respectively). Further studies however, may be required to

confirm the serine dephosphorylation observed equates to receptor ligand-independent activation, this could be achieved via the addition of a receptor functionality assay, such as, cell spreading or migration analysis.
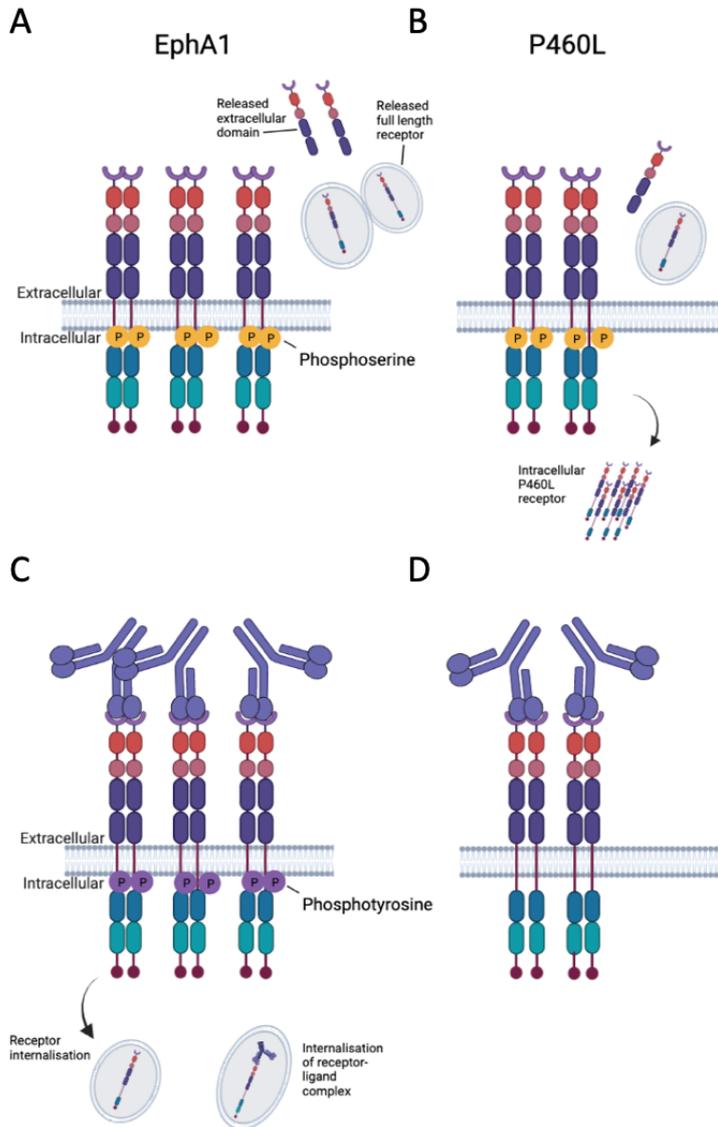
Figure 3.14: Summary of EphA1 and P460L receptor expression, activation and processing.
A. EphA1 receptor is mainly expressed at the cell membrane. Ligand independent release of soluble receptor, either by proteolytic cleavage of extracellular domain or release of full-length receptor into vesicles. Serine residues phosphorylated during ligand independent receptor activation. **B.** Reduced presence of the P460L receptor at the cell membrane with majority of expression seen within the cytosol. Decreased release of soluble receptor. **C.** Ligand activation by ephrinA1-Fc results in dephosphorylation of serine residues and subsequent phosphorylation of tyrosine residues followed by internalisation of either the EphA1 receptor or the EphA1-ephrinA1 signalling complex. **D**. Ligand activation of the P460L receptor does not cause significant receptor internalisation or receptor internalisation. Schematic created using BioRender.com

Chapter 3: Analysis of EphA1 receptor function and activation

It has been shown previously, within cancer metastasis, that proteolytic cleavage of ephrinA1 disrupts EphA1–ephrinA1 mediated cell contacts resulting in lung hyperpermeability, facilitating tumour cell entry. Therefore, reduction of endothelial cell surface EphA1 produced a downregulation of cell adhesion molecules (Ieguchi et al. 2014). This may translate into AD, explaining the potential mechanisms behind the P460L variant's association with disease progression; for example, the reduced level of receptor at the cell membrane may result in BBB hyperpermeability causing increased leukocyte migration and translocation into brain parenchyma. Increased peripheral immune cells within the brain may propagate AD-associated neuroinflammation and neurodegeneration (Kinney et al. 2018). Investigation into this mechanism could be done via analysis of EphA1 expression pattern within cells of the BBB and the consequences reduced membrane expression has on its function.

Chapter 4: Analysis of the regulatory ability of non-coding SNPs within the *EPHA1* risk loci

4.1 Introduction

EPHA1 disease associated variants or SNPs may have the potential to regulate expression of target genes, leading to an effect on an individual's Alzheimer's disease susceptibility. In this chapter, SNPs with the potential to regulate gene expression within the *EPHA1* locus will be identified for further investigation.

4.1.1 The limitations of GWAS

Despite allowing great progress in identifying genetic risk loci associated with AD, GWAS have several limitations which result in complex data interpretation. These issues arise due to the methods employed during GWAS as well as the genomic location of the SNPs themselves, resulting in difficulties assigning causality to a specific variant (Edwards et al. 2013).

Linkage disequilibrium (LD) describes the non-random association of nearby alleles at different loci within a population. SNPs that are inherited together within LD blocks which are highly correlated within ethnic backgrounds (Shifman et al. 2003). Therefore, it is possible to determine the common genome wide variations within an individual through genotyping a select population of SNPs. These are referred to as 'tagging' SNPs. GWAS methods exploit this, with modern studies using roughly 300-500 thousand tagging SNPs enabling cost effective studies with larger sample sizes (Kruglyak 1999). However, data interpretation issues arise when trying to identify the true disease associated variant together with their target gene(s) from this block of associated SNPs.

4.1.2 Methods to identify causal SNP/s within a disease associated locus

GWAS alone lacks the power needed to identify the causal variant linked to a complex trait or disease. Fine-mapping analysis, both statistical and functional, aims to prioritise genetic variants within GWAS data sets to help identify the potential causal variant(s) behind the disease associated signal (Broekema et al. 2020).

Multiple different models, often employed together in a stepwise fashion have been developed to enable statistical fine-mapping of GWAS loci. A common first approach is to interrogate the pairwise correlation (r^2) between the lead SNP from a GWAS data set and other SNPs at a particular locus (Schaid et al. 2018). This type of Heuristic fine-mapping approach triages SNPs based on LD structure. SNPs above a defined LD threshold ($r^2<0.8$ or 80%) are considered to be linked to the particular lead SNP analysed. All SNPs found to be within LD of the lead SNP are also potential candidate SNPs of interest and may even be the true causal variant over the GWAS lead SNP (Schaid et al. 2018). Approaches such as this allow an extra layer of sophistication over methods taking only GWAS p-values into account. The use of p-values alone may lead to misleading results due to being influenced by the power and effect size of the study, leading to fluctuations in p-values dependent on GWAS data set. (Spain and Barrett 2015).

Combining functional annotation from publicly available sources with statistical data is crucial for understanding the role of non-coding variants in gene regulation (Spain and Barrett 2015). The regulatory role of certain non-coding RNAs including the accessibility of open chromatin to regulatory factors has a strong effect on gene transcriptional levels (Degner et al. 2012). Therefore, genomic regions containing non-coding variants may possess gene regulatory ability, potentially due to the variant altering the efficiency of TF binding sites (Grubert et al. 2015). Analysis of DNA accessibility from transposase-accessible chromatin with high-throughput sequencing (ATAC-Seq) data could predict functionality of non-coding regions and disease associated variants (Broekema et al. 2020).

The ENCODE project, developed in 2004 details the genomic location of transcriptional regions, histone modifications and TF association sites (Consortium 2004). The advent of further consortia such as, FANTOM5 (functional annotation of the mammalian genome 5) and the NIH roadmap epigenomics consortium, allows overlay of functional elements, such as, enhancers, promoters and TF binding sites onto GWAS data (Consortium et al. 2014; Romanoski et al. 2015). Overlay of such functional annotation onto non-coding areas is essential to tease out alterations in gene expression which may underlying a causal variants' association with AD (Consortium 2012).

4.1.3 Methods to identify disease associated cell type

Studies have shown between 50-90% of disease associated variants exert effects on gene expression in a tissue specific manner (Dimas et al. 2009). The incorporation of expression quantitative trait loci (eQTL) with GWAS data sets allows the identification of SNPs that explain variations in gene expression within certain cell types (Nica and Dermitzakis 2013). Such cell type specific data could be used to determine possible functional mechanisms behind a variant's association with disease pathology (Tsompana and Buck 2014).

The analysis of AD associated GWAS signals in context with chromatin state and eQTL data sets has identified an enrichment of variants within immune cells, mainly monocytes, macrophages and microglia. This illustrates how the majority of GWAS loci potentially operate via affecting gene expression within these cell types, and therefore further supports their importance in AD pathology. These data can then be used to inform downstream functional assays within cell types of interest to allow further investigation of the biological pathways involved (Tansey et al. 2018).

4.1.4 Alzheimer's disease associated SNPs within the EphA1 locus

The *EPHA1* risk locus is extremely complex, containing multiple AD associated variants, reaching genome-wide association which have been identified through numerous GWAS data sets and meta-analysis. All of these variants, like most GWAS SNPs are non-coding,

Chapter 4: Analysis of the regulatory ability of non-coding SNPs within the *EPHA1* risk loci

lying either within intron regions of *EPHA1* itself or, as with the majority of the SNPs at this locus, within its neighbouring antisense gene, *EPHA1-AS1*. EphA1-AS1, like many lncRNAs is thought to have regulatory ability over either *EPHA1* itself or neighbouring genes (Amlie-Wolf et al. 2019).

The addition of functional annotation such as eQTLs to GWAS data has attempted to tease out the causal variant responsible for the disease association at the *EPHA1* locus. Such studies have identified multiple variants such as rs11767557 and rs11765305, the protective alleles of which co-localise with a whole blood eQTLs for increased expression of *EPHA1*, *EPHA1-AS1* and other neighbouring genes such as *TAS2R60* and *ZYX* (Liu et al. 2018; Amlie-Wolf et al. 2019). In addition, Schwartzentruber *et al* conducted SNP level fine-mapping analysis followed by co-localisation at the *EPHA1* locus and identified two further SNPs, rs12703526 (p=9.63x10^{-12}) and rs7810606 (p=1.13x10^{-11}), which colocalise with whole blood eQTLs for *EPHA1-AS1* and *ZYX* expression. Within this study they concluded that these SNPs have a 49% and 43% probability respectively of being the causal variant at this locus (Schwartzentruber et al. 2021). Various publications have also attempted to add functional data to identify the likely biological consequences of *EPHA1* SNPs identified through GWAS. For example, the protective A alleles of rs11771145 and rs10808026 have been noted to correlate with protection against right hippocampal atrophy and decreased grey matter cerebral blood flow respectively (Wang et al. 2015a; Chandler et al. 2019).

Difficulty arises however, when trying to identify the true causal SNP behind the AD-association or eQTLs due to the nature of their LD structure, as previously mentioned in section 4.1.1. As illustrated in figure 4.1, multiple lead SNPs within the *EPHA1* locus, highlighted through different GWAS analysis, reside within the same LD block (defined by an r^2<80%). Therefore, further analysis at an individual SNP level is necessary to identify which variant produces the AD-association signal seen at this locus and via what functional mechanism or pathway it influences disease pathology.
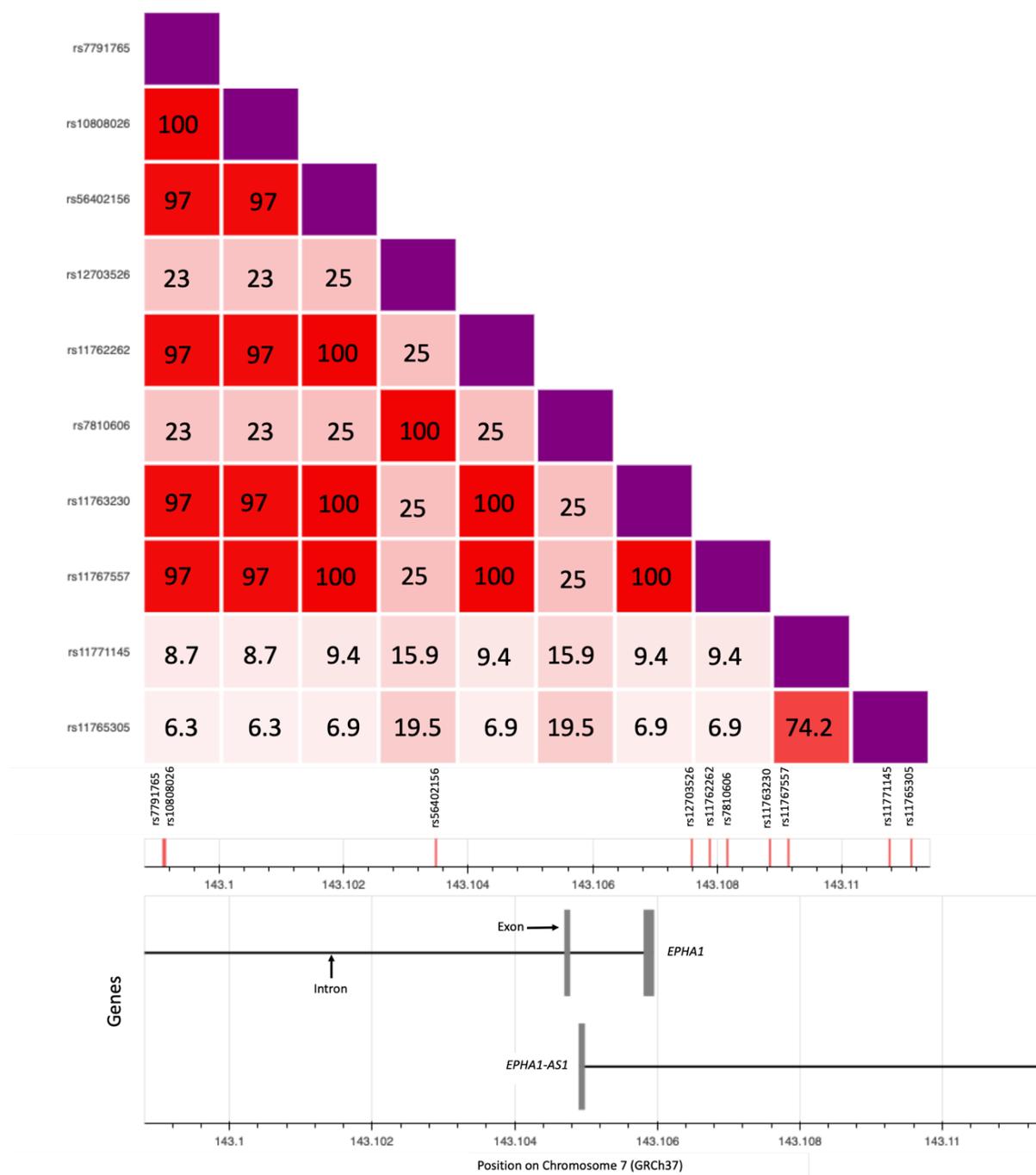
Figure 4.1: Linkage Disequilibrium heat map of *EPHA1* lead GWAS SNPs of interest.

Heat map showing the LD structure of lead GWAS SNPs based on p-value within the *EPHA1* locus along with other linked variants. Numbers within the grid represents the LD percentage score between two SNPs. Variants achieving an LD percentage score above 80% are considered linked. The genomic location of each variant within either *EPHA1* introns or *EPHA1-AS1* is shown along the bottom.

4.1.5 Consequences of non-coding disease associated variants

Differences in gene expression which drives genetic phenotypic variation is controlled by the ability of specific DNA motifs to bind TFs and other regulatory elements. It is therefore hypothesised that disease associated variants within non-coding regions alter an individual's genetic disease susceptibility via changing the consensus sequence of these transcription factor binding motifs and as a consequence, the affinity by which TFs can bind, if at all. This causes alterations to the molecular processes that control gene expression, resulting in changes to gene expression within individuals possessing such AD-associated variants (Maurano et al. 2012).

4.1.6 Hypothesis

EPHA1 AD-associated SNPs affect an individual's genetic disease susceptibility through their ability to regulate gene expression of target genes such as *EPHA1*, *EPHA1-AS1* and various other neighbouring genes such as *ZYX, TAS2R60 and FAM131B* via altering the binding affinity of regulatory proteins within specific cell types.

4.1.7 Overall Aim

The aim of this chapter is to identify the potential SNP(s) within the *EPHA1* locus which are responsible for this locus' association with Alzheimer's disease. To determine this, a SNPs genomic location and ability to form allele-specific interactions with regulatory proteins from nuclear extracts of disease-related cell types will be assessed.

4.1.8 Objectives:

1. Map the *EPHA1* locus SNPs to regions of open chromatin within cell types of interest, such as peripheral blood immune cells, microglia and neurons.

2. Identify any known TF binding sites or regulatory elements within these areas.

3. Identify *EPHA1* SNPs with the ability to bind regulatory proteins within microglial, monocyte and neuronal cell nuclear extracts using a LightShift® Chemiluminescent Electrophoretic Mobility-Shift Assay (EMSA).

4.2 Materials and Methods

4.2.1 Mapping EPHA1 SNPs to regions of Open Chromatin

The UCSC Genome Brower was used to interrogate the genomic location of the *EPHA1* locus SNPs prioritised by GWAS P value, fine mapping data and LD status (https://genome.ucsc.edu/). The full list of prioritised SNPs can be seen in Appendix II. ATAC-Seq open chromatin data from healthy untreated human monocytes (Novakovic et al. 2016), macrophages (Kang et al. 2017), microglia (Gosselin et al. 2017) and forebrain neurons (Fullard et al. 2017) were taken from published data deposited in the Cistrome Data Browser (http://cistrome.org/db/#/) and overlaid with SNP genomic locations.

4.2.2 Cell culture of the HMC3, SH-SY5Y and THP1 cell lines

HMC3 and SH-SY5Y cell lines were maintained in DMEM basal media (Gibco) containing 10% FBS (Sigma). Cultures were grown in a humidified incubator at 37°C in the presence of 5% CO_2. Cells were fed once every two days with fresh culture media, pre-warmed to 37°C. Once cell cultures had reached 70-80% confluency they were passaged. Spent media was removed and cell cultures washed with 1-5 mL PBS (Gibco) (depending on vessel size). Cells were removed from tissue culture vessel via incubation with 1-5 mL pre-warmed Trypsin/EDTA (0.05%/0.5 mM) (Gibco) (depending on vessel size), for 3 min or until cells detached upon mechanical agitation. An appropriate volume of pre-warmed media was added depending on vessel size in order to inactivate Trypsin/EDTA. Cells were centrifuged at 300 g for 5 min. The supernatant was removed and the cell pellet was resuspended in

appropriate volume of pre-warmed fresh media before being dispensed into fresh culture vessels, at a dilution of 1:4-1:10.

The human monocytic cell line, THP1 which is derived from an acute monocytic leukaemia patient is a suspension culture and maintained in RPMI basal medium (Gibco) containing 10% FBS. Cells were fed and passaged twice a week, for this the cell suspension was removed from the culture vessel and centrifuged at 300 g for 3 min. The spent media was removed and the cell pellet was resuspended in 1 mL of prewarmed media. A Trypan Blue (Hyclone) cell count was performed on a Luna Dual Florescence cell counter using Luna cell counting slides. Cell suspension was seeded into a fresh tissue culture vessel containing pre-warmed media at a cell concentration of $1X10^5$-$1x10^6$ cells per mL.

4.2.3 Nuclear protein extraction

Nuclear proteins were extracted from THP-1, HMC3 and SH-SY5Y cells using a NER Nuclear and Cytoplasmic extraction kit (ThermoFisher Scientific) following manufacturer's instructions.

Cells were removed from a 70-80 % confluent T75 flask via incubation with 5 mL pre-warmed Accutase (Gibco) for 3 min or until cells detached upon mechanical agitation. Cell suspension was centrifuged for 3 min at 300 g. Supernatant was removed and cell pellet allowed to dry prior to the addition of 500 µL ice cold Cytoplasmic Extraction Reagent I (CERI). Cell suspension was vortexed for 15 sec before being incubated on ice for 10 min. After which 55 µL ice cold Cytoplasmic Extraction Reagent II (CERII) was added, sample was vortexed for 5 sec and incubated for 1 min on ice. Samples were again vortexed for 15 sec before centrifugation at 16000 g for 5 min. The supernatant/cytoplasmic extract was discarded and pellet was resuspended in 250 µL Nuclear Extraction Reagent (NER) and incubated for 40 min, with a 15 sec vortex every 10 min. Sample was then centrifuged 1600 g for 10 min and supernatant/nuclear extract transferred to a fresh Eppendorf tube.

Chapter 4: Analysis of the regulatory ability of non-coding SNPs within the *EPHA1* risk loci

The protein concentration of the nuclear extract was quantitated using a BCA Protein Assay Kit (ThermoFisher Scientific) as per <u>manufacturer's protocol</u>.

4.2.4 Design and annealing of EPHA1 SNP Oligonucleotides

Complementary pairs of 50 nucleotide DNA oligonucleotides were designed for both the risk and protective allele of each of the 7 *EPHA1* SNPs prioritised via open chromatin data analysis (figure 4.3 and 4.4). These oligonucleotide pairs were created in duplicate, one containing a 5' Biotin tag and the other remaining unlabelled. A total of 32 pairs were created, 2 of which are shown in table 4.1, the exact location of the SNP allele is shown in red.

Table 4.1: Example of complementary DNA oligonucleotides created of the *EPHA1* SNPs

SNP	Allele	Oligonucleotide
rs11762262	Protective - C	TGTATACAGAGGGGCCATCTGACACACATTGCTCTGCACCAGAGGCCTCTT AAGAGGCCTCTGGTGCAGAGCAATGTGTGTCAGATGGCCCCTCTGTATACA
	Risk - T	TGTATACAGAGGGGCCATCTGACACATATTGCTCTGCACCAGAGGCCTCTT AAGAGGCCTCTGGTGCAGAGCAATATGTGTCAGATGGCCCCTCTGTATACA

Complementary oligonucleotides were annealed together via mixing of an equal concentration of both oligonucleotides and heating to 95 °C for 5 min followed by incremental cooling by 1 °C every 5 min until RT reached. The annealed oligonucleotide DNA was then diluted to 1 pmol/μL in H_2O for subsequent use in the EMSA.

4.2.5 LightShift® Chemiluminescent Electrophoretic Mobility-Shift Assay (EMSA)

The LightShift® Chemiluminescent EMS Assay (ThermoFisher Scientific) allowed the nuclear protein binding ability of the *EPHA1* SNPs to be assessed. The EMSA was conducted in sets of three 20 μL reactions for each SNP allele. The composition of each reaction is detailed in table 4.2. Reaction 1 acted as a negative control, containing only the Biotin-labelled DNA. Reaction 2 contained the DNA-protein binding reaction between the Biotin-labelled DNA and the nuclear protein extract. Reaction 3 acted as a positive competitor control, this

Chapter 4: Analysis of the regulatory ability of non-coding SNPs within the *EPHA1* risk loci

contained all the components of reaction 2 with the addition of a 200-fold excess of unlabelled DNA. The unlabelled DNA acted as a competitive inhibitor to out compete the labelled DNA binding for protein binding, this confirmed the specificity of the DNA-protein interaction to the specific SNP allele within the oligonucleotide sequence.

Table 4.2: Composition of three binding reactions required for each SNP allele.

Reagent	Reaction 1	Reaction 2	Reaction 3
H₂O	Make up to 20 µL	Make up to 20 µL	Make up to 20 µL
10X Binding Buffer	2 µL	2 µL	2 µL
Poly (dI•dC) 1µg/µl	1 µL	1 µL	1 µL
NP-40 (1%)	1 µL	1 µL	1 µL
Unlabelled DNA (1pmol/µl)	-	-	4 µL
Nuclear Protein Extract (HMC3 or THP-1)	-	10 µg	10 µg
Biotin labelled DNA	20 fmol	20 fmol	20 fmol

Reactions were incubated at RT for 20 min followed by the addition of 5 µL loading dye (5X) and mixed by gentle pipetting.

Samples were loaded into a pre-ran (100 V for 30 min) 5 % Mini-PROTEAN® Tris/Borate/EDTA (TBE) precast DNA gel (Bio-Rad) and electrophoresed in 0.5X TBE buffer at 100 V until the bromophenol blue loading dye had migrated to the bottom of the gel.

Once migration was complete, the gel was floated off its plastic casing within a bath of 0.5X TBE Buffer. Nylon membrane (ThermoFisher Scientific) which had previously soaked for 10 min in 0.5X TBE Buffer, was applied to one side of the gel and a sandwich created with blotting paper and sponges on each side, this was then placed into a transfer cassette. The transfer cassette was inserted into the electrophoresis unit containing fresh 0.5X TBE buffer prechilled to 10 °C and 100 V was applied for 30 min.

Following transfer, excess TBE buffer was blotted off the membrane using a paper towel and the DNA crosslinked onto it using a UV-light set to 120mJ/cm2 for 1 min.

The 4X Wash Buffer and Blocking Buffer, supplied within the EMSA Kit was preheated to 37°C to dissolve any precipitate. The membrane was blocked in 20 mL Blocking Buffer for 15 min at RT, with gentle shaking. After which Blocking Buffer was removed and replaced with 20 mL Conjugate/Blocking Buffer (66.7 μl Streptavidin – Horseradish Peroxidase Conjugate in 20 mL Blocking Buffer) and incubated for a further 15 min at RT with gentle shaking. The membrane was then washed 4 times for 5 min each in 20 mL 1X Wash Buffer at RT with gentle shaking. After washing, the membrane was incubated in 30 mL Substrate Equilibration buffer for 5 min at RT with gentle shaking. The Substrate Equilibration buffer was removed and excess blotted away with paper towel. The membrane was then incubated in Substrate Working Solution (a mix of 6 mL Luminol/Enhancer Solution and 6 mL Stable Peroxide Solution) for 5 min at RT without shaking.

Substrate Working Solution was removed and excess solution blotted away from the membrane using paper towel. The membrane was then wrapped in plastic wrap to prevent drying out and imaged using a Syngene G-Box imager. A schematic example of expected bands is shown in figure 4.2.

Results were quantitated as per western blot images outlined in the general Materials and Methods Chapter, Chapter 2, Section 2.5. The large mass of unbound biotin-labelled DNA at the bottom of the blot was used as a loading control for the normalisation of band intensity, this band used is shown in figure 4.2. Therefore, relative band densities of the DNA-protein interactions are reflective of the amount of Biotin labelled DNA loaded within each reaction, removing any effect that may arise from gel loading inconsistencies.
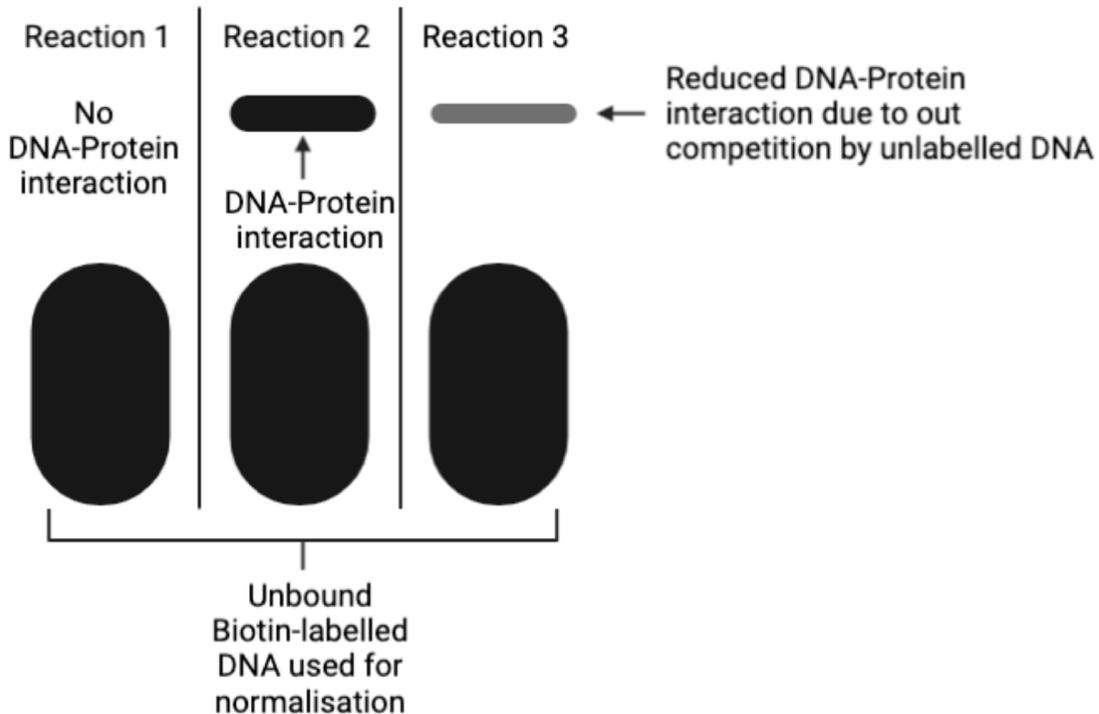
Figure 4.2: Schematic representation of expected bands on an EMSA gel.
Three binding reactions depicted. Reaction 1 shows the negative control containing only Biotin labelled DNA, producing only one large band of unbound labelled DNA at the bottom of the gel. Reaction 2 is the binding reaction containing labelled DNA and cell line nuclear extract. Smaller bands of DNA-protein interaction are expected above the unbound Biotin labelled DNA. Reaction 3 is the specificity control. This reaction lane should contain a smaller or less intense DNA-protein interaction band above the unlabelled DNA band. The unbound Biotin labelled DNA in each reaction lane was used as a loading control during quantitation of DNA-protein band relative density. Schematic created using Biorender.com

4.3 Results

4.3.1 Mapping EphA1 SNPs to region of Open Chromatin

As open chromatin strongly influences transcriptional activity, it was hypothesised that cell-type specific open chromatin regions (OCRs) may help prioritise potential functional non-coding risk SNPs, therefore, highlighting disease relevant sequences and cell types.

The genomic locations of the selected lead SNPs within the *EPHA1* locus (Appendix II) was mapped to OCRs within cell types of interest. For example, the peripheral immune cells, monocytes and macrophages and brain cells such as microglia and forebrain neurons.

Chapter 4: Analysis of the regulatory ability of non-coding SNPs within the *EPHA1* risk loci

When inspecting the OCRs from representative ATAC-Seq data taken from healthy control human monocytes (Novakovic et al. 2016), macrophages (Kang et al. 2017), microglia (Gosselin et al. 2017) and forebrain neurons (Fullard et al. 2017), a select number of *EPHA1* locus SNPs were identified to be located within regions of interest. This allowed the prioritisation of seven SNPs, depicted by red circles on the OCR maps of both peripheral immune cells and brain cells (Fig. 4.3) which potentially possess functional ability.

These seven prioritised SNPs; rs11767557, rs11762262, rs11763230, rs11771145, rs11765305, rs12703526, rs7810606, were taken forward for further analysis of their potential ability to alter gene expression. This was done through the analysis of their ability to bind regulatory proteins during an EMS assay.

Figure 4.3: Genomic location of *EPHA1* SNPs mapped to regions of ATAC-Seq open chromatin within peripheral immune and brain cells. Schematic created using the UCSC Genome browser overlaying the genomic locations of the 10 lead *EPHA1* SNPs (Appendix II) to ATAQ-Seq open chromatin data from control human monocytes (Novakovic et al. 2016), macrophages (Kang et al. 2017), control human microglia (Gosselin et al. 2017) and forebrain neurons (Fullard et al. 2017) imported from the Cistrome database. Seven prioritised *EPHA1* SNPs are circled in red.
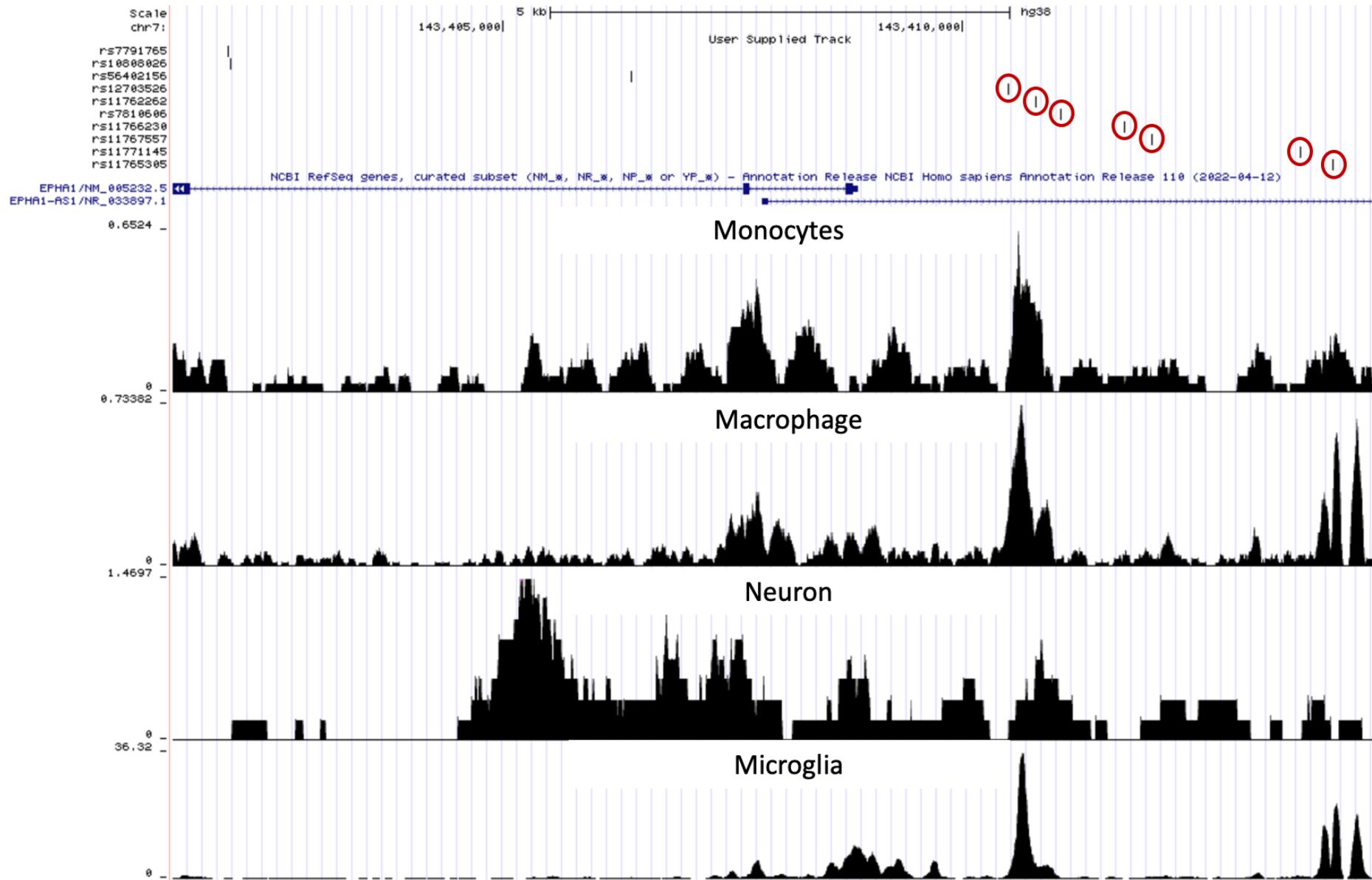
4.3.2 Bioinformatic analysis of regulatory elements and transcription factor binding sites at the EPHA1 risk loci

The non-coding genome contains regulatory elements and binding sites for TFs which allow control over gene expression. Therefore, a bioinformatic analysis was conducted to identify known regulatory elements or TF binding sites that overlap the *EPHA1* locus SNP locations. For this analysis the JASPAR Transcription Factor Binding Site Database and ENCODE Candidate Cis-Regulatory Elements tracks within the UCSC Genome browser were used. The results are shown in table 4.3, where N/A indicates no known TF binding site. Of the seven prioritised *EPHA1* SNPs located within regions of open chromatin, three were observed to lie within motifs of known transcription factor binding sites, rs11765305, rs11767557 and rs7810606. rs11762262 was noted to lie within a proximal enhancer region (ENCODE accession no. EH38E2598186).

Table 4.3: Bioinformatic analysis of *EPHA1* SNP location in relation to regulatory elements and transcription factor binding sites.

SNP	SNP Location (GRCh38/hg38 Assembly)	Transcription factors at this location	Location of transcription factor binding motif (GRCh38/hg38 Assembly)	Transcription factor motif consensus sequence at SNP location
rs11765305	7:143414019	CEBPB CEBPE CEBPG	7:143414019-143414028	C
rs11771145	7:143413669	N/A	N/A	N/A
rs11763230	7:143406388	N/A	N/A	N/A
rs11767557	7:143412046	ZSCAN31	7:143412031-143412049	G
rs7810606	7:143411065	SP8, SP9, KLF1, KLF6, KLF9, KLF10, KLF11, KLF16, KLF17, RREB1, NKX2-2 EGR1	7:143411055-143411066 7:143411055-143411066 7:143411057-143411067 7:143411057-143411067 7:143411053-143411068 7:143411056-143411066 7:143411056-143411066 7:143411056-143411066 7:143411055-143411069 7:143411047-143411066 7:143411062-143411075 7:143411052-143411065	C C C C C C A C C C C C
rs11762262	7:143410783	Proximal enhancer-like signal, CTCF-bound: EH38E2598186	7:143410614-143410938	N/A
rs12703526	7:143410495	N/A	N/A	N/A

4.3.3 The EPHA1 SNPs rs1176230 and rs11765305 form allele specific DNA-protein interactions within monocyte-like cells.

Biotinylated oligonucleotides containing either the risk or protective allele of all the prioritised *EPHA1* SNPs where incubated with nuclear extract from the monocyte cell line THP1. Only the SNPs rs11763230 and rs11765305 were able to form allele specific DNA-protein interactions.

Figure 4.4A, depicts a representative EMSA blot showing possible DNA-protein interactions, (highlighted by the white arrow heads) formed by each allele of the SNP rs11763230 on incubation with THP1 nuclear extract. These interactions were not seen within the respective negative control reactions (Fig. 4.4A, lane 1 and 4). On quantitation of these interactions, it was observed that the rs11763230 risk T allele was capable of binding a significantly increased amount of nuclear protein compared to the protective C allele as evident by the relative increase in band density seen in figure 4.4B (p=0.0133).

This DNA-protein interaction was noted to be specific to this allele's oligonucleotide sequence. The addition of a 200-fold excess of unlabelled DNA of the same sequence out-competed the biotinylated DNA for protein binding, producing the observed decrease in band intensity within the specificity control reaction (p=0.0110) (Fig 4.4A, lane 3 and 4.4C).
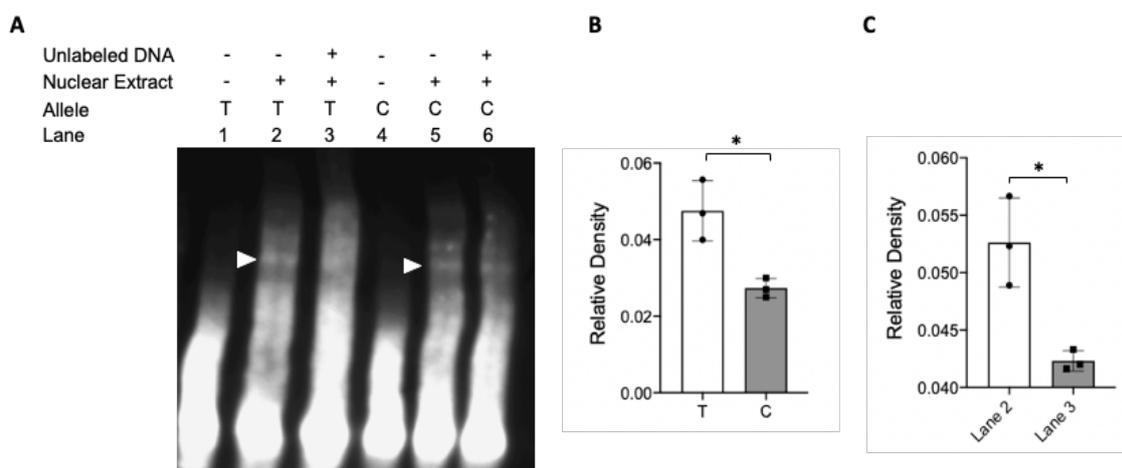
Figure 4.4: DNA-protein interactions of *EPHA1* SNP rs11763230 and THP1 nuclear extract.
A. A representative EMSA blot showing DNA-protein interactions formed on incubation of biotinylated oligonucleotides encompassing the risk T (lanes 1 to 3) and protective C allele (lanes 4 to 6) of the *EPHA1* SNP rs11763230 with nuclear protein extract from the THP1 cell line. Lanes 1 and 4 depicts the negative control reactions. Lanes 2 and 5 depict the binding reactions of the T and C allele respectively. The white arrow heads highlight DNA-protein interactions between nuclear proteins and the biotin-labelled DNA. Lanes 3 and 6 show the specificity control reactions for the T and C allele respectively, containing a 200-fold excess of unlabelled DNA of the same sequence. **B.** Quantitation of EMSA blots showing the relative density of DNA-protein interaction bands formed from both the risk T and protective C allele oligonucleotides (p=0.0133). **C.** Quantitation of DNA-protein bands formed within the binding reaction (lane 2) and the specificity reaction (lane 3) of the risk T allele (p=0.0110). Error bars indicate mean relative density ± SD of three independent data sets. Statistical test, t-test. *, p<0.05

Figure 4.5A depicts a representative EMSA blot showing DNA-protein interaction formed on incubation of the protective G allele of the SNP rs11765305 with nuclear proteins from THP1 cells. A possible allele specific interaction is depicted in figure 4.5A, lane 2, by the white arrow head which was not present within the negative control nor within the C allele binding reaction (Fig. 4.5A, lane 1 and 5). On quantitation, seen in figure 4.5B, it was observed that the DNA-protein band produced by the protective G allele oligonucleotide had a significantly higher relative density and therefore capacity to bind nuclear protein than that of the risk C allele (p=0.035).

This DNA-protein interaction was also noted to be specific to the G allele oligonucleotide sequence. The addition of a 200-fold excess of unlabelled DNA of the same sequence out-competed the biotinylated DNA for protein binding, causing the disappearance of the biotin labelled DNA-protein band (p=0.033) (Fig 4.5A, lane 3 and 4.5C).
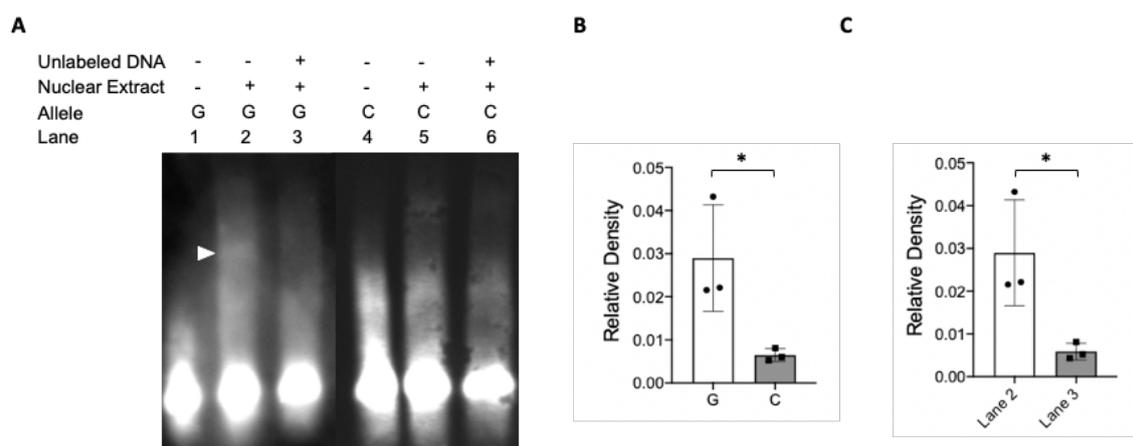
Figure 4.5: DNA-protein interactions of *EPHA1* SNP rs11765305 and THP1 nuclear extract.
A. A representative EMSA blot showing DNA-protein interaction band produced on incubation of biotinylated oligonucleotides encompassing the protective G and the risk C allele of the *EPHA1* SNP rs11765305 with nuclear extract protein from the THP1 cell line. Lanes 1 and 4 depict the negative control reaction. Lanes 2 and 5 depict the binding reactions of the G and C allele respectively. The white arrow head highlights the DNA-protein interaction between a nuclear protein and the biotin-labelled DNA. Lanes 3 and 6 show the specificity control reactions for the G and C allele respectively. **B.** Quantitation of EMSA blots showing relative band densities of the DNA-protein interaction formed from both the protective G and the risk C allele DNA oligonucleotides (p=0.035). **C.** Quantitation of DNA-protein bands formed within the binding reaction (lane 2) and the specificity reaction (lane 3) of the protective G allele (p=0.033). Error bars indicate mean relative density ± SD of three independent data sets. Statistical test, t-test. *, p<0.05

4.3.4 The EPHA1 SNPs rs7810606, rs11767557 and rs11763230 form allele specific DNA-protein interactions within a microglial cell line.

Biotinylated oligonucleotides containing either the risk or protective allele of all the prioritised *EPHA1* SNPs where incubated with nuclear extract from the human microglial cell line HMC3. Only the SNPs rs7810606, rs11767557 and rs11763230 were capable of forming allele specific DNA-protein interactions.

Figure 4.6A, depicts a representative EMSA blot showing two DNA-protein interactions formed by the risk T allele of the *EPHA1* SNP rs7810606, highlighted by the white arrow heads in lane 5. These interactions were not present within the negative control reaction (Fig. 4.6A, lane 4) or the protective C allele binding reaction (Fig. 4.6A, lane 2). On quantitation both the upper and lower bands where shown to have significantly greater relative densities (p=0.0092 and p=0.0012 respectively) and therefore DNA-protein binding capacity compared to that of the protective C allele (Fig. 4.6B and D respectively).

Both these DNA-protein interactions were noted to be specific to the oligonucleotide sequence encompassing the T allele. This was evident by the disappearance of the band signal within the specificity control reaction (Fig. 4.6A, lane 6) compared to the binding reaction (Fig. 4.6A, lane 5). Quantitation of the relative densities of both the upper and lower bands shows this reduction in band signal to be significant, as seen in figure 4.6C (p=0.005) and 4.6E (p=0.0006) respectively. This significant reduction in the relative band intensities is indicative of successful out competition of the biotin-labelled DNA by the excess unlabelled oligonucleotide of the same sequence, indicating the interactions observed were specific for the DNA sequence encompassing the risk T allele.
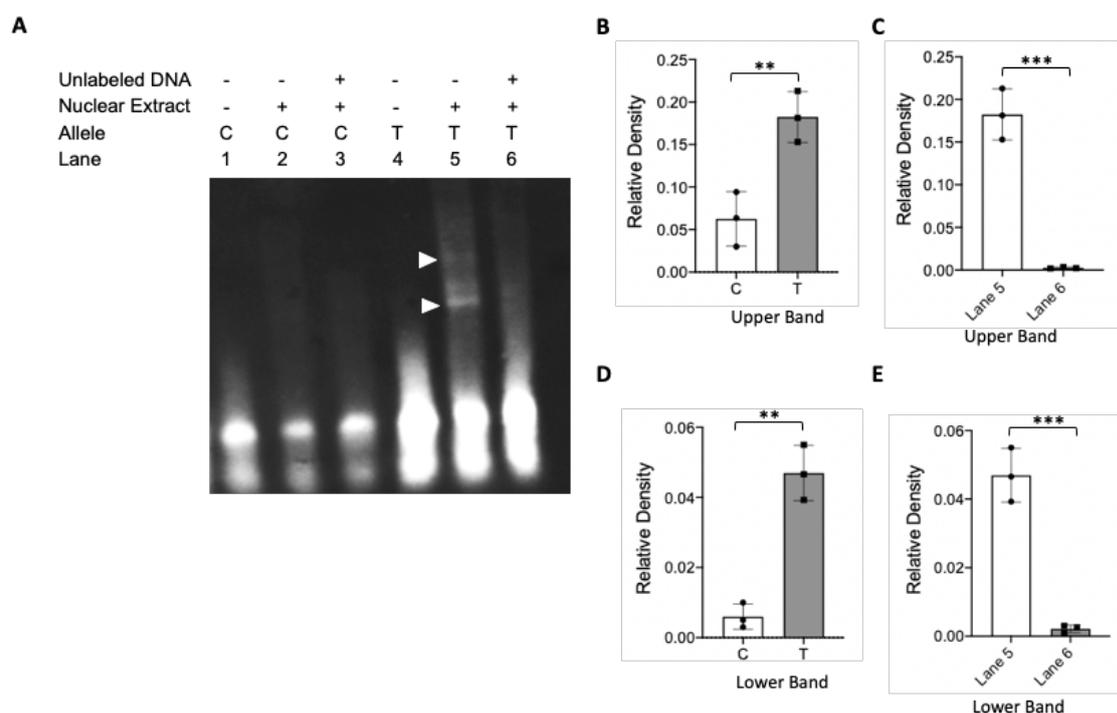
Figure 4.6: DNA-protein interactions of *EPHA1* SNP rs7810606 and HMC3 nuclear extract.
A. A representative EMSA blot showing incubation of biotinylated oligonucleotides encompassing the protective C and risk T allele of the *EPHA1* SNP rs7810606 with nuclear protein extract from the HMC3 cell line. Lanes 1 and 4 depicts the negative control reaction. Lanes 2 and 5 depict the binding reactions of the C and T allele respectively. The white arrow heads highlight two DNA-protein interactions formed between HMC3 nuclear proteins and the biotin-labelled T allele oligonucleotide. Lanes 3 and 6 show the specificity control reactions for the C and T allele oligonucleotides respectively. **B.** Quantitation of the relative band densities of the DNA-protein interaction forming the upper band (p=0.0092). **C.** Quantitation of the upper DNA-protein band relative densities formed within the binding reaction (lane 5) and the specificity reaction (lane 6) of the protective T allele (p=0.005). **D.** Quantitation of the relative band densities of the DNA-protein interaction forming the lower band (p=0.0012). **E.** Quantitation of the lower DNA-protein band formed within the binding reaction (lane 5) and the specificity reaction (lane 6) of the risk T allele (p=0.0006). Error bars indicate mean relative density ± SD of three independent data sets. Statistical test, t-test. **, p=0.01, ***, p=0.001.

Figure 4.7A, depicts a representative EMSA blot showing DNA-protein interactions formed on incubation of both the risk T and protective C allele of the SNP rs11767557 with nuclear proteins from HMC3 cells. DNA-protein interactions formed by both alleles are depicted by the white arrow heads. On quantitation, seen in figure 4.7B, it was observed that the DNA-protein band produced by the protective C allele oligonucleotide had a significantly higher relative density than that produced by the risk T allele (p=0.0062), indicating its ability to bind increased amounts of nuclear protein.

As shown in figure 4.7C, the DNA-protein interaction was noted to be specific for the DNA sequence encompassing the protective C allele due to the reduced relative density noted in the specificity control reaction (Fig. 4.7A, lane 6) compared to the binding reaction (Fig. 4.7A, lane 5) (p=0.0168).
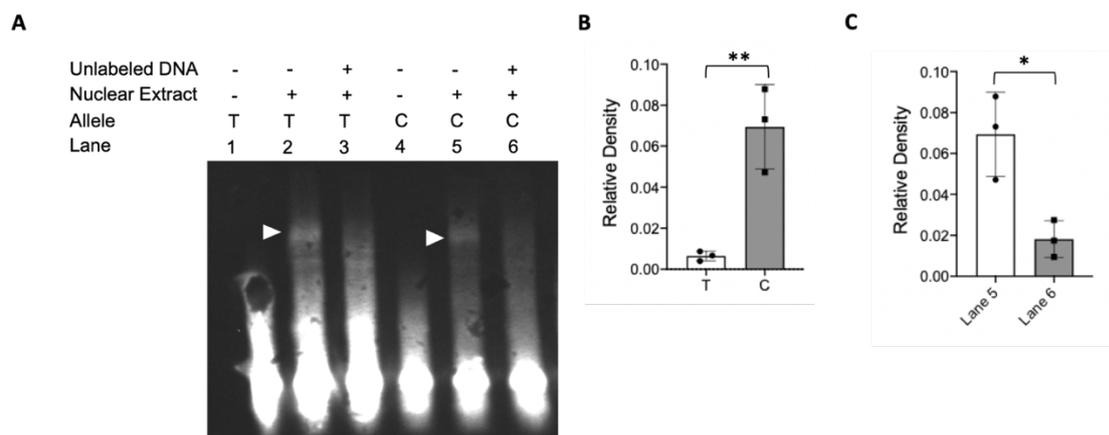
Figure 4.7: DNA-protein interactions of *EPHA1* SNP rs11767557 and HMC3 nuclear extract.
A. A representative EMSA blot showing the DNA-protein interactions formed on incubation of biotinylated oligonucleotides encompassing the risk T and protective C allele of the *EPHA1* SNP rs11767557 with nuclear extract protein from the HMC3 cell line. Lanes 1 and 4 depicts the negative control reaction. Lanes 2 and 5 depict the binding reactions of the T and C allele respectively. DNA-protein interaction bands formed by each allele are indicated by the white arrow heads. Lanes 3 and 6 show the specificity control reactions for the T and C allele oligonucleotides respectively. **B.** Quantitation comparison of the relative density of the DNA-protein bands produced by the risk T and protective C allele oligonucleotides (p=0.005). **C.** Quantitation of the relative densities of the DNA-protein bands produced from the binding reaction (lane 5) and the specificity reaction (lane 6) (p=0.0168). Error bars indicate mean relative density ± SD of three independent data sets. Statistical test, t-test, *, p=0.05, **, p=0.01.

Figure 4.8A, depicts a representative EMSA blot showing DNA-protein interactions, depicted by the white arrow heads, formed on incubation of both the risk T (lane 2) and protective C allele (lane 5) of the SNP rs11763230 with nuclear proteins from HMC3 cells. The DNA-protein interaction created was not observed within the negative control (Fig. 4.8A, lanes 1 and 4). On quantitation, seen in figure 4.8B, it was observed that the DNA-protein band produced by the protective C allele oligonucleotide had a significantly higher relative density and therefore capacity to bind nuclear protein than that of the risk T allele (p=0.0344).

This DNA-protein interaction was shown to be specific for the DNA sequence encompassing the protective C allele due to the reduced band intensity noted in the specificity control reaction (Fig. 4.8A, lane 6) compared to the binding reaction (Fig. 4.8A, lane 5). On quantitation this difference was observed to be significant, indicating the specificity of this DNA-protein interaction to that of the protective C allele (p=0.0157) (Fig. 4.8C).
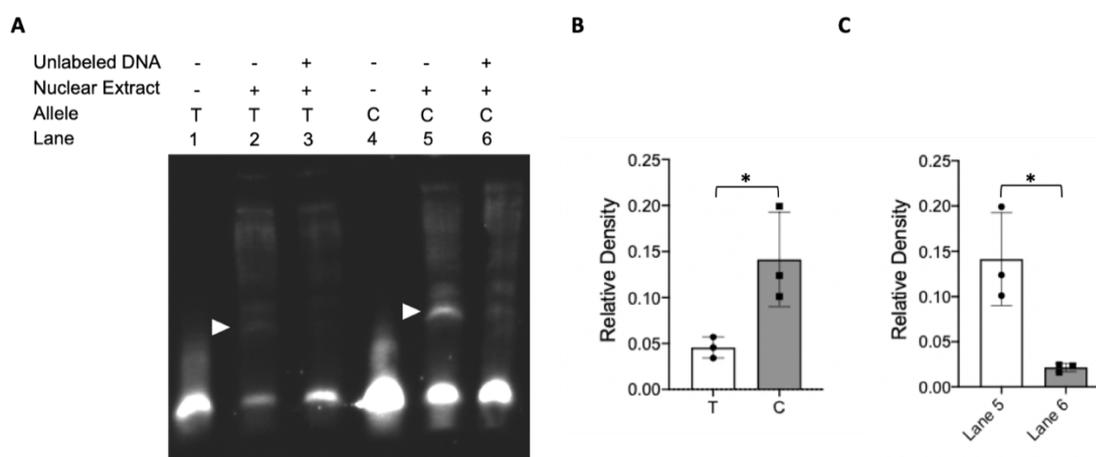
Figure 4.8: DNA-protein interactions of *EPHA1* SNP rs11763230 and HMC3 nuclear extract.
A. A representative EMSA blot showing DNA-protein interactions formed on incubation of biotinylated oligonucleotides encompassing the risk T and protective C allele of the *EPHA1* SNP rs11763230 with nuclear extract protein from the HMC3 cell line. Lanes 1 and 4 depicts the negative control reaction. Lanes 2 and 5 depict the binding reactions of the T and C allele respectively. DNA-protein interaction bands of each allele are indicated by the white arrow heads. Lanes 3 and 6 show the specificity control reactions for the T and C allele oligonucleotides respectively. **B.** Quantitation of DNA-protein band relative densities of the risk T and protective C allele DNA oligonucleotide (p=0.0344). **C.** Quantitation of the relative densities of the DNA-protein bands produced from the binding reaction (lane 5) and the specificity reaction (lane 6) (p=0.0157). Error bars indicate mean relative density ± SD of three independent data sets. Statistical test used, t-test. *, p=0.05.

4.3.4 The EPHA1 SNPs rs11763230 form allele specific DNA-protein interactions within a neuronal cell line

Biotinylated oligonucleotides containing either the risk or protective allele of all the prioritised *EPHA1* SNPs where incubated with nuclear extract from the human neuronal cell line SHSY5Y. Only the SNP rs11763230 had the capacity to form an allele specific DNA-protein interaction.

Figure 4.9A, depicts a representative EMSA blot showing two DNA-protein interactions formed by the risk T and protective C allele of the *EPHA1* SNP rs11763230, highlighted by the white arrow heads in figure 4.9A, lanes 2 and 5 respectively. These interactions were not present within the negative control reaction (Fig. 4.9A, lanes 1 and 4). On quantitation both the upper and lower DNA-protein interaction bands formed by the risk T allele where shown to have significantly greater relative densities (p=0.0005 and p=0.0333 respectively) and therefore DNA-protein binding capacity compared to that of the protective C allele (Fig. 4.9B and D respectively).

However, as depicted in figure 4.9C, on quantitation only the DNA-protein interaction producing the upper band seen in figure 4.9A, lane 2, was noted to have a significantly greater band relative density than that of its corresponding band within the specificity control reaction (p=0.0414) (Fig. 4.9A, lane 3). There was no significant difference noted in band relative densities between the binding reaction and specificity reaction producing the lower DNA-protein interaction (Fig. 4.9E).
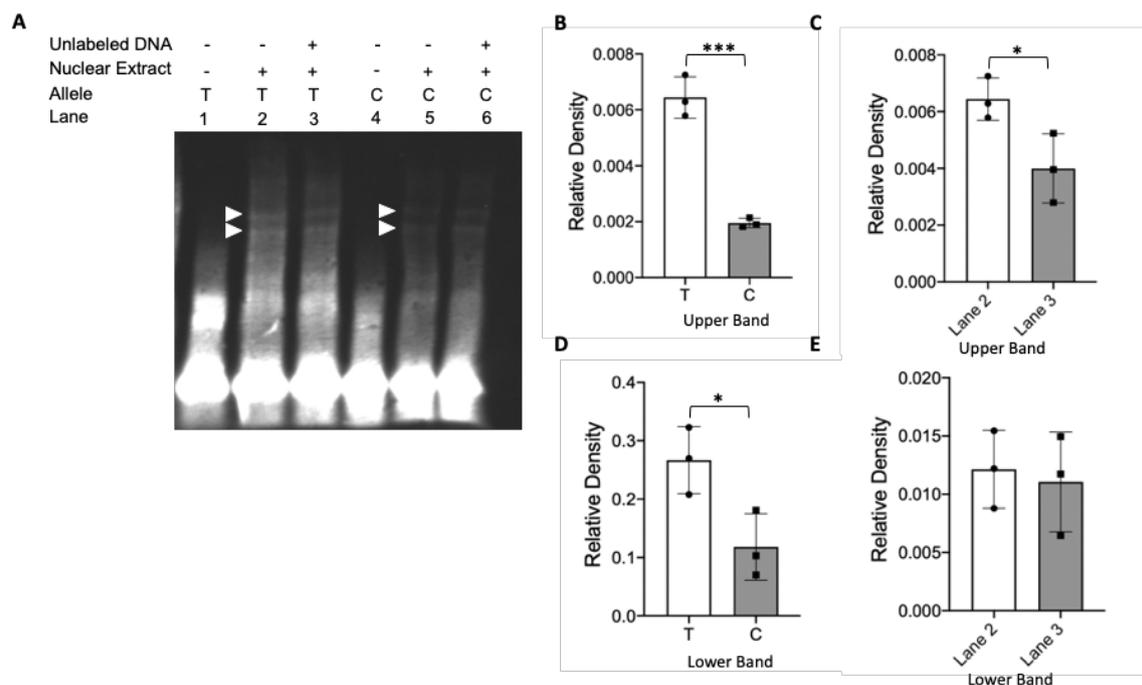
Figure 4.9: DNA-protein interactions of *EPHA1* SNP rs11763230 and SHSY5Y nuclear extract.
A. A representative EMSA blot showing DNA-protein interactions formed on incubation of biotinylated oligonucleotides encompassing the risk T and protective C allele of the *EPHA1* SNP rs11763230 with nuclear extract protein from the SHSY5Y cell line. Lanes 1 and 4 depicts the negative control reaction. Lanes 2 and 5 depict the binding reactions of the T and C allele respectively. DNA-protein interaction bands of each allele are indicated by the white arrow heads. Lanes 3 and 6 show the specificity control reactions for the T and C allele oligonucleotides respectively. **B.** Quantitation of the relative densities of the upper DNA-protein bands produced by the risk T and protective C allele oligonucleotides binding to SHSY5Y nuclear protein (p=0.0005). **C.** Quantitation of the relative densities of the upper DNA-protein bands produced from the binding reaction (lane 2) and the specificity reaction (lane 3) (p=0.0414). **D.** Quantitation of the relative densities of the lower DNA-protein bands produced by the risk T and protective C allele oligonucleotides binding to SHSY5Y nuclear protein (p=0.0333). **E.** Quantitation of the relative densities of the lower DNA-protein bands produced from the binding reaction (lane 2) and the specificity reaction (lane 3). Error bars indicate mean relative density ± SD of three independent data sets. Statistical test, t-test, *, p=0.05, ***, p=0.001.

4.3 Discussion

4.3.1 Prioritisation of GWAS EPHA1 locus SNPs

GWAS analysis has enabled great progress in identifying genetic risk loci and AD associated variants. However, further work is needed to convert this genetic data into biological information in terms gene expression alterations within different AD-relevant cell types in the brain and periphery and how this may contribute to disease pathology.

As with many AD risk loci, the non-coding nature of the variants at the *EPHA1* locus raises multiple issues when trying to ascertain the cell types and genes involved in the variant's association with AD. Fine-mapping and functional annotation allows the triaging of variants that may potentially be causal. However, the nature of LD structure (Fig. 4.1), resulting in multiple SNPs residing in the same inherited LD block as the statistical lead SNPs of GWAS datasets makes determining the exact variant behind the disease association extremely difficult.

Mapping lead GWAS *EPHA1* SNPs along with other linked variants to regions of open chromatin within disease relevant cell types such as monocytes, macrophage, microglia and forebrain neurons allowed the triaging of SNPs at this locus. As demonstrated in figure 4.3 the genomic location of various SNPs within the *EPHA1* locus reside within open chromatin regions from publicly available ATAC-Seq data from disease relevant cell lines, hinting at their potential regulatory ability over gene expression. From this schematic seven SNPs were prioritised: rs12703526, rs11762262, rs7810606, rs11763230, rs11767557, rs11771145 and rs11765305. This extra layer of information aims to negate the compounding variable LD structure, with the prioritised SNPs taken forward for further investigation into their potential ability to regulate gene expression via binding regulatory proteins such as TFs from the nuclear protein extracts of disease relevant cell types. Although residing just outside OCRs of the cell types investigated the SNP rs7810606 was still included in the selection of prioritised SNPs. This was due to this SNP previously being

Chapter 4: Analysis of the regulatory ability of non-coding SNPs within the *EPHA1* risk loci

identified within published fine mapping data as been able to regulate the gene expression of the gene *ZYX* (Schwartzentruber et al. 2021).

The methodology of variant selection, based on GWAS p-value and LD structure used during this chapter does have limitations which may have resulted in false negatives or variants being inadvertently ruled out of being causative. Additionally, open chromatin analysis to assess variants with likely gene regulatory ability was only conducted on control and untreated cells. It may potentially be the case that the role of common variants in gene regulation are only apparent during disease progression or under stress conditions such as chronic inflammation. Another possibility not considered here is that the sequence alteration produced by the variant itself that can cause alterations to chromatin structure and therefore wouldn't be identified in figure 4.3. Therefore, allele-specific ATAC-seq or the addition of further fine mapping approaches such as the overlay of eQTL data would allow the co-localisation of GWAS variants with gene expression change, aiding the identification of causal variants.

However, this approach still has limitations due to the often under powered nature of the data sets used. Therefore, high-medium throughput approaches enabling the analysis of hundreds of common variants at particular loci would allow the unbiased screening for potential causal variants. For example, the genome-engineering based interrogation of enhancers assay for transposase accessible chromatin (GenIE-ATAC) allows the screening of an individual variants effect on chromatin accessibility in an endogenous context. This approach would not only shed light on the effects of specific variants on the structure of chromatin and therefore gene expression but also aid the understanding on how sequence alterations introduced by the variants themselves effect the TF binding (Cooper et al. 2022).

Additionally, massively parallel reporter assays (MPRAs) allow large scale screens for variants that have the capacity to significantly alter gene expression. This is achieved through the cloning of a specific variant allele into a reporter plasmid to assess promoter or enhancer activity. MPRAs has enabled the creation of large publicly available databases such as the SNP-Survey of regulatory elements (SuRE) data browser, which contains the

regulatory information of 5.9 million SNPs, allowing the screening of the entire human genome at a high resolution and coverage within two different cell types (human HepG2 and K562) (van Arensbergen et al. 2017). The SNP-SuRE database identifies multiple variants at the *EPHA1* loci to be able to create regulatory ability within the cell types analysed, only two of which were present within our analysis (SNPs rs5640216 and rs1080826) (van Arensbergen et al. 2019). However, these variants were omitted from our further investigations due to not residing within regions of open chromatin. This potentially highlights the importance of such large-scale databases to screen all variants at a given locus for regulatory ability to reduce any false negatives.

Additionally, the advent of computational frameworks also enables further refinement of lists of GWAS SNPs in an unbiased and highly accurate manner. For example, the PAINTOR method (Probabilistic Annotation INTegratOR) allows the integration of GWAS summary statistics, LD structure and multiple functional annotations to give each variant at a particular locus a probability of being causal (Kichaev et al. 2014). The incorporation of such techniques and further fine mapping of *EPHA1* variants may be beneficial, identifying additional variants at this locus which require further functional characterisation.

4.3.2 Analysis of the prioritised EPHA1 SNPs' ability to bind regulatory nuclear proteins.

The prioritised *EPHA1* SNPs (Fig. 4.3, red circles) were analysed further via an EMS assay. This assay aimed to ascertain which SNPs had the ability to bind nuclear proteins from the AD-relevant cell types, monocytes (THP1), microglia (HMC3) and neurons (SHSY5Y). These data are summarised in table 4.4, detailing the prioritised SNPs, respective alleles and their ability to form allele-specific DNA-protein interactions within the specific cell types. The inability to form a DNA-protein interaction is noted by N/A.

Table 4.4: Summary of EMS assay data.

SNP	Risk Allele	Protective Allele	Allele capable of forming stronger DNA-protein interactions within specific cell types		
			Monocytes	Microglia	Neuron
rs11765305	C	G	G	N/A	N/A
rs11771145	G	A	N/A	N/A	N/A
rs11763230	T	C	T	C	T
rs11767557	T	C	N/A	C	N/A
rs7810606	C	T	N/A	T	N/A
rs11762262	T	C	N/A	N/A	N/A
rs12703526	G	T	N/A	N/A	N/A

These data however, alongside that of open chromatin region analysis supplies more evidence that certain variants within the *EPHA1* locus are associated with AD pathology through their ability to regulate gene expression via interacting with various nuclear proteins such as TFs in a cell type specific manner.

Incubation of biotin labelled oligonucleotides encompassing each allele of all seven prioritised *EPHA1* variants with nuclear extract from the monocyte line THP1 identified two SNPs (rs11763230 and rs11765305) which were capable of binding nuclear proteins in an allele specific manner. Quantitation of DNA-protein interactions formed showed the risk T allele of rs11763230 and the protective G allele of rs11765305 where capable of binding an increased amount of nuclear protein when compared to their alternative C alleles.

As noted in table 4.3 the SNP rs11763230 is not located within a region containing a known TF binding site. Therefore, the DNA-protein interaction formed between the oligonucleotide containing its risk T allele and a nuclear protein from both THP1 and SHSY5Y cells is a potentially novel interaction with an unknown TF or alternatively an artefact of the EMSA or non-specific binding. This is also the case for the protective C allele and a nuclear protein within HMC3 cells.

Chapter 4: Analysis of the regulatory ability of non-coding SNPs within the *EPHA1* risk loci

The *EPHA1* SNP rs11765305, on the other hand is located within a region of known binding sites for CCAAT-enhancer binding protein (*CEBP*) family members, including, *CEBPB, CEBPE* and *CEBPG*. It is therefore plausible that one of these TFs is responsible for the DNA-protein interaction observed during the EMSA (Fig. 4.5). These data are consistent with published data by Amlie-Wolf et al, who demonstrated via a luciferase expression assay that the protective G allele of rs11765305 is capable of creating a stronger binding site for the TF CEBPB (Amlie-Wolf et al. 2019). This enhancer binding protein has been identified to play important roles in regulating genes involved in immune and inflammatory responses, such as during CD4[+] T cell activation as well as genes involved in embryogenesis (Ruffell et al. 2009; Begay et al. 2018; Zhou et al. 2019). Amlie-Wolf et al concluded that rs11765305 enhances monocyte activity in an unidentified way, with its G allele increasing activity of an enhancer regulating gene within this genomic region.

When considering brain cells, three *EPHA1* SNPs (rs7810606, rs11767557 and rs11763230) were observed to potentially be able to form allele specific DNA-protein interactions with nuclear proteins from the microglial cell line, HMC3. The risk T allele of rs7810606, and the protective C alleles rs11767557 and rs11763230 were able to bind increased amounts of nuclear protein when compared to their alternate alleles (Fig 4.6, 4.7 and 4.8 respectively). In addition, the risk T allele of the SNP rs11763230 was capable to form an allele-specific DNA-protein interaction with nuclear proteins from the neuronal cell line, SHSY5Y. Therefore, within our analysis, this SNP demonstrates cell-type specific binding patterns, with the risk T allele conferring increased nuclear protein binding within monocytes (Fig 4.4) and neurons (Fig. 4.9) compared to the protective C allele within microglia (Fig 4.8). This demonstrates the potential for the alternative alleles to function differently, ultimately leading to differing gene expression changes within different cell types. This may explain the alterations in genetic susceptibility to AD seen within individuals based on which variant allele they carry.

As mentioned above, the risk T allele of the SNP rs7810606 is potentially capable of forming two allele specific interactions with microglial nuclear proteins (Fig 4.5). The location of this SNP resides within the binding motifs of multiple known TFs, such as those of the Krüppel-

like family (KLF) and specificity proteins (SP) (Tab. 4.3). The KLF/SP TFs found to bind at this site play key roles during processes such as erythrocyte maturation and differentiation, cell proliferation and immune suppression (Cao et al. 2009; McConnell and Yang 2010; Siatecka and Bieker 2011). Therefore, this leads to the possibility that this SNP functions in a similar way to that of rs11765305, where one allele is capable of forming a stronger binding interaction with TFs at this location. Potentially leading to allele-specific alterations in gene expression involving the immune response of microglial cells, leading to its AD-association.

Variant rs11763230 as mentioned previously is not located within regions containing any known TF binding sites. Therefore, as mentioned above this DNA-protein interaction has the potential to be a consequence of an interaction with an unknown monocyte nuclear protein or transcription factor or a result of non-specific binding during the EMSA.

Although informative, the use of databases such as JASPAR, a database of transcription factor binding profiles and ENCODE to investigate which *EPHA1* SNPs reside within known TF binding sites has limitations which may potentially result in false positive and negatives. These databases relay on the previous characterisation of a particular TF binding site, therefore, this approach may exclude SNPs residing within novel sites. Additionally, a SNP identified to be located within a TF binding site may simply be a region of homology and not the true binding site. Due to this it may be important to pair this database derived information with that produced from experimental techniques such as the EMS assay employed here or with an approach such as in vivo footprinting. This technique utilises ligation-mediated PCR to capture flanking regions of genomic DNA that are protected from Dnas1 cleavage by bound protein. In vivo footprinting is a beneficial technique allowing the investigation into DNA-protein binding within a living cell, allowing the tissue and condition dependent nature of binding events to also be investigated (Grange et al. 1997; Elnitski et al. 2006; Gorsche et al. 2014).

4.3.3 Limitations and outcomes of the Electrophoretic Mobility-Shift Assay.

The EMSA however, despite being a relatively quick and inexpensive assay is an artificial system with relatively short synthetic oligonucleotides, devoid of post translational modification or histone induced chromatin structure which influences DNA accessibility. Therefore, it cannot fully represent *in vivo* DNA binding to transcriptional complexes and regulatory proteins. That said, the use of the EMSA enables quick and simple identification of potentially functional variants from a relatively large list of SNPs of interest at a given locus. For example, the current data on which *EPHA1* variant is responsible for the AD association at this locus and via which gene it asserts its affects is often conflicting. The EMSA data lends weight to current GWAS and fine-mapping data implicating SNPs rs11767557, rs7810606 and rs11765305 as potential causal variants responsible for the AD-association at this locus.

Due to the nature of EMSA blots however, some caution is needed when interpreting these results. For example, some DNA-protein interactions were very clear, producing strong bands, noticeably reduced within the specificity control (Fig. 4.9) compared to other blots such as that seen within figure 4.4 which have fainter bands potentially resulting from non-specific binding between oligonucleotides and other proteins present within the lysate or an artefact produced due to variations during gel loading. A potential method to combat such variability could be the use of a multi-channel pipette during the loading procedure to minimise human error. A further optimisation step in order to produce clearer EMSA blots is the alteration of salt concentrations within the binding buffer. A salt concentration that is too high is a common cause of smeared or streaked bands during an EMSA. Reducing the salt concentration of the binding buffer will increase the electrostatic stabilisation of the DNA-protein complexes enabling better visualisation (Cozzolino et al. 2021). Therefore, optimising the KCL concentration within the EMSA binding buffer (10X Binding Buffer is 100mM Tris, 500mM KCl, 10mM DTT) may enable more accurate data interpretation due to producing clearer bands with less background noise. Additionally, the inclusion of further specificity controls would enable increased confidence in EMSA data interpretation. For

example, the of addition of excess alternate allele oligonucleotide as a competitor to disrupt the DNA-protein binding of interest as opposed to excess unlabelled oligonucleotide of the same allele would indicate if alterations in protein binding were a consequence of that specific base pair change. The inclusion of a scrambled oligonucleotide would also assess whether protein binding was specifically due to the variant in question or simple non-specific binding of DNA.

4.4 Conclusion

Interpreting genetic data from AD-association GWAS data sets and its conversion into functional biological information to allow the better understanding of signalling pathways and gene expression alterations is extremely complex due to multiple factors as previously discussed.

This chapter aimed to prioritise the multiple variants within the *EPHA1* risk locus in order to identify the causal SNP responsible for the AD-association. From the list of ten lead *EPHA1* SNPs, identified using GWAS p-value, published fine-mapping analysis and LD structure, seven SNPs were prioritised based on proximity to regions of open chromatin within cell types of interest (monocyte, microglia and neuron).

To further investigate an individual SNPs' potential to control gene expression, the ability to bind nuclear proteins in an allele specific manner from AD relevant cell lines was analysed by an EMSA. This concluded that four SNPs (rs11765305, rs11763230, rs11767557 and rs7810606) were able to bind nuclear proteins, within at least one of the cell types of interest. This ability to bind nuclear proteins during the EMSA hints at these SNPs being capable of influencing gene expression. These DNA-protein interactions appear to be cell type specific with all SNPs (with the exception of rs11763230) being able to form allele-specific DNA-protein interactions within one cell type. This correlates with evidence published on rs11675305 stating the cell type specific nature of this variants action, regulating *EPHA1* and *EPHA1-AS1* gene expression only within human monocytes (Amlie-

Wolf et al. 2019). Although the SNP rs11763230 was capable of forming DNA-protein interactions within all cell types analysed, the cell type specific nature of this SNPs action is conferred by the variants alleles, resulting in the alternate alleles conferring increased nuclear protein binding dependent on cell type. Therefore, not only are the *EPHA1* SNPs likely to alter gene expression in certain cell types but alternate alleles may influence gene expression changes differently within different cell types.

The SNPs rs11765305 and rs7810606 have been noted to reside within binding motifs of known TFs, therefore the potential signalling pathways may be hypothesised. However, despite possessing the ability to form DNA-protein interactions with nuclear proteins the SNP rs11763230 has no known TF binding sites overlapping its location. Identifying the proteins involved within these interactions via the use of mass spectrometry and an EMSA super shift would be of great interest due to the novelty of this DNA-protein interaction at this non-coding region.

The regulatory ability of the four SNPs effectively triaged within this chapter will be further analysed through the use of CRISPR/Cas9 technology to delete regions of non-coding DNA surrounding these variants. Alterations in gene expression of key genes of interest (*EPHA1*, *EPHA1-AS1* and *ZYX*) as a consequence of this DNA deletion will be quantitated using western blot and qRTPCR.

Chapter 5: The role of *EPHA1* SNPs in the regulation of *EPHA1, EPHA1-AS1* and *ZYX* expression patterns.

5.1 Introduction

As introduced in Chapter 1, the current challenge of functional genomics is identifying the regulatory elements which reside in the non-coding genome and their respective target gene/s. This understanding may shed light on how alterations and variations in non-coding DNA impact disease (Stadhouders et al. 2012).

5.1.1 Gene regulation by AD-associated GWAS SNPs

Multiple SNPs within the *EPHA1* locus with genome-wide significance have been associated with changes in *EPHA1* gene expression, colocalising with eQTLs in human whole blood (Rosenthal and Kamboh 2014; Amlie-Wolf et al. 2019). However, their target gene may not necessarily be the nearest gene to the SNP. This is evident in recent literature suggesting that non-coding SNPs within the *EPHA1* locus may impact expression of genes other than *EPHA1*. For example, SNP alleles at this locus such as rs11765305 (G) and rs11767557 (C) have been shown to colocalise with expression changes in the lncRNA, *EPHA1-AS1* within human whole blood (Liu et al. 2018; Amlie-Wolf et al. 2019).

An additional gene, *ZYX*, has been identified by both Schwartzentruber et al (Schwartzentruber et al. 2021) and Kunkle et al (Kunkle et al. 2019a) through the use of fine mapping analysis. Multiple *EPHA1* locus SNPs were identified to colocalise with eQTLs for *ZYX* expression changes within both human monocyte and microglial data sets.

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

5.1.2 Biology and function of EPHA1-AS1

The exact function of the *EPHA1* antisense lncRNA *EPHA1-AS1* is not fully understood. However, as a lncRNA it is proposed to have the ability to modulate gene expression either via binding additional regulatory proteins or target DNA sequences directly (Guttman et al. 2009).

Its proposed function can instead be computationally predicted based on available functional data from other genes which have correlating expression patterns. For example, the long non-coding DNA HUB (lncHUB) and the Kyoto Encyclopaedia of genes and genomes (KEGG) databases have been used to predict unknown lncRNA function based on genome-wide co-expression data of other genes. From these data it was hypothesised that *EPHA1-AS1s* function is associated with the immune response in some way, due to being linked to pathways and genes involved in B-cell receptor signalling, acute myeloid leukaemia and hematopoietic cell lineages (Kosoy et al. 2022).

Additionally, the use of RNA sequencing data has linked the potential function of *EPHA1-AS1* to AD pathology through the observed correlating expression of JAK2. JAK2 is part of the JAK2/STAT3 signalling pathway. Inactivation of this pathway by Aβ has been shown to cause memory and cognitive impairment (Chiba et al. 2009). *EPHA1-AS1s* function through this pathway can be seen during the actions of the protective minor allele of the *EPHA1* SNP rs11765305. Increased *EPHA1-AS1* expression as a result of this allele has been observed to correlate with an increase in JAK2/STAT3 signalling, potentially explaining this SNPs protective effect during AD pathology (Amlie-Wolf et al. 2019).

5.1.3 Biology and function of ZYX

The *ZYX* gene encodes for the 82 kDa Zyxin protein (ZYX), which is a Zinc binding phosphoprotein enriched in focal adhesions and associated with actin filaments, cell-cell and cell-extracellular matrix (ECM) adhesion contacts. ZYX is comprised of an N-terminal

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

Proline rich sequence, a nuclear export signal motif and 3 copies of a C-terminal LIM Zinc finger domain (Beckerle 1986).

ZYX has been implicated in multiple cell behaviours, such as cell motility, cell adhesion and stress fibre formation through its function as a mechano-transductor (Crawford et al. 1992). The force-dependent recruitment of ZYX to focal adhesions and cell junctions is crucial for cytoskeletal reorganisation involving the assembly and disassembly of actin fibres in response to mechanical stimuli (Hirata et al. 2008b,a; Smith et al. 2010). ZYX has also been observed to shuttle between focal adhesion sites and the cell nucleus to orchestrate expression changes of stretch-sensitive genes in response to mechanical stimuli, by acting as a transcription factor through its LIM domain (Nix and Beckerle 1997). Prolonged exposure to stretch stimuli, such as those present during hypertension or increased presence of proinflammatory cytokines, such as TNFα as seen during AD pathology, can trigger endothelial dysfunction (Wojtowicz et al. 2010). This is not only a consequence of the structural effects of losing ZYX at focal adhesions, but of the increased expression of ZYX-dependent genes, such as those that encode Interleukin-8, intracellular and vascular cell adhesion molecules (ICAM-1 and VCAM-1). Such genes lead to pro-inflammatory responses, as well as influencing processes such as cell apoptosis, proliferation and migration (Wojtowicz et al. 2010; Fang et al. 2019).

ZYX is an important component in tight junctions within the BBB and is critical for barrier integrity, protecting against leukocyte transmigration. This is particularly evident during the induction of meningitis by *Haemophilus influenzae* (*H. Influenzae*) infection. Human brain microvascular endothelial cells (HBMECs) treated with TNFα resulted in a 1.8-fold decrease in ZYX expression. Upon intranasal inoculation of dendritic cell-bound *H. Influenzae*, it was seen that ZYX-deficient mice had a significantly greater mortality rate and migration of the cell-bound *H. Influenzae* when compared to wild type mice. These findings suggest that TNFα produced during the primary infection causes a decrease in ZYX expression in the BBB, resulting in increased transmigration of peripheral blood cells containing *H. Influenzae* organisms, which induce meningitis (Miyazaki et al. 2014).

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

ZYX although primarily localised to focal adhesions, regulating cell adhesion and motility has also been shown to shuttle to the nucleus where it can function as a TF (Degenhardt and Silverstein 2001; Sadot et al. 2001). For example, in cancer cells ZYX has been noted to modulate the signalling of the Homeodomain-interacting protein kinase-2 (HIPK2) and P53 which is responsible for the DNA damage responses, such as, transient growth arrest, replicate senescence or apoptosis (Yee and Vousden 2005). ZYX has also been shown to be a target of Aβ, with Aβ causing a downregulation in ZYX expression. Taken together this data explains one way in which altered ZYX expression may contribute to AD pathology. The down regulation of ZYX effects the stability of HIPK2, also resulting in its downregulation. This has a knock-on effect as shown in figure 5.1, leading to the misfolding of p53 and therefore loss of its transcriptional activity and the mis-regulation of DNA damage responses. Within AD pathology such dysfunction may cause aberrant cell-cycle progression effecting neuronal plasticity (Stanga et al. 2010).
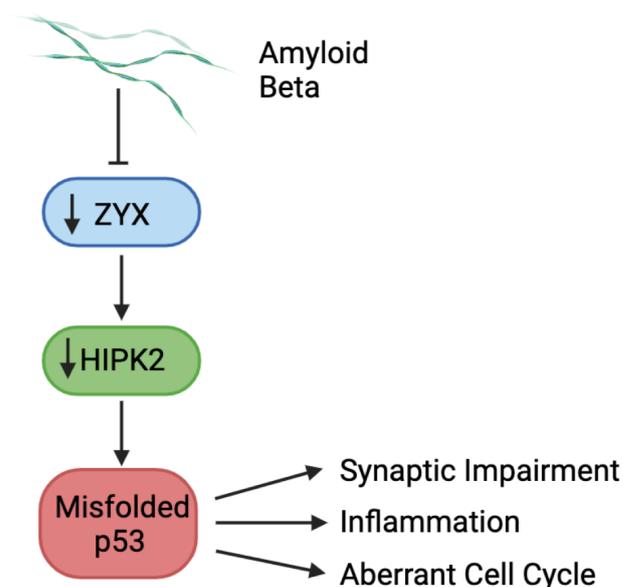
Figure 5.1: Downregulation of a ZYX signalling pathway by Amyloid Beta
Amyloid Beta causes the inhibition of the ZYX protein, this in turn causes downregulation of the protein HIPK2. HIPK2 downregulation causes a misfolding of p53 resulting in its inability to regulate DNA damage responses. There is evidence to suggest that the misfolding and therefore altered activity of the p53 protein contributes to AD pathology, for example, causing synaptic impairment, inflammation and aberrant cell cycle progression. Schematic created using BioRender.com

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX
expression patterns

5.1.4 Induced Pluripotent Stem Cells as a preclinical model of neurodegeneration

Investigations into the pathophysiology of neurodegeneration have used several different
preclinical models, ranging from post-mortem tissues, cellular-based models to mouse
models. Despite allowing valuable understanding into disease pathology in terms of cellular
and molecular alterations, the use of post mortem tissues and immortalised human cell
lines are limited due to their lack of manipulation ability or physiological relevance to the
age-related neurodegeneration of non-dividing brain cells. Additionally, such samples are
taken at the endpoint of disease therefore any brain alterations observed are potentially a
result of the disease and not the cause of pathology (Rowe and Daley 2019; Penney et al.
2020). Therefore, additional models are required to analyse alterations during the
preclinical stages of disease pathology to enable better characterisation of disease aetiology
or development of disease altering treatments.

To this end, genetically altered mice have become invaluable models for the successful
mimicking of age-related neurodegeneration. Despite this, such models rely on the artificial
overexpression of certain human proteins through the employment of common familial
Alzheimer's disease mutations (Oakley et al. 2006). However, mouse models such as this
often fail to fully encapsulate human neurodegenerative phenotypes. This is potentially due
to evolutionary differences between human and mouse protein functions and cellular
processes or fundamental differences in life-span or resilience to processes such as
oxidative stress and build-up of pathological protein aggregates (Seok et al. 2013).

The advent of iPSC technology and the vast array of methodologies to allow their directed
differentiation into any cell type of interest has created superior models in terms of
physiological relevance and ease of genetic manipulation for a range of neurodegenerative
diseases. Furthermore, the use of co-cultures or even three-dimensional cultures
incorporating multiple different cell types into one model allows the analysis of the complex
cell interactions occurring *in vivo* (Rowe and Daley 2019).

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

The reprogramming of patient derived cells into iPSCs circumvents the limitations of embryonic stem cells, allowing the modelling of physiologically relevant AD-related phenotypes. Such models have been shown to exhibit age-dependent accumulation of hyperphosphorylated Tau and Aβ aggregates and are indispensable for personalised medicine and investigation into AD polygenic risk (Ochalek et al. 2017). However, the genetic diversity arising from such models is often a hindrance when trying to understand the function of a specific gene or mutation's function and its involvement in disease pathology. To enable more controlled investigations of this manner, the ability to genetically manipulate iPSCs through the use of gene editing technologies, such as CRISPR-Cas9 technology, is revolutionary. Enabling the creation of targeted mutations within a consistent genetic background.

Different iPS cell lines however, have shown variation within phenotypes and differentiation ability. Therefore, the need for a common well-characterised cell line is required. An in-depth study looking into cell line characteristics such as morphology, growth-rate, genome integrity as well as gene editing and differentiation ability enabled the selection of a leading cell line, KOLF2.1J. This cell line, derived from the KOLF2-C1 cell line which is publicly available from the human induced-pluripotent stem cell initiative (HipSci) bank has under gone gene editing to correct the 19bp deletion within exon three of the ARID2 gene (Pantazis et al. 2022). This is due to whole genome sequencing of the KOLF1-C1 cell line identifying variations within 5 coding genes; *ARID2, COL3A1, ZNF398, UBE3C* and *CDC37* when compared to the parental KOLF2 cell line. Two of which (*ARID2* and *COL3A1*) have the potential for clinical significance. The widespread use of this well-preforming cell line would enable standardisation across the stem cell field to enable accurate data comparison and replication (Skarnes et al. 2019).

Therefore, although neurodegeneration models still remain far from perfect, iPSC-based models provide multiple advantages over previous models and may be pivotal in the understanding of AD pathophysiology (Rowe and Daley 2019; Penney et al. 2020).
The iPSC line employed during this thesis, was KOLF2-C1. Despite known gene variations, this cell line has been observed to have a stable karyotype for at least 25 passages within

cell culture (Skarnes et al. 2019). Macrophages derived from this cell line produced a morphology and gene expression profile similar to that seen within yolk-sac derived foetal macrophages allowing them to act as a good cell model for macrophage biology (Alasoo et al. 2015; Yeung et al. 2017).

5.1.5 CRISPR-Cas9 gene editing technology

Clustered regulatory interspersed short palindromic repeats (CRISPRs) were originally discovered in *E. coli* by Y.Ishino et al in 1987 (Ishino et al. 1987). Their function as part of the bacterial adaptive and inherited immune defence against invading foreign DNA through the use of targeted DNA cleavage was not fully identified until the early 2000s by Mojica and others (Bolotin et al. 2005; Mojica et al. 2005; Pourcel et al. 2005) Following this discovery, the CRISPR-Cas9 system first identified in *Streptococcus pyogenes* (spCas9) (Sapranauskas et al. 2011), has been utilised as a genome editing technology (Jiang et al. 2013). As shown in figure 5.2, CRISPR-Cas9 technology employs 20bp single-stranded RNA guides, complementary to the desired target DNA site. These guides are duplexed to a trans-activating crRNA (tracrRNA) scaffold which acts as a handle for the Cas9 protein (Scott et al. 2019). Such tracrRNAs can also be tagged with a fluorescent label to allow cell sorting of CRISPR-Cas9 edited cells. The RNA guides must also contain a three nucleotide protospacer adjacent motif (PAM) of an NGG sequence to enable a double stranded DNA cleavage event facilitated by the Cas9 nuclease (Bolotin et al. 2005; Sternberg et al. 2014).

The double strand DNA break formed is repaired in one of two ways, homology-directed repair or non-homologous end joining. Homology-directed repair (HDR) is a high-fidelity repair process, reconstructing the DNA break via the use of an exogenous DNA template. Repair via this process allows the insertion of precise site-specific mutations of interest. Alternatively, if the double-strand break is repaired via the more error prone non-homologous end joining (NHEJ) pathway, larger insertions or deletions are incorporated. Repair via this pathway allows the generation of gene insertions and deletions via either generating frameshift mutations causing the insertion of premature stop codons leading to their nonsense mediated decay (Fig. 5.2) (Sternberg et al. 2014).
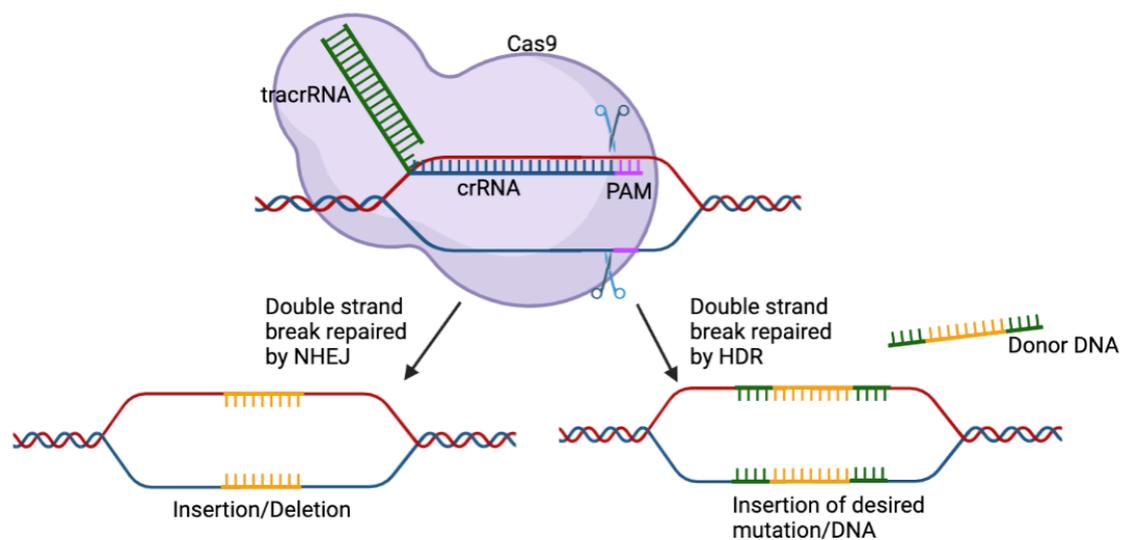
Figure 5.2: CRISPR-Cas9 gene editing system.
Diagram depicting CRISPR-Cas9 gene editing through targeting of the desired genomic DNA sequence via the crRNA CRISPR guide. The tracrRNA duplexed to the crRNA provides a handle of the cas9 nuclease allowing the production of a double strand break at the recognised PAM site. The double strand DNA break can then be repaired via one of two routes, NHEJ, creating insertions and deletions or HDR to allow insertion of a desired mutation or DNA sequence. Schematic created using BioRender.com

This RNA-guided CRISPR-Cas9 technology has multiple advantages over older gene editing tools employing TALE or Zinc finger nucleases. This is due to CRISPR-Cas9 technology allowing the targeting of different DNA sites of interest through simple RNA guide sequence alterations. The less labour-intensive methodology employed during CRISPR-Cas9 genome editing allows large scale genome editing (Becker 2021).

5.1.6 Hypothesis

It is hypothesised that the non-coding genome wide associated SNPs at the *EPHA1* locus may influence the pathology of AD through altering the expression of genes of interest which have been identified within the literature such as, *EPHA1*, *EPHA1-AS* and *ZYX* (Kunkle et al. 2019a; Schwartzentruber et al. 2021).

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

5.1.7 Overall Aim

The aim of this chapter is to understand the expression pattern of these genes of interest and identify those which may be influenced by the AD-associated SNPs at the *EPHA1* locus.

5.1.8 Objectives

1. Analyse single cell RNA sequencing data to determine the expression patterns of *EPHA1* and other potential genes of interest such as *EPHA1-AS1* and *ZYX* within different human immune and brain cell types.

2. Quantitate RNA levels of genes of interest in undifferentiated iPSC, iPSC-derived monocyte-like cells, microglia and cortical neurons via real time qRTPCR.

3. Analyse protein levels of genes of interest in undifferentiated iPSC, iPSC-derived monocyte-like cells, microglia and cortical neurons via western blot.

4. Analyse gene expression profiles following the deletion of the *EPHA1* SNPs of interest identified within Chapter 4.

5.2 Materials and Methods

5.2.1 Induced pluripotent stem cell culture

Cells were maintained in mTeSR^{TM} 1 (Stem cell Technologies) cell culture media. The cells were grown in a humidified incubator at 37°C in the presence of 5% CO_2 on Geltrex (1/100 dilution, ThermoFisher Scientific) coated culture vessels. Cells were fed every day with fresh mTeSR^{TM} 1, pre-warmed to 37°C.

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

Once cell cultures reached 70–80% confluency, they were passaged. Spent media was removed and cell cultures washed with 1-5 mL PBS (depending on vessel size). Cells were removed from the tissue culture vessels via incubation with ReLeSR™ (Stem Cell Technologies) for 3 min at 37°C. ReLeSR™ was removed and cells incubated for a further minute. Cells were then washed off the culture vessel with an appropriate volume of prewarmed mTeSR™ 1 and centrifuged at 300 g for 5 min. The supernatant was removed and the cell pellet was resuspended in an appropriate volume of pre-warmed fresh mTeSR™ 1 before being dispensed into fresh culture vessels pre-coated in Geltrex, at a dilution of 1:4-1:6.

5.2.2 Thawing and cryopreservation of induced pluripotent stem cells

Cells were stored in liquid nitrogen for long term storage. Cells were removed from liquid nitrogen and thawed rapidly in a 37°C water bath until a small ice crystal remained. Pre-warmed fresh mTeSR™ 1 media was added (5 mL) in a dropwise fashion and the suspension was centrifuged at 300 g for 5 min. The cell pellet was re-suspended in 1 mL of mTeSR™ 1 and seeded into 1 well of a 6-well plate (Greiner bio-one) which had been precoated with Geltrex. A complete media change was performed 24 h post thaw and cell cultures maintained as per method 5.2.1.

Cells were cryopreserved from 1 well of a 6-well plate when cultures reached 80% confluent. Cells were removed from culture vessel following the above-mentioned passaging procedure (section 5.2.1). The cell pellet was re-suspended in 2 mL cryopreservation media (mTeSR™ 1 containing 10% DMSO) and 1 mL aliquoted per cryovial. Cryovials were placed in a Mr Frosty freezing container, which provides 1°C/min cooling rate required for successful cryopreservation of cells. Cells were held at -80°C for 24 h before being placed in liquid nitrogen for long term storage.

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

5.2.3 iPSC-derived microglial cell differentiation

Media compositions used throughout the differentiation are detailed in table 5.1. Desired amount of iPSCs were cultured as per methods section 5.2.1 until 80% confluency was reached. Spent media was removed and cell cultures washed with 1 mL PBS. Cells were removed from the tissue culture vessels via incubation with Accutase (Gibco) for 3 min at 37°C and mechanically dissociated. The Accutase was neutralised via the addition of 1 mL prewarmed mTeSR^{TM} 1. Cell suspension was centrifuged at 300 g for 3 min and supernatant removed. A trypan blue cell count was preformed using a haemocytometer. A cell suspension of 1x10^{6} cells in 10 mL prewarmed Induction Media was created. This cell suspension was seeded into a round bottom 96 well plate (100 mL per well). The plate was then centrifuged for 3 min at 140 g in order to form cell aggregates at the bottom of the well.

Cells were incubated within this plate for 4 days with a 50% media change with prewarmed Induction Media performed on day 2. On day 4, cell aggregates/embryoid bodies were carefully removed from the 96 well plate using a 5 mL stripette and separated into wells of a 6 well plate (8 aggregates per well) in prewarmed Progenitor Differentiation Media.

Cultures were fed with warmed Progenitor Differentiation Media once a week. After 15-20 days, monocyte-like/microglial progenitor cells were produced from the embryoid bodies and released into the supernatant. The embryoid bodies were maintained in a humidified incubator at 37°C in the presence of 5% CO_2, carefully collecting spent media and replacing with fresh Progenitor Differentiation Media. Spent media was centrifuged at 300 g for 3 min to allow for the collection of microglial progenitor cells for maturation or use in assays.

For maturation, progenitor cells were plated at 10,000 cells per well of a 96 well plate in Maturation Media. Cells were cultured in a humidified incubator at 37 $^{\circ}$C in the presence of 5% CO_2 and fed once a week for 12 days with Maturation Media.

Table 5.1 Media composition for iPSC-derived microglial differentiation

Media	Basal Media	Supplements
Induction Media	mTeSR™ 1	Y27632 (10 mM) SCF (20 ng/mL) BMP4 (50 ng/mL) VEGF (50 ng/mL)
Progenitor Differentiation Media	XVIVO15	M-CSF (100 ng/mL) IL-3 (25 ng/mL) Beta Mercaptoethanol (50 nM) Glutamine (2 mM)
Maturation Media	DMEM/F12	N2 Supplement (1X) IL-34 (100 ng/mL) GM-CSF (10 ng/mL) Glutamine (2 mM) Beta Mercaptoethanol (50 nM)

5.2.4 Characterisation of iPSC-derived microglial progenitor cells via flow cytometry

Media containing non-adherent microglial progenitor cells were collected following the above-mentioned protocol. The cell pellet was washed via resuspension in 1 mL of PBS and centrifuged for 3 min at 300 g. The cell pellet was resuspended in 800 mL of 0.1% Bovine Serum Albumin (BSA) in PBS and split into four Eppendorf tubes. To one of the tubes, both the anti-CD11b and anti-CD45 antibodies were added. To the second, both the anti-CD14 and anti-CD34 antibodies were added. The third sample contained an isotype control for each fluorophore used and the final sample was left unstained. All Antibodies used are detailed in table 5.2.

Table 5.2: Flow cytometry antibodies for microglial progenitor characterisation.

Antibody	Species/Clonality	Dilution	Source	Catalogue No.
APC Anti-human CD11b	Mouse IgG1 monoclonal	5 ml Per test	Biolegend	301309
FITC Anti-human CD45	Mouse IgG1 monoclonal	5 ml Per test	Biolegend	304005
APC Anti-human CD14	Mouse IgG1 monoclonal	5 ml Per test	Biolegend	367117
PE/Cyanine7 Anti-human CD34	Mouse IgG2a monoclonal	5 ml Per test	Biolegend	343615
PE/Cyanine7 Isotype Control	Mouse IgG2a	1 mg	ThermoFisher Scientific	25-4724-81
FITC Isotype Control	Mouse IgG1	2.5 mg	ThermoFisher Scientific	31505
APC Isotype Control	Mouse IgG1	1 mg	ThermoFisher Scientific	MA5-18093

Samples were then incubated for 1 hr in the dark prior to centrifugation for 3 min at 300 g. The cell pellet was washed three times via centrifugation at 300 g for 3 min and resuspension in 100 mL 0.1% BSA in PBS, with a final resuspension volume of 200 mL 0.1% BSA in PBS. Samples were transferred to 5 mL round bottom FACS tubes (BD Falcon) and analysed using a BD Bioscience FACS Canto machine and FlowJo (Version 10).

An example of microglial progenitor flow cytometry data derived from the parental KOLF2-C1 cell line is shown in figure 5.3. The isotype control antibodies for each of the fluorophores show no staining (Fig. 5.3, A and B). The KOLF2-C1 cells demonstrate high expression of the hematopoietic cell marker, CD45 (98.8%), and the monocyte/macrophage markers CD11b (64.518%) and CD14 (49.01%) (Fig. 5.3, C and D), with 64.5% of cells co-expressing both CD11b and CD45. The hematopoietic stem cell marker CD34 was not detected (1.969%) (Fig 5.3, D).
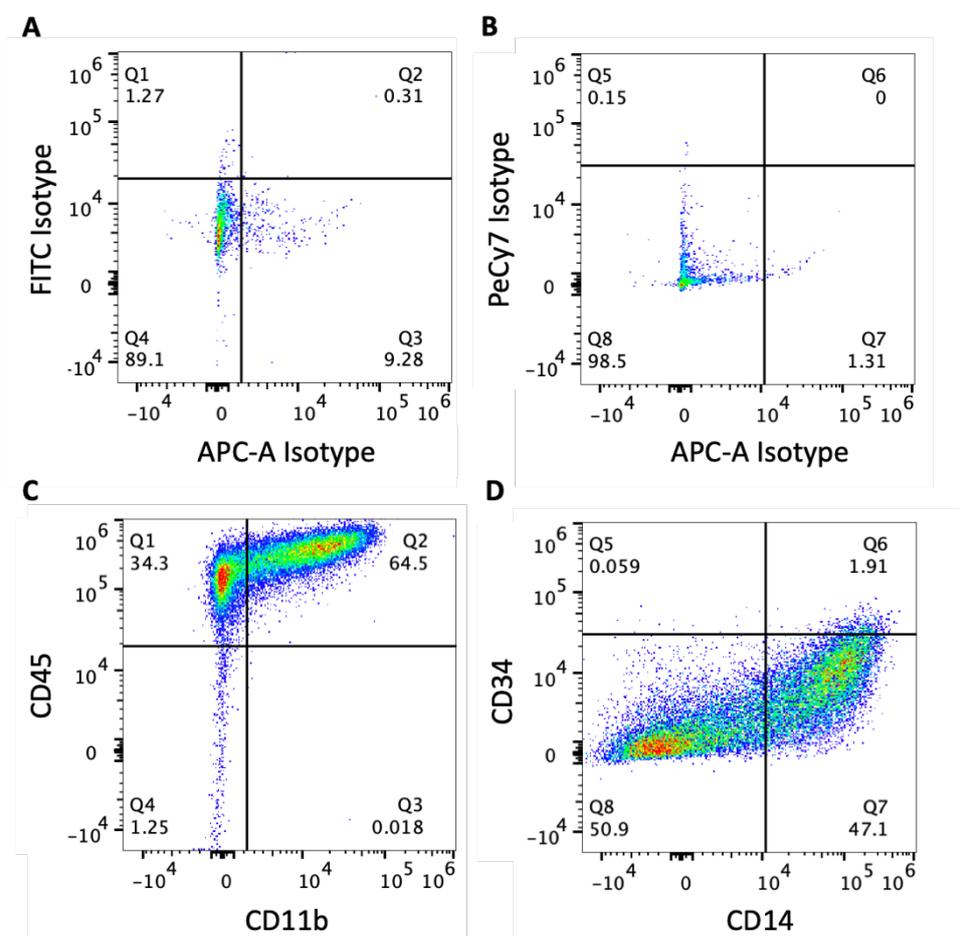
Figure 5.3: Microglial progenitor characterisation via flow cytometry.
Expression of monocyte/macrophage markers in control KOLF2-C1 cell line-derived
microglial progenitor cells **A.** FITC and APC-A Isotype control staining. **B.** PeCy7 and APC-A
isotype control staining. **C.** CD45 and CD11b expression. **D.** CD34 and CD14 expression.

5.2.5 Characterisation of iPSC-derived microglial cells via immunocytochemistry

Non-adherent microglial-progenitor cells were plated at a concentration of 300,000 cells per
cm^2 within a black, clear bottom 96 well plate (Nunc) and matured following above-
mentioned method (section 5.2.3). Matured microglial cells were then fixed in 4%
paraformaldehyde for 10 min at RT and washed three times in 1X PBS. The fixed cells were
incubated in blocking buffer (3% goat serum, 0.1% Triton-X in 1X PBS) for 1 hr at RT.
Blocking buffer was removed and cells were incubated overnight at 4°C in desired primary
antibodies. Following overnight incubation, cells were washed three times for 5 min in 1X
PBS. Secondary antibodies were added for 1 hr at RT. Finally, cells were incubated with

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

Hoechst (1:5000 in blocking buffer) for 15 min at RT before images taken using a Leica SP8 confocal microscope. All antibodies used are detailed in table 5.3 and were diluted in blocking buffer.

Table 5.3: Immunocytochemistry antibodies for mature microglia characterisation.

Antibody	Species/Clonality	Dilution	Source	Catalogue No.
Anti-human CX3CR1	Mouse polyclonal, IgG	1/100	Abcam	Ab167571
Anti-human IBA-1	Rabbit monoclonal, IgG	1/100	Abcam	Ab178846
Anti-human TMEM119	Rabbit Polyclonal, IgG	1/100	Abcam	Ab185333
Alexa Fluor 594 goat anti-rabbit IgG	N/A	1/400	Invitrogen	A11037
Alexa Fluor 488 goat anti-mouse IgG	N/A	1/400	Invitrogen	A11029

5.2.6 Design of CRISPR-Cas9 Guides to allow SNP block deletion and ZYX gene knockout

All reagents were purchased from Integrated DNA Technologies (IDT). CRISPR-Cas9 guides were designed using Deskgen CRISPR design tools. On arrival, lyophilised CRISPR guide oligonucleotides (Alt-R CRISPR-Cas9 crRNA, 2nmol) were resuspended in IDTE buffer (10 mM Tris, 0.1 mM EDTA) at a concentration of 200 mM and stored at -20 $^\circ$C.

The CRISPR-Cas9 guides were used to delete three regions of the non-coding sequence within the *EPHA1-AS1* gene containing the prioritised *EPHA1* locus SNPs identified in chapter 4, section 4.3.1. These were labelled SNP Blocks 1, 2 and 3 deletions. The deletion of the genomic region and SNPs residing within each block are detailed in table 5.4. Sequences of all guides used are detailed in table 5.5.

Table 5.4: Prioritised *EPHA1* SNPs within each SNP deletion block regions.

SNP Block deletion number	SNPs within this region	Genomic Region Deleted (GRCh38/hg38)
1	rs12703526 rs11762262 rs7810606	7:143410355 - 7:143411220
2	rs11763230 rs11767557	7:143411621 - 7:143412134
3	rs11765305 rs11771145	7:143413644 - 7:143414753

Table 5.5: Sequence of CRISPR-Cas9 guides

SNP Block deletion	CRISPR-Cas9 guide sequence
SNP Block 1 Right Guide	GTCTGAAAGTCTCTGCATAT
SNP Block 1 Left Guide	CATTTCGGTACTGTGAATGA
SNP Block 2 Right Guide	TTTGGTCCCCTAGGAACCAC
SNP Block 2 Left Guide	AATTGCTGGAGTTCTCCTGC
SNP Block 3 Right Guide	GAAGACACCCGATACTGTTT
SNP Block 3 Left Guide	TATCCTGACTTAAAACACCA
***ZYX* knockout Right Guide**	GTCTCCCGCGATCTCCGTTT
***ZYX* knockout Left Guide**	AGCGCGCACAGATGGGCCGG

An example of guide placement and SNP deleted region are shown in figure 5.4. Screening primers were also designed to enable characterisation of the resulting clones via end-point polymerise chain reaction (PCR) and PCR product sequencing. This allowed the identification of clones containing a homozygous deletion for the desired SNP region.

Figure 5.4 CRISPR-Cas9 guides and screening primer pairs designed to delete and characterise *EPHA1* SNP Block 1 deletion clones.
A schematic depicting the location of the two CRISPR-Cas9 guides (grey) designed within the non-coding region of *EPHA1-AS1* to allow the deletion of SNP Block 1, the respective PAM sequence of each guide (orange) and the genomic location of designed screening primers (purple arrows). Region of non-coding DNA deleted is shown in blue with the *EPHA1* SNPs located within this region shown in green underneath the DNA sequence. Schematic created in SnapGene.

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns
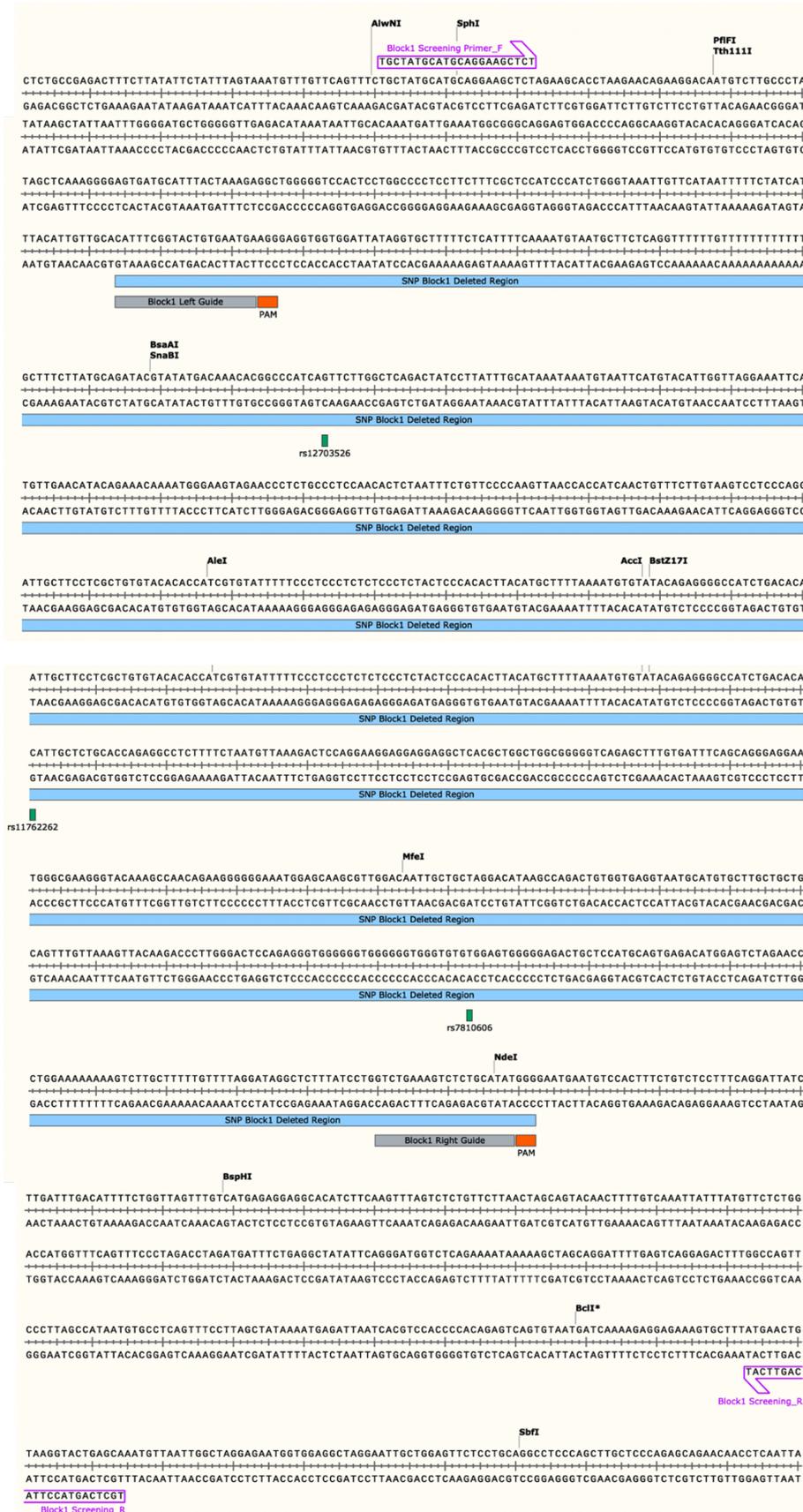
Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

CRISPR-Cas9 guides were also created to allow the knockout of the gene *ZYX*. These guides where designed to create a 160bp deletion within exon two, enabling a frameshift mutation to be produced, resulting in the knockout of the *ZYX* gene through nonsense mediated decay.

5.2.7 Nucleofection of CRISPR-Cas9 machinery into iPS cells

CRISPR-Cas9 was performed by nucleofection of ribonucleoprotein (RNP) CRISPR-Cas9 complexes. These complexes were formed separately for each of the two Alt-R CRISPR-Cas9 crRNA guides (IDT) via incubation of equimolar concentrations of crRNA and ATTO™ 550 labelled Alt-R® CRISPR-Cas9 tracrRNA (IDT) at 95 °C for 2 min. Alt-R® S.p. HiFi Cas9 Nuclease V3 (6.2 mg/mL) (IDT) was then added to each crRNA:tracrRNA complexes at a 1:1 ratio and incubated for 20 min at RT.

Nucleofection of CRISPR-Cas9 components into the control iPS cell line, KOLF2-C1, used the P3 Primary Cell 4D-Nucleofector™ X Kit S (Lonza) and the Amaxa 4D system (Lonza). A single cell suspension of 1 x10^6 iPS cells in mTeSR™ 1 was created following methods detailed in section 5.2.3. The cell suspension was centrifuged and cell pellet resuspended in Nucleofection Solution (78 mL P3 Primary cell Nucleofector™ solution and 22 mL Supplement 1). The cell suspension was transferred into a Nucleocurvette (Lonza) and both crRNA:tracrRNA and Cas9 complexes where added. The Nucleocurvette was placed inside the Amaxa 4D system and nucleofected using the CA137 programme. The cell suspension was then plated into one well of a six well plate pre-coated with Geltrex (1/100) in mTeSR™ 1 containing ROCK Inhibitor, Y27632 (10 mM) (Cell Guidance Systems).

5.2.8 Fluorescence-Activated Cell Sorting of CRISPR-Cas9 edited iPS cells

A complete media change with fresh pre-warmed mTeSR™ 1 without Y27623 was preformed 24 hr post transfection of the CRISPR-Cas9 machinery. After a further 24 hr, the cells were FACS. Cells were dissociated from the tissue culture vessel following methods detailed in section 5.2.3. The cell pellet was re-suspended in mTeSR™ 1 containing

Penicillin-Streptomycin (10,000 U/mL/10,000 mg/mL, Gibco) and Y27632 (10 mM). Cells were FAC sorted using a FACS Aria Fusion based on ATTO™ 550 fluorescent tracrRNA levels. The top 10% of fluorescent cells were re-plated at a concentration of 2000 cells per well of a pre-coated Geltrex (1/100) 6 well plate in mTeSR™ 1 containing Penicillin-Streptomycin, ensuring single cell plating to allow colony picking.

5.2.9 Clonal selection of CRISPR-Cas9 cells

Single cell cultures of iPS cells were cultured for approximately 1 week in mTeSR™ 1 containing Penicillin-Streptomycin or until the single cells had expanded into small but separate colonies to allow clonal picking. Clonal colonies were removed from the culture vessel via gently scraping with a P20 pipette tip to dislodge the colony then transferred gently into separate wells of a Geltrex (1/100) coated 96 well plate (Nunc) containing mTeSR™ 1 and Y27632 (10 mM) using a P200 pipette (45-96 colonies were picked per CRISPR-Cas9 reaction). Clonal cultures were then expanded into two replica 96 well plates for characterisation and cryopreservation following methods detailed on section 5.2.1.

5.2.10 DNA extraction and PCR analysis of CRISPR-Cas9 edited clones

Clonal DNA was extracted from one 96 well plate using QuickExtract™ DNA Extraction Solution (Lucigen). Cells were dissociated into 50 μL mTeSR™ 1 as per methods section 5.2.1, 50 μL QuickExtract™ was then added to each well. The cell suspension was then transferred to a 96 well PCR plate (ThermoFisher Scientific), sealed using adhesive covers (ThermoFisher Scientific) and placed in a thermal cycler at 65 °C for 6 min, followed by an incubation at 98 °C for 2 min. Extracted DNA was stored at -80 °C for long term storage.

The characterisation of CRISPR-Cas9 generated iPSC clones was performed via end point PCR using GoTaq Green Master Mix (Promega). Reaction components and thermal cycler parameters are detailed in tables 5.6 and 5.7 respectively. Screening primer sequence and PCR product sizes (full DNA length between the primers and DNA length on successful region deletion) for each SNP Block deletion and ZYX gene knock out are detailed in table

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

5.8. Dependent on number of homozygous clones identified either 45 or 96 clones were screened. The clones taken forward were chosen based on clarity of PCR band and sequence alignment data.

Table 5.6: PCR components

Component	Volume (µL)	Final concentration
GoTaq Green Master Mix	12.5	1X
Forward Primer	0.5	0.2 mM
Reverse Primer	0.5	0.2 mM
DNA Template	2	<250 ng
Nuclease free water	To 25 µL	N/A

Table 5.7: Thermal cycler parameters

Stage	Temperature (°C)	Time
Initial denaturing	95	2 min
Denaturing	95	30 sec
Annealing	SNP Block 1: 57 SNP Block 2: 59 SNP Block 3: 60 ZYX knockout: 56	30 sec
Extension	72	1 min
Final extension	72	5 min

Table 5.8: Screening primer sequences

Primer	Sequence	Size of full-length DNA product (bp)	Size of DNA product on deletion (bp)
SNP Block 1 Forward	TGCTATGCATGCAGGAAGCTCT	1504	699
SNP Block 1 Reverse	TGCTCAGTACCTTACAGTTCAT		
SNP Block 2 Forward	CGTCCACCCCACAGAGTCAG	699	187
SNP Block 2 Reverse	GCCAGACGCCAGAAGGGGAA		
SNP Block 3 Forward	AGAGATGTTAGCAAACACCT	1244	131
SNP Block 3 Reverse	CAACCGCCTCTATCTCAAAA		
ZYX gene knockout Forward	GGGGTCACCAAGGGGAGCTG	434	314
ZYX gene knockout Reverse	CATCTGCTCGGGACAGGGTG		

PCR products were then visualised along with a 1 Kb plus DNA ladder (NEW ENGLAND BioLabs) via gel electrophoresis on a 1% Agarose gel (Sigma) ran in TEA buffer for 1 hr at 100 V in the presence of SafeView nucleic acid stain (5 mL per 100 mL agarose gel) (NBS Biologicals). This allowed characterisation of the genotype of the CRISPR-Cas9 generated clones based on the number and size of PCR products produced.

5.2.11 Sanger sequencing characterisation of CRISPR-Cas9 edited clones

Following end-point PCR characterisation, selected clones of desired genotype where further characterised via Sanger sequencing to confirm correct deletion of desired region via CRISPR-Cas9 editing. CRISPR-Cas9 control clones were KOLF2-C1 cells which had under gone the CRISPR-Cas9 process but failed to delete the desired DNA regions.

To enable sequencing, a larger PCR was performed by scaling up the reaction components detailed in table 5.6 to a total volume of 50 μL. Following gel electrophoresis, the PCR product bands were excised from the gel using a transilluminator light box and scalpel. The clonal DNA was purified using a Monarch® DNA Gel Extraction Kit (NEW ENGLAND BioLabs) as per manufacturer's instructions. Purified DNA was eluted in 6 μL elution buffer followed by concentration and purity assessed using a nanodrop.

Clonal DNA was then sent for Sanger sequencing by Eurofins genomics. For this, 100 ng of clonal DNA, 2 μL of both forward and reverse screening primers (100 mM) (Table 5.8) was combined within 15 μL water within a 1.5 mL Eppendorf tube. Sequence data obtained was aligned to the reference *EPHA1-AS1* sequence containing using CRISP-ID software.

5.3 Results

5.3.1 Single cell RNA sequencing analysis of EPHA1, EPHA1-AS1 and ZYX expression

Single cell RNA sequencing data from the peripheral blood mononuclear cell (PBMC) 68K data base was retrieved from https://www.10xgenomics.com/resources/datasets and used to analyse the expression levels of *EPHA1*, *EPHA1-AS1* and *ZYX* within multiple different types mononuclear cell in human peripheral blood. The dataset is composed of 68,000 freshly processed peripheral blood mononuclear cells obtained from one donor. These cells were sequenced allowing subdivision into 10 different cell types based on marker gene expression. Gene expression results within each cell type were visualised on a violin plot, as seen in figure 5.5. The proportion of cells expressing each gene of interest as a percentage of the total number of cells within each cell type cluster is shown in table 5.9. Data analysis was conducted by You Zhou and Birong Zhang (Zhou laboratory, Cardiff University).

EPHA1 (Fig. 5.5A) and *EPHA1-AS1* (Fig. 5.5B) messenger RNA (mRNA) expression was only detectable within the natural killer cell cluster (NK cells), CD4[+] and CD8[+] T cells. However, *EPHA1* mRNA was only seen within a very small proportion of these cells, 0.13%, 0.89% and 1.08% respectively (Table. 5.9). Despite being detected in a slightly larger proportion of cells than observed for *EPHA1,* the gene *EPHA1-AS1* was still only expressed by a small subset of cells within these cell types; NK cells=0.23%, CD4[+] T cells=3.0%, CD8[+] T cells=5.01%.

ZYX gene expression on the other hand was detected in most of the cell types analysed with a higher proportion of cells expressing this gene, including slightly higher mRNA levels than those seen for *EPHA1* and *EPHA1-AS1* (Fig. 5.5C). Of particular note, 35.4% of CD16/FcγRIIIA+[+] monocytes, 36% of conventional dendritic cells and 62.6% of platelets and expressed *ZYX* mRNA (Table. 5.9). Although CD14[+] monocytes expressed *ZYX* at the highest level (Fig. 5.5C), only a small subset of these cells (10.5%) express the gene (Table. 5.9).
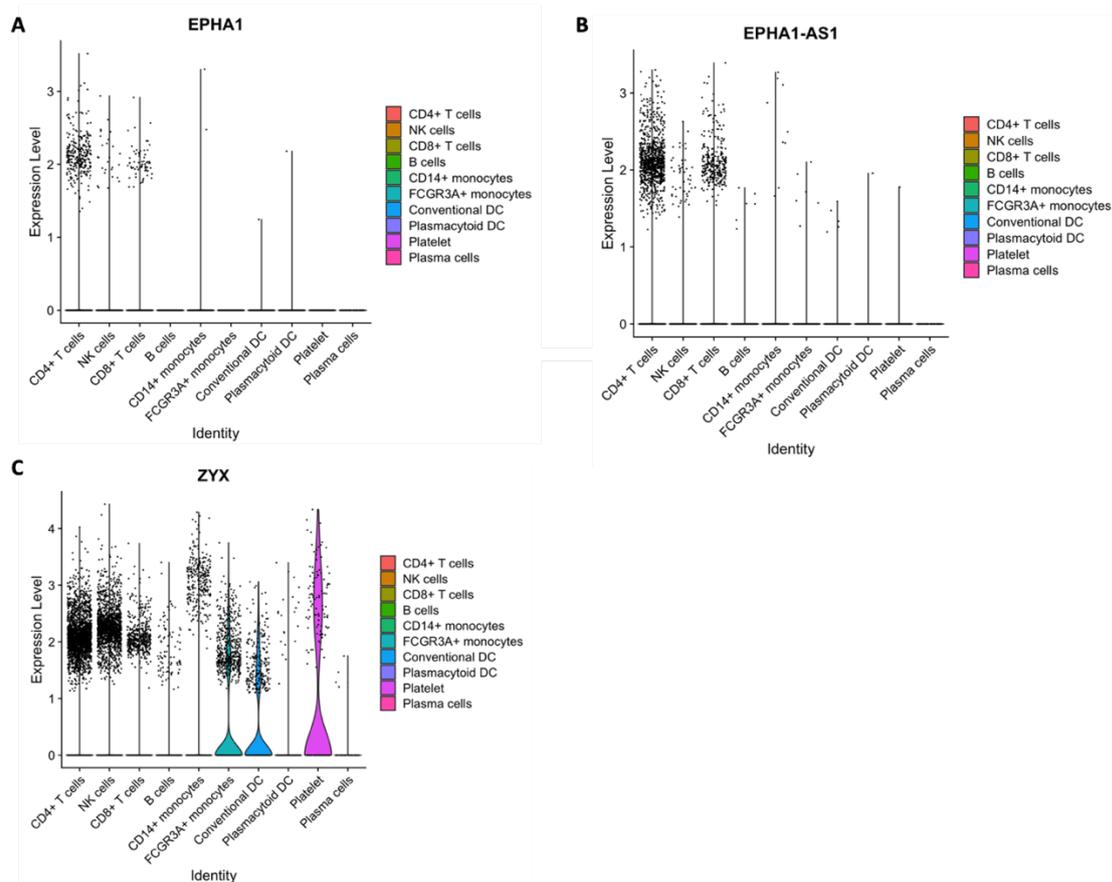
Figure 5.5: Violin plots showing single cell RNA sequencing data of *EPHA1, EPHA1-AS1* and *ZYX* expression within different cell types.

Expression levels of each gene of interest, *EPHA1, EPHA-AS1* and *ZYX* from single cell RNA sequencing data were plotted on violin plots. The total sequenced 68,000 cells were subdivided into 10 different cell types based on gene marker expression, plotted along the x-axis. The y-axis represents log10(raw counts +1). **A.** *EPHA1* mRNA expression levels within cell types analysed. **B.** *EPHA1-AS1* mRNA expression within cell types analysed. **C.** *ZYX* mRNA expression within cell types analysed. Data analysis conducted by You Zhou and Birong Zhang (Zhou laboratory, Cardiff University).

Table 5.9: Proportion of cells from each cell type expressing *EPHA1, EPHA1-AS1* or *ZYX*.

Cell Type	*EPHA1* (%)	*EPHA1-AS1* (%)	*ZYX* (%)
CD4[+] T cells	0.89	3.0	7.21
Natural killer cells	0.13	0.23	7.67
CD8[+] T cells	1.08	5.01	7.67
B cells	0	0.15	2.5
CD14[+] Monocytes	0.06	0.3	10.5
FCGR3A[+] Monocytes	0	0.38	35.4
Conventional Dendritic cells	0.1	0.62	36
Plasmacytoid Dendritic cells	0.27	0.27	6.4
Platelet cells	0	0.38	62.6
Plasma	0	0	5.6

The expression patterns of the genes *EPHA1, EPHA1-AS1, ZYX* and the ephrinA1 ligand (*EFNA1*) where further analysed using the EMBL-EBI data base, The Expression Atlas of Gene expression across species and biological conditions. Figure 5.6A details basal RNA-Seq data from 53 human tissue samples taken from the genotype-tissue expression (GTEx) project dataset within the Human protein atlas, allowing the comparison of gene expression across multiple different body regions.

EPHA1-AS1 was observed to have a very selective and low expression pattern, with expression only noted within 9 out of the 27 tissue regions analysed with levels remaining under 1 transcript per million (TPM). *EPHA1* on the other hand shows expression in almost all tissue regions analysed. However, most of this expression remains low (<5 TPM) with only the lower leg, suprapubic skin and vagina showing slightly higher expression as 35, 33 and 54 TPM respectively. The oesophagus mucosa displays the highest level of expression at 86 TPM. Expression of *EPHA1's* ligand ephrinA1 (*EFNA1*) was observed within all tissues analysed at fairly consistent levels of 15-60 TPM. The highest expression is noted within tissues such as the liver, lung and prostate gland (241, 161 and 127 TPM). Expression of *ZYX* was noted in all tissues at levels higher than the other genes analysed. For example, within

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

the blood and coronary artery (700 and 467 TPM respectively) and of particular note within brain regions such as the cerebellum, 136 TPM and cerebral cortex, 102 TPM.

Figure 5.6B, C and D depict more in-depth Uniform Manifold Approximation and Projection plots (UMAPs) detailing gene expression of *EPHA1, ZYX* and *EFNA1* within vascular cells.

Figure 5.6B shows the highest expression of *EPHA1* within a small proportion of T-cells (9 TPM), with low to no expression within other cell types (0-0.4 TPM). Correlating with the GTEx data, *ZYX* is expressed at high levels by a large proportion of cells within each cell type, with highest expression observed within the c-8 smooth muscle cells cluster (249.2 TMP) (Fig. 5.6C). Figure 5.6D shows a proportion of all vascular cell types analysed express *EFNA1* at varying levels, with the highest level of expression noted from cell within all the endothelial cell clusters, c-19 (138.8 TPM), c-16 (81.6 TPM), c-15 (69.8 TPM) and c-2 (76.4 TPM).

Figure 5.6E, F and G depict UMAPs of gene expression of the genes *EPHA1, ZYX* and *EFNA1* respectively from cell types within the brain. *EPHA1* and *EFNA1* are both expressed at low levels by only a small proportion of cells within each cell type. The highest of which is noted within excitatory neurons (c-36 and c-29 clusters, 6.7 TPM and 4.3 TPM respectively) and oligodendrocytes (c-12 (28.3 TPM) and c-37 (21.8 TPM) clusters) respectively (Fig. 5.6E and G). On the other hand, high expression of ZYX by a large proportion of cells within each cell type is observed (an average of 29 TPM) (Fig 5.6F).

Figure 5.6: Gene expression analysis of *EPHA1-AS1, EPHA1, ZYX* and *EFNA1* within different tissues and cell types.

A. Heat map detailing basal gene expression of *EPHA1-AS1, EPHA1, ZYX* and *EFNA1* within different tissue regions. **B.** UMAP of *EPHA1* gene expression within vascular cells. **C.** UMAP of *ZYX* gene expression within vascular cells. **D.** UMAP of *EFNA1* gene expression within vascular cells. **E.** UMAP of *EPHA1* gene expression brain cells. **F.** UMAP of *ZYX* gene expression brain cells. **G.** UMAP of *EFNA1* gene expression brain cells.

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

5.3.2 Analysis of EphA1 mRNA and protein levels

Following on from the single cell RNA sequencing data, mRNA and protein levels of EphA1 were analysed within cell lines of interest, such as undifferentiated KOLF2-C1 iPSCs and KOLF2-C1 iPSC-derived monocyte-like cells, microglia and cortical neurons. DNA and protein from KOLF2-C1-derived cortical neurons used throughout gene expression characterisation experiments were supplied by Dr Kimberley Marie Jones (Cardiff University).

EPHA1 mRNA transcript levels were measured by qRTPCR. Transcript expression fold changes were calculated using the 2-ΔΔCq method (as detailed in materials and methods chapter 2, section 2.7) with the EphA1-transfected HEK293 cell line used as the control for normalization. As shown in figure 5.7, *EPHA1* mRNA was not detectable in undifferentiated KOLF2-C1 iPSCs or any of the KOLF2-C1 iPSC-derived cell types of interest.

Levels of EphA1 protein expression were analysed via western blotting using an anti-EphA1 monoclonal antibody targeting the N-terminus of the protein (Fig 5.7B, black arrow head). The transfected EphA1 HEK293 cell line was used as a positive control, with alpha-tubulin staining as a loading control for overall protein levels (Fig. 5.7B, blue arrow head). As seen in figure 5.7B, no EphA1 protein expression was detected within undifferentiated KOLF2-C1 iPSCs or KOLF2-C1 iPSC-derived monocytes, microglia or cortical neurons (black arrow head).
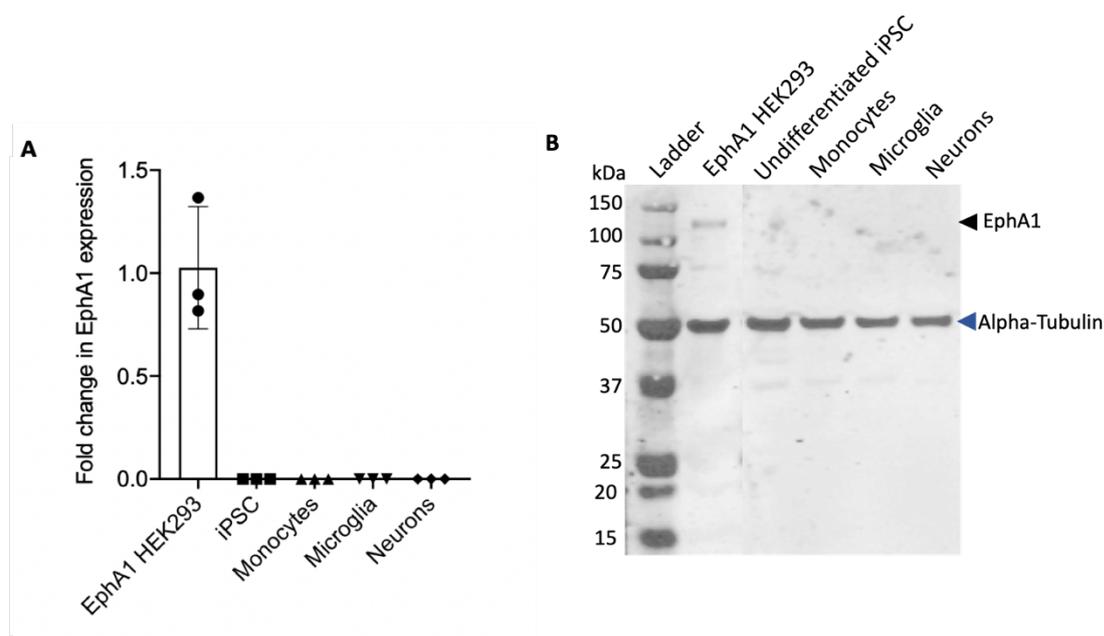
Figure 5.7: Quantitation of *EPHA1* expression in EphA1 HEK293 cells and KOLF2-C1 iPSC-derived cell lines.

A. *EPHA1* mRNA expression levels within undifferentiated KOLF2-C1 iPSCs and KOLF2-C1 iPSC-derived cell lines quantitated by qRTPCR. Fold changes are shown relative to the EphA1 transfected HEK293 cell line calculated using the 2-ΔΔCq method. **B.** Representative western blot of EphA1 protein expression within undifferentiated KOLF2-C1 iPSCs and KOLF2-C1 iPSC-derived cell lines analysed via western blot. Blots were probed with an anti-EphA1 monoclonal antibody recognizing the N-terminal domain of EphA1. EphA1-transfected HEK293 cells were used as a positive control. EphA1 depicted by black arrow head and alpha-tubulin loading control by the blue arrow head. Error bars indicate mean relative fold change values ± SD of three independent data sets.

5.3.3 Analysis of EphA1-AS1 and ZYX mRNA and protein levels

As mentioned above, additional genes have been identified through fine mapping analysis of AD-associated GWAS data sets that may be regulated by the non-coding SNPs at the *EPHA1* locus. Therefore, the expression levels of both *ZYX* and *EPHA1-AS1* mRNA within the cell types of interest were analysed via qRTPCR (Fig. 5.8A). *EPHA1-AS1* mRNA was not detected in the undifferentiated KOLF2-C1 iPS cells or any differentiated KOLF2-C1 iPSC-derived cell types (data not shown). However, expression of *ZYX* was seen in all cell types, with vastly increased expression within the iPSC-derived microglia (p=0.0024) and cortical neurons (p=<0.0001). *ZYX* transcript expression fold changes were calculated using the 2-ΔΔCq method using undifferentiated KOLF2-C1 iPS cells for normalization.

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

Following on from the identification of *ZYX* mRNA transcript expression within the cell types of interest, a western blot was performed to detect protein presence. ZYX protein expression was observed in undifferentiated KOLF2-C1 iPSC as well as KOLF2-C1 iPSC-derived monocyte-like cells, microglia and cortical neurons via western blotting. A representative blot is illustrated in figure 5.8, B, showing bands of ZYX protein at the expected size of 61 kDa, in all cell lines analysed (black arrow head).
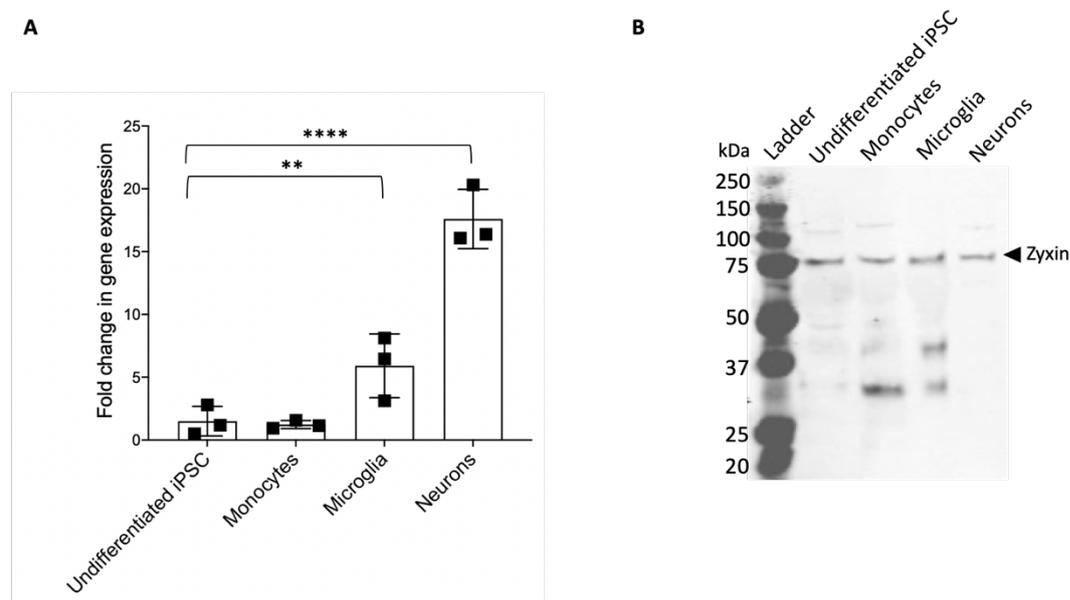
Figure 5.8: Quantitation of *ZYX* mRNA and protein expression in KOLF2-C1 iPSC and KOLF2-C1 iPSC-derived cell lines.

A. *ZYX* mRNA expression within undifferentiated KOLF2-C1 iPSC and KOLF2-C1 iPSC-derived monocyte-like cells, microglia and cortical neurons. Fold changes of iPSC-derived monocytes, microglia (p=0.0024) and neurons (p<0.0001) are shown relative to the undifferentiated KOLF2-C1 iPS cell line, calculated using the 2-ΔΔCq method. **B.** Representative western blot depicting ZYX protein expression within undifferentiated KOLF2-C1 iPSC and KOLF2-C1 iPSC-derived monocyte-like cells, microglia and cortical neuronal cells. Blots were probed with an anti-ZYX antibody. Data from three independent data sets, error bars indicate mean relative fold change values ± SD. Statistical test used, t-test, **, p<0.01. ****, P<0.0001

5.3.4 Generation of SNP Block deletion and ZYX gene knockout cell lines.

CRISPR-Cas9 technology was employed to delete three regions of non-coding DNA within the *EPHA1-AS1* gene containing the prioritised *EPHA1* SNPs (Table. 5.4) within the control

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

KOLF2-C1 iPS cell line. The *ZYX* gene was also targeted for deletion. Following CRISRP/Cas9 editing, clonal DNA was subject to end-point PCR screening and genotype assessed as per methods section 5.2.10 and 5.2.11. Representative blot images showing a small number of iPSC clones generated from all four CRISPR-Cas9 reactions are shown in figure 5.9.

As seen from figure 5.9A, the *ZYX* gene failed to be deleted within any of the CRISPR-Cas9 edited iPSC clones. One DNA product band on PCR screening is seen at 434bp (Fig. 5.9A, white arrow head) corresponding to the full length of DNA spanned by the ZYX screening primers (Table. 5.8).

The presence of only one DNA product band, again at the full DNA length spanned by the screening primers (699bp, Table. 5.8), was produced from all SNP Block 2 deletion iPSC clones following CRISPR-Cas9 editing (Fig. 5.9C, white arrow head). This indicates all clones screened were unedited and still contained the non-coding sequence and *EPHA1* SNPs.

Only CRISPR-Cas9 edited iPSC clones from SNP Blocks 1 and 3 deletion reactions showed successful deletion of the desired region of non-coding DNA. Cell lines containing SNP Block 1 (Fig. 5.9B, black arrow head) and 3 deletions (Fig. 5.9D, black arrow head) produced one DNA product on PCR screening of 699bp and 131bp respectively. These band sizes produced from the respective screening primers are to be expected upon homozygous deletion of the desired non-coding region following CRISPR-Cas9 editing (Table. 5.8).
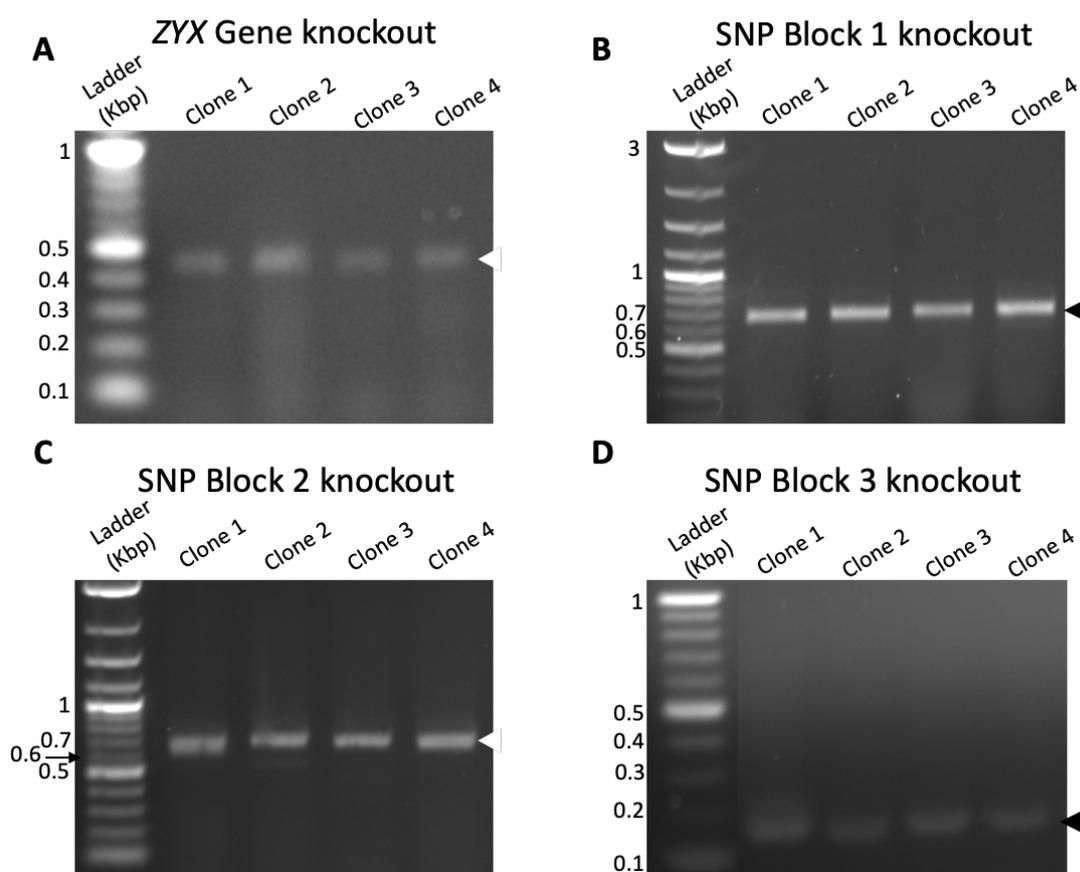
Figure 5.9: End-point PCR screening of CRISPR-Cas9 edited iPSC clones.
Representative images of PCR screening blots showing four clones of each intended deletion region. Samples were run alongside a 1 Kb DNA plus ladder, only the main band sizes of interest are labelled for clarity. **A.** IPSC clones generated from the *ZYX* gene knock out CRISPR-Cas9 reaction produced one DNA product band at 434bp in length. **B.** IPSC clones from CRISPR-Cas9 editing of SNP Block 1 deletion produced one DNA product band at 699bp in length. **C.** IPSC clones from CRISPR-Cas9 editing of SNP Block 2 deletion produced one DNA product band at 699bp in length. **D.** IPSC clones from CRISPR-Cas9 editing of SNP Block 3 deletion produced one DNA product band at 131bp in length.

5.3.5 Sanger sequencing of SNP Block deletion and control clone cell lines.

Clonal DNA from three selected homozygous SNP Block 1 and 3 deletion clones along with three CRISPR-Cas9 control clones were sequenced as per the above method. The CRISPR-Cas9 control clones were taken from unsuccessful CRISPR-Cas9 reactions to delete SNP Block 1. The sequence alignment of DNA from a CRISPR-Cas9 control clone to that of the reference genome sequence of *EPHA1-AS1* can be seen in figure 5.10. Sequence alignment can be seen just after the SNP Block 1 screening primers sequences (Fig. 5.10, black boxes)

and throughout the SNP Block 1 region that was expected to be deleted (Fig. 5.10, red box). The exact position of Cas9 nuclease cleavage depicted via the blue arrow head. This alignment was repeated using screening primers for the SNP Block 3 deletion region (Table. 5.8), illustrating the CRISPR-Cas9 control clones used contained both SNP Block 1 and 3 regions (sequence alignments from remaining two control clones showing no deletion of the Block 1 region along with those showing retainment of the Block 3 deletion region in all three clones are depicted in Appendix III).

Sequence alignment of DNA from a SNP Block 1 deletion iPSC clone can be seen in figure 5.11. No alignment is observed after the Cas9 nuclease cut site and throughout the intended deleted region (Fig. 5.11, red box), but some alignment either side following the Block 1 screening primer location can be seen (Fig. 5.11, black box). This alignment was again conducted on clonal DNA taken from the three SNP Block 3 deletion iPS cell lines using SNP Block 3 screening primers (Table. 5.8) confirming deletion of the expected SNP Block 3 region along with retention of the region within the control clones (Appendix III).
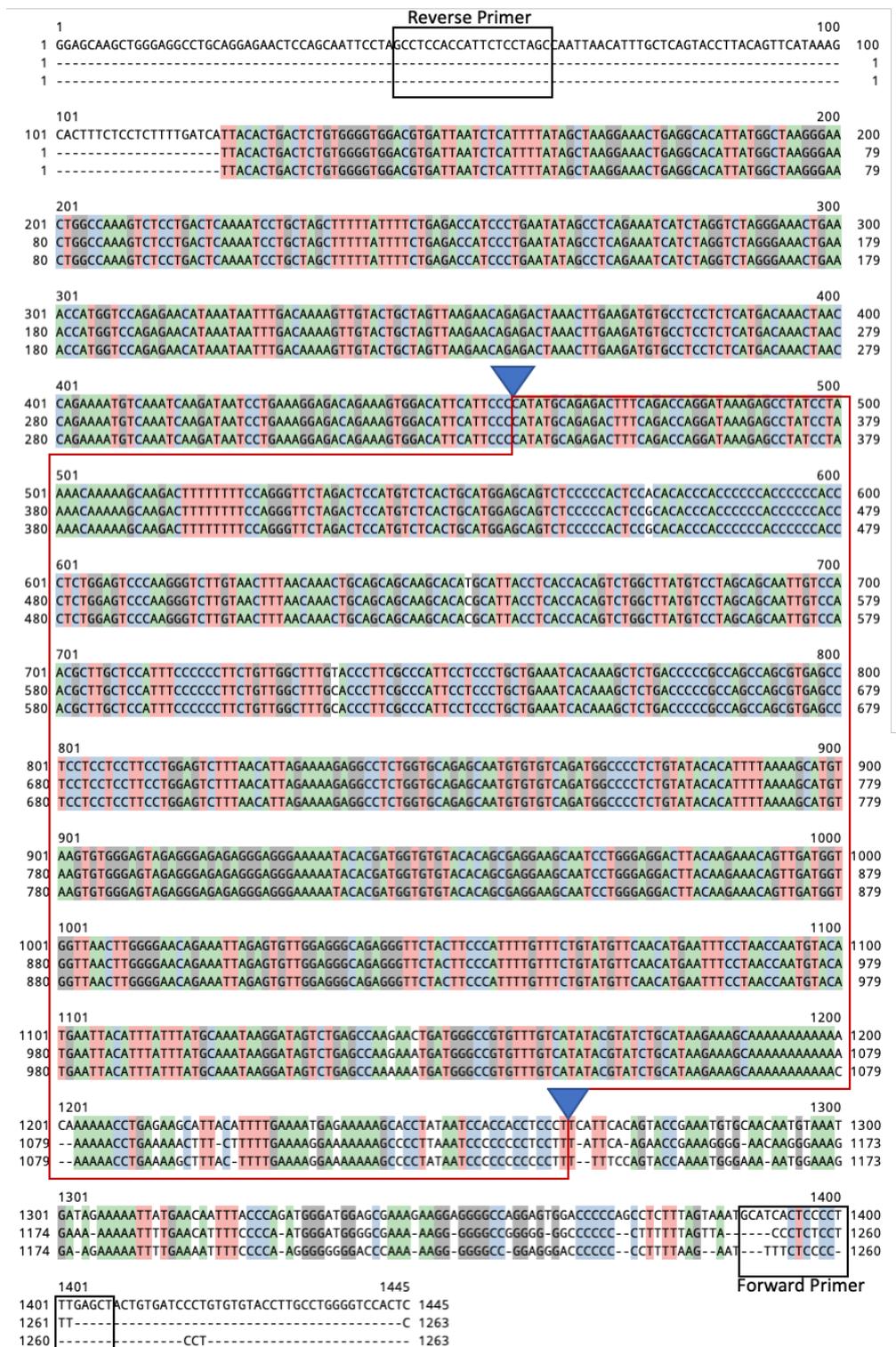
Figure 5.10: Characterisation of a CRISPR-Cas9 generated control clone.
Sequence alignment of an iPS cell CRISPR-Cas9 control clonal DNA with the *EPHA1-AS1* reference sequence surrounding the SNP Block 1 deleted region. Forward and reverse SNP Block 1 screening primer locations depicted within the black box. Expected SNP Block 1 deleted region outlined in red, exact Cas9 nuclease cut sites are illustrated via the blue arrow head. Alignment schematic created using CRISP-ID V1v1.
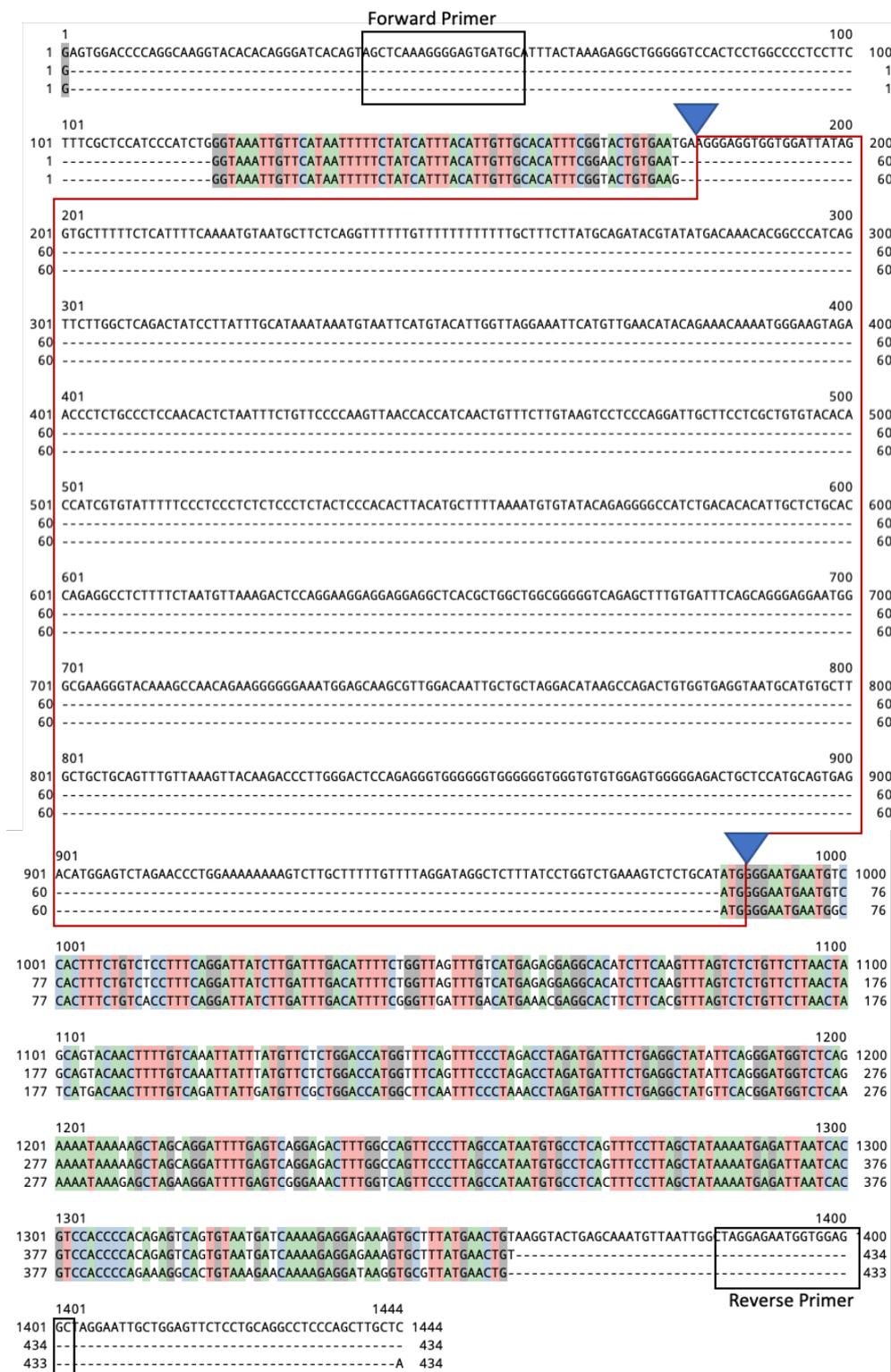
Figure 5.11: Characterisation of homozygous deletion CRISPR-Cas9 generated clone.
Sequence alignment of clonal DNA taken from a SNP Block 1 deletion iPS cell line with the *EPHA1-AS1* reference sequence surrounding the SNP Block 1 region. Forward and reverse SNP Block 1 screening primer locations depicted within the black box. Expected SNP Block 1 deleted region outlined in red, exact Cas9 nuclease cut sites are illustrated via the blue arrow head. Alignment schematic created using CRISP-ID V1v1.

5.3.6 Morphological differences between the CRISPR-Cas9 control and SNP Block deletion cell lines.

Following screening, three homozygous deletion iPSC clones from both SNP Block 1 and 3, along with three CRISPR-Cas9 control iPSC clones, were cultured as per methods section 5.2.1. Upon culturing, morphological differences between the SNP Block deletions and the unedited control cell lines became apparent.

Figure 5.12 shows a schematic containing representative phase contrast images of a CRISPR-Cas9 control and a SNP Block 1 deletion cell line. The CRISPR-Cas9 control cell line (Fig. 5.12Ai and Aii) displays typical iPSC morphology such as compact, round cells with well-defined nuclei. However, this was not noted within the SNP Blocks 1 and 3 deletion cell cultures. As depicted in figure 5.12Bi, these cultures appeared to be a mixture of spontaneously differentiated cells with no presence of iPSC colonies. Displaying a large, flat and elongated morphology typical of fibroblast-like cell cultures as seen in the enlarged image of figure 5.12Bii.

Both the CRISPR-Cas9 control and SNP Blocks 1 and 3 deletion clonal cell lines were subjected to microglial differentiation, as per methods section 5.2.3. The CRISPR-Cas9 control cell lines successfully formed the necessary embryoid bodies (EBs) for the development of monocyte-like cell producing factories. As shown in figure 5.12C, these EBs/monocyte factories form large EBs with internal cystic structures capable of creating and blebbing off monocyte-like cells into the cell culture media (Fig, 5.12Ei, Eii). These cells were then harvested and successfully matured into microglial cells (Fig. 5.12G).

This was not observed within the SNP Blocks 1 and 3 deletion cell cultures. These cell lines either failed to form cell aggregates upon microglial differentiation induction or formed small EBs which did not develop internal structures and proceeded to disintegrate in culture, resulting in spontaneously differentiated adherent cells (Fig. 5.12D). These cultures therefore failed to produce any large round monocyte-like cells as seen in figure 5.12Fi.

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

Small non-adherent cells and cell debris observed as seen in the enlarged image of figure 5.12Fii.
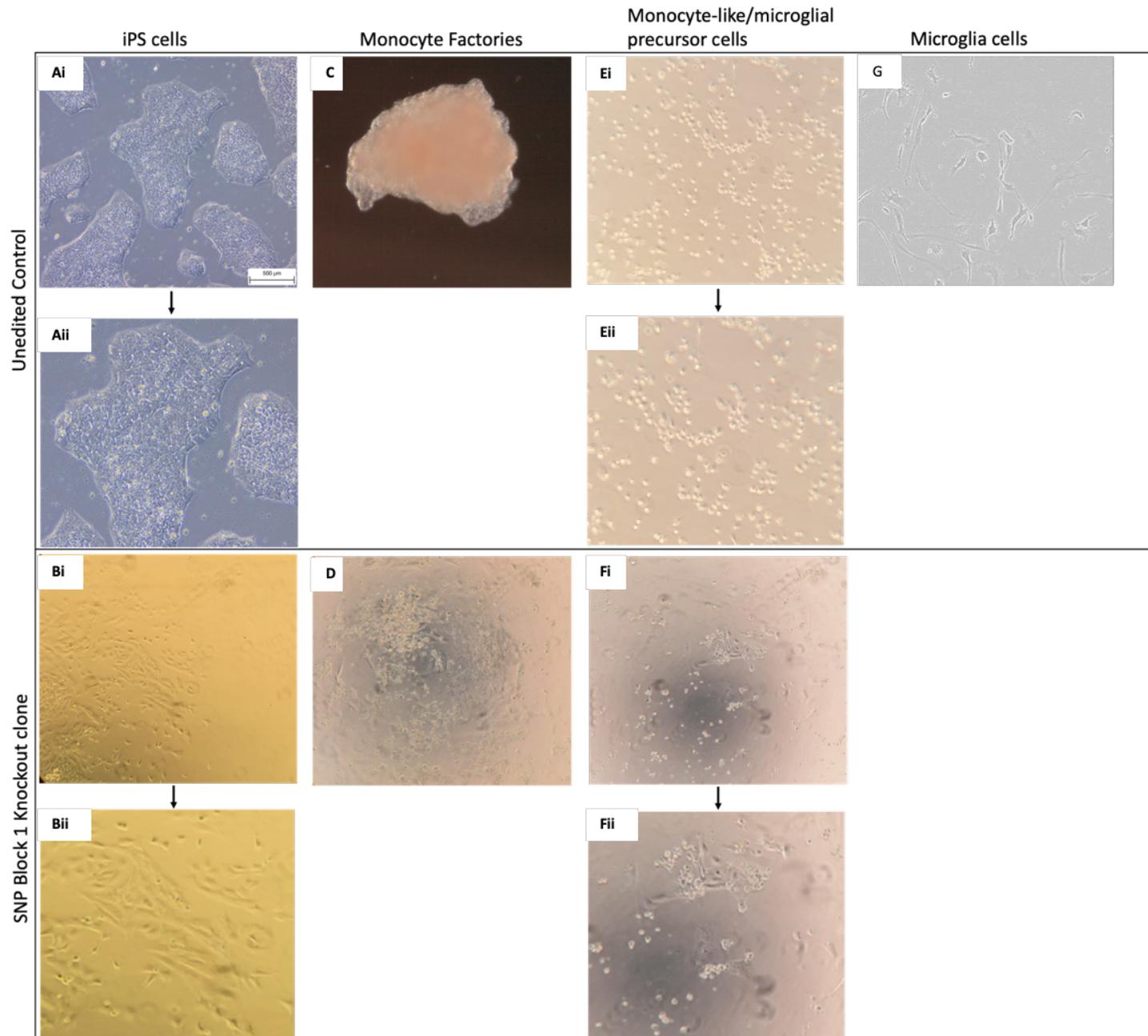
Figure 5.12: Morphological differences between CRISPR-Cas9 control and SNP Block 1 deletion cell cultures.
A schematic showing representative phase contrast cell culture images from both a CRISPR-Cas9 control and SNP Block 1 deletion clonal cell line.
Ai. IPSC culture of the CRISPR-Cas9 control clone showing typical iPSC morphology. **Aii.** An enlarged image of Ai. **Bi.** IPSC culture of the SNP Block 1 deletion clone depicting a mixture of spontaneously differentiated cells. **Bii.** An enlarged image of Bi showing large, flat cell morphology. **C.** A CRISPR-Cas9 control clone-derived monocyte producing factory. **D.** A disintegrated embryoid body and adhered spontaneously differentiated cells produced on microglial differentiation induction of a SNP Block 1 deletion clone. **Ei.** Monocyte-like cells produced from the unedited control clone derived-monocyte factories. **Eii.** An enlarged image of Ei. **Fi.** Cell debris and spontaneously differentiated cells formed by the SNP Block 1 deletion clone during microglial differentiation. **Fii.** An enlarged image of Fii showing flat cell morphology. **G.** Microglial cells matured from a CRISPR-Cas9 control clone.
Images taken on Zeiss microscope at x10 magnification, scale bar representative of all images.

150

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

5.3.7 Gene expression alterations within SNP Block1 and 3 deletion cell lines.

ZYX gene expression was analysed within the CRISPR-Cas9 edited SNP Block deletion cell lines via qRTPCR and western blot (Fig. 5.13).

As depicted in figure 5.13A, on quantitation via qRTPCR, both SNP Block 1 and 3 deletion cell lines expressed a significantly increased level of *ZYX* gene mRNA transcript ranging from 2 to 4-fold when compared to the CRISPR-Cas9 control clones, p=0.0005 and p=<0.0001 respectively.

An increase in ZYX protein expression was noted by western blot analysis. Figure 5.13B shows a representative western blot image containing cell lysates from three CRISPR-Cas9 control cell lines and three SNP Block 1 and 3 deletion cell lines. More intense bands were observed within all six SNP Block deletion cell lysates at the expected size of the ZYX protein, 61 kDa (Fig. 5.13B, black arrow head). GAPDH was used as a loading control (Fig. 5.13B, blue arrow head), this was used during quantitation of blot images as per methods chapter, chapter 2, section 2.5. On quantitation, ZYX protein expression within the SNP Block 1 and 3 cell lysates was observed to be significantly greater than that present within the CRISPR-Cas9 control cell lines, p=0.0086 and p=0.0023 respectively (Fig. 5.13C).
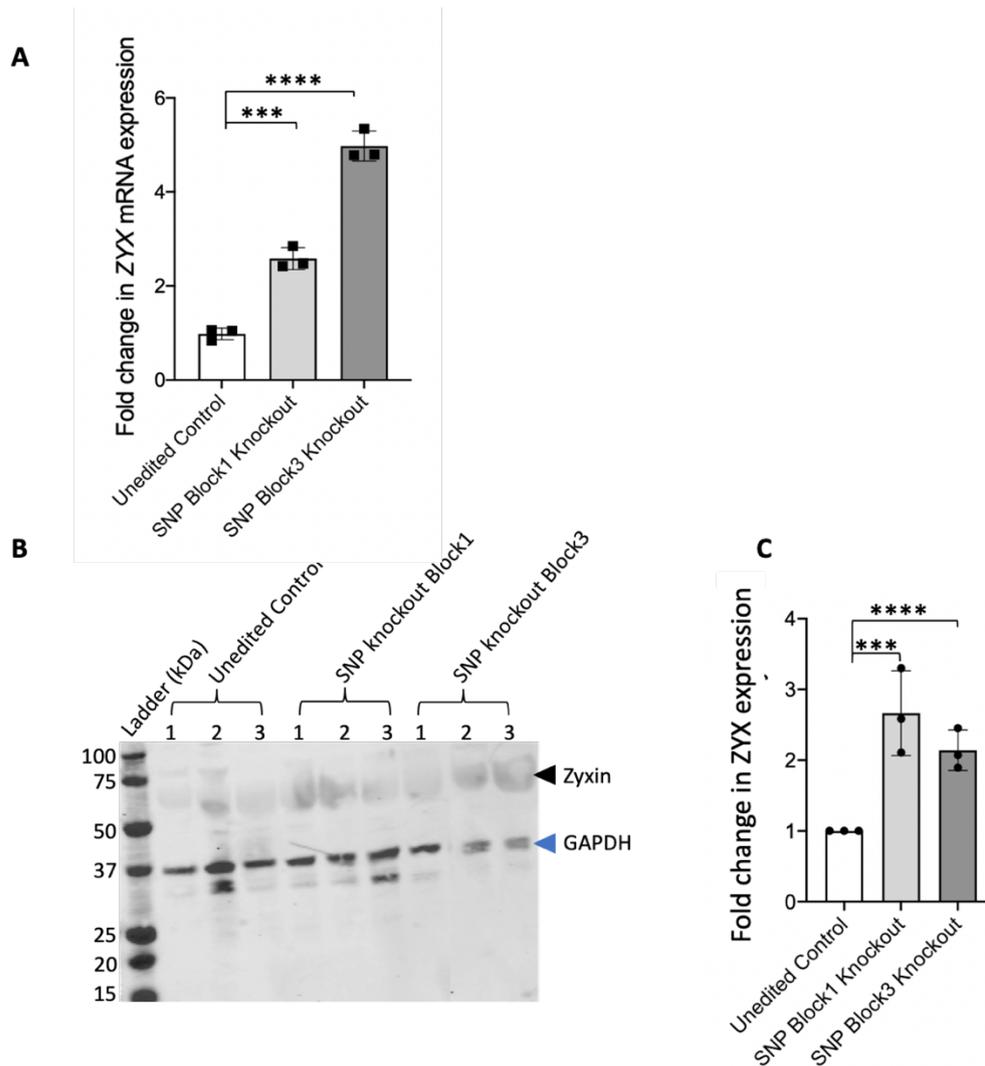
Figure 5.13: Alterations in *ZYX* mRNA transcript and protein levels within SNP Block 1 and 3 deletion cell lines.

A. Quantitation of qRTPCR data showing *ZYX* mRNA expression levels within CRISPR-Cas9 control and SNP Block 1 and 3 deletion cell lines ($p=0.005$ and $p<0.0001$ respectively). **B.** A representative western blot image depicting ZYX protein expression levels within CRISPR-Cas9 control and SNP Block 1 and 3 deletion cell lysates ($p=0.0086$ and $p=0.0023$ respectively). Black arrow head depicts ZYX protein band at expected size of 61 kDa. Blue arrow head depicts GAPDH loading control bands at expected size of 36 kDa. **C.** Quantitation of western blotting images. Data from 3 independent clonal iPS cell lines. Error bars represent mean ±SD. Statistical test used, t-test. ***, $p<0.001$. ****, $p<0.0001$.

Following on from the observed morphological alterations on culturing the iPSC lines derived from the CRISPR-Cas9 deletion of both SNP Blocks, the loss of pluripotency was analysed via qRTPCR and western blot. This was achieved by analysing expression of the transcription factor NANOG (Fig. 5.14). This TF controls both self-renewal and pluripotency

of embryonic stem cells, therefore can be used as a marker to determine the stemness of cell lines (Ling et al. 2012).

Figure 5.14A shows qRTPCR quantitation of *NANOG* mRNA levels within SNP Block 1 and 3 cell lines. These data show significantly decreased *NANOG* levels in both SNP Block 1 and 3 deletion iPS cell lines when compared to the CRISPR-Cas9 control cell lines (p=0.0052, p=0.0006 respectively).

The western blot from figure 5.13B was stripped and re-stained for NANOG expression as per methods detailed in chapter 2, section 2.5. From this re-stained blot (Fig. 5.14B), a reduction in NANOG band intensity was observed in all six of the SNP Block deletion cell lysates at the expected size of 35 kDa (Fig 5.13B, red arrow head). However, quantitation of the western blot images (Fig, 5.14C) showed only the SNP Block 3 deletion to produce a significant reduction in NANOG protein levels compared to the CRISPR-Cas9 control cell lysates (P=0.003).
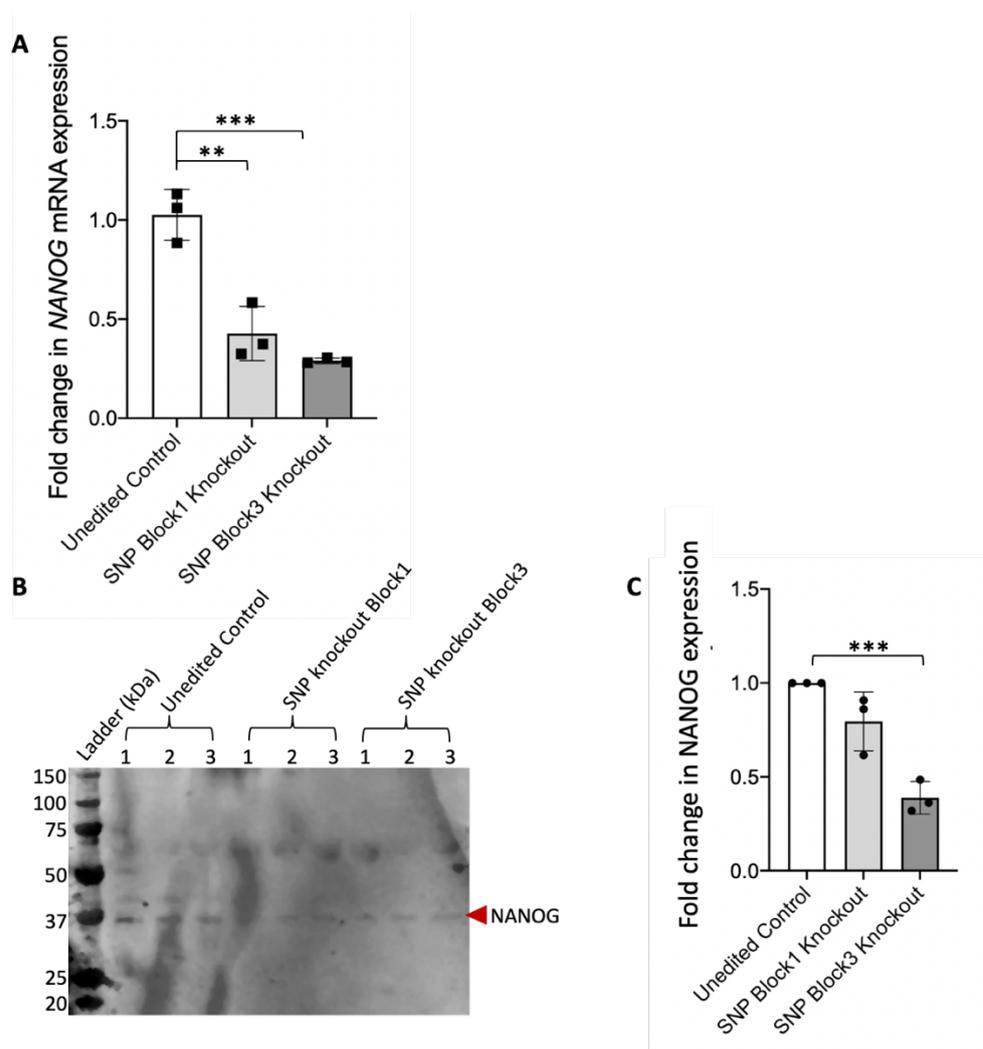
Figure 5.14: Alterations in *NANOG* mRNA transcript and protein levels within SNP Block 1 and 3 deletion cell lines.
A. Quantitation of qRTPCR data showing *NANOG* mRNA expression levels within CRIPSR-Cas9 control and SNP Block 1 and 3 deletion cell lines (p=0.0052 and p=0.0006 respectively).
B. A representative western blot image depicting NANOG protein expression levels within CRIPSR-Cas9 control and SNP Block 1 and 3 deletion cell lysates (p=0.003). Red arrow head depicts NANOG protein bands at expected size of 35kDa. **C.** Quantitation of western blotting images. Data from 3 independent clonal iPS cell lines. Error bars represent mean ±SD. Statistical test used, t-test. **, p<0.01. ***, p<0.001. ****, p<0.0001.

The expression levels of the *CDH1* gene encoding for the epithelial cadherin (E-cadherin) protein were also analysed within the SNP Block deletion cell lines, due to this gene's crucial involvement in cell-cell adhesion during early mammalian development and its role in maintaining iPSC pluripotency.

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

Figure 5.15 shows the quantitation of qRTPCR data from the SNP Block 1 and 3 deletion iPSC lines. These cell lines were noted to have significantly reduced levels of *CDH1* mRNA transcript when compared to the CRIPSR-Cas9 control cell lines (p=0.0207 and p=0.0255 for both SNP Block 1 and 3 deletion cell lines respectively).
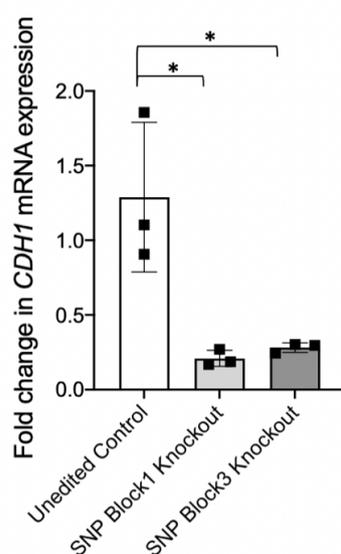
Figure 5.15: Alterations in *CDH1* mRNA transcript levels within SNP Block 1 and 3 deletion iPS cell lines. Quantitation of qRTPCR data showing *CDH1* mRNA expression levels within CRIPSR-Cas9 control and SNP Block 1 and 3 deletion cell lines (p=0.0207 and p=0.0255 respectively). Data points from 3 independent clonal iPS cell lines. Error bars represent mean ±SD. Statistical test used, t-test. *, p<0.05.

5.4 Discussion

Although the role of some antisense lncRNAs have been linked to the regulatory control of their sense genes, their exact functional role remains elusive (Hung and Chang 2010). Therefore, *EPHA1-AS1* may possess the ability to regulate the expression of other surrounding genes, such as *ZYX* along with *EPHA1* via acting as a site of regulatory protein binding. In addition to this, as previously mentioned, certain AD-associated SNPs within the *EPHA1* locus reside within regions of non-coding DNA containing a known enhancer or TF binding sites. The functional range of which these enhancers or TF binding sites often makes identification of their target gene/s extremely complex.

Teasing out the functional role of such non-coding genes is increasingly important in terms of understanding how non-coding variations influence gene regulation and ultimately affect AD pathology (Idda et al. 2018). Therefore, the aim of this chapter was to identify the

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

expression patterns of key genes of interest, such as *EPHA1, EPHA1-AS1* and *ZYX* within cell types linked to AD, such as monocytes, microglia and cortical neurons. This was in conjunction with trying to ascertain if any of the prioritised *EPHA1* SNPs have the capacity to alter such expression patterns.

5.4.1 Analysis of EPHA1, EPHA1-AS1 and ZYX expression levels within cell types of interest.

Analysis of public single cell RNA sequencing datasets from human peripheral blood cells conducted by You Zhou and Birong Zhang (Zhou laboratory, Cardiff University) indicated that the *EPHA1* mRNA gene is only expressed by a small population of NK cells, CD4[+] and CD8[+] T cells (Fig. 5.5 and Tab. 5.9). *EPHA1-AS1* had a very similar expression pattern as *EPHA1*, with expression again only noted within a small subset of NK cells, CD4[+] and CD8[+] T cells (Fig. 5.5B). Although the proportion of cells within each subtype expressing *EPHA1-AS1* was slightly higher than that for *EPHA1* as seen in table 5.9, still only a very small number of cells within this group expressed this gene. However, this low expression of mRNA transcript may not be fully representative of the protein level within these cell types due to the unknown stability of the EphA1 protein. High protein stability would result in low transcription levels. Therefore, flow cytometry or western blot analysis of human peripheral blood samples may yield more information of *EPHA1s* exact expression levels within these cell types.

Gene expression patterns of this nature can also be found using other large databases such as the EMBL-EBI Single Cell Expression Atlas and The Human Protein Atlas. For example, within these databases *EPHA1* and *EPHA1-AS1* are seen to be expressed at low levels within a small subset of tissues. Of interesting note, the ephrinA1 ligand (EFNA1) is expressed at highest levels within endothelial cells. This potentially correlates with evidence that the EphA1 receptor expressed on peripheral immune cells is activated by ligands expressed on endothelial cells, for instance those within the BBB.

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

Correlating with this and data presented within this thesis, single cell expression analysis from The Human Protein Atlas show very little to no expression of *EPHA1* and *EFNA1* within both vascular and brain cell types analysed. Although the highest expression of each gene can be noted within a small subset of excitatory neurons and oligodendrocytes respectively, expression levels remain low 4-28.3 TPM). This is in contrary to *ZYX* expression which is expressed in all cell types analysed at levels 4 - 500-fold higher than *EPHA1.*

Quantitation of mRNA transcript levels of both *EPHA1* and *EPHA1-AS1* and protein expression of EphA1 within KOLF2-C1 iPSC-derived monocyte-like cells correlated with these single cell RNA-Seq data, with no expression of either gene detected through qRTPCR or western blot analysis (Fig. 5.7). This lack of notable expression also extended to other KOLF2 iPSC-derived AD relevant cell types, such as microglia and cortical neurons (Fig. 5.7). However, as seen in the RNA-Seq data it may be the case that EphA1 is expressed albeit by a small proportion of cells at a level below that which is detectable via qPCR or western blot analysis.

The observed lack of expression of *EPHA1* and *EPHA1-AS*1 within cell types linked to AD pathology suggests despite potentially being regulated by the non-coding SNPs at the *EPHA*1 locus, there may be additional genes involved in the association found between these SNPs and AD pathology during genome wide studies. However, only a few cell types thought to be involved in AD pathology were analysed, it is possible that *EPHA1* and *EPHA1-AS*1 expression can be seen within alternative cell types, for example astrocytes, oligodendrocytes or cells within the BBB, such as, endothelial cells or pericytes. Leading to *EPHA1* and *EPHA1-AS1* potentially exerting their effect on AD pathology through altering the integrity of the BBB as previously described in Chapter 3.

In contrast to *EPHA1* and *EPHA1-AS1*, the gene *ZYX,* whose expression changes have also been noted to co-localise with multiple SNPs within the *EPHA1* locus (Kunkle et al. 2019b), was expressed widely across all analysed human white blood cell subtypes. The largest proportion of cells expressing *ZYX* was noted within CD16/FcγRIIIA⁺ monocytes (35.4%), conventional dendritic cells (36%) and platelet cells (62.6%), as seen in table 5.7, with the

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

highest level of *ZYX* expression observed within CD14$^+$ monocytes (Fig. 5.5C). This high expression levels were also observed in The Human Protein Atlas database, showing *ZYX* expression in all vascular and brain cell types analysed. This potentially lends weight to the hypothesis that the *ZYX* gene is also involved in AD pathology at this locus.

On analysis of other AD-relevant cell types derived from KOLF2-C1 iPSCs, *ZYX* mRNA transcript expression was noted throughout. Expression levels were seen to be significantly increased in microglia and cortical neurons from undifferentiated iPSCs (Fig. 5.7A). This was confirmed with protein expression analysis within each cell type (Fig. 5.7B).

As previously discussed AD risk genes cluster around certain biological pathways, for example many risk loci such as; *CD2AP, EPHA1, ABI3, CELF1* and *FERMT2* have all been associated with the Clatherin-mediated endocytosis pathway or the regulation of cytoskeletal machinery (Satoh et al. 2017; Caspers et al. 2020). There is evidence of dysregulated cytoskeletal dynamics and defects in downstream focal adhesion signalling during AD pathology as a result of pathogenetic species such as Aβ oligomers and hyperphosphorylated Tau protein (Kang et al. 2011; Hodges 2021). This is consistent with the biological pathways within which ZYX is thought to function.

The ZYX protein has been implicated in multiple cell behaviours centred around motility and adhesion, through its role in cytoskeletal reorganisation of actin filaments (Crawford et al. 1992). A precedent has also been set during *H. Influenzae* infection for ZYX to influence the permeability of the BBB via its effects on tight junction structure (Miyazaki et al. 2014). Taken together, the expression of *ZYX* in AD-relevant cell lines within the human brain and peripheral blood, it is possible that expression changes within this gene observed as a result of non-coding SNPs within the *EPHA1* locus is responsible for their association with AD pathology.

5.4.2 Generation and troubleshooting of SNP Block deletions and ZYX knockout cell lines.

As detailed above, CRISPR-Cas9 gene editing technology provides an amazing resource to alter the genome of iPSCs. This ability to delete non-coding DNA regions as well as creating gene knock outs allows functional investigation and direct comparison to unedited CRISPR-Cas9 control clones from the same starting iPS cell line. This chapter aimed to delete the *ZYX* gene to allow further investigation of its function within AD relevant cell types and to attempt to tease out how any alterations in its expression levels may influence AD pathology.

However, as seen in figure 5.9A, CRISPR-Cas9 editing using the *ZYX* knockout left and right guides detailed in table 5.5 failed to produce any homozygous *ZYX* gene knockout clones. The location of these guides are shown in figure 5.16 as grey boxes, designed within the second exon of the *ZYX* gene and should have led to the deletion of an 119bp DNA region (Fig. 5.16, purple box), resulting in the creation of a frameshift mutation and ultimately nonsense mediated decay and *ZYX* gene deletion. Following the initial unsuccessful CRISPR-Cas9 editing, an alternative guide (Fig. 5.16, *ZYX* Knockout Alternative Guide, grey box) was designed and used in conjunction with the existing *ZYX* Knockout Left Guide. This was an attempt to ascertain whether gene knock out failure was due to guide design or inefficiency. However, this CRISPR-Cas9 editing reaction again failed to produce homozygous gene knock out clones (data not shown).
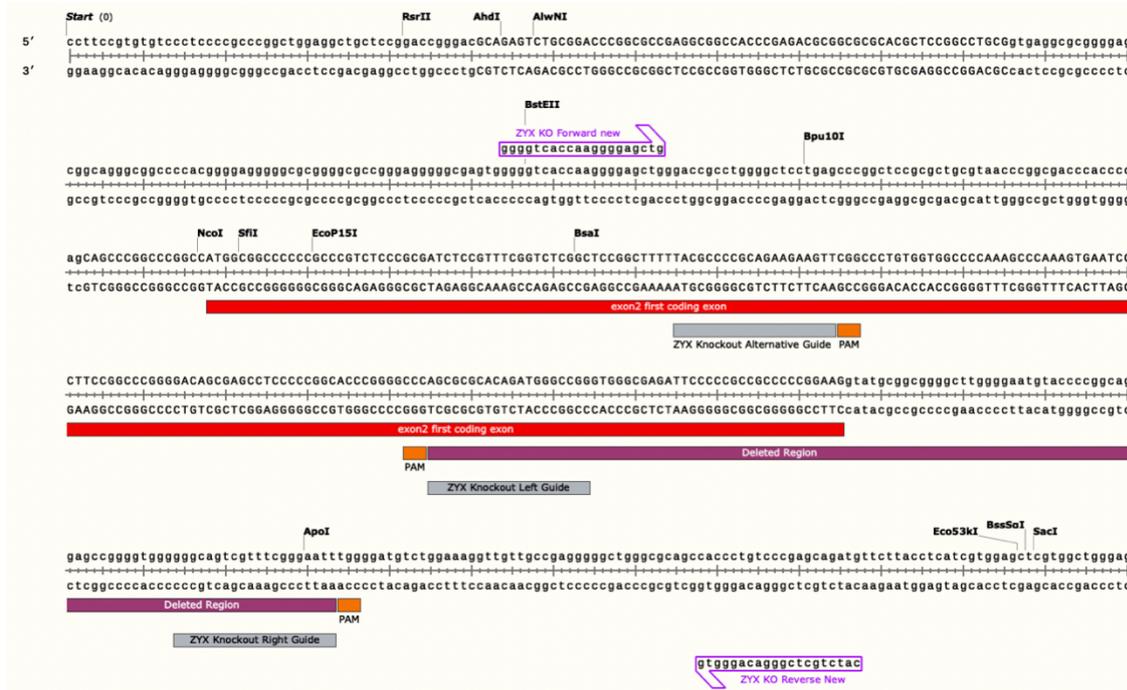
Figure 5.16: Genomic location of *ZYX* CRISPR-Cas9 guides and deleted region.
A schematic of a portion of *ZYX* gene genomic sequence showing the location of exon 2 (red box). The location of the original *ZYX* Knockout Left and Right Guides including the Alternative Guide created are depicted as grey boxes. Respective PAM sequences are shown in orange. The region of deleted DNA is depicted by a purple box. *ZYX* knockout end-point PCR screening primers are shown as purple arrows. Schematic created in SnapGene.com.

The deletion of the non-coding DNA region SNP Block 2 also failed to produce any homozygous deletion clones upon CRISPR-Cas9 editing (Fig. 5.9C). Neither failed deletion/knock out reactions were followed up due to time constraints. However, potential optimisation experiments could be carried out to troubleshoot this CRISPR-Cas9 editing failure of both gene and non-coding DNA region. Such experiments include further trials of alternative CRISPR guide design, potentially targeting an alternative exon in the case of the *ZYX* gene deletion. The nucleofection parameters may also be modified to attempt to improve the transfection efficiency of the CRISPR-Cas9 machinery into the iPSCs.

A T7E1 (T7 endonuclease 1) assay could also be employed to further assess the editing efficiency of the CRISPR-Cas9 sgRNAs. This cost effective and simple technique employs the structure selective T7 endonuclease to recognise and cleave heteroduplexes produced by aberrant NHEJ events created by ineffective and suboptimal CRISPR-Cas9 cleavage events.

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

Analysis of AD-associated variants is inherently difficult due to any alterations produced as a result of individual SNPs likely to be very small, this may be compounded by the variability often observed within iPSC differentiations (Sentmanat et al. 2018). Therefore, this thesis only attempted to characterise the *EPHA1* AD-associated variants via producing homozygous deletions of sections of the non-coding genome containing multiple SNPs of interest. Due to the nature of common AD-associated variants contributing small effect sizes to AD pathology and therefore to cell phenotype, homozygous deletions were chosen in the first instance as these genotypes where thought to produce the strongest phenotype within the iPS cells. This would allow cell characterisation and assay development in order to best quantitate any phenotype alterations produced as a consequence of variant deletion. This would then allow more detailed experiments looking into specific variants, potentially with heterozygous cell lines allowing allele-specific gene expression alterations to be analysed.

5.4.3 Spontaneous differentiation and gene expression alterations within SNP Block 1 and 3 deletion iPS cell lines.

Following CRISPR-Cas9 genome editing, clonal cell lines with deletions of SNP Block 1 and 3 non-coding DNA regions were successfully created. These cell lines were cultured and alterations in the expression of genes of interest (*EPHA1, EPHA-AS1* and *ZYX*) analysed. This aimed to determine if a SNP within one of these blocks is capable of regulatory ability that maybe linked to AD pathology (Tab. 5.2).

Upon cell culture, it was noted that all cell lines containing the SNP block deletions spontaneously differentiated (Fig. 5.12B) when compared to the CRISPR-Cas9 control cell lines, isolated from the same gene editing experiments (Fig. 5.12A). Spontaneously differentiated cells formed a heterogenous culture of large, flat cells or long projection containing cells (Fig. 5.12Bii). In conjunction with the morphological alterations observed, the cell lines were unable to form EBs (Fig. 5.12D) or produce monocyte-like cells (Fig. 5.12F) compared to CRISPR-Cas9 controls, which were capable of differentiation (Fig. 5.12C, E and G). It was therefore hypothesised that the deletion of SNP Block 1 and 3 regions of

non-coding DNA (Tab. 5.4), disrupted the pluripotency of these cell lines resulting in a lack of stemness which was not observed within the CRISPR-Cas9 control cell lines.

On qRTPCR and western blot analysis of the SNP Block deletion cell lines for alterations in expression of genes of interest, no alteration in the lack of *EPHA1* or *EPHA1-AS1* expression was observed (data not shown). However, SNP Block 1 and 3 deletion cell lines showed a significant increase in both *ZYX* mRNA transcript (Fig. 5.13A) and protein expression levels (Fig. 5.13C) when compared to the CRISPR-Cas9 control cell lines.

However, further experiments are required to confirm if the alterations in ZYX expression observed and the lack of stemness is indeed a consequence of the SNP Block 1 and 3 deletion genotype and not a result of spontaneous differentiation through another means, for example, a result of the CRISPR-Cas9 and cell culturing processes. To investigate this the CRISPR-Cas9 control clone cell lines could be allowed to spontaneously differentiate in culture prior to reanalysis of ZYX expression levels. This would remove the variable of the lack of stemness and allow confirmation that it is indeed the genotype resulting in the raised ZYX expression levels

ZYX is best characterised for its functions in cytoskeletal remodelling, regulating cell behaviour such as motility and adhesion. However, recently an additional role in the regulation of pluripotency genes has been identified. ZYX acts to inhibit the activity of pluripotency genes responsible for maintaining embryonic stem cell status, including *POU5F3/OCT4*. Parshina et al, observed that ZYX down regulation within zebrafish embryos and human HEK293 cell cultures resulted in increased mRNA expression of pluripotency genes such as *NANOG, KLF-4* and *POU5F3/OCT4* (Parshina et al. 2020).

Following on from this, the expression levels of the pluripotency genes *NANOG* and *CDH1* were analysed. Decreased *NANOG* (Fig. 5.14A) and *CDH1* (Fig. 5.15) mRNA levels were noted in both SNP Block deletion cell lines when compared to their CRISPR-Cas9 control cell lines. However, the decrease in NANOG protein expression was only observed to be significant within SNP Block 3 deletion cell lines (Fig. 5.14C), which is at odds with the qRTPCR data.

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

Such discrepancies in observed expression levels between mRNA transcript and protein could be due to the quantitation of the western blot images. Potential future repetition of such blots to obtain clearer images may provide more consistent quantitation results. Lack of complete western blot stripping following ZYX and GAPDH staining may also contribute to the inconsistences in results due to NANOG being a similar sized protein to that of GAPDH. This lack of complete protein stripping from the original blot (Fig. 5.13B) can be observed in figure 5.14B due to the remaining presence of a faint band corresponding to the ZYX protein. Alternatively, the stability of the NANOG protein may play a factor in explaining its prolonged presence despite the reduction in mRNA levels observed. To this end the qRTPCR data may provide more accurate and consistent results for analysis of expression patterns of these genes of interest.

The decreased NANOG expression levels observed within the SNP block deletion cell lines is of particular importance as this homeodomain protein is essential for embryonic stem cell identity, along with the SOX2 and POU5f3/OCT4 proteins. NANOG has been shown to play a crucial role in not only maintaining pluripotency but also preventing differentiation. A NANOG deficiency within the inner cell mass of a mouse blastocyst results in a failure of these cells to generate the epiblast or primitive ectoderm due to loss of pluripotency. This deficiency will result in the differentiation of cells into an extraembryonic endoderm lineage only. Such cells have a flat morphology with multiple elongated projections, similar to the morphology observed within the SNP block deletion iPSC lines (Fig. 5.12B).

The *CDH1* gene encodes for the membrane-spanning cell adhesion molecule epithelial-cadherin (E-cadherin) and plays a crucial role in adhesion junctions and cell-cell contact formation. However, an additional role in the generation and maintenance of iPS cell pluripotency has been identified (Bedzhov et al. 2013). It was noted that in the absence of E-cadherin, mouse fibroblasts fail to reprogram effectively (Redmer et al. 2011). Moreover, it was observed that E-cadherin expression could generate reprogrammed cells even in the absence of the Yamanaka factor *OCT4*.

Chapter 5: The role of EPHA1 SNPs in the regulation of EPHA1, EPHA1-AS1 and ZYX expression patterns

The dual roles of *CDH1* gene in cell adhesion and pluripotency regulation may explain some of the observed cell phenotypes within the SNP block deletion cell lines. Significantly decreased *CDH1* mRNA expression was identified within both SNP block deletion cell lines when compared to their CRISPR-Cas9 control cell lines (Fig 5.15). Such gene expression deficiency may influence iPSC morphology in multiple ways, through the cells inability to form the tight adherent cell colonies essential for maintaining pluripotency as well as the loss of stemness and subsequent differentiation observed. The downregulation of E-cadherin within these cell lines may also explain the inability to form cell aggregates and the loss of EB integrity due to its requirement for cell compaction during mammalian development.

Taken together these data potentially implements *ZYX* in the control of stem cell maintenance and differentiation. As previously outlined, Zyxin has the ability to shuttle to the nucleus and function as a TF to control expression of pluripotency genes and those involved in stem cell differentiation. This suggests that alterations within the genome that disrupts the homeostatic expression of ZYX for example, causing its increased expression as seen here. results in a subsequent decrease of pluripotency genes such as OCT4 (Parshina et al. 2020).

5.4 Conclusion

This chapter concludes, based on the single cell RNA-Seq gene expression data analysed and observations from KOLF2-C1 undifferentiated and -derived monocyte-like, microglia and neuronal cells, that it is likely that *ZYX*, and not *EPHA1* or *EPHA1-AS1*, is behind the GWAS AD association of SNPs at the *EPHA1* locus.

The deletion of two DNA regions containing the prioritised *EPHA1* locus SNPs via CRISPR-Cas9 gene editing technology (SNP Block 1 and 3 deletion) illustrates that these non-coding regions of DNA can influence gene expression patterns. However, these SNP blocks contain three and two GWAS SNPS respectively, therefore further investigation will be required to

confirm which exact SNP is responsible for these gene expression alterations. To allow the independent analysis of each SNP, base alteration of each SNP within these blocks would have to be undertaken during separate CRISPR-Cas9 editing reactions.

As ZYX is implemented in cell motility and BBB integrity, it is therefore hypothesised that alterations in *ZYX* expression levels could explain the association between non-coding SNPs identified at the *EPHA1* locus during GWAS and AD pathology. This is possible that BBB hyperpermeability produced as a consequence of altered ZYX expression may cause increased leukocyte migration and translocation into brain parenchyma where they propagate AD-associated neuroinflammation and neurodegeneration (Kinney et al. 2018).

These data also suggested an alternative role for *ZYX* within the control of pluripotency, supporting recent published data in human HEK293 cells showing that ZYX knockdown causes an increase in cell pluripotency genes (Parshina et al. 2020). The lack of pluripotency observed via the decrease in expression of the pluripotency markers, *NANOG* and *CDH1*, within the SNP Block deletion cell lines highlights the potential role of ZYX expression during stem cell maintenance.

Chapter 6: Discussion

6.1 General discussion

Multiple GWAS have identified *EPHA1* as a risk gene of LOAD with numerous SNPs within this locus have reached genome wide significance. Aberrant Eph-ephrin signalling has been identified in a variety of diseases including cancers and neurological disorders such as Amyotrophic Lateral Sclerosis (Mo et al. 2013). EphA1 is of interest as a risk gene in AD due to its signalling being implicated in immune processes and its role in modulating dynamic cytoskeletal rearrangements, allowing processes such as cell migration, immune cell signalling and regulation as well as neuronal stem cell proliferation and neurite outgrowth (Aasheim et al. 2005; Holen et al. 2010; Yang et al. 2018).

This thesis aimed to shed light on aspects of EphA1 receptor biology, such as, subcellular location, potential proteolytic processing and both ligand dependent and ligand independent activation methods, with the view to identify how signalling through this receptor may influence an individual's susceptibility to AD.

The EphA1 variant rs20217856 (P460L), producing a Proline to Leucine substitution at amino acid 460, causes this receptor variant to segregate perfectly with AD within a Hispanic Caribbean family (Vardarajan et al. 2015). This rare receptor variant was employed to make a phenotypic comparison to the wild type receptor during this thesis, allowing any alterations in receptor biology to be analysed via western blot, flow cytometry and ELISA. These data taken together offer insights into alterations to expression levels, subcellular localisation and ligand activation produced as a result of the P460L variant. Such alterations may potentially be attributed to its association with AD pathology.

The vast majority of AD-associated SNPs identified through GWAS at the *EPHA1* loci are within *EPHA1s* neighbouring non-coding antisense gene, *EPHA1-AS1*. As discussed

in Chapter 4 numerous aspects of GWAS methodology, SNP inheritance structure and their genomic location makes assigning causality to the particular extremely difficult (Edwards et al. 2013). Interpreting how these SNPs contribute to disease aetiology from genetic data will enable greater understanding into the pathology of AD and the inheritance of LOAD.

Therefore, this thesis also aimed to analyse the vast number of non-coding, AD-associated SNPs at the *EPHA1* risk locus, with the aim of determining which SNP/s are likely the causal variant/s linking this locus to disease pathology. This analysis was based on the hypothesis that AD-associated SNPs affect disease pathology via allele specific alterations disrupting TF binding motif DNA sequences and therefore gene regulation.

The methodology employed during this investigation included bioinformatic approaches such as the mapping of SNP location onto regions of open chromatin and TF binding site location data. Functional data investigating DNA-protein interactions between SNP DNA and cell type specific nuclear proteins and gene expression analysis upon deletion of SNP containing genomic regions were analysed.

6.2 Expression and activation of the EphA1 receptor

As discussed in Chapter 3, the EphA1 receptor was observed to be predominantly expressed on the cell surface membrane of EphA1 HEK293 cells. Upon activation by the ephrinA1 ligand, receptor internalisation with the absence of receptor C-terminal degradation was observed. Similarly, to other members of the Eph RTK family, this thesis demonstrates release of a soluble EphA1 product into the cell culture media of both EphA1 and P460L HEK293 cell cultures with the variant, however producing significantly less (Mancia and Shapiro 2005; Lisabeth et al. 2013). The amount of soluble EphA1 receptor released under homeostatic and ligand activated conditions where comparable. Levels of soluble receptor where also not altered upon addition of a broad spectrum MMP inhibitor (GM6001) or a γ-secretase inhibitor (DAPT), indicating the protease responsible for this receptor processing was not an MMP inhibited by GM6001 nor γ-secretase.

However, further investigation into MMP processing may be warranted as, despite being a broad spectrum MMP inhibitor, GM6001 does not inhibit ADAM12 with great efficiency. ADAM12 was identified as a binding partner of EphA1 during a yeast two-hybrid screen conducted by Leguchi et al (Ieguchi et al. 2014). Therefore, further investigation into this protease's potential to regulate EphA1 using a more specific ADAM12 inhibitor or a small interfering (siRNA) to knock down ADAM12 expression may prove insightful into EphA1 receptor regulation. The role of other proteases, such as serine proteases have also been linked to the regulation of Eph receptors and may also warrant investigation into their potential involvement in EphA1 receptor processing (Janes et al. 2011). For example, ephrinB3 has been identified as a potential substrate for membrane-associated serine proteases such as Rhomboids (Pascall and Brown 2004). Additionally, secreted serine proteases such as Neuropsin has been implemented on the stress-induced cleavage of EphB2 as part of its regulation of stress-related neuroplasticity (Attwood et al. 2011).

Alternatively, the observed release of soluble EphA1 receptor may be due to membrane blebbing and release of the receptor into vesicles. For example, the stimulus-induced exocytosis of large dense-core vesicles (LDCVs) within the CNS has been associated with multiple receptors, with LDCVs been observed to contain multiple different types of receptors including preassembled receptor signalling complexes (Zhao et al. 2011). The release of full length EphA1 receptor in this manner could be investigated via the use of a western blot on either concentrated or V5-tag purified EphA1 HEK293 cell culture media to determine if the full-length receptor is the released product observed by ELISA.

Remaining possibilities not investigated here to explain the altered subcellular localisation of EphA1 variants, is that the P460L receptor causes a processing or trafficking issue of the receptor protein, preventing its correct migration to, and insertion into, the cell membrane. Another possibility is that the P460L receptor is more rapidly recycled between the membrane and intracellular compartments. The theory of a processing or trafficking issue could be further investigated through the use of the 'cell paint' staining technique and the Perkin Elmer Opera Phenix high content screener (as optimised by Andrew Lloyd, PhD student, Cardiff University). This will allow staining of various cell compartments (Golgi apparatus, mitochondria, nucleus, endothelial reticulum, actin and plasma membrane) to

determine if the P460L EphA1 receptor variant is co-localised, and therefore held, within one of these subcellular compartments. Whether the ephrinA1 ligand is also internalised as part of the receptor-ligand complex, which has been noted in other Eph receptors (Pitulescu and Adams 2010) can also be studied on the Phenix via ephrinA1 staining. Rapid P460L variant receptor recycling could potentially be analysed using imaging flow cytometry via labelling cell surface EphA1 receptor with the N-terminus EphA1 antibody and following its internalisation into the cytosolic compartment over a time course experiment.

Ligand independent activation of the EphA1 receptor is observed through phosphoserine-EphA1 expression under basal conditions. Transient dephosphorylation of serine residues of EphA1 was noted on 1 min ligand incubation, this preceded phosphorylation of tyrosine residues after 5 min ligand incubation. The rare coding P460L variant of the EphA1 receptor that has been associated with AD shows alterations in receptor expression location and activation. The P460L variant causes the majority of receptor expression to be shifted from the cell surface membrane into the cytosol. Analysis of receptor activation by quantitation of phosphotyrosine-EphA1 levels indicate altered receptor activation in response to ligand activation. Unlike the EphA1 receptor the P460L receptor variant shows no significant increase in tyrosine phosphorylation. Ligand-independent receptor activation remains unaltered by the P460L variant.

The consequences of EphA-ephrinA signalling on cell migration appears to be extremely cell type dependent. However, using the identified role of EphA1s closest homolog, EphA2 in the regulation of the migration of T-cells, a role for EphA1 in the regulation of immune cell trafficking across the BBB can be hypothesised along with how alterations to this regulation as a result of the P460L receptor variant may affect AD pathology (Fig.6.1) (Sharfe et al. 2008).

EphA-ephrinA signalling plays a role in T-cell trafficking and migration into lymph nodes through the regulation of integrin-mediated cell adhesions. It has been observed that increased activation of EphA2 inhibits T-cell interactions with endothelial cells, as well as the suppression of EphA2 activity significantly increased T-cell-cell interactions, promoting migration into peripheral tissues.

Therefore, it can be hypothesis that the activation of EphA1 via forward signalling, described in figure 3.1 causes the inhibition of immune cell adhesion (Fig.6.1A). Decreased membrane receptor expression in combination with no observed ligand activation response produced as a consequence of the P460L receptor variant may result in increased immune cell adhesion to endothelial cells. This is potentially mediated through the EphA receptors ability to modulate conformational changes within integrins causing interactions with adhesion molecules such as VCAM-1 and ICAM-1 on the endothelial cells of the BBB (Fig.6.1B). This increased immune cell adhesion to the BBB may promote cell trafficking across the BBB and exacerbate the chronic inflammation observed during AD pathology.
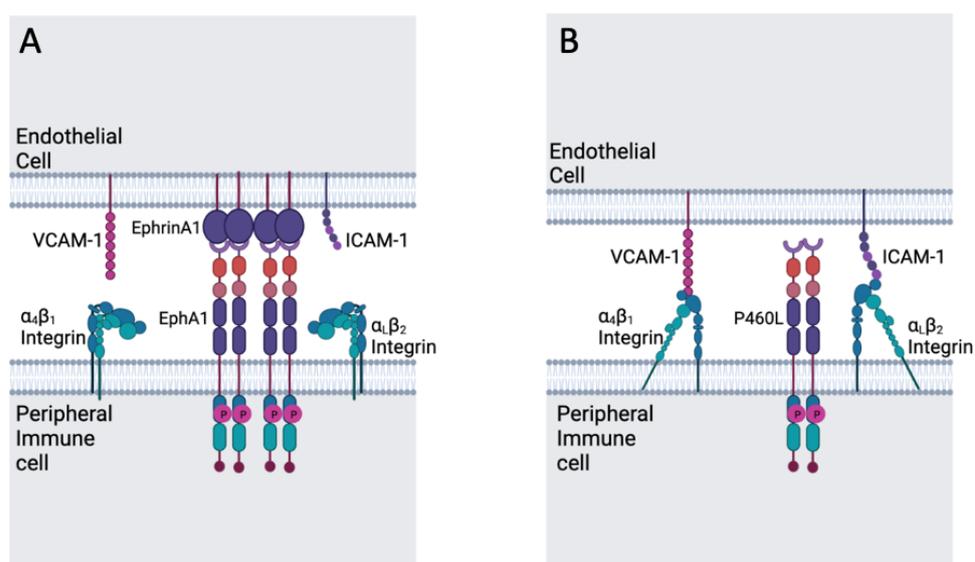
Figure 6.1: Proposed hypothesis of EphA1 regulation of peripheral immune cell trafficking across the BBB.
A. EphA1 activation blocks integrin-mediated cell adhesion of immune cells and endothelial cells. **B.** Reduction of EphA1 receptor membrane expression as a consequence of the P460L variant results causes immune cell adhesion to endothelial cells via integrins, ICAM and VCAM1 adhesion molecules. This causes increased trafficking of immune cells across the BBB, exacerbating AD pathology through perpetuating inflammation. Schematic created in BioRender.com

Structural model predictions and electropositive calculations suggest that EphA1 P460 lies within an electropositive patch of the FN2 domain, partially embedded within the membrane. Therefore, could suggest that the P460L variant may play a role in stabilizing the EphA1 ectodomain onto the lipid bilayer through enhanced hydrophobic contacts (Kim 2021). Increased stabilisation may result in the P460L receptor variant being resistant to

protease processing, this hypothesis potentially correlates with data from this thesis showing the P460L variant to produce significantly reduced levels of soluble receptor than the wild type EphA1 receptor (Fig. 3.11A and B).

It was suggested that this enhanced stabilisation of the EphA1 receptor at the cell membrane as a consequence of the P460L variant promotes clustering and receptor activation independently of ligand binding. It was therefore hypothesised by Lefort et al that the P460L variant results in the EphA1 receptor being constitutively active. This was demonstrated by increased levels of phosophotyrosine-EphA1 observed during western blot analysis of P460L compared to the WT receptor. Lefort et al also noted only a 60% increase in phosophotyrosine-EphA1 after ligand stimulation, they suggested that this was due to the majority of the tyrosine on the P460L receptor already being phosphorylated (Kim 2021). This increased basal level of tyrosine phosphorylation potentially explains the lack of increased levels of phosophotyrosine-EphA1 observed on ligand stimulation of our P460L HEK293 cells.

However, on looking at the ratios of phosphotyrosine to total EphA1 protein to allow comparison between the cell lines, no significant difference was noted between P460L and EphA1 HEK293 cell lysate. Indicating that increased basal activity of the P460L receptor compared to the EphA1 receptor was not observed during this thesis. This discrepancy may be potentially due to differences in methodology between this thesis and the analysis conducted by Lefort et al. For example, within their studies a Myc-Flag-tagged EphA1 expression vector in conjunction with immunoprecipitation of the EphA1 receptor was employed prior to western blotting (Kim 2021). Whereas, the Flp-In system was used within this thesis to allow receptor expression within HEK293 in conjunction with a pan-tyrosine antibody during western blotting. The use of immunoprecipitation may allow for the concentration of the EphA1 receptor and therefore a more sensitive read out over the use of total cell lysates, allowing small changes in receptor activation to become apparent.

However, when analysing receptor activation, the use of the FLP-In system to create isogenic HEK293 cell lines containing each receptor variant maybe advantageous. This is due to the Flp-In system being designed to allow the insertion of different cDNAs into the same

genomic location, enabling protein expression to be controlled by the same promoter. The use of expression vectors as those employed by Lefort et al may lead to an overexpression of one variant over another. For example, this may lead to an overexpression of P460L receptor variant in relation to the EphA1 receptor depending on location of genomic insertion. Overexpression of this variant in addition to its inherent increased stability within the cell membrane may result in the creation of an increased number of auto-activated receptor clusters above levels otherwise observed, resulting in inflated instances of ligand-independent activation. Active receptor dimers under ligand-independent conditions have been noted (Artemenko et al. 2008).

6.3 Identification of the causal SNP/s behind the Alzheimer's disease association of EPHA1 loci.

The main goal of this thesis was to provide functional data to supplement the genomic GWAS data surrounding the AD risk locus *EPHA1*. Described in Chapter 4, this thesis identified seven SNPs that reside within regions of open chromatin, three of which lie within known TF binding motifs and one within an CTCF-bound enhancer-like signal. From the SNPs that reside within these regions of interest four were also observed to be able to form allele specific DNA-protein interactions during EMSA analysis (rs11765305, rs11763230, rs11767557, rs7810606). These data, taken together imply that the DNA sequence alterations as a result of these AD-associated SNPs may disrupt gene regulation via effecting TF binding affinity.

For example, the SNP rs11765305 is noted to reside within a binding motif of the TF CEBPB. Its protective G allele was observed during the EMSA to potentially be able to form enhanced DNA-protein binding within monocytes. This data correlates with data published by Amlie-Wolf et al who noted this variant's protective G allele created a stronger binding site for CEBPB (Amlie-Wolf et al. 2019). As discussed in Chapter 4, the CEBPB TF plays a role in the regulation of immune responses. CEBPB is upregulated in mature monocytes and has been noted to be required for the survival of lymphocyte antigen 6C low (ly6C^{low}) monocytes (Tamura et al. 2017). This subset of monocytes also expresses low levels of C-C

chemokine receptor 2 (CCR2) but high levels of CXC3C chemokine receptor 1 (CX3CR1). These CX3CR1^{high}CCR2^{low}Ly6C^{low} monocytes are considered to be patrolling, anti-inflammatory monocytes which survey the vasculature and repair tissues (Kratofil et al. 2017). There is evidence for this monocyte subset being important within AD pathology as their depletion within APP/PS1 mice leads to increased Aβ depositions within the vasculature with increased cognitive impairment (Theriault et al. 2015). CEPB has also been shown to be involved in the regulation of the JAK2/STAT3 signalling axis, with upregulated signalling through this pathway being observed to correlate with high CEBPB levels (Murray 2007). JAK2/STAT3 signalling has been implicated in AD pathology and memory impairment due to its inactivation caused by Aβ. JAK2/STAT3 is also an activator of the protein Humanin which has been identified as neuroprotective, reducing many aspects of AD pathology in vitro such as Aβ build up (Chiba et al. 2009).

The role of the CEBPB TF in the regulation of genes involved in the survival and maintenance of anti-inflammatory monocytes and the regulation of JAK2/STAT3 signalling could explain how alterations in its binding affinity influence AD (Chiba et al. 2009; Wang et al. 2021). For example, as described within this thesis and within published data, the alteration of CEBPB's binding motif by the G allele of the SNP rs11765305 leads to a greater affinity and more binding of proteins to this locus (Amlie-Wolf et al. 2019). This may result in its observed protective phenotype against AD pathology via potentially increasing the anti-inflammatory CX3CR1^{high}CCR2^{low}Ly6C^{low} monocyte subset and leading to increased signalling through the neuroprotective JAK2/STAT3 axis. The DNA-protein binding data presented in this thesis and published functional data from the SNP rs11765305 sets the precedent that DNA sequence alterations within TF binding motifs as a result of AD-associated variants have the ability to alter protein binding affinity and as a consequence influence gene expression (Amlie-Wolf et al. 2019).

As shown in Chapter 4, the risk T allele of the SNP rs7810606 results in increased DNA-protein binding within microglial cells. This variant lies within the binding motif of multiple TFs such as those within the KLF family which have been identified to be involved in immune response and cell survival (Sweet et al. 2018). Also, at this location is the early growth

response 1 (EGR1) TF (Table. 4.3). The EGR1 TF has been identified to accelerate Aβ accumulation within the AD brain through its activation of BACE1 (Qin et al. 2016). It is therefore likely that sequence alterations produced by the risk T allele of rs7810606 results in enhanced deposition of Aβ through creating a stronger binding site for the TF EGR1, increasing BACE1 activity and therefore processing of APP through the amyloidogenic pathway leading to increased AD pathology as a consequence.

Linkage disequilibrium however, may still be a compounding variable during this *EPHA1* variant analysis via EMSA and the assignment of causality. The SNPs rs11763230 and rs11767557 for example have both been identified as potential causal variants at this locus, due to their capacity to form enhanced DNA-protein interactions in an allele specific manner within the cell types analysed. However, these two SNPs are considered to be linked (Fig 4.1), having an LD percentage score of 100% therefore, difficulty would arise when saying for certain which SNP is behind any potential gene expression alteration. However, as noted in figure 4.5 and 4.10 the risk T allele of rs11763230 is capable of forming stronger DNA-protein interactions within monocytes and neurons, whereas it is the protective C allele which is able to bind increased protein within microglia (Fig. 4.9). The ability to potentially regulate gene expression in more than one cell type in an allele specific manner may lend weight to this variant being more important in conferring the AD genetic susceptibility of an individual than its linked partner, rs11767557.

In addition, despite SNP rs11767557 residing within a known TF binding site (ZSCAN31) the role of this TF is relatively unknown therefore the consequence of DNA sequence alterations produced as a result of this SNP remain elusive. The SNP rs11763230 on the other hand is located just outside the binding motif of the TF SIX3. As previously mentioned, alterations to the DNA sequence may also alter gene expression via effecting the binding of regulatory protein complexes despite not residing directly in a TF binding motif itself. *SIX3* is part of the *sine oculis* homeobox family of TFs (Oliver et al. 1995). These TFs have been shown to play key roles in cell survival and proliferation, with SIX3 knockout mice failing to form telencephalic or optic vesicles. SIX3 overexpression also appears to drive metastasis in cancer cells, increasing EMT, cell migration and tumour invasion via indirectly regulating Cadherin-1 (Mo et al. 2013). This implication of the TF SIX3 within roles such as progenitor

cell proliferation and cell migration along with it being identified as a site of DNA methylation (a common epigenetic marker gene repression) within AD hippocampus implies that alterations within this regulatory region that alter the binding affinity of proteins may play a role in AD pathology (Moore et al. 2013; Altuna et al. 2019).

Taken together the bioinformatic and protein binding data presented in Chapter 4 of this thesis successfully triaged the *EPHA1* AD-associated GWAS SNPs down to the four potentially causal SNPs (rs11767557, rs1176230, rs7810606 and rs11765305). Further functional analysis via CRISPR-Cas9 deletion of SNP-containing blocks of non-coding DNA, followed by gene expression analysis was undertaken to determine the impact of these SNPs on gene regulation.

Two blocks of non-coding sequences were deleted from the control KOLF2-C1 iPS cell line to create homozygous deletion cell lines of the SNPs of interest (Tab. 5.4). SNP Block 1 contains the SNPs rs11762262, rs12705526 and rs7810606 whist SNP Block 3 contains the SNPs rs11765305 and rs11771145. Upon deletion of these blocks of SNPs the resulting homozygous deletion iPS cell lines showed phenotypic and morphological alterations. Both deletion cell lines showed little-to-no stem cell like morphology, appearing elongated and flat with characteristics of spontaneously differentiated cultures.

Alterations to protein and gene expression in these deletion iPS cell lines were analysed via western blot and qRTPCR. *EPHA1* and *EPHA1-AS1* were not detected, which strengthens evidence that these genes are not the sole causal genes behind this loci's AD-association. However, both SNP Block 1 and 3 deletion cell lines showed a significant increase in *ZYX* expression, with SNP Block 3 showing the larger increase. The increase in expression observed suggests that the SNPs within these sections of non-coding DNA are capable of affecting *ZYX* gene expression via altering TF binding motif sequences and therefore affinity by which TFs can bind to the DNA. The finding of this thesis correlate with the fine mapping analytic conducted by both Schwartzentruber and Kunckle et al indicating *ZYX* as the risk gene behind the AD-association at this locus (Kunkle et al. 2019b; Schwartzentruber et al. 2021).

Chapter 6: Discussion

As discussed in Chapter 5, although a cytoskeletal protein, *ZYX* has the ability to regulate gene expression of pluripotency genes such as *POU5F3/OCT4* via impacting mRNA stability through the binding of the Y-box factor (Ybx1) (Fig 6.2). Ybx1 functions as a chaperone to protect mRNA transcripts from degradation (Evdokimova et al. 2001). It has also been shown that inhibition of the ZYX protein within human HEK293 cells produces a significant increase in mRNA levels of the pluripotency genes *NANOG, OCT4* and *KLF-4 (Parshina et al. 2020).* This correlates with the observed data within this thesis showing increased *ZYX* expression resulting in a loss of pluripotency as observed in the CRISPR-Cas9 edited cell lines. Additionally, subsequent gene expression analysis of SNP Block 1 and 3 deletion cell shows significant reductions in the pluripotency genes *CDH1* and *NANOG.* Reductions in the expression of such genes required to maintain iPSC pluripotency likely contribute the cell morphological changes and spontaneous differentiation observed during cell line culturing.
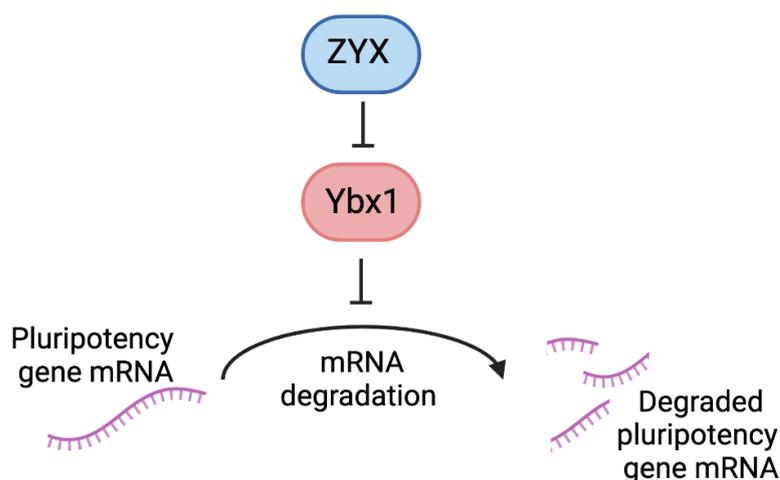
Figure 6.2: Regulation of pluripotency genes by ZYX.
ZYX regulates the expression of a range of pluripotency genes via the inhibition of the mRNA stabilising protein Ybx1. This causes the inhibition of pluripotency genes via mRNA degradation. Schematic created in BioRender.com

The gene expression and EMSA data together suggest that the SNPs responsible are rs7810606 and rs11765305 since these are the only SNPs capable of binding nuclear proteins in AD-related cell types of interest in an allele specific manner.

As mentioned above, the risk T allele of the SNP rs7810606 alters the TF binding motifs at this locus, such as EGR1. The increased binding of EGR1 as a result of this risk allele will

cause exacerbation of Aβ deposition through the activities of the BACE1 enzyme (Qin et al. 2016). As described in Chapter 5, this is therefore likely to result in the exacerbated Aβ-dependent decrease in *ZYX* expression and observed AD-related phenotypes produced as a consequence such as, aberrant DNA damage response, endothelial dysfunction and increased pro-inflammatory response (Crone et al. 2011; Lanni et al. 2013).

In contrast it is the protective G allele of the SNP rs11765305 which results in a higher affinity binding site for TFs such as CEBPB found at this locus (Amlie-Wolf et al. 2019). As mentioned above, this is likely to result in the AD pathology protection attributed to this SNP through enhanced binding of CEBPB. A higher affinity of CEBPB binding results in increased JAK2/STAT3 signalling and regulation of the immune response resulting in an increased population of patrolling, anti-inflammatory monocytes (Tamura et al. 2017; Yang et al. 2022b). This correlates with the cell type specificity of this SNPs allele seen during the EMSA, resulting in enhanced nuclear protein binding only within the monocyte cell line THP1.

6.5 Final Conclusions

Multiple GWAS have identified numerous SNPs to be associated with AD. However, identification of which SNP at any given risk loci is responsible for the disease is confounded by multiple factors, as previously described (Edwards et al. 2013).

The most pertinent finding of this thesis is that certain non-coding variants within the *EPHA1* locus have the capacity to alter expression of the *ZYX* gene through alterations of the DNA sequences at TF binding sites and therefore, TF binding to these motifs. Moreover, this thesis also identified *ZYX* as another risk gene operating at this locus. This is consistent with recent published data noting strong chromatin interactions between the *ZYX* promoter and AD risk variants within monocytes, macrophages and microglia, albeit at different eQTL profiles suggesting cell type specificity but implementing the same causal gene (Podlesny-Drabiniok et al. 2020; Novikova et al. 2021).

Chapter 6: Discussion

Despite *EPHA1* mRNA or protein not being of a sufficient level to allow detection within the AD relevant iPSC-derived cell types analysed, nor its expression altered as a consequence of the deletion of the AD-associated SNPs rs11765305 and rs7810606, data observed within this thesis in addition to that recently published suggest that the EphA1 receptor does still play a role during disease pathology. This is evident due to the alterations in receptor biology that is produced as a consequence of the AD-associated P460L receptor variant. The EphA1 receptor has also been shown to have enhancer interactions with AD risk variants within monocyte-derived macrophage. This therefore aligns with the discovery of the *EPHA1* risk loci being identified as an AD-associated locus containing multiple signals on network analysis and quantitative gene prioritisation. Taken together these data implement more than one gene acting at this loci to convey its association with Alzheimer's Disease pathology (Schwartzentruber et al. 2021).

The amalgamation of bioinformatics data overlaying SNP locations to that of TF binding motifs and open chromatin regions has allowed the successful triaging of numerous AD-associated SNPs at the *EPHA1* locus. This thesis strongly suggests that the two SNPs rs11765305 and rs7810606 as those responsible for the AD-association at this locus. Located within SNP block 3 and 1 respectively, these SNPs are the only variants within their respective deleted regions to be capable of binding regulatory nuclear proteins in an allele specific manner. Therefore, the resulting increase in *ZYX* expression within the CRISPR-Cas9 deletion cell lines must be a consequence of their genotype.

However, it was noted that the SNPs sr11767557 and rs11763230 also have the capacity to bind nuclear proteins within cell types of interest during the EMSA. Due to the failure of the CRISPR-Cas9 process to delete this section of non-coding DNA, further work will be required to identify whether these two SNPs residing within the deleted Block 2 non-coding DNA are also capable of controlling gene regulation of *ZYX* or indeed *EPHA1* and *EPHA1-AS1*. It should be noted however that despite this observed protein binding ability the SNP rs11763230 does not reside within any known TF binding motifs. Additionally, rs11767557 only resides within the binding site of the TF ZSCAN31 whose role remains largely unknown. This is in contrast to the SNPs rs11765305 and rs7810606 which are located within known TF binding motifs which have been shown to be involved in aspects of immune cell

regulation as described above (Ruffell et al. 2009; McConnell and Yang 2010). Therefore, alterations to the DNA sequence and subsequently the binding affinity of the TFs found at these locations produced by the variants rs11765305 and rs7810606 have a far greater likelihood of being the causal SNPs behind the AD-association at the *EPHA1* locus.

This thesis has identified alterations in EphA1 receptor biology as a consequence of the P460L variant, suggesting that this gene although only observed to be expressed at low levels within vascular and brain cells receptor signalling maybe linked to disease pathology.

Altered receptor activation was observed within this thesis, with the P460L receptor showing no significant response to ephrinA1 ligand addition. This may be a consequence of the tyrosine residues already being phosphorylated due to increased basal activity as described above. As described in Chapter 3, ligand induced forward signalling of the EphA1 receptor leads to an inhibition of cell migration, therefore, the reduced response of the P460L receptor variant observed may lead to increased cell migration and invasion. For example, increased migration of peripheral immune cells across the BBB and into the brain, perpetuating chronic inflammation noted during AD pathology. Altered or disrupted receptor processing was also noted, leading to decreased receptor insertion into and therefore, expression at the cell surface membrane in combination with reduced basal soluble receptor release. However, it is not known whether this soluble form of the receptor released is the full-length receptor or an N-terminal domain following cleavage by a protease not investigated during this thesis.

EPHA1 expression was only observed at low levels within a small population of CD8 and CD4+ T cells along with NK cells in single cell RNA-Seq data analysed by Birong Zhang. This was confirmed by data noted within The Human Protein Atlas, showing very low *EPHA1* expression within a small subset of vascular and brain cells. A lack of enrichment of both *EPHA1* and its antisense gene, *EPHA1-AS1* was also noted in numerous BBB cells such as endothelial and smooth muscle cells as well as pericytes (Yang et al. 2022a). In contrast, high expression levels of *ZYX* was noted in all of vascular and brain cell types analysed.

However, this thesis only looked in cells under homeostatic conditions which may not be a physiologically relevant model of cell behaviour for a disease such as AD. Indeed, studies have shown an increase of *EPHA1* within reactive astrocytes, neurons and oligodendrocytes (Wang and Ye 2021). However, single cell RNA-Seq data showing alterations in *EPHA1, EPHA1-AS1* and *ZYX* expression during disease conditions such as AD or inflammation are not present within the databases analysed. Therefore, to fully understand the role that the EphA1 receptor plays during AD pathology, and how reductions in membrane expression and release of soluble receptor may contribute to disease pathology it be important to study expression level alterations under inflammatory conditions. The role of the P460L variant within AD pathology could be explained if this increase in *EPHA1* expression by brain cells during inflammation is protective. For example, as described in Chapter 3 ligand independent EphA1 activity promotes cell migration and invasion.

Therefore, it is plausible that the increased *EPHA1* expression in reactive astrocytes causes ligand-independent EphA1 receptor activation resulting in increased cell migration to and clearance of Aβ deposits, as these cells along with microglia play important roles in Aβ clearance and degradation. This hypothesis would explain how reduced cell surface receptor expression on astrocytes as a consequence of the P460L variant may promote AD pathology through reduced Aβ clearance.

6.6 Future Directions

6.6.1 The role of EPHA1 within Alzheimer's disease

With the observed increase in *EPHA1* expression being noted in T cells, reactive astrocytes and neurons the role of this receptor during disease pathology may only be apparent under inflammatory conditions (Nestor et al. 2007). Therefore, the addition of an inflammatory cytokine found to be increased during AD, such as TNFα or IL-1β, prior to gene expression analysis or receptor characterisation may yield pertinent findings. Alternatively, the addition of Aβ oligomers may allow a physiologically relevant cell inflammation response.

6.6.2 The role of ZYX within Alzheimer's disease

As shown in Chapter 5, *ZYX* appears to play a role in the regulation of stem cell pluripotency. IPS cell lines in which disease-associated SNPs have been deleted via CRISPR-Cas9 editing resulted increased *ZYX* expression. This appeared to correlate with a loss of stemness, both in iPSC appearance and behaviour in addition to the decreased expression of the pluripotency genes *NANOG* and *CDH1*. Confirmation of this gene's involvement in the regulation of pluripotency could be obtained through the use of a *ZYX* expression vector. Transfection into a control iPS cell line would allow the consequences of *ZYX* over expression on cell pluripotency and alterations in expression of pluripotency genes to be analysed without subjecting the cells to external stressors such as CRISPR-Cas9 editing which has the potential to affect the stemness of sensitive IPS cell cultures.

The role that *ZYX* plays in relation to its involvement in cytoskeletal rearrangements and cell motility aimed to be investigated through the creation of a *ZYX* knockout IPS cell line. This would allow the analysis of the potential roles of *ZYX* expression alterations as a result of the AD-associated SNPs as described in Chapter 5. Such as, increased peripheral blood cell migration across the BBB through loss of *ZYX* at focal adhesions or through the increased expression of *ZYX*-dependent genes which may perpetuate chronic inflammation (Nix and Beckerle 1997; Cattaruzza et al. 2004; Hirata et al. 2008a). However, due to time constraints the failed creation of homozygous *ZYX* and SNP Block 2 deletion cell lines via CRISPR-Cas9 technology were not followed up. A potential explanation for the failure to generate knockout cell lines lies with the inability to create viable cells upon the deletion of either the *ZYX* gene or SNP Block 2 DNA. This may be of particular importance for SNP Block 2, since deletion of SNP blocks 1 and 3 showed evidence of limited iPS cell pluripotency. A potential method to circumvent this issue would be the employment of a drug-inducible CRISPR-Cas9 system. This would allow control over the timing of any genome editing, allowing the iPS cell lines to be differentiated partly or fully into the cell type of interest before gene or DNA region deletion induction (Sun et al. 2019). The use of siRNA to allow efficient *ZYX* knockdown may be an alternative time and cost-effective method to allow *ZYX* knockdown in cell types of interest (Liu et al. 2020).

6.6.3 Analysis of AD-associated EPHA1 locus SNPs

Further investigation into the identity of the bound nuclear proteins within DNA-protein interactions demonstrated in this thesis may prove beneficial in understanding the signalling pathways via which they act and the potential consequences this may have in terms of gene regulation and AD pathology. In order to address this, the EMSA could be paired with mass spectrometry. Combination of a fast and simple assay such as the EMSA with the high-sensitivity and resolution of mass spectrometry would enable the unbiased identification of protein interactors. Such analysis would require the excision of the DNA-protein band of interest and the protein components identified via tryptic digestion and mass spectrometry analysis of the resulting protein mixture (Cozzolino et al. 2021). A confirmation super shift EMSA reaction may then be performed. A super shift EMSA includes an extra control reaction containing an antibody against the suspected protein forming the DNA-protein interaction. If this protein is involved a further larger band will be noted due to the increase size of the DNA-protein-antibody complex formed (Koh and Jeong 2016).

Bibliography

Aasheim, H. C., Delabie, J. and Finne, E. F. 2005. Ephrin-A1 binding to CD4+ T lymphocytes stimulates migration and induces tyrosine phosphorylation of PYK2. *Blood* 105(7), pp. 2869-2876. doi: 10.1182/blood-2004-08-2981

Alasoo, K. et al. 2015. Transcriptional profiling of macrophages derived from monocytes and iPS cells identifies a conserved response to LPS and novel alternative transcription. *Sci Rep* 5, p. 12524. doi: 10.1038/srep12524

Altuna, M. et al. 2019. DNA methylation signature of human hippocampus in Alzheimer's disease is linked to neurogenesis. *Clin Epigenetics* 11(1), p. 91. doi: 10.1186/s13148-019-0672-7

Amlie-Wolf, A. et al. 2019. Inferring the Molecular Mechanisms of Noncoding Alzheimer's Disease-Associated Genetic Variants. *J Alzheimers Dis* 72(1), pp. 301-318. doi: 10.3233/JAD-190568

Arnold, S. E., Hyman, B. T., Flory, J., Damasio, A. R. and Van Hoesen, G. W. 1991. The topographical and neuroanatomical distribution of neurofibrillary tangles and neuritic plaques in the cerebral cortex of patients with Alzheimer's disease. *Cereb Cortex* 1(1), pp. 103-116. doi: 10.1093/cercor/1.1.103

Arriagada, P. V., Growdon, J. H., Hedley-Whyte, E. T. and Hyman, B. T. 1992. Neurofibrillary tangles but not senile plaques parallel duration and severity of Alzheimer's disease. *Neurology* 42(3 Pt 1), pp. 631-639. doi: 10.1212/wnl.42.3.631

Artemenko, E. O., Egorova, N. S., Arseniev, A. S. and Feofanov, A. V. 2008. Transmembrane domain of EphA1 receptor forms dimers in membrane-like environment. *Biochim Biophys Acta* 1778(10), pp. 2361-2367. doi: 10.1016/j.bbamem.2008.06.003

Arvanitakis, Z., Capuano, A. W., Leurgans, S. E., Bennett, D. A. and Schneider, J. A. 2016. Relation of cerebral vessel disease to Alzheimer's disease dementia and cognitive function in elderly people: a cross-sectional study. *Lancet Neurol* 15(9), pp. 934-943. doi: 10.1016/S1474-4422(16)30029-1

Atapattu, L., Lackmann, M. and Janes, P. W. 2014. The role of proteases in regulating Eph/ephrin signaling. *Cell Adh Migr* 8(4), pp. 294-307. doi: 10.4161/19336918.2014.970026

Augustinack, J. C., Schneider, A., Mandelkow, E. M. and Hyman, B. T. 2002. Specific tau phosphorylation sites correlate with severity of neuronal cytopathology in Alzheimer's disease. *Acta Neuropathol* 103(1), pp. 26-35. doi: 10.1007/s004010100423

Bibliography

Baik, S. H. et al. 2014. Migration of neutrophils targeting amyloid plaques in Alzheimer's disease mouse model. *Neurobiol Aging* 35(6), pp. 1286-1292. doi: 10.1016/j.neurobiolaging.2014.01.003

Bannister, A. J. and Kouzarides, T. 2011. Regulation of chromatin by histone modifications. *Cell Res* 21(3), pp. 381-395. doi: 10.1038/cr.2011.22

Bartova, E., Krejci, J., Harnicarova, A., Galiova, G. and Kozubek, S. 2008. Histone modifications and nuclear architecture: a review. *J Histochem Cytochem* 56(8), pp. 711-721. doi: 10.1369/jhc.2008.951251

Bauer, J. et al. 1991. Interleukin-6 and alpha-2-macroglobulin indicate an acute-phase state in Alzheimer's disease cortices. *FEBS Lett* 285(1), pp. 111-114. doi: 10.1016/0014-5793(91)80737-n

Baumgart, M., Snyder, H. M., Carrillo, M. C., Fazio, S., Kim, H. and Johns, H. 2015. Summary of the evidence on modifiable risk factors for cognitive decline and dementia: A population-based perspective. *Alzheimers Dement* 11(6), pp. 718-726. doi: 10.1016/j.jalz.2015.05.016

Beauchamp, A., Lively, M. O., Mintz, A., Gibo, D., Wykosky, J. and Debinski, W. 2012. EphrinA1 is released in three forms from cancer cells by matrix metalloproteases. *Mol Cell Biol* 32(16), pp. 3253-3264. doi: 10.1128/MCB.06791-11

Becker, S. B., J. 2021. TALE and TALEN genome editing technologies. *Gene and Genome Editing* 2,

Beckerle, M. C. 1986. Identification of a new protein localized at sites of cell-substrate adhesion. *J Cell Biol* 103(5), pp. 1679-1687. doi: 10.1083/jcb.103.5.1679

Bedzhov, I., Alotaibi, H., Basilicata, M. F., Ahlborn, K., Liszewska, E., Brabletz, T. and Stemmler, M. P. 2013. Adhesion, but not a specific cadherin code, is indispensable for ES cell and induced pluripotency. *Stem Cell Res* 11(3), pp. 1250-1263. doi: 10.1016/j.scr.2013.08.009

Begay, V., Baumeier, C., Zimmermann, K., Heuser, A. and Leutz, A. 2018. The C/EBPbeta LIP isoform rescues loss of C/EBPbeta function in the mouse. *Sci Rep* 8(1), p. 8417. doi: 10.1038/s41598-018-26579-y

Bell, A. C., West, A. G. and Felsenfeld, G. 1999. The protein CTCF is required for the enhancer blocking activity of vertebrate insulators. *Cell* 98(3), pp. 387-396. doi: 10.1016/s0092-8674(00)81967-4

Bell, R. D., Winkler, E. A., Sagare, A. P., Singh, I., LaRue, B., Deane, R. and Zlokovic, B. V. 2010. Pericytes control key neurovascular functions and neuronal phenotype in the adult brain and during brain aging. *Neuron* 68(3), pp. 409-427. doi: 10.1016/j.neuron.2010.09.043

Bibliography

Bellenguez, C., Grenier-Boley, B. and Lambert, J. C. 2020. Genetics of Alzheimer's disease: where we are, and where we are going. *Curr Opin Neurobiol* 61, pp. 40-48. doi: 10.1016/j.conb.2019.11.024

Bolotin, A., Quinquis, B., Sorokin, A. and Ehrlich, S. D. 2005. Clustered regularly interspaced short palindrome repeats (CRISPRs) have spacers of extrachromosomal origin. *Microbiology (Reading)* 151(Pt 8), pp. 2551-2561. doi: 10.1099/mic.0.28048-0

Braak, E., Braak, H. and Mandelkow, E. M. 1994. A sequence of cytoskeleton changes related to the formation of neurofibrillary tangles and neuropil threads. *Acta Neuropathol* 87(6), pp. 554-567. doi: 10.1007/BF00293315

Braak, H. and Braak, E. 1991. Neuropathological stageing of Alzheimer-related changes. *Acta Neuropathol* 82(4), pp. 239-259. doi: 10.1007/BF00308809

Braak, H. and Braak, E. 1997. Frequency of stages of Alzheimer-related lesions in different age categories. *Neurobiol Aging* 18(4), pp. 351-357. doi: 10.1016/s0197-4580(97)00056-0

Bradbury, M. W. 1984. The structure and function of the blood-brain barrier. *Fed Proc* 43(2), pp. 186-190.

Brantley-Sieders, D. M., Caughron, J., Hicks, D., Pozzi, A., Ruiz, J. C. and Chen, J. 2004. EphA2 receptor tyrosine kinase regulates endothelial cell migration and vascular assembly through phosphoinositide 3-kinase-mediated Rac1 GTPase activation. *J Cell Sci* 117(Pt 10), pp. 2037-2049. doi: 10.1242/jcs.01061

Broach, J. R., Guarascio, V. R. and Jayaram, M. 1982. Recombination within the yeast plasmid 2mu circle is site-specific. *Cell* 29(1), pp. 227-234. doi: 10.1016/0092-8674(82)90107-6

Broekema, R. V., Bakker, O. B. and Jonkers, I. H. 2020. A practical view of fine-mapping and gene prioritization in the post-genome-wide association era. *Open Biol* 10(1), p. 190221. doi: 10.1098/rsob.190221

Brundin, P., Melki, R. and Kopito, R. 2010. Prion-like transmission of protein aggregates in neurodegenerative diseases. *Nat Rev Mol Cell Biol* 11(4), pp. 301-307. doi: 10.1038/nrm2873

Calsolaro, V. and Edison, P. 2016. Neuroinflammation in Alzheimer's disease: Current evidence and future directions. *Alzheimers Dement* 12(6), pp. 719-732. doi: 10.1016/j.jalz.2016.02.010

Cao, Z. et al. 2009. Kruppel-like factor KLF10 targets transforming growth factor-beta1 to regulate CD4(+)CD25(-) T cells and T regulatory cells. *J Biol Chem* 284(37), pp. 24914-24924. doi: 10.1074/jbc.M109.000059

Bibliography

Carrano, A., Hoozemans, J. J., van der Vies, S. M., van Horssen, J., de Vries, H. E. and Rozemuller, A. J. 2012. Neuroinflammation and blood-brain barrier changes in capillary amyloid angiopathy. *Neurodegener Dis* 10(1-4), pp. 329-331. doi: 10.1159/000334916

Carrieri, C. et al. 2012. Long non-coding antisense RNA controls Uchl1 translation through an embedded SINEB2 repeat. *Nature* 491(7424), pp. 454-457. doi: 10.1038/nature11508

Caspers, S. et al. 2020. Pathway-Specific Genetic Risk for Alzheimer's Disease Differentiates Regional Patterns of Cortical Atrophy in Older Adults. *Cereb Cortex* 30(2), pp. 801-811. doi: 10.1093/cercor/bhz127

Cattaruzza, M., Lattrich, C. and Hecker, M. 2004. Focal adhesion protein zyxin is a mechanosensitive modulator of gene expression in vascular smooth muscle cells. *Hypertension* 43(4), pp. 726-730. doi: 10.1161/01.HYP.0000119189.82659.52

Chandler, H. L., Wise, R. G., Murphy, K., Tansey, K. E., Linden, D. E. J. and Lancaster, T. M. 2019. Polygenic impact of common genetic risk loci for Alzheimer's disease on cerebral blood flow in young individuals. *Sci Rep* 9(1), p. 467. doi: 10.1038/s41598-018-36820-3

Chen, F. et al. 2018. Activation of EphA4 induced by EphrinA1 exacerbates disruption of the blood-brain barrier following cerebral ischemia-reperfusion via the Rho/ROCK signaling pathway. *Exp Ther Med* 16(3), pp. 2651-2658. doi: 10.3892/etm.2018.6460

Chen, G. F., Xu, T. H., Yan, Y., Zhou, Y. R., Jiang, Y., Melcher, K. and Xu, H. E. 2017. Amyloid beta: structure, biology and structure-based therapeutic development. *Acta Pharmacol Sin* 38(9), pp. 1205-1235. doi: 10.1038/aps.2017.28

Chiba, T., Yamada, M., Sasabe, J., Terashita, K., Shimoda, M., Matsuoka, M. and Aiso, S. 2009. Amyloid-beta causes memory impairment by disturbing the JAK2/STAT3 axis in hippocampal neurons. *Mol Psychiatry* 14(2), pp. 206-222. doi: 10.1038/mp.2008.105

Chin, C. L. et al. 2019. A human expression system based on HEK293 for the stable production of recombinant erythropoietin. *Sci Rep* 9(1), p. 16768. doi: 10.1038/s41598-019-53391-z

Clapier, C. R. and Cairns, B. R. 2009. The biology of chromatin remodeling complexes. *Annu Rev Biochem* 78, pp. 273-304. doi: 10.1146/annurev.biochem.77.062706.153223

Claussnitzer, M. et al. 2015. FTO Obesity Variant Circuitry and Adipocyte Browning in Humans. *N Engl J Med* 373(10), pp. 895-907. doi: 10.1056/NEJMoa1502214

Clavaguera, F. et al. 2009. Transmission and spreading of tauopathy in transgenic mouse brain. *Nat Cell Biol* 11(7), pp. 909-913. doi: 10.1038/ncb1901

Consortium, E. P. 2004. The ENCODE (ENCyclopedia Of DNA Elements) Project. *Science* 306(5696), pp. 636-640. doi: 10.1126/science.1105136

Bibliography

Consortium, E. P. 2012. An integrated encyclopedia of DNA elements in the human genome. *Nature* 489(7414), pp. 57-74. doi: 10.1038/nature11247

Consortium, F. et al. 2014. A promoter-level mammalian expression atlas. *Nature* 507(7493), pp. 462-470. doi: 10.1038/nature13182

Coon, K. D. et al. 2007. A high-density whole-genome association study reveals that APOE is the major susceptibility gene for sporadic late-onset Alzheimer's disease. *J Clin Psychiatry* 68(4), pp. 613-618. doi: 10.4088/jcp.v68n0419

Cooper, S., Schwartzenruber, J., Coomber, E. L., Wu, Q. and Bassett, A. 2022. Screening for functional regulatory variants in open chromatin using GenIE-ATAC. *bioRxiv*,

Coulthard, M. G. et al. 2001. Characterization of the Epha1 receptor tyrosine kinase: expression in epithelial tissues. *Growth Factors* 18(4), pp. 303-317. doi: 10.3109/08977190109029118

Cozzolino, F., Iacobucci, I., Monaco, V. and Monti, M. 2021. Protein-DNA/RNA Interactions: An Overview of Investigation Methods in the -Omics Era. *J Proteome Res* 20(6), pp. 3018-3030. doi: 10.1021/acs.jproteome.1c00074

Crawford, A. W., Michelsen, J. W. and Beckerle, M. C. 1992. An interaction between zyxin and alpha-actinin. *J Cell Biol* 116(6), pp. 1381-1393. doi: 10.1083/jcb.116.6.1381

Crone, J., Glas, C., Schultheiss, K., Moehlenbrink, J., Krieghoff-Henning, E. and Hofmann, T. G. 2011. Zyxin is a critical regulator of the apoptotic HIPK2-p53 signaling axis. *Cancer Res* 71(6), pp. 2350-2359. doi: 10.1158/0008-5472.CAN-10-3486

Cruz Hernandez, J. C. et al. 2019. Neutrophil adhesion in brain capillaries reduces cortical blood flow and impairs memory function in Alzheimer's disease mouse models. *Nat Neurosci* 22(3), pp. 413-420. doi: 10.1038/s41593-018-0329-4

Darling, T. K. and Lamb, T. J. 2019. Emerging Roles for Eph Receptors and Ephrin Ligands in Immunity. *Front Immunol* 10, p. 1473. doi: 10.3389/fimmu.2019.01473

Das, G., Yu, Q., Hui, R., Reuhl, K., Gale, N. W. and Zhou, R. 2016. EphA5 and EphA6: regulation of neuronal and spine morphology. *Cell Biosci* 6, p. 48. doi: 10.1186/s13578-016-0115-5

De Luca, V., Orfei, M. D., Gaudenzi, S., Caltagirone, C. and Spalletta, G. 2016. Inverse effect of the APOE epsilon4 allele in late- and early-onset Alzheimer's disease. *Eur Arch Psychiatry Clin Neurosci* 266(7), pp. 599-606. doi: 10.1007/s00406-015-0663-4

de Vries, H. E., Blom-Roosemalen, M. C., van Oosten, M., de Boer, A. G., van Berkel, T. J., Breimer, D. D. and Kuiper, J. 1996. The influence of cytokines on the integrity of the blood-brain barrier in vitro. *J Neuroimmunol* 64(1), pp. 37-43. doi: 10.1016/0165-5728(95)00148-4

Bibliography

Deane, R. et al. 2008. apoE isoform-specific disruption of amyloid beta peptide clearance from mouse brain. *J Clin Invest* 118(12), pp. 4002-4013. doi: 10.1172/JCI36663

Degenhardt, Y. Y. and Silverstein, S. 2001. Interaction of zyxin, a focal adhesion protein, with the e6 protein from human papillomavirus type 6 results in its nuclear translocation. *J Virol* 75(23), pp. 11791-11802. doi: 10.1128/JVI.75.23.11791-11802.2001

Degner, J. F. et al. 2012. DNase I sensitivity QTLs are a major determinant of human expression variation. *Nature* 482(7385), pp. 390-394. doi: 10.1038/nature10808

Deplancke, B., Alpern, D. and Gardeux, V. 2016. The Genetics of Transcription Factor DNA Binding Variation. *Cell* 166(3), pp. 538-554. doi: 10.1016/j.cell.2016.07.012

Derrien, T. et al. 2012. The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. *Genome Res* 22(9), pp. 1775-1789. doi: 10.1101/gr.132159.111

DiCarlo, G., Wilcock, D., Henderson, D., Gordon, M. and Morgan, D. 2001. Intrahippocampal LPS injections reduce Abeta load in APP+PS1 transgenic mice. *Neurobiol Aging* 22(6), pp. 1007-1012. doi: 10.1016/s0197-4580(01)00292-5

Dimas, A. S. et al. 2009. Common regulatory variation impacts gene expression in a cell type-dependent manner. *Science* 325(5945), pp. 1246-1250. doi: 10.1126/science.1174148

Diniz, B. S. and Teixeira, A. L. 2011. Brain-derived neurotrophic factor and Alzheimer's disease: physiopathology and beyond. *Neuromolecular Med* 13(4), pp. 217-222. doi: 10.1007/s12017-011-8154-x

Doane, A. S. and Elemento, O. 2017. Regulatory elements in molecular networks. *Wiley Interdiscip Rev Syst Biol Med* 9(3), doi: 10.1002/wsbm.1374

Donahue, J. E. et al. 2006. RAGE, LRP-1, and amyloid-beta protein in Alzheimer's disease. *Acta Neuropathol* 112(4), pp. 405-415. doi: 10.1007/s00401-006-0115-3

Dreyfuss, G., Matunis, M. J., Pinol-Roma, S. and Burd, C. G. 1993. hnRNP proteins and the biogenesis of mRNA. *Annu Rev Biochem* 62, pp. 289-321. doi: 10.1146/annurev.bi.62.070193.001445

Duthey, B. 2013. Alzheimer Disease and other Dementias. *Priority Medicines for Europe and the World*
"A Public Health Approach to Innovation",

Edison, P. et al. 2008. Microglia, amyloid, and cognition in Alzheimer's disease: An [11C](R)PK11195-PET and [11C]PIB-PET study. *Neurobiol Dis* 32(3), pp. 412-419. doi: 10.1016/j.nbd.2008.08.001

Bibliography

Edwards, S. L., Beesley, J., French, J. D. and Dunning, A. M. 2013. Beyond GWASs: illuminating the dark road from association to function. *Am J Hum Genet* 93(5), pp. 779-797. doi: 10.1016/j.ajhg.2013.10.012

Elnitski, L., Jin, V. X., Farnham, P. J. and Jones, S. J. 2006. Locating mammalian transcription factor binding sites: a survey of computational and experimental techniques. *Genome Res* 16(12), pp. 1455-1464. doi: 10.1101/gr.4140006

Escott-Price, V., Shoai, M., Pither, R., Williams, J. and Hardy, J. 2017. Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease. *Neurobiol Aging* 49, pp. 214 e217-214 e211. doi: 10.1016/j.neurobiolaging.2016.07.018

Escott-Price, V. et al. 2015. Common polygenic variation enhances risk prediction for Alzheimer's disease. *Brain* 138(Pt 12), pp. 3673-3684. doi: 10.1093/brain/awv268

Evdokimova, V., Ruzanov, P., Imataka, H., Raught, B., Svitkin, Y., Ovchinnikov, L. P. and Sonenberg, N. 2001. The major mRNA-associated protein YB-1 is a potent 5' cap-dependent mRNA stabilizer. *EMBO J* 20(19), pp. 5491-5502. doi: 10.1093/emboj/20.19.5491

Fagiolo, U. et al. 1993. Increased cytokine production in mononuclear cells of healthy elderly people. *Eur J Immunol* 23(9), pp. 2375-2378. doi: 10.1002/eji.1830230950

Fang, Y., Wu, D. and Birukov, K. G. 2019. Mechanosensing and Mechanoregulation of Endothelial Cell Functions. *Compr Physiol* 9(2), pp. 873-904. doi: 10.1002/cphy.c180020

Fillit, H., Ding, W. H., Buee, L., Kalman, J., Altstiel, L., Lawlor, B. and Wolf-Klein, G. 1991. Elevated circulating tumor necrosis factor levels in Alzheimer's disease. *Neurosci Lett* 129(2), pp. 318-320. doi: 10.1016/0304-3940(91)90490-k

Fullard, J. F. et al. 2017. Open chromatin profiling of human postmortem brain infers functional roles for non-coding schizophrenia loci. *Hum Mol Genet* 26(10), pp. 1942-1951. doi: 10.1093/hmg/ddx103

Gale, N. W. et al. 1996. Eph receptors and ligands comprise two major specificity subclasses and are reciprocally compartmentalized during embryogenesis. *Neuron* 17(1), pp. 9-19. doi: 10.1016/s0896-6273(00)80276-7

Ghafouri-Fard, S., Dashti, S. and Taheri, M. 2020. The HOTTIP (HOXA transcript at the distal tip) lncRNA: Review of oncogenic roles in human. *Biomed Pharmacother* 127, p. 110158. doi: 10.1016/j.biopha.2020.110158

Giannakopoulos, P. et al. 2003. Tangle and neuron numbers, but not amyloid load, predict cognitive status in Alzheimer's disease. *Neurology* 60(9), pp. 1495-1500. doi: 10.1212/01.wnl.0000063311.58879.01

Godbout, J. P., Chen, J., Abraham, J., Richwine, A. F., Berg, B. M., Kelley, K. W. and Johnson, R. W. 2005. Exaggerated neuroinflammation and sickness behavior in aged mice following

activation of the peripheral innate immune system. *FASEB J* 19(10), pp. 1329-1331. doi: 10.1096/fj.05-3776fje

Goh, L. K. and Sorkin, A. 2013. Endocytosis of receptor tyrosine kinases. *Cold Spring Harb Perspect Biol* 5(5), p. a017459. doi: 10.1101/cshperspect.a017459

Goldeck, D. et al. 2013. Enhanced Chemokine Receptor Expression on Leukocytes of Patients with Alzheimer's Disease. *PLoS One* 8(6), p. e66664. doi: 10.1371/journal.pone.0066664

Gong, Y. et al. 2003. Alzheimer's disease-affected brain: presence of oligomeric A beta ligands (ADDLs) suggests a molecular basis for reversible memory loss. *Proc Natl Acad Sci U S A* 100(18), pp. 10417-10422. doi: 10.1073/pnas.1834302100

Gorsche, R., Jovanovic, B., Gudynaite-Savitch, L., Mach, R. L. and Mach-Aigner, A. R. 2014. A highly sensitive in vivo footprinting technique for condition-dependent identification of cis elements. *Nucleic Acids Res* 42(1), p. e1. doi: 10.1093/nar/gkt883

Gosselin, D. et al. 2017. An environment-dependent transcriptional network specifies human microglia identity. *Science* 356(6344), doi: 10.1126/science.aal3222

Grange, T., Bertrand, E., Espinas, M. L., Fromont-Racine, M., Rigaud, G., Roux, J. and Pictet, R. 1997. In vivo footprinting of the interaction of proteins with DNA and RNA. *Methods* 11(2), pp. 151-163. doi: 10.1006/meth.1996.0401

Grubert, F. et al. 2015. Genetic Control of Chromatin States in Humans Involves Local and Distal Chromosomal Interactions. *Cell* 162(5), pp. 1051-1065. doi: 10.1016/j.cell.2015.07.048

Gucciardo, E., Sugiyama, N. and Lehti, K. 2014. Eph- and ephrin-dependent mechanisms in tumor and stem cell dynamics. *Cell Mol Life Sci* 71(19), pp. 3685-3710. doi: 10.1007/s00018-014-1633-0

Guttman, M. et al. 2009. Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. *Nature* 458(7235), pp. 223-227. doi: 10.1038/nature07672

Halliday, M. R., Rege, S. V., Ma, Q., Zhao, Z., Miller, C. A., Winkler, E. A. and Zlokovic, B. V. 2016. Accelerated pericyte degeneration and blood-brain barrier breakdown in apolipoprotein E4 carriers with Alzheimer's disease. *J Cereb Blood Flow Metab* 36(1), pp. 216-227. doi: 10.1038/jcbfm.2015.44

Harold, D. et al. 2009. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. *Nat Genet* 41(10), pp. 1088-1093. doi: 10.1038/ng.440

Hattori, M., Osterfield, M. and Flanagan, J. G. 2000. Regulated cleavage of a contact-mediated axon repellent. *Science* 289(5483), pp. 1360-1365. doi: 10.1126/science.289.5483.1360

Bibliography

Hawkes, C. A. and McLaurin, J. 2009. Selective targeting of perivascular macrophages for clearance of beta-amyloid in cerebral amyloid angiopathy. *Proc Natl Acad Sci U S A* 106(4), pp. 1261-1266. doi: 10.1073/pnas.0805453106

Hayden, E. Y. and Teplow, D. B. 2013. Amyloid beta-protein oligomers and Alzheimer's disease. *Alzheimers Res Ther* 5(6), p. 60. doi: 10.1186/alzrt226

Heintzman, N. D. et al. 2009. Histone modifications at human enhancers reflect global cell-type-specific gene expression. *Nature* 459(7243), pp. 108-112. doi: 10.1038/nature07829

Heneka, M. T. et al. 2015. Neuroinflammation in Alzheimer's disease. *Lancet Neurol* 14(4), pp. 388-405. doi: 10.1016/S1474-4422(15)70016-5

Heneka, M. T., Sastre, M., Dumitrescu-Ozimek, L., Dewachter, I., Walter, J., Klockgether, T. and Van Leuven, F. 2005. Focal glial activation coincides with increased BACE1 activation and precedes amyloid plaque deposition in APP[V717I] transgenic mice. *J Neuroinflammation* 2, p. 22. doi: 10.1186/1742-2094-2-22

Hickman, S. E., Allison, E. K. and El Khoury, J. 2008. Microglial dysfunction and defective beta-amyloid clearance pathways in aging Alzheimer's disease mice. *J Neurosci* 28(33), pp. 8354-8360. doi: 10.1523/JNEUROSCI.0616-08.2008

Himanen, J. P. et al. 2010. Architecture of Eph receptor clusters. *Proc Natl Acad Sci U S A* 107(24), pp. 10860-10865. doi: 10.1073/pnas.1004148107

Hirai, H., Maru, Y., Hagiwara, K., Nishida, J. and Takaku, F. 1987. A novel putative tyrosine kinase receptor encoded by the eph gene. *Science* 238(4834), pp. 1717-1720. doi: 10.1126/science.2825356

Hirata, H., Tatsumi, H. and Sokabe, M. 2008a. Mechanical forces facilitate actin polymerization at focal adhesions in a zyxin-dependent manner. *J Cell Sci* 121(Pt 17), pp. 2795-2804. doi: 10.1242/jcs.030320

Hirata, H., Tatsumi, H. and Sokabe, M. 2008b. Zyxin emerges as a key player in the mechanotransduction at cell adhesive structures. *Commun Integr Biol* 1(2), pp. 192-195. doi: 10.4161/cib.1.2.7001

Hjorthaug, H. S. and Aasheim, H. C. 2007. Ephrin-A1 stimulates migration of CD8+CCR7+ T lymphocytes. *Eur J Immunol* 37(8), pp. 2326-2336. doi: 10.1002/eji.200737111

Hnisz, D., Day, D. S. and Young, R. A. 2016. Insulated Neighborhoods: Structural and Functional Units of Mammalian Gene Control. *Cell* 167(5), pp. 1188-1200. doi: 10.1016/j.cell.2016.10.024

Hodges, A. K. P., T.M. Collier, D. Cousins, O. Pocock, J.M 2021. Pathways linking Alzheimer's disease risk genes expressed highly in microglia. *Neuroimmunol Neuroinflammation*,

Bibliography

Holen, H. L., Nustad, K. and Aasheim, H. C. 2010. Activation of EphA receptors on CD4+CD45RO+ memory cells stimulates migration. *J Leukoc Biol* 87(6), pp. 1059-1068. doi: 10.1189/jlb.0709497

Hollingworth, P. et al. 2011. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. *Nat Genet* 43(5), pp. 429-435. doi: 10.1038/ng.803

hUB, D. S. 2018. Statistics about dementia. *Alzheimer's Research UK*,

Hung, T. and Chang, H. Y. 2010. Long noncoding RNA in genome regulation: prospects and mechanisms. *RNA Biol* 7(5), pp. 582-585. doi: 10.4161/rna.7.5.13216

Hung, T. et al. 2011. Extensive and coordinated transcription of noncoding RNAs within cell-cycle promoters. *Nat Genet* 43(7), pp. 621-629. doi: 10.1038/ng.848

Idda, M. L., Munk, R., Abdelmohsen, K. and Gorospe, M. 2018. Noncoding RNAs in Alzheimer's disease. *Wiley Interdiscip Rev RNA* 9(2), doi: 10.1002/wrna.1463

Ieguchi, K. et al. 2014. ADAM12-cleaved ephrin-A1 contributes to lung metastasis. *Oncogene* 33(17), pp. 2179-2190. doi: 10.1038/onc.2013.180

International, A. s. D. 2015. Dementia Statistics. *The Global Voice on Dementia*,

Irizarry, M. C., McNamara, M., Fedorchak, K., Hsiao, K. and Hyman, B. T. 1997. APPSw transgenic mice develop age-related A beta deposits and neuropil abnormalities, but no neuronal loss in CA1. *J Neuropathol Exp Neurol* 56(9), pp. 965-973. doi: 10.1097/00005072-199709000-00002

Ishino, Y., Shinagawa, H., Makino, K., Amemura, M. and Nakata, A. 1987. Nucleotide sequence of the iap gene, responsible for alkaline phosphatase isozyme conversion in Escherichia coli, and identification of the gene product. *J Bacteriol* 169(12), pp. 5429-5433. doi: 10.1128/jb.169.12.5429-5433.1987

Jack, C. R., Jr. et al. 2010. Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade. *Lancet Neurol* 9(1), pp. 119-128. doi: 10.1016/S1474-4422(09)70299-6

Janes, P. W. et al. 2011. Eph receptor function is modulated by heterooligomerization of A and B type Eph receptors. *J Cell Biol* 195(6), pp. 1033-1045. doi: 10.1083/jcb.201104037

Jansen, I. E. et al. 2019. Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet* 51(3), pp. 404-413. doi: 10.1038/s41588-018-0311-9

Jendrzejewski, J. et al. 2012. The polymorphism rs944289 predisposes to papillary thyroid carcinoma through a large intergenic noncoding RNA gene of tumor suppressor type. *Proc Natl Acad Sci U S A* 109(22), pp. 8646-8651. doi: 10.1073/pnas.1205654109

Bibliography

Ji, K., Akgul, G., Wollmuth, L. P. and Tsirka, S. E. 2013. Microglia actively regulate the number of functional synapses. *PLoS One* 8(2), p. e56293. doi: 10.1371/journal.pone.0056293

Jiang, W., Bikard, D., Cox, D., Zhang, F. and Marraffini, L. A. 2013. RNA-guided editing of bacterial genomes using CRISPR-Cas systems. *Nat Biotechnol* 31(3), pp. 233-239. doi: 10.1038/nbt.2508

Jin, M., Shepardson, N., Yang, T., Chen, G., Walsh, D. and Selkoe, D. J. 2011. Soluble amyloid beta-protein dimers isolated from Alzheimer cortex directly induce Tau hyperphosphorylation and neuritic degeneration. *Proc Natl Acad Sci U S A* 108(14), pp. 5819-5824. doi: 10.1073/pnas.1017033108

Kamboh, M. I. 2018. A Brief Synopsis on the Genetics of Alzheimer's Disease. *Curr Genet Med Rep* 6(4), pp. 133-135. doi: 10.1007/s40142-018-0155-8

Kang, D. E., Roh, S. E., Woo, J. A., Liu, T., Bu, J. H., Jung, A. R. and Lim, Y. 2011. The Interface between Cytoskeletal Aberrations and Mitochondrial Dysfunction in Alzheimer's Disease and Related Disorders. *Exp Neurobiol* 20(2), pp. 67-80. doi: 10.5607/en.2011.20.2.67

Kang, K. et al. 2017. Interferon-gamma Represses M2 Gene Expression in Human Macrophages by Disassembling Enhancers Bound by the Transcription Factor MAF. *Immunity* 47(2), pp. 235-250 e234. doi: 10.1016/j.immuni.2017.07.017

Kang, M. J. et al. 2014. HuD regulates coding and noncoding RNA to induce APP-->Abeta processing. *Cell Rep* 7(5), pp. 1401-1409. doi: 10.1016/j.celrep.2014.04.050

Khalil, A. M. et al. 2009. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci U S A* 106(28), pp. 11667-11672. doi: 10.1073/pnas.0904715106

Kichaev, G. et al. 2014. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. *PLoS Genet* 10(10), p. e1004722. doi: 10.1371/journal.pgen.1004722

Kim, S., Yu, N. K. and Kaang, B. K. 2015. CTCF as a multifunctional protein in genome regulation and gene expression. *Exp Mol Med* 47, p. e166. doi: 10.1038/emm.2015.33

Kim, Y., Lasso, G.,Patel, H., Vardarajan, B., Santa-Maria, I, Lefort, R. 2021. Alzheimer's disease-associated P460L mutation in ephrin receptor type A1 (EphA1) leads to dysregulated Rho-GTPase signaling. *bioRxiv preprint*

Kinney, J. W., Bemiller, S. M., Murtishaw, A. S., Leisgang, A. M., Salazar, A. M. and Lamb, B. T. 2018. Inflammation as a central mechanism in Alzheimer's disease. *Alzheimers Dement (N Y)* 4, pp. 575-590. doi: 10.1016/j.trci.2018.06.014

Bibliography

Kino, T., Hurt, D. E., Ichijo, T., Nader, N. and Chrousos, G. P. 2010. Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. *Sci Signal* 3(107), p. ra8. doi: 10.1126/scisignal.2000568

Kisler, K. et al. 2017. Pericyte degeneration leads to neurovascular uncoupling and limits oxygen supply to brain. *Nat Neurosci* 20(3), pp. 406-416. doi: 10.1038/nn.4489

Knoll, B. and Drescher, U. 2004. Src family kinases are involved in EphA receptor-mediated retinal axon guidance. *J Neurosci* 24(28), pp. 6248-6257. doi: 10.1523/JNEUROSCI.0985-04.2004

Koh, K. H. and Jeong, H. 2016. Electrophoretic Mobility Shift Assay (EMSA) and Supershift Assay of Cytochrome P450 2B6 in Response to Estrogen. *Methods Mol Biol* 1366, pp. 41-51. doi: 10.1007/978-1-4939-3127-9_5

Kojro, E. and Fahrenholz, F. 2005. The non-amyloidogenic pathway: structure and function of alpha-secretases. *Subcell Biochem* 38, pp. 105-127. doi: 10.1007/0-387-23226-5_5

Kok, E., Haikonen, S., Luoto, T., Huhtala, H., Goebeler, S., Haapasalo, H. and Karhunen, P. J. 2009. Apolipoprotein E-dependent accumulation of Alzheimer disease-related lesions begins in middle age. *Ann Neurol* 65(6), pp. 650-657. doi: 10.1002/ana.21696

Kosoy, R. et al. 2022. Genetics of the human microglia regulome refines Alzheimer's disease risk loci. *Nat Genet* 54(8), pp. 1145-1154. doi: 10.1038/s41588-022-01149-1

Kratofil, R. M., Kubes, P. and Deniset, J. F. 2017. Monocyte Conversion During Inflammation and Injury. *Arterioscler Thromb Vasc Biol* 37(1), pp. 35-42. doi: 10.1161/ATVBAHA.116.308198

Krstic, D. et al. 2012. Systemic immune challenges trigger and drive Alzheimer-like neuropathology in mice. *J Neuroinflammation* 9, p. 151. doi: 10.1186/1742-2094-9-151

Kruglyak, L. 1999. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nat Genet* 22(2), pp. 139-144. doi: 10.1038/9642

Kunkle, B. W. et al. 2019a. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* 51(3), pp. 414-430. doi: 10.1038/s41588-019-0358-2

Kunkle, B. W. et al. 2019b. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Abeta, tau, immunity and lipid processing. *Nat Genet* 51(3), pp. 414-430. doi: 10.1038/s41588-019-0358-2

Lambert, J. C. et al. 2013. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 45(12), pp. 1452-1458. doi: 10.1038/ng.2802

Bibliography

Lambert, M. P. et al. 1998. Diffusible, nonfibrillar ligands derived from Abeta1-42 are potent central nervous system neurotoxins. *Proc Natl Acad Sci U S A* 95(11), pp. 6448-6453. doi: 10.1073/pnas.95.11.6448

Lambert, S. A. et al. 2018. The Human Transcription Factors. *Cell* 175(2), pp. 598-599. doi: 10.1016/j.cell.2018.09.045

Lane, C. A., Hardy, J. and Schott, J. M. 2018. Alzheimer's disease. *Eur J Neurol* 25(1), pp. 59-70. doi: 10.1111/ene.13439

Langst, G. and Manelyte, L. 2015. Chromatin Remodelers: From Function to Dysfunction. *Genes (Basel)* 6(2), pp. 299-324. doi: 10.3390/genes6020299

Lanni, C. et al. 2013. Zyxin is a novel target for beta-amyloid peptide: characterization of its role in Alzheimer's pathogenesis. *J Neurochem* 125(5), pp. 790-799. doi: 10.1111/jnc.12154

Lasagna-Reeves, C. A., Castillo-Carranza, D. L., Sengupta, U., Sarmiento, J., Troncoso, J., Jackson, G. R. and Kayed, R. 2012. Identification of oligomers at early stages of tau aggregation in Alzheimer's disease. *FASEB J* 26(5), pp. 1946-1959. doi: 10.1096/fj.11-199851

Lee, D. C. et al. 2010. LPS- induced inflammation exacerbates phospho-tau pathology in rTg4510 mice. *J Neuroinflammation* 7, p. 56. doi: 10.1186/1742-2094-7-56

Lee, S. H. et al. 2013. Estimation and partitioning of polygenic variation captured by common SNPs for Alzheimer's disease, multiple sclerosis and endometriosis. *Hum Mol Genet* 22(4), pp. 832-841. doi: 10.1093/hmg/dds491

Lewis, J. et al. 2001. Enhanced neurofibrillary degeneration in transgenic mice expressing mutant tau and APP. *Science* 293(5534), pp. 1487-1491. doi: 10.1126/science.1058189

Li, K. et al. 2010. A noncoding antisense RNA in tie-1 locus regulates tie-1 function in vivo. *Blood* 115(1), pp. 133-139. doi: 10.1182/blood-2009-09-242180

Li, Z. et al. 2014. The long noncoding RNA THRIL regulates TNFalpha expression through its interaction with hnRNPL. *Proc Natl Acad Sci U S A* 111(3), pp. 1002-1007. doi: 10.1073/pnas.1313768111

Ling, G. Q., Chen, D. B., Wang, B. Q. and Zhang, L. S. 2012. Expression of the pluripotency markers Oct3/4, Nanog and Sox2 in human breast cancer cell lines. *Oncol Lett* 4(6), pp. 1264-1268. doi: 10.3892/ol.2012.916

Lisabeth, E. M., Falivelli, G. and Pasquale, E. B. 2013. Eph receptor signaling and ephrins. *Cold Spring Harb Perspect Biol* 5(9), doi: 10.1101/cshperspect.a009159

Litterst, C. et al. 2007. Ligand binding and calcium influx induce distinct ectodomain/gamma-secretase-processing pathways of EphB2 receptor. *J Biol Chem* 282(22), pp. 16155-16163. doi: 10.1074/jbc.M611449200

Bibliography

Liu, C. C., Liu, C. C., Kanekiyo, T., Xu, H. and Bu, G. 2013. Apolipoprotein E and Alzheimer disease: risk, mechanisms and therapy. *Nat Rev Neurol* 9(2), pp. 106-118. doi: 10.1038/nrneurol.2012.263

Liu, C. H. and Wu, P. S. 2006. Characterization of matrix metalloproteinase expressed by human embryonic kidney cells. *Biotechnol Lett* 28(21), pp. 1725-1730. doi: 10.1007/s10529-006-9147-y

Liu, G. et al. 2018. Alzheimer's Disease rs11767557 Variant Regulates EPHA1 Gene Expression Specifically in Human Whole Blood. *Journal of Alzheimer's Disease* 61(3), pp. 1077-1088. doi: 10.3233/JAD-170468

Liu, Q. et al. 2020. Increased Expression of Zyxin and Its Potential Function in Androgenetic Alopecia. *Front Cell Dev Biol* 8, p. 582282. doi: 10.3389/fcell.2020.582282

Liu, T., Huang, Y., Chen, J., Chi, H., Yu, Z., Wang, J. and Chen, C. 2014. Attenuated ability of BACE1 to cleave the amyloid precursor protein via silencing long noncoding RNA BACE1AS expression. *Mol Med Rep* 10(3), pp. 1275-1281. doi: 10.3892/mmr.2014.2351

Loffek, S., Schilling, O. and Franzke, C. W. 2011. Series "matrix metalloproteinases in lung health and disease": Biological role of matrix metalloproteinases: a critical balance. *Eur Respir J* 38(1), pp. 191-208. doi: 10.1183/09031936.00146510

Long, Y., Wang, X., Youmans, D. T. and Cech, T. R. 2017. How do lncRNAs regulate transcription? *Sci Adv* 3(9), p. eaao2110. doi: 10.1126/sciadv.aao2110

Low, Y. H., Asi, Y., Foti, S. C. and Lashley, T. 2021. Heterogeneous Nuclear Ribonucleoproteins: Implications in Neurological Diseases. *Mol Neurobiol* 58(2), pp. 631-646. doi: 10.1007/s12035-020-02137-4

Luger, K., Dechassa, M. L. and Tremethick, D. J. 2012. New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? *Nat Rev Mol Cell Biol* 13(7), pp. 436-447. doi: 10.1038/nrm3382

Lyman, M., Lloyd, D. G., Ji, X., Vizcaychipi, M. P. and Ma, D. 2014. Neuroinflammation: the role and consequences. *Neurosci Res* 79, pp. 1-12. doi: 10.1016/j.neures.2013.10.004

Lynch, M. A. 2010. Age-related neuroinflammatory changes negatively impact on neuronal function. *Front Aging Neurosci* 1, p. 6. doi: 10.3389/neuro.24.006.2009

Mancia, F. and Shapiro, L. 2005. ADAM and Eph: how Ephrin-signaling cells become detached. *Cell* 123(2), pp. 185-187. doi: 10.1016/j.cell.2005.10.004

Manda-Handzlik, A. and Demkow, U. 2019. The Brain Entangled: The Contribution of Neutrophil Extracellular Traps to the Diseases of the Central Nervous System. *Cells* 8(12), doi: 10.3390/cells8121477

Manolio, T. A. et al. 2009. Finding the missing heritability of complex diseases. *Nature* 461(7265), pp. 747-753. doi: 10.1038/nature08494

Maston, G. A., Evans, S. K. and Green, M. R. 2006. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* 7, pp. 29-59. doi: 10.1146/annurev.genom.7.080505.115623

Maurano, M. T. et al. 2012. Systematic localization of common disease-associated variation in regulatory DNA. *Science* 337(6099), pp. 1190-1195. doi: 10.1126/science.1222794

McConnell, B. B. and Yang, V. W. 2010. Mammalian Kruppel-like factors in health and diseases. *Physiol Rev* 90(4), pp. 1337-1381. doi: 10.1152/physrev.00058.2009

McKhann, G. M. et al. 2011. The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease. *Alzheimers Dement* 7(3), pp. 263-269. doi: 10.1016/j.jalz.2011.03.005

Mendez, M. F. 2017. Early-Onset Alzheimer Disease. *Neurol Clin* 35(2), pp. 263-281. doi: 10.1016/j.ncl.2017.01.005

Mendez, M. F. 2019. Early-onset Alzheimer Disease and Its Variants. *Continuum (Minneap Minn)* 25(1), pp. 34-51. doi: 10.1212/CON.0000000000000687

Mercer, T. R., Dinger, M. E. and Mattick, J. S. 2009. Long non-coding RNAs: insights into functions. *Nat Rev Genet* 10(3), pp. 155-159. doi: 10.1038/nrg2521

Merilahti, J. A. M. and Elenius, K. 2019. Gamma-secretase-dependent signaling of receptor tyrosine kinases. *Oncogene* 38(2), pp. 151-163. doi: 10.1038/s41388-018-0465-z

Merilahti, J. A. M., Ojala, V. K., Knittle, A. M., Pulliainen, A. T. and Elenius, K. 2017. Genome-wide screen of gamma-secretase-mediated intramembrane cleavage of receptor tyrosine kinases. *Mol Biol Cell* 28(22), pp. 3123-3131. doi: 10.1091/mbc.E17-04-0261

Merlini, M., Kirabali, T., Kulic, L., Nitsch, R. M. and Ferretti, M. T. 2018. Extravascular CD3+ T Cells in Brains of Alzheimer Disease Patients Correlate with Tau but Not with Amyloid Pathology: An Immunohistochemical Study. *Neurodegener Dis* 18(1), pp. 49-56. doi: 10.1159/000486200

Meziane, H., Dodart, J. C., Mathis, C., Little, S., Clemens, J., Paul, S. M. and Ungerer, A. 1998. Memory-enhancing effects of secreted forms of the beta-amyloid precursor protein in normal and amnestic mice. *Proc Natl Acad Sci U S A* 95(21), pp. 12683-12688. doi: 10.1073/pnas.95.21.12683

Miao, H. et al. 2009. EphA2 mediates ligand-dependent inhibition and ligand-independent promotion of cell migration and invasion via a reciprocal regulatory loop with Akt. *Cancer Cell* 16(1), pp. 9-20. doi: 10.1016/j.ccr.2009.04.009

Miao, H. and Wang, B. 2012. EphA receptor signaling--complexity and emerging themes. *Semin Cell Dev Biol* 23(1), pp. 16-25. doi: 10.1016/j.semcdb.2011.10.013

Miao, H. et al. 2001. Activation of EphA receptor tyrosine kinase inhibits the Ras/MAPK pathway. *Nat Cell Biol* 3(5), pp. 527-530. doi: 10.1038/35074604

Micheau, O. and Tschopp, J. 2003. Induction of TNF receptor I-mediated apoptosis via two sequential signaling complexes. *Cell* 114(2), pp. 181-190. doi: 10.1016/s0092-8674(03)00521-x

Misra, A., Chakrabarti, S. S. and Gambhir, I. S. 2018. New genetic players in late-onset Alzheimer's disease: Findings of genome-wide association studies. *Indian J Med Res* 148(2), pp. 135-144. doi: 10.4103/ijmr.IJMR_473_17

Miyazaki, Y., Yusa, T., Matsuo, S., Terauchi, Y. and Miyazaki, S. 2014. Zyxin modulates the transmigration of Haemophilus influenzae to the central nervous system. *Virulence* 5(6), pp. 665-672. doi: 10.4161/viru.29786

Mo, M. L. et al. 2013. Down-regulation of SIX3 is associated with clinical outcome in lung adenocarcinoma. *PLoS One* 8(8), p. e71816. doi: 10.1371/journal.pone.0071816

Mojica, F. J., Diez-Villasenor, C., Garcia-Martinez, J. and Soria, E. 2005. Intervening sequences of regularly spaced prokaryotic repeats derive from foreign genetic elements. *J Mol Evol* 60(2), pp. 174-182. doi: 10.1007/s00239-004-0046-3

Montagne, A. et al. 2015. Blood-brain barrier breakdown in the aging human hippocampus. *Neuron* 85(2), pp. 296-302. doi: 10.1016/j.neuron.2014.12.032

Moon, O. R. 2020. *Engineering homing properties of cancer-specific T lymphocytes in adoptive cell therapy.* Doctor of Philosophy, Cardiff University.

Moore, L. D., Le, T. and Fan, G. 2013. DNA methylation and its basic function. *Neuropsychopharmacology* 38(1), pp. 23-38. doi: 10.1038/npp.2012.112

Mormino, E. C. et al. 2016. Polygenic risk of Alzheimer disease is associated with early- and late-life processes. *Neurology* 87(5), pp. 481-488. doi: 10.1212/WNL.0000000000002922

Morris, K. V. and Mattick, J. S. 2014. The rise of regulatory RNA. *Nat Rev Genet* 15(6), pp. 423-437. doi: 10.1038/nrg3722

Muoio, V., Persson, P. B. and Sendeski, M. M. 2014. The neurovascular unit - concept review. *Acta Physiol (Oxf)* 210(4), pp. 790-798. doi: 10.1111/apha.12250

Bibliography

Murray, P. J. 2007. The JAK-STAT signaling pathway: input and output integration. *J Immunol* 178(5), pp. 2623-2629. doi: 10.4049/jimmunol.178.5.2623

Musunuru, K. et al. 2010. From noncoding variant to phenotype via SORT1 at the 1p13 cholesterol locus. *Nature* 466(7307), pp. 714-719. doi: 10.1038/nature09266

Nagano, T., Mitchell, J. A., Sanz, L. A., Pauler, F. M., Ferguson-Smith, A. C., Feil, R. and Fraser, P. 2008. The Air noncoding RNA epigenetically silences transcription by targeting G9a to chromatin. *Science* 322(5908), pp. 1717-1720. doi: 10.1126/science.1163802

Naj, A. C. et al. 2011. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. *Nat Genet* 43(5), pp. 436-441. doi: 10.1038/ng.801

Nakagawa, Y. and Chiba, K. 2014. Role of microglial m1/m2 polarization in relapse and remission of psychiatric disorders and diseases. *Pharmaceuticals (Basel)* 7(12), pp. 1028-1048. doi: 10.3390/ph7121028

Nation, D. A. et al. 2019. Blood-brain barrier breakdown is an early biomarker of human cognitive dysfunction. *Nat Med* 25(2), pp. 270-276. doi: 10.1038/s41591-018-0297-y

Nelson, A. R., Sweeney, M. D., Sagare, A. P. and Zlokovic, B. V. 2016. Neurovascular dysfunction and neurodegeneration in dementia and Alzheimer's disease. *Biochim Biophys Acta* 1862(5), pp. 887-900. doi: 10.1016/j.bbadis.2015.12.016

Nestor, M. W., Mok, L. P., Tulapurkar, M. E. and Thompson, S. M. 2007. Plasticity of neuron-glial interactions mediated by astrocytic EphARs. *J Neurosci* 27(47), pp. 12817-12828. doi: 10.1523/JNEUROSCI.2442-07.2007

Ni Chasaide, C. and Lynch, M. A. 2020. The role of the immune system in driving neuroinflammation. *Brain Neurosci Adv* 4, p. 2398212819901082. doi: 10.1177/2398212819901082

Nica, A. C. and Dermitzakis, E. T. 2013. Expression quantitative trait loci: present and future. *Philos Trans R Soc Lond B Biol Sci* 368(1620), p. 20120362. doi: 10.1098/rstb.2012.0362

Nimmerjahn, A., Kirchhoff, F. and Helmchen, F. 2005. Resting microglial cells are highly dynamic surveillants of brain parenchyma in vivo. *Science* 308(5726), pp. 1314-1318. doi: 10.1126/science.1110647

Nix, D. A. and Beckerle, M. C. 1997. Nuclear-cytoplasmic shuttling of the focal contact protein, zyxin: a potential mechanism for communication between sites of cell adhesion and the nucleus. *J Cell Biol* 138(5), pp. 1139-1147. doi: 10.1083/jcb.138.5.1139

Novakovic, B. et al. 2016. beta-Glucan Reverses the Epigenetic State of LPS-Induced Immunological Tolerance. *Cell* 167(5), pp. 1354-1368 e1314. doi: 10.1016/j.cell.2016.09.034

Bibliography

Novikova, G. et al. 2021. Integration of Alzheimer's disease genetics and myeloid genomics identifies disease risk regulatory elements and genes. *Nat Commun* 12(1), p. 1610. doi: 10.1038/s41467-021-21823-y

O'Brien, R. J. and Wong, P. C. 2011. Amyloid precursor protein processing and Alzheimer's disease. *Annu Rev Neurosci* 34, pp. 185-204. doi: 10.1146/annurev-neuro-061010-113613

Oakley, H. et al. 2006. Intraneuronal beta-amyloid aggregates, neurodegeneration, and neuron loss in transgenic mice with five familial Alzheimer's disease mutations: potential factors in amyloid plaque formation. *J Neurosci* 26(40), pp. 10129-10140. doi: 10.1523/JNEUROSCI.1202-06.2006

Ochalek, A. et al. 2017. Neurons derived from sporadic Alzheimer's disease iPSCs reveal elevated TAU hyperphosphorylation, increased amyloid levels, and GSK3B activation. *Alzheimers Res Ther* 9(1), p. 90. doi: 10.1186/s13195-017-0317-z

Ogbourne, S. and Antalis, T. M. 1998. Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem J* 331 (Pt 1), pp. 1-14. doi: 10.1042/bj3310001

Okello, A. et al. 2009. Microglial activation and amyloid deposition in mild cognitive impairment: a PET study. *Neurology* 72(1), pp. 56-62. doi: 10.1212/01.wnl.0000338622.27876.0d

Oliver, G., Mailhos, A., Wehr, R., Copeland, N. G., Jenkins, N. A. and Gruss, P. 1995. Six3, a murine homologue of the sine oculis gene, demarcates the most anterior border of the developing neural plate and is expressed during eye development. *Development* 121(12), pp. 4045-4055. doi: 10.1242/dev.121.12.4045

Ong, C. T. and Corces, V. G. 2014. CTCF: an architectural protein bridging genome topology and function. *Nat Rev Genet* 15(4), pp. 234-246. doi: 10.1038/nrg3663

Pantazis, C. B. et al. 2022. A reference human induced pluripotent stem cell line for large-scale collaborative studies. *Cell Stem Cell* 29(12), pp. 1685-1702 e1622. doi: 10.1016/j.stem.2022.11.004

Paresce, D. M., Ghosh, R. N. and Maxfield, F. R. 1996. Microglial cells internalize aggregates of the Alzheimer's disease amyloid beta-protein via a scavenger receptor. *Neuron* 17(3), pp. 553-565. doi: 10.1016/s0896-6273(00)80187-7

Parshina, E. A. et al. 2020. Cytoskeletal Protein Zyxin Inhibits the Activity of Genes Responsible for Embryonic Stem Cell Status. *Cell Rep* 33(7), p. 108396. doi: 10.1016/j.celrep.2020.108396

Pasquale, E. B. 2010. Eph receptors and ephrins in cancer: bidirectional signalling and beyond. *Nat Rev Cancer* 10(3), pp. 165-180. doi: 10.1038/nrc2806

Bibliography

Penke, B., Bogar, F. and Fulop, L. 2017. beta-Amyloid and the Pathomechanisms of Alzheimer's Disease: A Comprehensive View. *Molecules* 22(10), doi: 10.3390/molecules22101692

Pennacchio, L. A., Bickmore, W., Dean, A., Nobrega, M. A. and Bejerano, G. 2013. Enhancers: five essential questions. *Nat Rev Genet* 14(4), pp. 288-295. doi: 10.1038/nrg3458

Penney, J., Ralvenius, W. T. and Tsai, L. H. 2020. Modeling Alzheimer's disease with iPSC-derived brain cells. *Mol Psychiatry* 25(1), pp. 148-167. doi: 10.1038/s41380-019-0468-3

Perenthaler, E., Yousefi, S., Niggl, E. and Barakat, T. S. 2019. Beyond the Exome: The Non-coding Genome and Enhancers in Neurodevelopmental Disorders and Malformations of Cortical Development. *Front Cell Neurosci* 13, p. 352. doi: 10.3389/fncel.2019.00352

Pietronigro, E. C., Della Bianca, V., Zenaro, E. and Constantin, G. 2017. NETosis in Alzheimer's Disease. *Front Immunol* 8, p. 211. doi: 10.3389/fimmu.2017.00211

Pizza, V., Agresta, A., D'Acunto, C. W., Festa, M. and Capasso, A. 2011. Neuroinflammation and ageing: current theories and an overview of the data. *Rev Recent Clin Trials* 6(3), pp. 189-203. doi: 10.2174/157488711796575577

Placanica, L., Chien, J. W. and Li, Y. M. 2010. Characterization of an atypical gamma-secretase complex from hematopoietic origin. *Biochemistry* 49(13), pp. 2796-2804. doi: 10.1021/bi901388t

Podlesny-Drabiniok, A., Marcora, E. and Goate, A. M. 2020. Microglial Phagocytosis: A Disease-Associated Process Emerging from Alzheimer's Disease Genetics. *Trends Neurosci* 43(12), pp. 965-979. doi: 10.1016/j.tins.2020.10.002

Pourcel, C., Salvignol, G. and Vergnaud, G. 2005. CRISPR elements in Yersinia pestis acquire new repeats by preferential uptake of bacteriophage DNA, and provide additional tools for evolutionary studies. *Microbiology (Reading)* 151(Pt 3), pp. 653-663. doi: 10.1099/mic.0.27437-0

Prolla, T. A. 2002. DNA microarray analysis of the aging brain. *Chem Senses* 27(3), pp. 299-306. doi: 10.1093/chemse/27.3.299

Qin, X., Wang, Y. and Paudel, H. K. 2016. Early Growth Response 1 (Egr-1) Is a Transcriptional Activator of beta-Secretase 1 (BACE-1) in the Brain. *J Biol Chem* 291(42), pp. 22276-22287. doi: 10.1074/jbc.M116.738849

Qiu, C., Kivipelto, M. and von Strauss, E. 2009. Epidemiology of Alzheimer's disease: occurrence, determinants, and strategies toward intervention. *Dialogues Clin Neurosci* 11(2), pp. 111-128.

Qiu, T., Liu, Q., Chen, Y. X., Zhao, Y. F. and Li, Y. M. 2015. Abeta42 and Abeta40: similarities and differences. *J Pept Sci* 21(7), pp. 522-529. doi: 10.1002/psc.2789

Quina, A. S., Buschbeck, M. and Di Croce, L. 2006. Chromatin structure and epigenetics. *Biochem Pharmacol* 72(11), pp. 1563-1569. doi: 10.1016/j.bcp.2006.06.016

Rabinovici, G. D. 2019. Late-onset Alzheimer Disease. *Continuum (Minneap Minn)* 25(1), pp. 14-33. doi: 10.1212/CON.0000000000000700

Ransohoff, R. M. 2016. A polarizing question: do M1 and M2 microglia exist? *Nat Neurosci* 19(8), pp. 987-991. doi: 10.1038/nn.4338

Redmer, T., Diecke, S., Grigoryan, T., Quiroga-Negreira, A., Birchmeier, W. and Besser, D. 2011. E-cadherin is crucial for embryonic stem cell pluripotency and can replace OCT4 during somatic cell reprogramming. *EMBO Rep* 12(7), pp. 720-726. doi: 10.1038/embor.2011.88

Reiter, F., Wienerroither, S. and Stark, A. 2017. Combinatorial function of transcription factors and cofactors. *Curr Opin Genet Dev* 43, pp. 73-81. doi: 10.1016/j.gde.2016.12.007

Rinn, J. L. and Chang, H. Y. 2012. Genome regulation by long noncoding RNAs. *Annu Rev Biochem* 81, pp. 145-166. doi: 10.1146/annurev-biochem-051410-092902

Romanoski, C. E., Glass, C. K., Stunnenberg, H. G., Wilson, L. and Almouzni, G. 2015. Epigenomics: Roadmap for regulation. *Nature* 518(7539), pp. 314-316. doi: 10.1038/518314a

Rosenberger, A. F. et al. 2016. Protein Kinase Activity Decreases with Higher Braak Stages of Alzheimer's Disease Pathology. *J Alzheimers Dis* 49(4), pp. 927-943. doi: 10.3233/JAD-150429

Rosenthal, S. L. and Kamboh, M. I. 2014. Late-Onset Alzheimer's Disease Genes and the Potentially Implicated Pathways. *Curr Genet Med Rep* 2, pp. 85-101. doi: 10.1007/s40142-014-0034-x

Rossi, B., Angiari, S., Zenaro, E., Budui, S. L. and Constantin, G. 2011. Vascular inflammation in central nervous system diseases: adhesion receptors controlling leukocyte-endothelial interactions. *J Leukoc Biol* 89(4), pp. 539-556. doi: 10.1189/jlb.0710432

Rowe, R. G. and Daley, G. Q. 2019. Induced pluripotent stem cells in disease modelling and drug discovery. *Nat Rev Genet* 20(7), pp. 377-388. doi: 10.1038/s41576-019-0100-z

Ruffell, D., Mourkioti, F., Gambardella, A., Kirstetter, P., Lopez, R. G., Rosenthal, N. and Nerlov, C. 2009. A CREB-C/EBPbeta cascade induces M2 macrophage-specific gene expression and promotes muscle injury repair. *Proc Natl Acad Sci U S A* 106(41), pp. 17475-17480. doi: 10.1073/pnas.0908641106

Bibliography

Ruitenberg, A., den Heijer, T., Bakker, S. L., van Swieten, J. C., Koudstaal, P. J., Hofman, A. and Breteler, M. M. 2005. Cerebral hypoperfusion and clinical onset of dementia: the Rotterdam Study. *Ann Neurol* 57(6), pp. 789-794. doi: 10.1002/ana.20493

Ryu, J. K. and McLarnon, J. G. 2009. A leaky blood-brain barrier, fibrinogen infiltration and microglial reactivity in inflamed Alzheimer's disease brain. *J Cell Mol Med* 13(9A), pp. 2911-2925. doi: 10.1111/j.1582-4934.2008.00434.x

Sackmann, C. and Hallbeck, M. 2020. Oligomeric amyloid-beta induces early and widespread changes to the proteome in human iPSC-derived neurons. *Sci Rep* 10(1), p. 6538. doi: 10.1038/s41598-020-63398-6

Sadot, E., Geiger, B., Oren, M. and Ben-Ze'ev, A. 2001. Down-regulation of beta-catenin by activated p53. *Mol Cell Biol* 21(20), pp. 6768-6781. doi: 10.1128/MCB.21.20.6768-6781.2001

Sapranauskas, R., Gasiunas, G., Fremaux, C., Barrangou, R., Horvath, P. and Siksnys, V. 2011. The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli. *Nucleic Acids Res* 39(21), pp. 9275-9282. doi: 10.1093/nar/gkr606

Satoh, J. I., Kino, Y., Yanaizu, M., Tosaki, Y., Sakai, K., Ishida, T. and Saito, Y. 2017. Microglia express ABI3 in the brains of Alzheimer's disease and Nasu-Hakola disease. *Intractable Rare Dis Res* 6(4), pp. 262-268. doi: 10.5582/irdr.2017.01073

Saunders, A. M. et al. 1993. Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* 43(8), pp. 1467-1472. doi: 10.1212/wnl.43.8.1467

Schaid, D. J., Chen, W. and Larson, N. B. 2018. From genome-wide associations to candidate causal variants by statistical fine-mapping. *Nat Rev Genet* 19(8), pp. 491-504. doi: 10.1038/s41576-018-0016-z

Schmechel, D. E. et al. 1993. Increased amyloid beta-peptide deposition in cerebral cortex as a consequence of apolipoprotein E genotype in late-onset Alzheimer disease. *Proc Natl Acad Sci U S A* 90(20), pp. 9649-9653. doi: 10.1073/pnas.90.20.9649

Schwartzentruber, J. et al. 2021. Author Correction: Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk genes. *Nat Genet* 53(4), pp. 585-586. doi: 10.1038/s41588-021-00822-1

Scilabra, S. D., Pigoni, M., Pravata, V., Schatzl, T., Muller, S. A., Troeberg, L. and Lichtenthaler, S. F. 2018. Increased TIMP-3 expression alters the cellular secretome through dual inhibition of the metalloprotease ADAM10 and ligand-binding of the LRP-1 receptor. *Sci Rep* 8(1), p. 14697. doi: 10.1038/s41598-018-32910-4

Scott, T., Urak, R., Soemardy, C. and Morris, K. V. 2019. Improved Cas9 activity by specific modifications of the tracrRNA. *Sci Rep* 9(1), p. 16104. doi: 10.1038/s41598-019-52616-5

Bibliography

Seals, D. F. and Courtneidge, S. A. 2003. The ADAMs family of metalloproteases: multidomain proteins with multiple functions. *Genes Dev* 17(1), pp. 7-30. doi: 10.1101/gad.1039703

Selkoe, D. J. and Hardy, J. 2016. The amyloid hypothesis of Alzheimer's disease at 25 years. *EMBO Mol Med* 8(6), pp. 595-608. doi: 10.15252/emmm.201606210

Sengupta, A., Kabat, J., Novak, M., Wu, Q., Grundke-Iqbal, I. and Iqbal, K. 1998. Phosphorylation of tau at both Thr 231 and Ser 262 is required for maximal inhibition of its binding to microtubules. *Arch Biochem Biophys* 357(2), pp. 299-309. doi: 10.1006/abbi.1998.0813

Sentmanat, M. F., Peters, S. T., Florian, C. P., Connelly, J. P. and Pruett-Miller, S. M. 2018. A Survey of Validation Strategies for CRISPR-Cas9 Editing. *Sci Rep* 8(1), p. 888. doi: 10.1038/s41598-018-19441-8

Seok, J. et al. 2013. Genomic responses in mouse models poorly mimic human inflammatory diseases. *Proc Natl Acad Sci U S A* 110(9), pp. 3507-3512. doi: 10.1073/pnas.1222878110

Serrano-Pozo, A., Frosch, M. P., Masliah, E. and Hyman, B. T. 2011. Neuropathological alterations in Alzheimer disease. *Cold Spring Harb Perspect Med* 1(1), p. a006189. doi: 10.1101/cshperspect.a006189

Shankar, G. M. et al. 2008. Amyloid-beta protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory. *Nat Med* 14(8), pp. 837-842. doi: 10.1038/nm1782

Sharfe, N., Nikolic, M., Cimpeon, L., Van De Kratts, A., Freywald, A. and Roifman, C. M. 2008. EphA and ephrin-A proteins regulate integrin-mediated T lymphocyte interactions. *Mol Immunol* 45(5), pp. 1208-1220. doi: 10.1016/j.molimm.2007.09.019

Shifman, S., Kuypers, J., Kokoris, M., Yakir, B. and Darvasi, A. 2003. Linkage disequilibrium patterns of the human genome across populations. *Hum Mol Genet* 12(7), pp. 771-776. doi: 10.1093/hmg/ddg088

Siatecka, M. and Bieker, J. J. 2011. The multifunctional role of EKLF/KLF1 during erythropoiesis. *Blood* 118(8), pp. 2044-2054. doi: 10.1182/blood-2011-03-331371

Singh, D. R. et al. 2015. Unliganded EphA3 dimerization promoted by the SAM domain. *Biochem J* 471(1), pp. 101-109. doi: 10.1042/BJ20150433

Skarnes, W. C., Pellegrino, E. and McDonough, J. A. 2019. Improving homology-directed repair efficiency in human stem cells. *Methods* 164-165, pp. 18-28. doi: 10.1016/j.ymeth.2019.06.016

Bibliography

Smith, M. A., Blankman, E., Gardel, M. L., Luettjohann, L., Waterman, C. M. and Beckerle, M. C. 2010. A zyxin-mediated mechanism for actin stress fiber maintenance and repair. *Dev Cell* 19(3), pp. 365-376. doi: 10.1016/j.devcel.2010.08.008

Spain, S. L. and Barrett, J. C. 2015. Strategies for fine-mapping complex traits. *Hum Mol Genet* 24(R1), pp. R111-119. doi: 10.1093/hmg/ddv260

Spitale, R. C., Tsai, M. C. and Chang, H. Y. 2011. RNA templating the epigenome: long noncoding RNAs as molecular scaffolds. *Epigenetics* 6(5), pp. 539-543. doi: 10.4161/epi.6.5.15221

Stadhouders, R., van den Heuvel, A., Kolovos, P., Jorna, R., Leslie, K., Grosveld, F. and Soler, E. 2012. Transcription regulation by distal enhancers: who's in the loop? *Transcription* 3(4), pp. 181-186. doi: 10.4161/trns.20720

Stalder, A. K. et al. 2005. Invasion of hematopoietic cells into the brain of amyloid precursor protein transgenic mice. *J Neurosci* 25(48), pp. 11125-11132. doi: 10.1523/JNEUROSCI.2545-05.2005

Stanga, S., Lanni, C., Govoni, S., Uberti, D., D'Orazi, G. and Racchi, M. 2010. Unfolded p53 in the pathogenesis of Alzheimer's disease: is HIPK2 the link? *Aging (Albany NY)* 2(9), pp. 545-554. doi: 10.18632/aging.100205

Sternberg, S. H., Redding, S., Jinek, M., Greene, E. C. and Doudna, J. A. 2014. DNA interrogation by the CRISPR RNA-guided endonuclease Cas9. *Nature* 507(7490), pp. 62-67. doi: 10.1038/nature13011

Stewart, C. R. et al. 2010. CD36 ligands promote sterile inflammation through assembly of a Toll-like receptor 4 and 6 heterodimer. *Nat Immunol* 11(2), pp. 155-161. doi: 10.1038/ni.1836

Sugiyama, N., Gucciardo, E., Tatti, O., Varjosalo, M., Hyytiainen, M., Gstaiger, M. and Lehti, K. 2013. EphA2 cleavage by MT1-MMP triggers single cancer cell invasion via homotypic cell repulsion. *J Cell Biol* 201(3), pp. 467-484. doi: 10.1083/jcb.201205176

Sun, N. et al. 2019. Development of drug-inducible CRISPR-Cas9 systems for large-scale functional screening. *BMC Genomics* 20(1), p. 225. doi: 10.1186/s12864-019-5601-9

Sweet, D. R., Fan, L., Hsieh, P. N. and Jain, M. K. 2018. Kruppel-Like Factors in Vascular Inflammation: Mechanistic Insights and Therapeutic Potential. *Front Cardiovasc Med* 5, p. 6. doi: 10.3389/fcvm.2018.00006

Takata, F. et al. 2011. Brain pericytes among cells constituting the blood-brain barrier are highly sensitive to tumor necrosis factor-alpha, releasing matrix metalloproteinase-9 and migrating in vitro. *J Neuroinflammation* 8, p. 106. doi: 10.1186/1742-2094-8-106

Bibliography

Tamura, A. et al. 2017. C/EBPbeta is required for survival of Ly6C(-) monocytes. *Blood* 130(16), pp. 1809-1818. doi: 10.1182/blood-2017-03-772962

Tansey, K. E., Cameron, D. and Hill, M. J. 2018. Genetic risk for Alzheimer's disease is concentrated in specific macrophage and microglial transcriptional networks. *Genome Med* 10(1), p. 14. doi: 10.1186/s13073-018-0523-8

Taylor, H., Campbell, J. and Nobes, C. D. 2017. Ephs and ephrins. *Curr Biol* 27(3), pp. R90-R95. doi: 10.1016/j.cub.2017.01.003

Tessarz, P. and Kouzarides, T. 2014. Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol* 15(11), pp. 703-708. doi: 10.1038/nrm3890

Theriault, P., ElAli, A. and Rivest, S. 2015. The dynamics of monocytes and microglia in Alzheimer's disease. *Alzheimers Res Ther* 7(1), p. 41. doi: 10.1186/s13195-015-0125-2

Thomas, P. and Smart, T. G. 2005. HEK293 cell line: a vehicle for the expression of recombinant proteins. *J Pharmacol Toxicol Methods* 51(3), pp. 187-200. doi: 10.1016/j.vascn.2004.08.014

Thomas, T., Thomas, G., McLendon, C., Sutton, T. and Mullan, M. 1996. beta-Amyloid-mediated vasoactivity and vascular endothelial damage. *Nature* 380(6570), pp. 168-171. doi: 10.1038/380168a0

Togo, T. et al. 2002. Occurrence of T cells in the brain of Alzheimer's disease and other neurological diseases. *J Neuroimmunol* 124(1-2), pp. 83-92. doi: 10.1016/s0165-5728(01)00496-9

Tsompana, M. and Buck, M. J. 2014. Chromatin accessibility: a window into the genome. *Epigenetics Chromatin* 7(1), p. 33. doi: 10.1186/1756-8935-7-33

van Arensbergen, J., FitzPatrick, V. D., de Haas, M., Pagie, L., Sluimer, J., Bussemaker, H. J. and van Steensel, B. 2017. Genome-wide mapping of autonomous promoter activity in human cells. *Nat Biotechnol* 35(2), pp. 145-153. doi: 10.1038/nbt.3754

van Arensbergen, J. et al. 2019. High-throughput identification of human SNPs affecting regulatory element activity. *Nat Genet* 51(7), pp. 1160-1169. doi: 10.1038/s41588-019-0455-2

Vardarajan, B. N. et al. 2015. Rare coding mutations identified by sequencing of Alzheimer disease genome-wide association studies loci. *Ann Neurol* 78(3), pp. 487-498. doi: 10.1002/ana.24466

Varnum, M. M. and Ikezu, T. 2012. The classification of microglial activation phenotypes on neurodegeneration and regeneration in Alzheimer's disease brain. *Arch Immunol Ther Exp (Warsz)* 60(4), pp. 251-266. doi: 10.1007/s00005-012-0181-2

Bibliography

Walker, D. G., Dalsing-Hernandez, J. E., Campbell, N. A. and Lue, L. F. 2009. Decreased expression of CD200 and CD200 receptor in Alzheimer's disease: a potential mechanism leading to chronic inflammation. *Exp Neurol* 215(1), pp. 5-19. doi: 10.1016/j.expneurol.2008.09.003

Walker-Daniels, J., Riese, D. J., 2nd and Kinch, M. S. 2002. c-Cbl-dependent EphA2 protein degradation is induced by ligand binding. *Mol Cancer Res* 1(1), pp. 79-87.

Wang, H. F. et al. 2015a. Effect of EPHA1 genetic variation on cerebrospinal fluid and neuroimaging biomarkers in healthy, mild cognitive impairment and Alzheimer's disease cohorts. *J Alzheimers Dis* 44(1), pp. 115-123. doi: 10.3233/JAD-141488

Wang, P. and Ye, Y. 2021. Astrocytes in Neurodegenerative Diseases: A Perspective from Tauopathy and alpha-Synucleinopathy. *Life (Basel)* 11(9), doi: 10.3390/life11090938

Wang, S. et al. 2021. C/EBPbeta regulates the JAK/STAT signaling pathway in triple-negative breast cancer. *FEBS Open Bio* 11(4), pp. 1250-1258. doi: 10.1002/2211-5463.13138

Wang, W. Y., Tan, M. S., Yu, J. T. and Tan, L. 2015b. Role of pro-inflammatory cytokines released from microglia in Alzheimer's disease. *Ann Transl Med* 3(10), p. 136. doi: 10.3978/j.issn.2305-5839.2015.03.49

Wojtowicz, A., Babu, S. S., Li, L., Gretz, N., Hecker, M. and Cattaruzza, M. 2010. Zyxin mediation of stretch-induced gene expression in human endothelial cells. *Circ Res* 107(7), pp. 898-902. doi: 10.1161/CIRCRESAHA.110.227850

Xu, W. et al. 2015. Meta-analysis of modifiable risk factors for Alzheimer's disease. *J Neurol Neurosurg Psychiatry* 86(12), pp. 1299-1306. doi: 10.1136/jnnp-2015-310548

Yamazaki, Y. and Kanekiyo, T. 2017. Blood-Brain Barrier Dysfunction and the Pathogenesis of Alzheimer's Disease. *Int J Mol Sci* 18(9), doi: 10.3390/ijms18091965

Yamazaki, Y., Zhao, N., Caulfield, T. R., Liu, C. C. and Bu, G. 2019. Apolipoprotein E and Alzheimer disease: pathobiology and targeting strategies. *Nat Rev Neurol* 15(9), pp. 501-518. doi: 10.1038/s41582-019-0228-7

Yang, A. C. et al. 2022a. A human brain vascular atlas reveals diverse mediators of Alzheimer's risk. *Nature* 603(7903), pp. 885-892. doi: 10.1038/s41586-021-04369-3

Yang, C. N. et al. 2011. Mechanism mediating oligomeric Abeta clearance by naive primary microglia. *Neurobiol Dis* 42(3), pp. 221-230. doi: 10.1016/j.nbd.2011.01.005

Yang, J. S., Wei, H. X., Chen, P. P. and Wu, G. 2018. Roles of Eph/ephrin bidirectional signaling in central nervous system injury and recovery. *Exp Ther Med* 15(3), pp. 2219-2227. doi: 10.3892/etm.2018.5702

Bibliography

Yang, Y., Li, Y., Li, X. B., Li, T. X., Qi, J., Zhou, X. and Li, P. 2022b. EGR1 Enhances Lymphangiogenesis via SOX18-Mediated Activation of JAK2/STAT3 Pathway. *Comput Math Methods Med* 2022, p. 6448724. doi: 10.1155/2022/6448724

Ye, F. and Zhang, M. 2013. Structures and target recognition modes of PDZ domains: recurring themes and emerging pictures. *Biochem J* 455(1), pp. 1-14. doi: 10.1042/BJ20130783

Yee, K. S. and Vousden, K. H. 2005. Complicating the complexity of p53. *Carcinogenesis* 26(8), pp. 1317-1322. doi: 10.1093/carcin/bgi122

Yeung, A. T. Y. et al. 2017. Exploiting induced pluripotent stem cell-derived macrophages to unravel host factors influencing Chlamydia trachomatis pathogenesis. *Nat Commun* 8, p. 15013. doi: 10.1038/ncomms15013

Yokokura, M. et al. 2011. In vivo changes in microglial activation and amyloid deposits in brain regions with hypometabolism in Alzheimer's disease. *Eur J Nucl Med Mol Imaging* 38(2), pp. 343-351. doi: 10.1007/s00259-010-1612-0

Yoon, J. H. et al. 2012. LincRNA-p21 suppresses target mRNA translation. *Mol Cell* 47(4), pp. 648-655. doi: 10.1016/j.molcel.2012.06.027

Yoshiyama, Y. et al. 2007. Synapse loss and microglial activation precede tangles in a P301S tauopathy mouse model. *Neuron* 53(3), pp. 337-351. doi: 10.1016/j.neuron.2007.01.010

Zenaro, E. et al. 2015. Neutrophils promote Alzheimer's disease-like pathology and cognitive decline via LFA-1 integrin. *Nat Med* 21(8), pp. 880-886. doi: 10.1038/nm.3913

Zeng, T. et al. 2019. BACE1-AS prevents BACE1 mRNA degradation through the sequestration of BACE1-targeting miRNAs. *J Chem Neuroanat* 98, pp. 87-96. doi: 10.1016/j.jchemneu.2019.04.001

Zhang, Q. et al. 2020. Risk prediction of late-onset Alzheimer's disease implies an oligogenic architecture. *Nat Commun* 11(1), p. 4799. doi: 10.1038/s41467-020-18534-1

Zhang, T., Cooper, S. and Brockdorff, N. 2015. The interplay of histone modifications - writers that read. *EMBO Rep* 16(11), pp. 1467-1481. doi: 10.15252/embr.201540945

Zhong, N. and Weisgraber, K. H. 2009. Understanding the association of apolipoprotein E4 with Alzheimer disease: clues from its structure. *J Biol Chem* 284(10), pp. 6027-6031. doi: 10.1074/jbc.R800009200

Zhou, J. et al. 2019. Anti-inflammatory Activity of MTL-CEBPA, a Small Activating RNA Drug, in LPS-Stimulated Monocytes and Humanized Mice. *Mol Ther* 27(5), pp. 999-1016. doi: 10.1016/j.ymthe.2019.02.018

Bibliography

Zhou, T. et al. 2017. Microglia Polarization with M1/M2 Phenotype Changes in rd1 Mouse Model of Retinal Degeneration. *Front Neuroanat* 11, p. 77. doi: 10.3389/fnana.2017.00077

Zisch, A. H. and Pasquale, E. B. 1997. The Eph family: a multitude of receptors that mediate cell recognition signals. *Cell Tissue Res* 290(2), pp. 217-226. doi: 10.1007/s004410050926

Appendix

Appendix I: EphA1 sequence, domain location and location of P460L variant

>AAD43440.1 EPH receptor A1 (Homo Sapiens)

MERRWPLGLGLVLLLCAPLPPGARAKEVTLMDTSKAQGELGWLLDPPKDGWSEQQQILNG
TPLYMYQDCPMQGRRDTDHWLRSNWIYRGEEASRVHVELQFTVRDCKSFPGGAGPLGCKE
TFNLLYMESDQDVGIQLRRPLFQKVTTVAADQSFTIRDLVSGSVKLNVERCSLGRLTRRG
LYLAFHNPGACVALVSVRVFYQRCPETLNGLAQFPDTLPGPAGLVEVAGTCLPHARASPR
PSGAPRMHCSPDGEWLVPVGRCHCEPGYEEGGSGEACVACPSGSYRMDMDTPHCLTCPQQ
STAESEGATICTCESGHYRAPGEGPQVACTGPPSAPRNLSFSASGTQLSLRWEPPADTGG
RQDVRYSVRCSQCQGTAQDGGPCQPCGVGVHFSPGARGLTTPAVHVNGLEPYANYTFNVE
AQNGVSGLGSSGHASTSVSISMGHAESLSGLSLRLVKKEPRQLELTWAGSRPRSPGANLT
YELHVLNQDEERYQMVLEPRVLLTELQPDTTYIVRVRMLTPLGPGPFSPDHEFRTSPPVS
RGLTGGEIVAVIFGLLLGAALLLGILVFRSRRAQRQRQQRQRDRATDVDREDKLWLKPYV
DLQAYEDPAQGALDFTRELDPAWLMVDTVIGEGEFGEVYRGTLRLPSQDCKTVAIKTLKD
TSPGGQWWNFLREATIMGQFSHPHILHLEGVVTKRKPIMIITEFMENGALDAFLREREDQ
LVPGQLVAMLQGIASGMNYLSNHNYVHRDLAARNILVNQNLCCKVSDFGLTRLLDDFDGT
YETQGGKIPIRWTAPEAIAHRIFTTASDVWSFGIVMWEVLSFGDKPYGEMSNQEVMKSIE
DGYRLPPPVDCPAPLYELMKNCWAYDRARRPHFQKLQAHLEQLLANPHSLRTIANFDPRM
TLRLPSLSGSDGIPYRTVSEWLESIRMKRYILHFHSAGLDTMECVLELTAEDLTQMGITL
PGHQKRILCSIQGFKD

Yellow = Recognition site of the EphA1 ELISA antibody on human recombinant EphA1
Turquoise = EphA1 receptor transmembrane domain
Red = Cytoplasmic domain of EphA1 receptor
Pink = Residue 460 which is altered to a Leucine within the P460L receptor variant

The human EphA1 ELIS kit from R&D systems will recognise the extracellular domain/ amino acid 26-547 (yellow) of the human EphA1 receptor.

Appendix II: Selected *EPHA1* locus lead SNPs based on GWAS p-value

SNP	Position	Gene	Effect Allele	Non-Effect Allele	P Value
rs11767557	7:143412046	*EPHA1-AS*	T	C	1.56E^{-08}
rs10808026	7:143402040	*EPHA1*	A	C	3.06E^{-08}
rs11766230	7:143411748	*EPHA1-AS*	C	T	2.15E^{-08}
rs56402156	7:143406388	*EPHA1*	G	A	2.26E^{-08}
rs11762262	7:143410783	*EPHA1-AS*	C	T	2.14E^{-08}
rs7791765	7:143402014	EPHA1	T	G	3.09E^{-08}
rs11771145	7:143413669	*EPHA1-AS*	A	G	4.82E^{-06}
rs11765305	7:143414019	*EPHA1-AS*	C	G	6.75E^{-06}
rs12703526	7:143410495	*EPHA1-AS*	T	G	1.5E-05
rs7810606	7:143411065	*EPHA1-AS*	T	C	1.62E-05

Appendix III: Characterisation of CRISPR-Cas9 generated control clones

Sequence alignment showing retention of Block 1 non-coding DNA in second control clone

Appendix

Sequence alignment showing retention of Block 1 non-coding DNA in third control clone

Appendix

Sequence alignment showing retention of Block 3 non-coding DNA in first control clone

```
1                                                                            100
1   TAATAATAGTTACCACTTATTACAGCACCTACGCAGGCTGCCGCACTTCACATACATCGTTTCATTGACTTCTCACACCAACCCTCGGAGGAAGGTATCA   100
1   T---------------------------------------------------------------------------------------------------   1
1   T---------------------------------------------------------------------------------------------------   1
1   T---------------------------------------------------------------------------------------------------   1

101                                                                          200
101 TTAGTCTCATTTTATAGGTGGAAAAATCTGAGACACCGGGAATTAAAGTAGTTCAAAGTCACTCAACTGCAGAACCAGGATTTAAACCTAGTTCTGTCTG   200
1   -----------------------------------------------------------------------------------------------------   1
1   -----------------------------------------------------------------------------------------------------   1
1   -----------------------------------------------------------------------------------------------------   1
                                                                     Forward Primer
201                                                                          300
201 ACTCCCAGTCCTGTGGCCTAACCTCTATTCAGGGTTTCAAGAAGCAGAGATGTTAGCAAACACCTCACTTTTGTTTTACTACCCCGCCCCCCACCTGAGA   300
1   -------------------------------------------------------------------------TTACTACCCCGCCCCCCACCTGAGA   26
1   -------------------------------------------------------------------------TTACTACCCCGCCCCCCACCTGAGA   26
1   -------------------------------------------------------------------------TTACTACCCCGCCCCCCACCTGAGA   26

301                                                                          400
301 GTAATATCCTGACTTAAAACACCACGGAGTGGATTTGCCTGTTTTTAAACTTTATGAAACGGAATCATCTGATATGCATTCTTTGGTGTCCGAGTTCTTT   400
27  GTAATATCCTGACTTAAAACACCACGGAGTGGATTTGCCTGTTTTTAAACTTTATGAAACGGAATCATCTGATATGCATTCTTTGGTGTCCGAGTTCTTT   126
27  GTAATATCCTGACTTAAAACACCACGGAGTGGATTTGCCTGTTTTTAAACTTTATGAAACGGAATCATCTGATATGCATTCTTTGGTGTCCGAGTTCTTT   126
27  GTAATATCCTGACTTAAAACACCACGGAGTGGATTTGCCTGTTTTTAAACTTTATGAAACGGAATCATCTGATATGCATTCTTTGGTGTCCGAGTTCTTT   126

401                                                                          500
401 CATTTACCGTTACGTTTGGAAGATGTATTCCTTGGTTGTGTGTGGATTCAGTCTCATTGCTGTATGGTATCCCACTGTATAAAGAACACCTTATTTTTAA   500
127 CATTTACCGTTACGTTTGGAAGATGTATTCCTTGGTTGTGTGTGGATTCAGTCTCATTGCTGTATGGTATCCCACTGTATAAAGAACACCTTATTTTTAA   226
127 CATTTACCGTTACGTTTGGAAGATGTATTCCTTGGTTGTGTGTGGATTCAGTCTCATTGCTGTATGGTATCCCACTGTATAAAGAACACCTTATTTTTAA   226
127 CATTTACCGTTACGTTTGGAAGATGTATTCCTTGGTTGTGTGTGGATTCAGTCTCATTGCTGTATGGTATCCCACTGTATAAAGAACACCTTATTTTTAA   226

501                                                                          600
501 AAACTTCTTGACCTGTGGTTAAACAGGTTTAGGTTCTACCTTTGAATTGCTACATTGAGCGACTTTCATGGTTTATCCGGCTTATCTGTGTCTCTGAGAG   600
227 AAACTTCTTGACCTGTGGTTAAACAGGTTTAGGTTCTACCTTTGAATTGCTACATTGAGCGACTTTCATGGTTTATCCGGCTTATCTGTGTCTCTGAGAG   326
227 AAACTTCTTGACCTGTGGTTAAACAGGTTTAGGTTCTACCTTTGAATTGCTACATTGAGCGACTTTCATGGTTTATCCGGCTTATCTGTGTCTCTGAGAG   326
227 AAACTTCTTGACCTGTGGTTAAACAGGTTTAGGTTCTACCTTTGAATTGCTACATTGAGCGACTTTCATGGTTTATCCGGCTTATCTGTGTCTCTGAGAG   326

601                                                                          700
601 GTTTAATCTTTCCAAACAAAACGTACTTCCTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTCAGGTGGAGTCTCGCTCTGTCGCCCAGGCTGGA   700
327 GTTTAATCTTTCCAAACAAAACGTACTTCCTTTTTTTTTTTTG-------------------------AGGGGGAGTCTCGCTCTGTCCCCCAGGCTGGA   400
327 GTTTAATCTTTCCAAACAAAACGTACTTCCTTTTTTTTTTTTTG------------------------AGGGGGAGTCTCGCTCTGTCCCCCAGGCTGGA   400
327 GTTTAATCTTTCCAAACAAAACGTACTTCCTTTTTTTTTTTTTG------------------------AGGGGGAAGCCCCCCTCTGCCGCCCAGGCGGGA   400

701                                                                          800
701 GTGCACTGGCGCAATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTACGTCATTCTCCTGCCTCAGCCTCCCGAATAGCTGGGACTACAGGCGCCCACC   800
401 GGGCACGGGCGCAATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTACGTCATTCCCCTGCCTCACCCTCCCAAATAGCTGGGACTACAGGCCCCCACC   500
401 GGGCACGGGCGCAATCTCGGCTCACTGCAAGCTCCGCCTCCCGGGTTACGTCATTCCCCTGCCTCACCCTCCCAAATAGCTGGGACTACAGGCCCCCACC   500
401 GTGAATTGGCCAAATCTCGGCTCACGGAAAGCCCCCCCCCCCGGGTTACGTCTTTCTCCGGCCTCAGCCCCCCGAAAAGTGGGAATTACGGGGGCCCACC   500

801                                                                          900
801 ACCACGCCCGGCTAATTTTTTTGTATTTTTAGTAGAGACGGGGTTTCACCGTGTTAGCCAGGATGGTCTCCATCTCCTGACCTCGTGATCTGCCCGCCTCG   900
501 ACCACCCCCGGCTAATTTTTTTGTATTTTTAGAAAAGACGGGGTTTCACCGTGTTAGCCAGGAGGGTCTCCATCTCCGGACCTCGGGATCGGCCCGCCTCG   600
501 ACCACCCCCGGCTAATTTTTTTGTATTTTTAGAAAAGACGGGGTTTCACCGTGTTAGCCAGGAGGGTCTCCATCTCCGGACCTCGGGATCGGCCCGCCTCG   600
501 ACCCCGCCGGGCTATTTTTTTGAATTTTAAGTAAAAAGGGGTTTCCCCCGGGTTACCCGGGATGGTCCCCTCCCCCTGACCCCGGGATTTGCCCCCCTCC   600

901                                                                          1000
901 GCCTCCCAAAGTGCTGGGATTACAGGCATGAGCCACTGCGCCCGGCCGTACTTCCTTTTTTAAACTAGACTTTTGCTAACCCAAATGATTCTTTTTTTTT   1000
601 GCCTCCCAAAGTGGTGGGATTACAGGCAAGAGCCACGGCGCCCGGCCGTCCTTCCTTTTTTAAACTAAACTTTTGCTAACCCAAAGGATTCTTTTTTTTC   700
601 GCCTCCCAAAGCGGTGGGATTACAGGCAAGAGCCACGGCGCCCGGCCGTCCTTCCTTTTTTAAACTAAACTTTTGCTAACCCAAAGGATTCTTTTTTTTC   700
601 GCCTCCCAAAAGGGCGGGAATTACGGGTATGAGCCATTGCGCCGGCCGTACTTCCTTTTTTAAACAAAACTTTTGCAAACCAAAATGATTTTTTTTTTTC   700

1001                                                                         1100
1001 CTTTTTTGAGATGGAGTCTCACTCTGTTGCCCAGGCTGAAGTGCAGTGGAATGATCTCGGC-TCACTGCAACCTCTGCCTCCCAGGTTCAAGCGATTCTC   1099
701  CTTTTTTAAAAGGGACTCTCACTCTGTTGCCCAGGTTGAAATGCATAGCAAAGTTCCCGGCATCT-TGAAACCTCTGTCTCCCAGGTCTAAGCGATCCTC   799
701  CTTTTTTGAAAAGGGACTCTCACTCTGTTGCCCAGGTTGAAATGCATAGCAAAGTTCCCGGCATC-TTGAAACCTCTGTCTCCCAGGTCTAAGCGATCCTC   799
701  CTTTTTCGAGATGAAGCCTCACTCTGTTGCCCAAGGCTGAAATCTATAGAAAAGATCCCGACTCA-CTGAAACCTATGCCACCGAGGTTTGAGCGATTCTC   799

1101                                                                         1200
1100 CTGCT----TCAGCCTCCTGAGTAG-CTGGGATT--ACAGGTACCTGCCA-CCATGCCCGGCTAATTTTTGTATTTTTAGTAGAGACGGGGTTTCACCATG   1192
800  CTGCTCGCTC--CCTCCCGAGTAGACTGGGAATAAACAGGTCCCTGGCACCCATGCCCTGATAATTTGCGTGTTTTAAA-ACAGA-GGGGTTTC-CCATG   894
800  CTGCTCGCTC--CCTCCCGAGTAGACTGGGAATAAACAGGTCCCTGGCACCCATGCCCTGATAATTTGCGTGTTTTAAA-ACAGA-GGGGTTTC-CCATG   894
800  CTGCTCATC--CCTCCCGAGAAGACTGGGCTGTAACGGGTCCCTCGCACCCGTCCCCGGATAATTTTCTGTTTTAGA-AATGA-GGAGTATC-CCATG   894

1201                                                                         1300
1193 TTGGCCAGGCTGGTCTCAAACTCCTGATCTCAAGTAATCTGCCCGTCTTGGCCTCCCAAAGTGCTGAA-AT-TAC-AGGAGTGAGCCAT-GACACCCAGC   1288
895  TTGCCCACGCCGGTCTCAAAATCCTGATTCCAAGAAAATGCCCCCTCTAGTCCCCCCTACATGGTTAACATCTACTGGGAGTGGCTCATAAACCCCCAGC   994
895  TTGCCCACGCTGGTCTCAAAATCCTGATTCCAAGAAAATGCCCCCTCTAGTCCCCCCTACATGGTTAACATCTACTGGGAGTGGCTCATAAACCCCCAGC   994
895  TTGGCCAGGCGGGCCCCAAACTCCCGATCCCAAGTAATGGCCAGTCTAGTCCCCCCAAAATGGTTAAAATTTACTGGAAGTGGCCCATGATCCCCCGC   994

1501                                                                         1600
1416 ---------GGAAGACA-------CCCGATACTGTTTA---------------------GGA--ATCT---GTCTTCTT------GCTAGAAAT-------   1462
1185 AATTTAAATAAAAAACAAAAAAATCCC----CTGTGAAAAAATCCCCCAGGAAAAAAGGGA--ATTT---TTTTTTTTTTTTCCCCAAAAAATTTTTTTA   1275
1185 AATTTAAATAAAAAACAAAAAAAGCCC----CTTTGGAAAAAACCCCCAGGAAAAAAGGG--GATTT---TTTATTTGTTTTCCCCAAAAAATTTTTTTA   1275
1186 CTTTTTTATAAAACACAAAAAAA--CC----CTCAGTTAAAATCTTCCAGGAGAAAAGGGACCATTTCCCTTTCTTT---TGCCCCAAAAAGTTTTTATA   1276

1601                                                                         1700
1462 -------------------------------------------------AG-AA-------CTG-------------TTTT----   1473
1276 TTTGCCCCCCTT--TTTTTTTTAAAGGGGGGGTTCCG--GCCCTTTCC--CCCTTACGGGGGGGGGAA-AACCCCCCCCCTGGGGGGGGGGGGGAAATTTT----   1364
1276 TTTGCCCCCCTT--TTTTTTTTAAAGGGGGGGTTC--CCGCCTTTTCC--CCCTAACGGGGGGGGGAA-AACCCCCCCCCGGGGGGGGGGGGGAAATTTTA----   1365
1277 TGTGCCCC---TTAATATTTTTAGAGGGG--TTCTGTCGCATTTTCCCTCCCCA---GGGGGGGGAACAACCCCCCCCTATGGGGGGGGA--AAATTATACCC   1367
                                                    Reverse Primer
1701                                                                         1797
1474 GAGATAGAGGCGG----------------------TTGGAAG---------------------TTTTCTC-------------------   1500
1365 AAAAGGGGGGGGGGGGGGGGGCCCCTTTTTTTTTTTTTTTCTCAGAAGGTT---CAAAAAAAAAAAAATATTTTTTTTTTCCCTCCCTTT--TTTTTTTGGG-CCC   1455
1365 -AATTGGGGGGGGGGGGGCCCCTTTTTTTTTTTTTTTC-TC--AGGGTGTGCAAAAAAAAAAAAATATTTTTTTTTTTCCCTCCCTTT--TTTTTTTGGG-GCC   1455
1368 GA--TGGGGGGGGGGGGGGC---TTTTTTTTTTTTTTCTTC--AAGTTATGCCAAAAAAAAAAAAATATATTTTTT--CCCCCACTTTTCTTTTTTTTGGGAACC   1455
```

214

Appendix

Sequence alignment showing retention of Block 3 non-coding DNA in second control clone

Appendix

Sequence alignment showing retention of Block 3 non-coding DNA in third control clone

Sequence alignment showing deletion of Block 1 non-coding DNA in second clone

Sequence alignment showing deletion of Block 1 non-coding DNA in third clone

Appendix

Sequence alignment showing deletion of Block 3 non-coding DNA in first clone

Appendix

Sequence alignment showing deletion of Block 3 non-coding DNA in second clone

```
         1                                                                                                  100
    1 ATCTTGCAAATAAACCTGTACAGGAGGACACAAAAATAAAAAACATGTATCCTTAAATTGCATGGTGCCACCAGAAGCTTTTGCTTGTACACTTTCTCAT 100
    1 ATCTT----------------------------------------------------------------------------------------------- 5
    1 ATCTT----------------------------------------------------------------------------------------------- 5
    1 ATCTT----------------------------------------------------------------------------------------------- 5

       101                                                                                                 200
  101 ATTTGTACTTTTCTCATATACTATGACCAACAACATTATCATCATTTCCCCCATTTTGCAGATGAGGAAGCGGAGTCTCAAAACAGCTAGGTGATTTGCC 200
    5 --------------------------------------------------------------------------------------------------- 5
    5 --------------------------------------------------------------------------------------------------- 5
    5 --------------------------------------------------------------------------------------------------- 5

       201                                                                                                 300
  201 CAAGATCACTTGGCTATGAAGTGACAATATTAAAATCTTCTAAAATATAACTTTGCTTTTGACACTTCAGAGAAAACTTCCAACCGCCTCTATCTCAAAA 300
    5 -----------------------------------------CTAAAATATAACTTTGCTTTTGACACTTCAGAGAAAACTTCCAACCGCCTCTATCTCAAAA 66
    5 -----------------------------------------CTAAAATATAACTTTGCTTTTGACACTTCAGAGAAAACTTCCAACCGCCTCTATCTCAAAA 66
    5 -----------------------------------------CTAAAATATAACTTTGCTTTTGACACTTCAGAGAAAACTTCCAACCGCCTCTATCTCAAAA 66

       301                                                                                                 400
  301 CAGTTCTATTTCTAGCAAGAAGACAGATTCCTAAAACA-GTATCGGGTGTCTTCCTCTTAGCAAACTGACTGGAAGAAGGGAGGACATGCAGGGTTTAAGG 399
   67 CAGTTCTATTTCTAGCAAGAAGACAGATTCCTAAATG-TT------TATATTACTCATA----------TGG------------------------GG 123
   67 CAGTTCTATTTCTAGCAAGAAGACAGATTCCTAAATGTTT-------ATATTCCTC------------------------ATATG---------TGG 123
   67 CAGTTCTATTTCTAGCAAGAAGACAGATTCCTAAATGTTTA------ATATTC-TC------------------------ACGTG---------AGG 123

       401                                                                                                 500
  400 GGGAGGGTGTATAATTGGGACCTGTGGCACTGCTTACAGACCGCCTCTGCATACTTTTAATCAATTAAAAAGTTGTTAGGAGGCTGGGTGTCATG-GCTC 498
  124 GGCGGGGGGTA------GGA---------------------------------------AAACAAATGAGACGT--------------GTGTGATGAGCTC 165
  124 GGCGGGGGGGTAGA--A------------------------------------------AAACAAATGAGACGTG------------TGTGATAAGCTC 165
  124 GG--GGCGGTAGAATA------------------------------------------AGAGAAGTGAGATGTT------------TGTGATCAGATC 165

       501                                                                                                 600
  499 ACT-CCTGTAATTTCAGCACTTTGGGAGGCCAAGACGGGCAGATTACTTGAGATCAGGAGTTTGAGACCAGCCTGGCCAACATGGTGAAACCCCGTCTCT 597
  166 TCTGCCT----TCTCAGC-CT------------------------CTTGATACC------TTG---------------------------------- 193
  166 TCTACCT----TCTCGAGC--------------------------------------------------------------------------CTCT 183
  166 GCTTTCT----TCTATAC--------------------------------------------------------------------------CTCT 183

       601                                                                                                 700
  598 ACTAAAAATACAAAAATTAGCCGGGCATGGTGGCAGGTACCTGTAATCCCAGCTACTCAGGAGGCTGAAGCAGGAGAATCGCTTGAACCTGGGAGGCAGA 697
  193 --------TACA-----------------------------CCAG---------------------------------TTGA-------------- 205
  183 ----------------------GGA-----------------TA--CACTA-TAC-----------------------------------GCCAG- 201
  184 A--AAAA---------------GGGT-----------------TAATCACAG-GACT-----------------------------------GC---- 207

       701                                                                                                 800
  698 GGTTGCAGTGAGCCGAGATCATTCCACTGCACTTCAGCCTGGGCAACA-GAGTGAGACTCCATCTCAAAAAAGAAAAAAAAGAATCATTTGGGTTAGCA 796
  205 -------GTGAG-------------ACT---------------AACACAAGTCAGACAC----------------------------TTT--------- 232
  201 ---TTAAGTGAGACAA---------------------------ACACAAGTCAGACAC----------------------------TTT--------- 232
  207 ---------GAGAGGC-------------------------ATACAAGTCAGACAC------------------------------TTT--------- 232

       801                                                                                                 900
  797 AAAGTCTAGTTTAAAAAAGGAAGTACGGCCGGGCGCAGTGGCTCATGCCTGTAATCCCAGCACTTTGGGAGGCCGAGGCGGGCAGATCACGAGGTCAGGA 896
  232 ----------TTT---------------------------CTCAT-------------------------------------------------- 240
  232 ----------TTT---------------------------CTCAT-------------------------------------------------- 240
  232 ----------TTT---------------------------TTCAT-------------------------------------------------- 240

       901                                                                                                1000
  897 GATGGAGACCATCCTGGCTAACACGGTGAAACCCCGTCTCTACTAAAAATACAAAAAATTAGCCGGGCGTGGTGGTGGGCGCCTGTAGTCCCAGCTATTC 996
  240 ----------ATCCTC-----------------CTCTTCTTATCATA-------TT----------------------------------TCTT- 266
  240 ----------ATCCTC-----------------CTCTTCTTATCATA-------TT----------------------------------TCTT- 266
  240 ----------ATCCCC-----------------TTCCTATTCTCATA-------TT----------------------------------TCTT- 266

       1001                                                                                               1100
  997 GGGAGGCTGAGGCAGGAGAATGACGTAACCCGGGAGGCGGAGCTTGCAGTGAGCCGAGATTGCGCCAGTGCACTCCAGCCTGGGCGACAGAGCGAGACTC 1096
  266 --------------------------------------------------------------------------------------------------- 266
  266 --------------------------------------------------------------------------------------------------- 266
  266 --------------------------------------------------------------------------------------------------- 266

       1101                                                                                               1200
 1097 CACCTGAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGGAAGTACGTTTTGTTTGGAAAGATTAAACCTCTCAGAGACACAGATAAGCCGGATA 1196
  266 ----------------------------------------------TTTTGTTTT-------------TCT-----------------GCATA 283
  266 ----------------------------------------------TTTTGTTTT--------------TCTG---------------- 279
  266 ----------------------------------------------TTTTGTTTT--------------CTAG---------------- 279

       1201                                                                                               1300
 1197 AACCATGAAAGTCGCTCAATGTAGCAATTCAAAGGTAGAACCTAAACCTGTTTAACCACAGGTCAAGAAGTTTTTAAAAATAAGGTGTTCTTTATACAGT 1296
  283 --------------------------------TATTCCTGGTT---------------------------TGCTCGTT------- 302
  279 ---CAT--------------------------ATATTCCTGGTT---------------------------TGCTCGTT------- 302
  279 ---CAT--------------------------CTATTCCTGGTG---------------------------TGCCGGTC------- 302

       1301                                                                                               1400
 1297 GGGATACCATACAGCAATGAGACTGAATCCACACACAACCAAGGAATACATCTTCCAAACGTAACGGTAAATGAAAGAACTCGGACACCAAAGAATGCAT 1396
  302 ----TGCCTT--------------------------------GT----------------------TGCAT 315
  302 ----TGCCTT--------------------------------GT----------------------TGCAT 315
  302 ----TGTCTT--------------------------------GT----------------------TGCAT 315

       1401                                                                                               1500
 1396 -ATCAGATGATTCCGTTTCATAAAGTTTAAAAACAGGCAAATCCACTCCGTGGTGTTTTAAGTCAGGATATTACTCTCAGGTGGGGGGCGGGGTAGTAAA 1495
  316 CATAAGATAAT--------ATTA------------------TCCCCCCC--------------------ATTGCT--------GGCCC---------- 349
  316 CATAAGATAAT--------ATTA------------------TCCCCCCC--------------------ATTGCT--------GGCCC---------- 349
  316 CATAAGAAAAT--------ACTA------------------TCCTCCCC--------------------ATTGCT--------GGTCC---------- 349

       1501 1505
 1496 ACAAA 1500
  349 ----T 350
  349 ----T 350
  349 ----T 350
```

Appendix

Sequence alignment showing deletion of Block 3 non-coding DNA in third clone