The University of Nottingham

UNITED KINGDOM · CHINA · MALAYSIA

Identification of Functional Variants in the Alzheimer's Disease Candidate Gene *ABCA7*

Naomi Susan Clement BMedSci (Hons)

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

March 2016

Mischief managed.

Mooney, Wormtail, Padfoot and Prongs

Harry Potter and the Prisoner of Azkaban

Abstract

Late onset Alzheimer's disease (LOAD) is the commonest form of dementia, affecting approximately 850,000 patients in the UK alone, predicted to exceed one million by 2025. The cause of LOAD is complex, but several large Genome Wide Association Studies have highlighted 21 genetic loci associated with this devastating disease and the ATP-Binding Cassette Protein, family A, member 7 (*ABCA7*) is one of these genetic loci. However, the exact reasons behind this association are still unknown, focusing work on identifying functional, pathogenic mutations within this locus.

A total of 240 exonic variations within *ABCA7* were therefore annotated in order to identify ones potentially altering the functionality of ABCA7. A total of five variants were predicted to be damaging by *in silico* annotation tools: rs3752233; rs59851484; rs3752237; rs114782266 and a novel mutation at genomic position 19:1056958. These were genotyped in the ARUK DNA Bank resource and three (rs59851484, rs3752239 and 19:1056958) showed tentative association with LOAD. However, lack of power in this study prevented any definitive associations from being formed. A further two variants were examined within functional cell assays. rs881768 had been predicted to affect the splicing of the ABCA7 protein and appeared to do so within minigene cellular assays. However, this did not appear to be the case when RNA from brain tissue harbouring this variation was examined. rs2020000 was examined through the dual luciferase assays, with the minor allele seeming to down regulate the reporter protein by approximately 30% ($p<0.02$) in these *in vitro* assays.

Functional variations within the *ABCA7* locus do play a role in LOAD risk and improvements within functional databases and annotation programmes will assist in identifying these causative mutations, in order to put a halt to LOAD, as well as other destructive complex disorders.

Publications and Presentations

Publications

Clement, N., Medway, C., Carrasquillo, M. M., Guetta-Baranes, T. et al., Mapping regulatory variants in the AD candidate gene *ABCA7*. Manuscript in preparation.

Clement, N., Braae, A., Turton, J., Lord, J. et al., November 2016. Investigating splicing variants uncovered by next-generation sequencing the Alzheimer's disease candidate genes, *CLU*, *PICALM*, *CR1*, *ABCA7*, *BIN1*, the *MS4A* locus, *CD2AP*, *EPHA1* and *CD33*. J. Alzheimer's Dis. Park.

Boden, K., Barber, I., ***Clement N***., et al., October 2016. Methylation profiling of key genes associated with sporadic early onset Alzheimer's disease. Submitted to J. of Alzheimer's Dis.

Sassi, C., Guerreiro, R., …, ***Clement N.***…., et al., October 2016. *ABCA7* p.G215S as potential protective factor for Alzheimer's disease. Neurobiol. Aging. doi:10.1016/j.neurobiolaging.2016.04.004

Barber, I.S., Braae, A., ***Clement, N.***, et al., September 2016. Mutation analysis of sporadic early-onset Alzheimer's disease using the NeuroX array. Neurobiol. Aging.

Brookes K., Patel T., … ***Clement, N.*** …, et al., May 2016. Identifying Polymorphisms in the Alzheimer's Related APP Gene Using the Minion Sequencer. Next Generat Sequenc & Applic 3:125. doi:10.4172/2469-9853.1000125

Barber, I.S., García-Cárdenas, J.M., …, ***Clement, N.***…., et al., March 2016. Screening exons 16 and 17 of the amyloid precursor protein gene in sporadic early-onset Alzheimer's disease. Neurobiol. Aging 39, 220.e1-7. doi:10.1016/j.neurobiolaging. 2015.12.011

Presentations

"Examining Predicted Splice Variants in Late Onset AD Risk Loci." Alzheimer's Research UK Conference 2015. Poster presentation, 9^{th}-11^{th} March 2015, London, UK.

"Mapping Regulatory Variants in the AD Candidate Gene *ABCA7*." Alzheimer's Research UK Conference 2016. Poster presentation, 7^{th}-9^{th} March 2016, Manchester, UK.

"Mapping Regulatory Variants in the AD Candidate Gene *ABCA7*." Alzheimer's Association International Conference 2016. Poster presentation, 24^{th}-28^{th} 2016, Toronto, Canada.

"Mapping Regulatory Variants in Late Onset AD Candidate Genes." Alzheimer's Research UK Midlands Network Conference 2016. Oral Presentation, 13^{th} September 2016, Loughborough, UK.

"Mapping Regulatory Variants in the AD Candidate Gene *ABCA7*." East Midlands Student Research Conference 2016. Oral Presentation, 5^{th} November 2016, Leicester, UK. Awarded Best Oral Breakout Presentation.

"Mapping Regulatory Variants in the AD Candidate Gene *ABCA7*." Genetics Society 2016 Autumn Meeting. Poster presentation, 10th-11th November 2016, London, UK.

Grants

Glomax 96 Microplate Luminometer, awarded to ***Naomi Clement*** and Professor Kevin Morgan. Equipment Grant (£2000) received from ARUK East Midlands Network. January 2016.

Acknowledgements

First of all, I would like to acknowledge my supervisors Professors Kevin Morgan and Mohammad Ilyas for their support and advice, especially Kevin for his endless advice, positivity and (sometimes) humorous stories!

I am also indebted to the Jean Shanks Foundation for funding me throughout my PhD, as well as ARUK for their support of the Human DNA and Brain Bank hosted in the Human Genetics laboratory at the University of Nottingham. And of course, my gratitude to all the patients and families who have agreed to donate their tissues as, without them, none of this work would be possible.

My thanks also go to the students whose work has contributed in part to this thesis: Sara Garin; Akili Mata and Ahmed Alahmad (MSc. Students) whose dissertation works form part of Chapter 3. My appreciation also to the Molecular Diagnostics lab for analysing all of the Sanger sequencing performed throughout this project.

Tamar Guetta-Baranes: I really couldn't have done this without your constant advice, teaching and, of course, the regular escape to your office. Thank you for everything. To all the other members of the Human Genetics Group as well, thanks for providing such a wonderful atmosphere in which to work: Anne Braae; Keeley Brookes; Tulsi Patel and Sally Chappell all provided help, advice and, when needed, just a chat. Particular thanks to Keeley and Anne for their comments on much of this thesis. I would also like to thank James Turton for performing much of the ground work for this thesis as well as introducing me to the wonderful world that is human genetics.

My unreserved thanks and love to my family whom, again, I couldn't have done this without. Thanks to Mum and Dad for providing me so many opportunities throughout my life and allowing me to take this one. My total gratitude also goes to Emma, Nana and Helen for always being there for me, no matter what. I love you all unconditionally. Mentions as well, to the friends who sometimes knew not to ask questions and just provide more wine: Abigail; Jack and Adam.

And finally to Rob, for being my rock. I don't know what I did to deserve you but I am so glad you pushed me to do this, stuck with me throughout the roller coaster this thesis was and believed in me, even when I didn't myself. I hope one day I can even partially return the favour.

Table of Contents

Table of Figures

Table of Tables

List of Abbreviations

$(NH_4)_2SO_4$	Ammonium Sulphate
µl	Microliter
µm	Micromole
ABC	ATP-Binding Cassette
ABCA1	ATP-Binding Cassette Protein, Family A, Member 1
ABCA7	ATP-Binding Cassette Protein, Family A, Member 7
ACh	Acetylcholine
AChE	Acetylcholinesterase
AD	Alzheimer's Disease
ADP	Adenosine Diphosphate
AICD	Amyloid Precursor Protein Intracellular Domain
AKAP9	A-Kinase Anchoring Protein 9
AMP^r	Ampicillin Resistance gene
apoA-I	Apolipoprotein A-1
APOE	Apolipoprotein E
APP	Amyloid Precursor Protein
APPs	Secreted APP Ectodomain
ARUK	Alzheimer's Research UK
ATP	Adenosine Triphosphate
Aβ	Amyloid-Beta
BACE1	Beta-Secretase 1
BDGP	Berkeley Drosophila Genome Project
BIN1	Myc Box-Dependant-Interacting Protein 1
bp	Base Pair
BRCA1	Breast Cancer Gene 1
CASS4	Cas Scaffolding Protein Family Member 4
CD2AP	Cluster of Differentiation 2-Associated Protein
CD33	Sialic Acid Binding Immunoglobulin-Like-Lectin-3
cDNA	Complementary DNA
CED-7	Cell Death Abnormality Gene, Member 7
CELF1	CUG Triplet Repeat RNA Binding Protein Elav-Like, Member 1
CERAD	Consortium to Establish a Registry for Alzheimer's Disease
CFTR	Cystic Fibrosis Transmembrane Conductase Regulator (also known as *ABCC7*)
ChIP-seq	Chromatin Immunoprecipitation with Parallel DNA Sequencing
CI	Confidence Interval

CLU	Clusterin (Apolipoprotein J)
CNN2	Calporin 2
CNS	Central Nervous System
CO_2	Carbon Dioxide
CpG	5’ - C - Phosphate - G - 3’
CR1	Complement Receptor 1
CRISP	Comprehensive Read Analysis for Identification of Single Nucleotide Polymorphisms from Pooled Sequencing
CSF	Cerebral Spinal Fluid
CT	Computed Tomography
CV	Coefficient of Variation
dH_2O	Distilled Water
DMEM	Dulbecco’s Modified Eagle Medium
DMSO	Dimethyl Sulfoxide
DNA	Deoxyribonucleic Acid
DNase I	Deoxyribonuclease I
dNTPs	Deoxynucleotide
DSG2	Desmoglein 2
ECD	Extracellular Domain
EDTA	Ethylenediaminetetraacetic acid
eGWAS	Expression Genome Wide Association Study
EMBL-EBI	European Molecular Biology Laboratory - European Bioinformatics Institute
EMEM	Eagle’s Modified Essential Medium
EMSA	Electrophoretic Mobility Assay
ENCODE	Encyclopaedia of DNA Elements
EOAD	Early Onset Alzheimer’s Disease
EOFAD	Early Onset Familial Alzheimer’s Disease
EPHA1	Erythropoietin-Producing Human Hepatocellular Carcinoma 1
eQTL	Expression Quantitative Trait Loci
ESE	Exonic Splicing Enhancer
ESS	Exonic Splicing Silencer
EVS	Exome Variant Server
ExoSAP-IT	Exonuclease I - Shrimp Alkaline Phosphatase
FBS	Foetal Bovine Serum
FERMT2	Fermitin Family Member 2
fMRI	Functional Magnetic Resonance Imagine
FRET	Florescence Resonance Energy Transfer
FTD	Frontal Temporal Dementia

g	Gram/ Gravity
GP	General Practitioner
GRN	Granulin
GWAS	Genome Wide Association Study
HA-1	Minor Histocompatibility Protein
HDL	High Density Lipoprotein
HEK	Human Embryonic Kidney Cells
hESC	Human Embryonic Stem Cells
HH1	Hydrophobic Loop
HIV	Human Immunodeficiency Virus
HLA-DRB	Human Leukocyte Antigen - D Related Beta Chain
HUVEC	Human Umbilical Vein Endothelial Cells
IGAP	International Genomics of Alzheimer's Project
INDEL	Insertion - Deletion
INPP5D	Inositol Polyphosphate - 5 - Phosphatase D
iPSC	Induced Pluripotent Stem Cells
IQR	Interquartile Range
ISE	Intronic Splicing Enhancer
ISS	Intronic Splicing Silencer
KASP	Competitive Allele Specific Polymerase Chain Reaction
kbp	Kilo Base Pair
kDa	Kilo Daltons
LARII	Luciferase Assay Reagent II
LCL	Lymphoblastoid Cell Lines
LD	Linkage Disequilibrium
LDL	Low Density Lipoprotein
LOAD	Late Onset Alzheimer's Disease
LOD	Log of the Likelihood Odds Ratio
LRP1	Low Density Lipoprotein Receptor - Related Protein 1
luc+	Luciferase gene
LXR	Liver X Receptor
M	Molar
MAF	Minor Allele Frequency
MAPT	Microtubule Associated Protein Tau
Mbp	Mega Base Pair
MCI	Mild Cognitive Impairment
MEF2C	Myocyte Enhancer Factor 2 C
mg	Milligram

$MgCl_2$	Magnesium Chloride
miRNA	Micro Ribonucleic Acid
ml	Millilitre
mM	Micro Molar
mm	Millimetre
MRI	Magnetic Resonance Imaging
mRNA	Messenger Ribonucleic Acid
MS4A	Membrane-Spanning 2-Domain Family, Subfamily A
MUT	Mutant
NA	Not Available
NBD	Nuclear Binding Domain
NFT	Neurofibrillary Tangle
ng	Nano Gram
NGRL	National Genomics Reference Laboratories
NGS	Next Generation Sequencing
NICE	National Institute for Health and Clinical Excellence
NINCDS -ADRDA	National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association
nl	Nanoliter
NMD	Nonsense Mediated Decay
NMES	Nucleoside Diphosphatase Kinase Family Member 8
NSAIDs	Non-Steroidal Anti-Inflammatory Drugs
NTC	Non Template Control
OR	Odds Ratio
PBS	Phosphate Buffered Saline
PCA	Principal Component Analysis
PCR	Polymerase Chain Reaction
PICALM	Phosphatidylinositol Binding Clathrin Assembly Protein
PLD3	Phospholipase D Family Member 3
pmol	Picomolar
pre-mRNA	Precursor - Messenger Ribonucleic Acid
PRS	Polygenic Risk Score
PSEN	Presenilin
PTC	Premature Termination Codon
PTK2B	Protein Tyrosine Kinase 2 Beta
qPCR	Quantitative Polymerase Chain Reaction
RIN3	Ras and Rab Interactor 3
RNA	Ribonucleic Acid

RNase	Ribonuclease
RNA-seq	Ribonuclease - Sequencing
rpm	Rotations Per Minute
RT-PCR	Reverse Transcriptase Polymerase Chain Reaction
RXR	Retinoid X Receptor
SD	Standard Deviation
siRNA	Small Interfering Ribonucleic Acid
SOC	Super Optimal Broth with Catabolite Repression
SORL1	Sortilin-Related Receptor L
SR	Serine Arginine
SREBP1	Sterol Regulatory Element-Binding Transcription Factor 1
SRSF	Serine/Arginine Rich Splicing Factors
SV40	Simian Vacuolating Virus 40
TAE	Tris Base, Acetic Acid and EDTA Buffer Solution (for components see Section 3.2.3)
TE	Tris and EDTA Buffer Solution (for components see Section 4.2.13)
TF	Transcription Factor
TFBS	Transcription Factor Binding Site
Tm	Annealing Temperature
TMD	Transmembrane Domain
TREM2	Triggering Receptor Expressed on Myeloid Cells 2
U	Unit (of Enzymes)
UCSC	University of California, Santa Cruz
UK	United Kingdom
UNC5C	Unc-5 Netrin Receptor C
UPF	Regulator of Nonsense Transcripts (Protein Family)
US(A)	United States (of America)
UTR	Untranslated Region
UV	Ultraviolet
V	Volts
VCF	Variant Call File
VEP	Variant Effect Predictor
WT	Wild-Type
XBP1	X-Box Binding Protein 1
ZCWPW1	Zinc-Finger CW-Type and PWWP Domain Containing 1

1 Introduction

1.1 Alzheimer's Disease

In 1906 Alois Alzheimer first defined Alzheimer's disease (AD) in a 51-year-old patient of his (Auguste D) who exhibited symptoms of worsening cognitive impairment, delusions and impaired social functions. On post-mortem examination he identified neurological amyloid plaques and neurofibrillary tangles, now known as the hallmarks of this disease (Alzheimer et al., 1995). These neuropathological markers of AD are thought to cause neurotoxicity, inflammation and neuronal dysfunction, thus leading to cognitive impairment (Eikelenboom et al., 2006). Initially the neuronal and synaptic damage is in the parahippocampal regions, the area of the brain which is predominantly responsible for forming new memories. However, pathology subsequently spreads, eventually causing cortical atrophy and ventricular enlargement, commonly reducing total brain mass by up to 35% as shown in Figure 1.1 (Alves et al., 2012; Farfara et al., 2008).

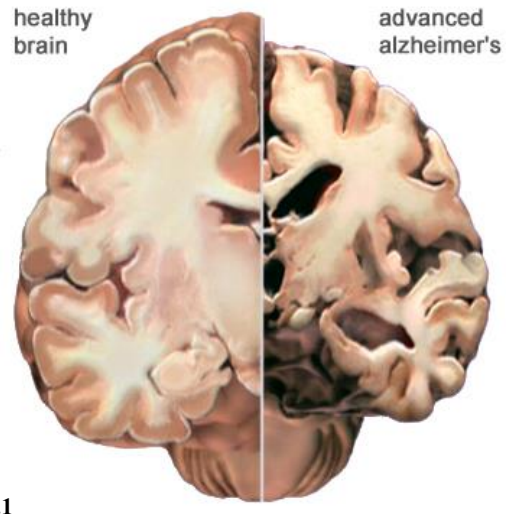


Figure 1.1
Comparison between a healthy brain and a brain with Alzheimer's disease. In AD the neural cortex shrinks, especially in the hippocampal region which is responsible for forming new memories. The ventricles (fluid filled cavities in the brain) consequently enlarge. Taken from https://www.alz.org/braintour/images/alzheimer_brain.jpg

AD is the commonest cause of dementia, attributing 77% of the demented population (Alzheimer's Association, 2015). This accounted for an estimated 820,000 people living with dementia in the UK in 2008, 32% of all UK residents over the age of 85 and one third of Americans over the age of 85. This figure is expected to double every 20 years, reaching 131.5 million worldwide by 2050, partly due to life expectancies continuing to raise (Alzheimer's Association, 2015; Ferri et al., 2005; Lewis and Torgerson, 2016).

The risk of AD in individuals over the age of 85 is 11% in males and 17% in females (Genin et al., 2011). Those diagnosed at the age of 80 experience a reduction in lifespan of 39%, increasing to 67% if they are identified earlier at the age of 65 (Brookmeyer et al., 2002). In the UK, these patients cost the economy £23 billion a year (on average $47,752 per patient), half of which is met by unpaid carers, commonly family members, as patients lose the activities of daily living. The global cost of AD is likely to exceed US $1 trillion by 2018 (Wimo et al., 2016). This demonstrates that AD has not only a large impact on the family members caring for these patients, but also on the national – and international - economy. It is an increasingly common cause of death: where the number of deaths due to other major diseases, such as heart disease and stroke, has decreased significantly, deaths due to AD have increased by 71% between 2000 and 2013. However it has the lowest number of compounds progressing to therapeutic trials, showing the lack of pharmaceutical progress made towards treating this disease (Alzheimer's Association, 2015; Lewis and Torgerson, 2016).

1.1.1 Alzheimer's Disease Pathological "Hallmarks"

AD is often very difficult to distinguish clinically from other forms of dementia and cannot be definitively diagnosed until post-mortem upon identification of amyloid plaques and tau neurofibrillary tangles within the brain (see Figure 1.2).

Unfortunately, therefore, a diagnosis of only probably or possible AD can be made during a patient's lifetime, based on symptoms and cognitive assessments. This makes diagnosing, treating and managing suspected AD extremely problematic and, consequently, a vast area of ongoing research is on trying to image these markers, as well as other biomarkers, in live patients.

AD, like many neurodegenerative diseases, is characterised by the aggregation of certain misfolded proteins. The two proteins distinguishing AD are summarised in text below and in pictorially in Figure 1.2.

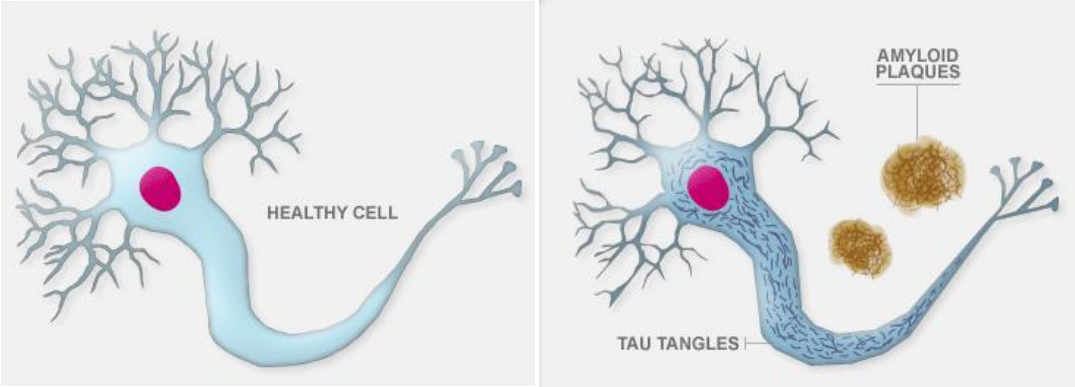


Figure 1.2
A healthy neuronal cell (left) and a neuronal cell from an individual with AD pathology (right), showing the extracellular amyloid plaques and intracellular tau neurofibrillary tangles. Taken from http://www.alzheimersresearchuk.org/brain-tour/

1.1.1.1 Aβ Plaques

For many years, it has been accepted that pathological amyloid plaques (Aβ) are the primary event in the AD disease process (Hardy and Allsop, 1991). These plaques are formed from the proteolysis of Amyloid Precursor Protein (APP) by α-, β- and γ-secretases (Bu, 2009) producing the hydrophobic Aβ peptide. This, in combination with other proteins, including apolipoprotein E (apoE) and apolipoprotein J, also known as clusterin, form the pathological plaques (Liao et al., 2004).

APP is a protein located in cellular membranes, expressed in a variety of tissues but especially in neuronal synapses. APP's most defined role is in synapse formation and transmission (Priller et al., 2006) although its primary function still remains unknown (Turner et al., 2003). This postulated role in synapse function is questioned by the lack of expression it has in synaptic-rich tissues when examined in expression databases (for example, Protein Atlas) (Uhlen et al., 2010). The full role of this protein, therefore, remains to be established.

When APP is cleaved, it can follow either the amyloidogenic or non-amyloidgenic pathway depending on which secretase enzymes are involved (summarised in Figure 1.3) (Thinakaran and Koo, 2008). It is the amyloidgenic pathway, involving the β- and γ-secretases, which generates Aβ: either Aβ40 or Aβ42 depending on where γ-secretase cleaves APP. Aβ42 is more likely to aggregate and form plaques and therefore a higher proportion of this is likely to be damaging, to both neurones and synapses (Holtzman et al., 2012). Aβ can exist in both soluble and insoluble forms but studies in rat brains have shown it to be the soluble Aβ dimers that are neurotoxic, disrupting learned behaviour and inhibiting long-term potentiation (Shankar et al., 2008).

There are several theories as to why Aβ accumulates and causes plaques. These include: increased production; decreased clearance; a higher Aβ42:Aβ40 ratio; altered Aβ metabolism or even a mixture of all of these. Higher levels of β-secretase (coded for by the *BACE1* gene) have also been known to increase the likelihood of the amyloidgenic pathway being followed (see Figure 1.3). Aβ mainly aggregates in the post-synaptic compartment of synapses, resulting in both synaptic and dendritic loss (Minati et al., 2009; Shankar et al., 2008). Recently, Aβ42 has also been implicated in gene transcription, altering the expression of some of the other disease modifying

genes, perhaps playing a part in AD pathology through this mechanism (Barucker et al., 2014).

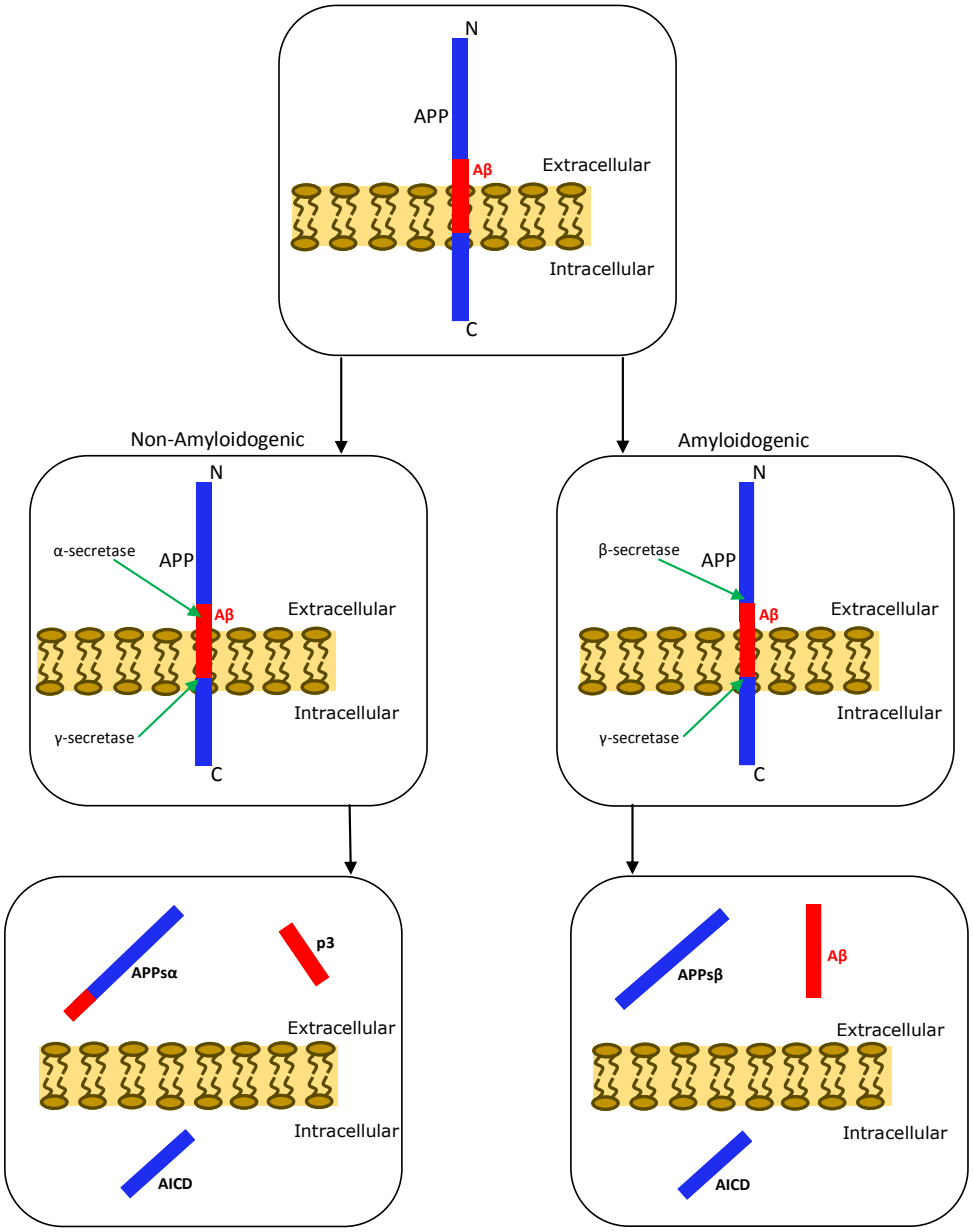


Figure 1.3
Amyloid Precursor Protein (APP) proteolysis and the two alternate pathways. In the non-amyloidogenic pathway, α– and γ-secretase cleave APP to form the secreted APP ectodomain (APPs) α, p3 and APP Intracellular, Cytoplasmic Domain (AICD). In the amyloidogenic pathway, β– and γ-secretase cleave APP to form Aβ, APPsβ and, again, AICD. Aβ, when in combination with other proteins, then goes on to form the neuropathological plaques. Modified from Thinakaran and Koo 2008, Martins *et al* 2009 and Blennonw *et al* 2010.

1.1.1.2 Neurofibrillary Tangles

Tau is a microtubule binding protein, stabilising neuronal microtubules as well as encouraging axonal growth and activity, and is the main component of these pathological neurofibrillary tangles (NFTs) (Roy et al., 2005). It is coded for by the microtubule associated tau (*MAPT*) gene and it is functionally modulated by its phosphorylation status. When the balance of phosphorylated to unphosphorylated protein is disrupted, hyperphosphorylation of the protein occurs causing it to accumulate and form paired helical filaments. These go on to form β-pleated sheets, which aggregate within the cell cytoplasm, blocking nutrient transport and (see Figure 1.2) forming the neurotoxic tangles (Mandelkow and Mandelkow, 1998). The quantity of NFTs within an AD brain correlates with the disease severity, whilst Aβ load does not (Arriagada et al., 1992). Neuronal studies have also shown that NFTs need to be present in order for Aβ plaques to be neurotoxic (Rapoport et al., 2002).

It is not entirely defined as to why tau hyperphosphorylates. Some theories include abnormal developmental regulation, upregulated protein kinases (Tolnay and Probst, 1999) and abnormal oxidation (Mandelkow and Mandelkow, 1998). Rare variants in the *MAPT* gene have been linked disease, including several other neurodegenerative disorders such as frontal temporal dementia (FTD) (Jin et al., 2012; Neumann et al., 2009). *MAPT* has also, recently, been genetically linked with AD when a previously existing dataset (the International Genomics of Alzheimer's Project (IGAP) consortium GWAS data) was meta-analysed, association signals were accounted for between a locus near the *MAPT* gene and AD risk (Jun et al., 2016).

1.1.2 Amyloid Cascade Hypothesis

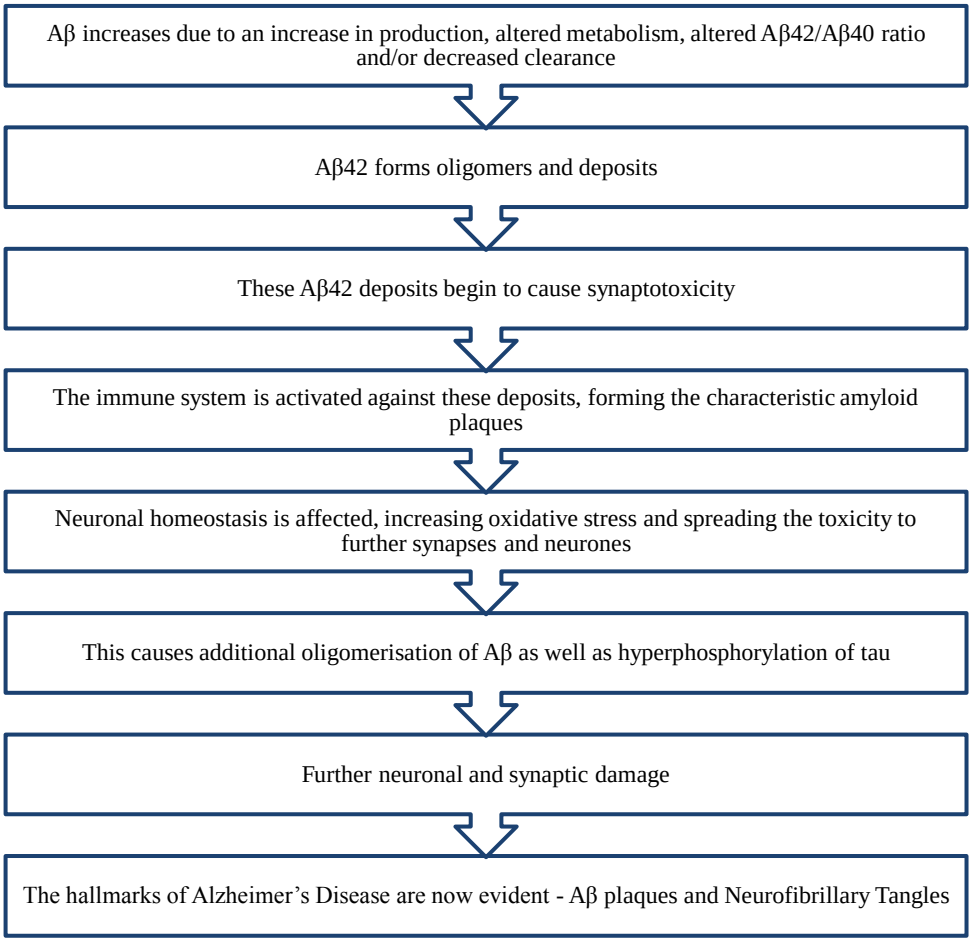


Figure 1.4
Flow chart summarising the Amyloid Cascade Hypothesis, where the formation of Aβ oligomers are the initiating event in Alzheimer's Disease pathology. Modified from Haass and Selkoe (2007).

The suggestion that Aβ plaques are the initiating event in AD pathology – known as the amyloid cascade hypothesis – was based on evidence that genetic alterations to the *APP* gene cause a severe, early onset form of AD, while similar alterations to *MAPT* do not (Hardy and Allsop, 1991; M. Lee et al., 2014). The fact that individuals with Trisomy 21 (commonly known as Down's syndrome), where the *APP* gene is located, also supports this hypothesis with these individuals developing age-related Aβ plaques (Hartley et al., 2014; Musiek and Holtzman, 2015). It has been suggested that the oxidative stress and ionic homeostasis disrupted by Aβ may create an imbalance of

phosphatases and kinases, stimulating the hyperphosphorylation of tau (Haass and Selkoe, 2007). This hypothesis is summarised in Figure 1.4.

Although these two signature markers are descriptive of AD cases, it is not clear whether they are the direct cause of the consequential neuronal loss. Alongside them, AD brain frequently exhibit activated immune cells (Eikelenboom et al., 2006), mitochondrial dysfunction (Reddy, 2011), cerebral inflammation (Wyss-Coray, 2006) and vascular involvement (Marchesi, 2011). However, protein aggregates, specifically Aβ, may also contribute to this chronic inflammation of the neurones itself by activating microglia (Fan et al., 2014). This then goes on to accelerate the protein aggregates spread throughout the brain, as well as upregulating the expression of APP and tau, most likely aggravating the pathology (Guo and Lee, 2014; M. Lee et al., 2014).

1.1.3 Signs & Symptoms

Commonly, initial presentation of AD patients will be to primary care facilities, such as General Practitioners (GPs), with gradual worsening of memory loss. Dementia is, generally, a progressive and mostly irreversible condition with widespread loss of mental function. However, AD is, more specifically, defined as memory loss in combination with loss of activities of daily living (Alves et al., 2012). The pathophysiology of AD may, in fact, begin up to 20 years before symptoms begin to manifest and, on average, is fatal four to eight years after an AD diagnosis (Alzheimer's Association, 2015; Holtzman et al., 2012).

As the disease progresses, patients may experience some, or all, of the following: memory loss; language impairment; disorientation; challenges in planning and problem solving; changes in personality; difficulties with activities of daily living; self-neglect; psychiatric symptoms or generally developing out-of-character

behaviour. The rate of disease progression varies from individual to individual and more of these symptoms may worsen or increase in number until patients are largely unable to care for themselves. They therefore require full-time care in order to complete the activities of daily living, frequently becoming dysphagic, aphasic and incontinent as well as the symptoms previously mentioned (Alzheimer's Association, 2015).

The reasoning behind the time lag between pathology occurring and symptoms becoming apparent is thought to be the idea of "cognitive reserve." This implies that, patients with a higher number of neurons to start with, being able to withstand higher levels of neurological Aβ plaques before cognition begins to decline due to a higher number of "reserve" neurones. This implies they will be able to increase their neurological function (assessable on functional magnetic resonance imaging (fMRI)) in proportion to the amount of Aβ plaques present, possibly providing an early diagnostic tool (Beecham et al., 2014; Elman et al., 2014).

Advancement of AD often results in individual's having difficulties moving, leaving them vulnerable to infections, most commonly pneumonia. This, and other acute conditions, are frequently the contributing factors to AD-related deaths with a very blurred distinction between "death *with* dementia" and death *from* dementia" (Alzheimer's Association, 2015).

1.1.4 Diagnosis

As mentioned before, presentation of dementia is commonly to the GP with memory loss. However, only 50% of AD patients are diagnosed correctly in the general community, outside of specialist centres. Therefore patients showing signs of Mild Cognitive Impairment (MCI – memory loss without loss of the activities of daily living) should be referred to memory assessment services (Mayeux et al., 2011).

Within these dedicated facilities, diagnosis of dementia caused by AD can increase to 95% accuracy when monitored longitudinally (Alves et al., 2012; Villemagne et al., 2008).

A comprehensive assessment of all suspected AD patients should take place, including history taking (both from the patient and from a close family member), cognitive and mental state assessment, physical examination and a review of current medications to rule out drug induced cognitive impairment, as well as any condition which may present with dementia-like symptoms (Alzheimer's Association, 2015; Villemagne et al., 2008). The cognitive examination commonly involves the Mini Mental State Examination (MMSE) which examines attention and concentration, orientation, short and long-term memory, praxis, language and executive function. However, other standardised devices can also be used including the 6-item Cognitive Impairment Test (6-CIT), the GP Assessment of Cognition (GPCOG) and the 7-minute screen.

Baseline measurements should also be performed in order to rule out other causes of dementia. These can include routine haematology, biochemistry tests, thyroid function tests, serum vitamin B_{12} and folate levels as well as syphilis or human immunodeficiency virus (HIV) serology and midstream urine tests if the patient history suggests these may be necessary.

Following assessment of demented patients, they can be classified as Probable AD, Possible AD or Unlikely AD. These categories are defined in the National Institute of Neurological and Communicative Disorders and Stroke and the Alzheimer's Disease and Related Disorders Association (NINCDS-ARDA) Alzheimer's criteria. Possible AD is a dementia syndrome with an atypical onset but no other co-morbidity which is likely to have caused the dementia. Probable AD is cognitive impairments being

apparent in two or more areas of cognition that are progressive. The age of onset is typical for AD patients and there are no other diseases present which are likely to have caused the syndrome. Definite (or confirmed) AD can only be diagnosed with histopathological evidence of AD upon post-mortem (McKhann et al., 1984).

Neurological imaging may also be utilised in order to exclude other cerebral pathologies. Magnetic Resonance Imaging (MRI) is the preferred modality but computed tomography (CT) can also be used. More specific functional scanning, such as perfusion scans, can also be used to help differentiate from other dementia subtypes. In current guidelines (NICE 2014), cerebrospinal fluid (CSF) examination is not recommended as routine investigation for dementia. However, it may increase the accuracy of diagnosis if it is unsure, as discussed below. Amyloid position emission tomography (PET) scans have also recently become a very good predictor of true AD suffers, detecting early stages of amyloid deposition, providing a good indicator of cognitive decline although this is yet to be recommended by NICE (Dubois et al., 2016).

1.1.5 Biomarkers

As previously mentioned, pathology of Alzheimer's disease can exist for many years prior to becoming symptomatic. However, most of the successful disease-modifying therapeutic agents are likely to be their most effective very early on in the disease process. Therefore, a method needs to be identified for detecting AD patients at this point, perhaps when only MCI is apparent or even pre-symptomatically. Accuracy of AD diagnosis is also a problem, with diagnosis based purely on clinical criteria offering a sensitivity of only 70.9-87.3% and a specificity of 44.3-70.8% depending on weather the diagnosis is of probable or possible AD (Beach et al., 2012). Consequently, biomarkers are required in order to assist in AD diagnosis, monitor

reactions to any treatments given whilst also improving the quality of patient cohorts in AD clinical trials (Blennow et al., 2010).

Due to CSF being in direct contact with the Central Nervous System (CNS), the principle site of AD pathology, it is logical that its contents are influenced by the neurological environment (Kauwe et al., 2014). Therefore, levels of AD related proteins in the CSF may report on the patients AD status. CSF biomarkers for AD include total tau, phosphorylated tau and $A\beta_{1-42}$ and, in combination, these do increase the sensitivity and specificity of an AD diagnosis, therefore influencing patient management, when compared to a diagnosis based on clinical symptoms alone (Duits et al., 2014; Molinuevo et al., 2014). So far, they have been proven to reliably diagnose AD patients (with a sensitivity and specificity of 80-90% and increasing diagnostic confidence from 84% to 89%) as well as predicting future AD patients in MCI cohorts (with an accuracy of >80%). However, they are not as successful in discriminating AD cases from other forms of dementia, with the exception of FTD (Ewers et al., 2014). Until recently these methodologies were also ineffective in predicting AD patients from the pre-symptomatic patient groups (Blennow et al., 2010). However, a recent study has identified that the ratio of CSF Aβ42 to Aβ40 to be more sensitive to the level of cortical atrophy then Aβ42 levels alone, especially in the preclinical stages. However, hippocampal volumes can be predicted by Aβ42 levels alone (Lindberg et al., 2016). This Aβ42:Aβ40 ratio may, therefore, provide an early diagnostic tool of cortical atrophy. Unfortunately, obtaining CSF through a lumbar puncture is an invasive process and may not even be possible in a community health care setting.

A further suggested biomarker is imaging. This includes imaging of Aβ within the brain itself via position emission tomography, performing fMRI as well as measuring the level of cortical atrophy, again through MRI (Simmons et al., 2009; Small et al.,

1999). However, all of these modalities are rather expensive and repeat measurements, as with CSF sampling, are problematic. In contrast, Blood and plasma are a lot more accessible and repeat sampling is comparatively easy as well as being available in a community setting.

Until recently, no plasma biomarkers had been identified which correlated to AD status. Due to the blood-brain barrier, proteins present in the CNS are not reflected in the plasma. This is clear in the literature where conflicting evidence on plasma levels of both tau and Aβ levels show that they are not reliable biomarkers (Rosén et al., 2013). Other markers have been looked at, including inflammatory markers due to the known role of inflammatory processes in the pathophysiology of AD (Lueg et al., 2014). However, more recently the idea of a panel of molecular markers has gained ground. Due to the fact that AD is a multifactorial disease, being influenced by both genetics and environment, as well as lifestyle factors, one single molecule may not be able to provide the biomarker capabilities required (Richens et al., 2014). This concept has been proven recently with two studies putting forward suggested panels of plasma proteins, both published in 2014. One presents a set of ten lipids which predict conversion of MCI to AD within 2 to 3 years with an accuracy of 90% (Mapstone et al., 2014). The other proposes a different panel of ten plasma proteins, with only clusterin and apoE being identified as being related to AD previously. This study again predicts disease progression as well as disease severity with an accuracy of 87% (Hye et al., 2014).

Current guidelines for diagnosing AD primarily utilise neurological imaging, with CSF biomarkers if further modalities are required to increase the accuracy of diagnosis. MRI is the preferred assessment to detect cortical changes, although CT scanning may also be used if there are existing contraindications. However, due to recent clinical studies, it is looking increasingly likely that plasma protein panels will

become the future in diagnosing AD and assessing disease progression and severity as well as the outcome of any treatments given.

1.1.6 Treatments

As the primary cause of AD is still largely a mystery, it is difficult to develop successful treatments. AD is the only one of the top ten causes of death to not have a disease-altering treatment (Ridge et al., 2016). Many current therapies target symptoms of the disease, with the aim of delaying disease progression, primarily through increasing the amount of neurotransmitters present neurologically. The current National Institute for Health and Care Excellence (NICE) guidelines (2014) recommend acetylcholinesterase (AChE) inhibitors for mild to moderate AD patients, assessed on their cognition score. The three AChE inhibitors commonly used in AD are donepezil, galantamine and rivastigmine while memantine, an N-methyl-D-aspartate (NMDA) receptor antagonist, is recommended in more severe cases or if patients cannot tolerate AChE inhibitors. However, both of these lines of therapy do not target the cause of the disease itself, merely providing symptomatic relief and not altering the progression of the pathology. AChE inhibitors aim to increase the amount of acetylcholine (ACh) available at synapses by inhibiting its degradation, as there is a characteristic loss of cholinergic neurons in the basal forebrain on AD patients (Davies and Maloney, 1976). Alternatively, memantine, works by blocking the action of glutamine, which is seen to be increased in AD patients and is thought to contribute to neuronal dysfunction (Reisberg et al., 2003). However, these drugs only provide a slight improvement in the cognitive function of AD patients and have no effect on disease progression itself (Rang, 2008).

If any of these treatments are utilised, NICE recommends beginning on the lowest dose possible, reviewing therapies as often as possible, including any apparent side

effects, only continuing therapies if they are making a marked improvement, as well as always treating non-cognitive symptoms in combination.

As mentioned previously, all of these therapies provide only symptomatic relief, whilst the ideal AD treatment would be one that prevents further neurodegeneration, idyllically targeting the pathological pathways in AD: the amyloid cascade and tau hyperphosphorylation (Hanenberg et al., 2014).

Inhibitors of both β- and γ-secretase have been developed. β-secretase inhibitors (also known as BACE1 inhibitors) have shown much promise and examples include CTS-21166, which has been successful in Phase I clinical trials, and nilvadipine (an established anti-hypertensive medication) which, not only inhibits *BACE1* expression, but also lowers tau phosphorylation and reduces inflammatory markers in AD model mice (Panza et al., 2009; Paris et al., 2014). In late 2014 another BACE1 inhibitor, AZD3293, began in Phase II/III clinical trials, scheduled to be completed by May 2019. However, γ-secretase inhibitors, such as semagacestat, have shown much less promise. Semagacestat reached Phase III clinical trials before they were halted due to the treatment group showing worsening cognitive function in comparison to the control group (Samson, 2010). γ-secretase is involved in other signalling pathways as well as Aβ formation suggesting it is not, perhaps, such a promising therapeutic target as other molecular mechanisms.

In 1999 the idea of Aβ immunisation was first suggested. AD mouse models immunised with the Aβ protein, incredibly, had Aβ plaque formation reversed. However, this has not been replicated in human trials as a neuroinflammatory effects were found in human test subjects (Schenk et al., 1999) although treatment with the humanised anti-amyloid IgG monoclonal antibody Solanezumab has shown more success in clinical trials. In a preliminary Phase II results presented in 2015, the

EXPEDITION trial saw individuals suffering from mild AD who began Solanezumab treatment earlier had a slowing of both cognitive and functional decline (by 34% and 18% respectively) in comparison to those who were either in the placebo test group or who started it later as part of the delayed-start study design (Liu-Seifert et al., 2015; Siemers et al., 2015). However, Phase III trial results were reported in November 2016 which presented no significant slowing of memory and cognitive decline in mild AD cases in comparison to placebo intervention (McCartney, 2015). A further Phase III trial is planned involving prodromal AD patients with results expected in 2021. In spite of these setbacks, Solanezumab remains the most promising disease modifying treatment for AD to date.

Immunotherapy against tau has also been suggested and tested in mouse models. Development of these treatments has been more problematic due to tau being located intracellularly and therefore it is not as easy for antibodies to access (Mably et al., 2014). However, in 2015, three monoclonal antibodies against tau were reported to reduce tau pathology in mice and showed, in cell lines, to modify cellular uptake of tau as well as alter the clearance of it via microglia cells (Funk et al., 2015). Therapeutic antibodies may therefore be a potential therapy in the future. However, their exact functionality still needs to be determined as well as to ascertain as to whether they are therapeutically active *in vivo*.

There have also been several studies concentrating on modifying risk factors for the disease, for example, non-steroidal anti-inflammatories (NSAIDs), taken for other co-morbidities during life, lower the chances of getting AD in later life (Stewart et al., 1997) as do statins (cholesterol reducing drugs) (Zamrini et al., 2004). NSAIDs specifically have been shown to reduce neuronal inflammation, reducing both *MAPT* and *APP* expression due to a negative feedback loops, as well as increasing the amount of APP cleaved by β-secretase rather than α-secretase (M. Lee et al., 2014).

Further work on the immune system has shown that exogenously injected exosomes are taken into microglial cells, increasing the clearance of Aβ and, although injecting exosomes is perhaps not a practical therapy for AD patients, it does provide a novel therapeutic target (Yuyama et al., 2014). Work into non-pharmacological therapies has also recently shown promise but have not been shown to alter the course of AD. Exercise and cognitive activity approaches have shown promise in improving quality of life (Sink KM et al., 2015). However, additional research is needed into these therapies to better evaluate their effectiveness (Alzheimer's Association, 2015). Recently a highly personalised therapeutic approach has been suggested, tailoring the optimization of dozens of metabolic parameters in individual patients, through both therapeutic interventions and lifestyle changes including serum vitamin B_{12} levels, fasting insulin levels, GI health and sleep quality among many others. However, this programme is not only extremely demanding on both patients and health care professionals, it has also only been trialled in ten patients, all appearing to exhibit improvements, if subjectively (Bredesen, 2014).

Thus far, no disease-modifying drugs for AD have made it past clinical trials due to various adverse side effects or non-significant findings, although there is at least one promising Phase III trial continuing (Siemers et al., 2015). All conclusive studies so far, including those involving NSAIDs, have shown that drugs are required to be taken years in advance of symptoms becoming apparent, making treating AD a real problem. Currently, all that can be done is to treat this disease symptomatically, utilising AChE inhibitors and memantine.

1.1.7 Types of Alzheimer's Disease

AD is divided into early onset and late onset depending if symptoms are first exhibited before or after the age of 65.

1.1.7.1 Early Onset Alzheimer's Disease

Early onset AD (EOAD) is thought to only cause 1% or less of the total AD cases (Alzheimer's Association, 2015) and itself can be divided into Early Onset Familial AD (EOFAD) or Early Onset Sporadic AD (Antonell et al., 2013).

Phenotypically and pathologically, EOFAD and late onset AD are very similar. However, EOFAD has very defined genetic causes, with mutations in the genes coding for *APP*, presenilin (*PSEN*) 1 and 2 being found in 80% of EOFAD patients (Goldman et al., 2011). These mutations are inherited in an autosomal dominant fashion with age of onset often being very young and close to fully penetrant. All of these proteins are involved in APP metabolism: PSEN1 and 2 are part of the activated γ-secretase complex (Alves et al., 2012), increasing the amount of Aβ formed, or altering the Aβ42:Aβ40 ratio. Both of these result in plaques accumulating at an earlier age. The majority of mutations in *APP* are clustered around the α-, β- and γ-secretases cleavage sites (Shewale, 2012). In total, 24 *APP*, 185 *PSEN1* and 15 *PSEN2* mutations have been mapped (Schellenberg and Montine, 2012; Tanzi and Bertram, 2005) with mutations in *MAPT* and progranulin (*GRN*) also linked to EOFAD (Jin et al., 2012). In addition, *APP* is located on chromosome 21, causing a 70% increased risk of developing EOAD in patients with trisomy 21 (Hardy and Higgins, 1992). Symptoms frequently present during patients' third decade due to the increased production of Aβ as a consequence of carrying this extra *APP* gene (Hartley et al., 2014).

On the other hand, 40% of EOAD cases are sporadic with no previous family history, and the causes of these are largely unknown. This implies there is a complex interaction of genetic and environmental factors or there may even be genetic risk loci that have yet to be identified (Alves et al., 2012; Campion et al., 1999).

1.1.7.2 Late Onset Alzheimer's Disease

Compared to EOFAD the causes of Late Onset AD (LOAD) are less well defined, despite it being far more common, as it is a significantly more complex disorder with considerable environment interactions which contribute to the phenotype of any complex disease (Buil et al., 2014). In 2013 Shewale *et al* split the causes of LOAD into three domains: genetics; epigenetics and environment, all of which can alter the phenotypic expression and therefore the pathology of AD (Shewale, 2012). Each of these three domains will be discussed below.

Genetics

It has been estimated, through both monozygotic and dizygotic twin studies as well as sequencing projects, that LOAD is somewhere between 37 and 80% heriditable and, therefore, genetics must play a large part in a complex inheritance pattern (Ebbert et al., 2014; Medway and Morgan, 2014; Morgan, 2011). In the 1990s the first step in unravelling the genetics behind LOAD was taken when it was found that carriers of the ε4 allele of the gene coding for apolipoprotein E (*APOE*) had a far greater chance of developing LOAD whereas carriers of the ε2 allele had a much lower risk (Farrer LA et al., 1997; Saunders et al., 1993). The *APOE* haplotypes are defined by two non-synonymous variants (at amino acid positions 112 and 158), resulting in the ε2, ε3 or ε4 haplotypes, summarised in Table 1.1. This link was seen in familial genetic linkage studies and was also replicated in APOE knock-out, APOE allele-specific and humanized mice (Roses and Saunders, 2006). Other, rarer, variants have since been identified, some of which further sub-define the haplotypes and alter LOAD disease risk yet again (Medway et al., 2014).

Table 1.1

APOE haplotypes, defined by allelic combinations. rs429358 is a C to T change coding for a cysteine to arginine substitution at amino acid position 112 with a minor allele frequency (MAF) of 0.15 and rs7412 is a T to C change resulting in a arginine to cysteine replacement at amino acid position 158 with a MAF of 0.07. ε3 is taken to provide a risk level for LOAD of 1.0. ε4 is a major risk factor with odds ratios of 3 and 12 for carriers of one or two copies respectively. ε2 has a protective effect irrespective of the presence of a ε4 allele, decreasing LOAD risk by a factor of 4. ε1 is the combination of the two minor alleles and is, therefore, very rarely seen. Adapted from Farrer *et al* (1997) and Corder *et al*. (1994).

			MAF
rs429358	C/T	Cys112Arg	0.15
rs7412	T/C	Arg158Cys	0.07

		rs7412					
		Major (T)	**Haplotype**	**Frequency**	**Minor (C)**	**Haplotype**	**Frequency**
rs429358	**Major (C)**	TC	ε3	77.9%	CC	ε4	13.7%
	Minor (T)	TT	ε2	8.4%	CT	ε1	-

It has been established that carriers of the ε4 *APOE* haplotype have an additive risk, with the risk of LOAD increasing by an odds ratio of 2.84 with each additional ε4 carried. This risk haplotype also decreases the age of onset and the average survival of the disease, as well as doubling the amount of Aβ stained neurologically post-mortem (Corder et al., 1993; Strittmatter et al., 1993). It is now known that the ε4 isoform also affects brain function much earlier in the carrier's life span, decades prior to any clinical symptoms becoming obvious, associating it with both altered neurological and metabolic function as well as anatomical differences (Filippini et al., 2009; Michaelson, 2014). It is thought that between 40 and 60 percent of people diagnosed with LOAD carry at least one copy of the *APOE* ε4 gene. However, carrying this ε4 haplotype does not guarantee that an individual will develop LOAD (Alzheimer's Association, 2015).

Human apoE is a 299 amino acid glycoprotein which is highly expressed in the liver, as well as in the brain, where it is primarily expressed in astrocytes and microglia. It is known to function as a receptor during the endocytosis of lipoprotein particles, as well as being involved in cholesterol release, and therefore plays a role in neurological synaptic maintenance. It is also the chief apolipoprotein within the CNS (Kim et al., 2009).

Various reasons for *APOE's* association with AD have been suggested, the main theories falling into the following categories with different apoE haplotypes: affecting synaptic plasticity and repair (Teter, 2004); altering cholesterol transport and its lipid binding capabilities (Hirsch-Reinshagen et al., 2008; Martins et al., 2009); handling Aβ slightly differently resulting in modified clearance, deposition or aggregation (Bales et al., 1999; Shibata et al., 2000); altered formation of Aβ due to APP processing changing (Bu, 2009; Kim et al., 2009) or this even altering NFT formation (Holtzman et al., 2012). The variants dictating the isoform of apoE will alter the structure of the protein, therefore it is not surprising that its function is altered. However, it is discovering how that changed function impacts on LOAD pathology that still needs to be ultimately defined. It is estimated that *APOE* accounts for 27.3% of the Population Attributable Risk for LOAD which, although accounting for the strongest genetic risk factor, still leaves a substantial portion for others (Medway and Morgan, 2014).

Epigenetics

Epigenetics generally refers to alteration of gene regulation by something other than the DNA sequence (Laird, 2010). One of these alterations is methylation of DNA, primarily in cytosine guanine dinucleotide repeat rich regions of the genome, known as CpG islands (Bock et al., 2006). Methylation has been shown to play an important role in gene expression and silencing, chromosome stability, differentiation, cellular identity and, consequently, human disease (Roadmap Epigenomics Consortium et al., 2015). Several of the genes involved in AD (*APP*, *PSEN1*, and *BACE1*) have a proportionally high GC content and are therefore potential targets for methylation regulation (Shewale, 2012).

Recently the idea that epigenetics can cause phenotypical changes is gaining ground and an increasing number of studies are being done in this area. Hypermethylation

patterns (specifically CpG methylation) of several AD-related genes are seen in AD patients when compared to age-matched controls (Yu et al., 2015). This strengthens the notion that methylation alterations are associated with the LOAD pathology (Chouliaras et al., 2013). One recent report has also associated the level of methylation in the CpG islands of *ABCA7* and *BIN1* – known genes associated with LOAD –demonstrating that they are significantly associated with the level of neurological pathology (De Jager et al., 2014; Lord and Cruchaga, 2014). This has been repeated in the *SORL1*, *ABCA7*, *BIN1*, *HLA-DRB5* and *SLC24A4* genes, with altered neurological DNA methylation in the these genes being associated with pathological AD (Yu et al., 2015). The *APP* gene has also been shown to be hypermethylated in AD patients, altering downstream gene expression in combination with transcription factors (Lunnon et al., 2014). This indicates that epigenetic changes can also alter disease pathology and studies have shown that these epigenetic changes can manifest as an alteration in cognition and, specifically, memory formation (Day and Sweatt, 2011; Zovkic et al., 2013). Epistatic interactions between variants in different genes identified through genome wide association studies (discussed in more detail later) have also been identified, suggesting that the genetic causes of LOAD are not as straight forward as carrying a variant or not, but the haplotype carried may be of more importance (Gusareva et al., 2014).

Further areas of epigenetics include histone modification, chromatin remodelling and noncoding ribonucleic acid (RNA) interactions. Histone modification may also be involved in AD as the AICD peptide produced intracellularly by APP proteolysis (see Figure 1.3) has been shown to recruit peptides involved in histone modification and, consequently, gene expression (Abel and Zukin, 2008).

Environment

There are many environmental factors which increase the risk of LOAD, far too many for the scope of this thesis. It is known that oxidative stressors (caused by things like traumatic brain injury, inflammation, sleep apnoea etc.) can cause epigenetic changes impacting on the pathogenesis of AD. However, many other environmental factors also have an impact such as diet, chemical exposure, smoking, body mass index, folate intake and lack of exercise can alter an individual's chances of developing LOAD. Whether this is because they alter epigenetics, or for other reasons, it is still unknown (Shewale, 2012).

By far the largest risk factor for developing LOAD is age, with the risk of disease approximately doubling every 5 years after the age of 65 (Schellenberg and Montine, 2012). Alongside this, genetics appears to be the next largest risk factor, particularly in carriers of the apolipoprotein ε4 allele, as discussed above (Martins et al., 2009). However, many other hormonal, environmental and lifestyle factors have also been identified as contributing to an increased risk of LOAD. The first of these identified is that of vascular and metabolic disorders, suggesting that hypertension, especially variable systolic blood pressure, hypercholesterolemia, coronary artery disease and obesity all increase the risk of AD in later life (Kivipelto et al., 2005; Skoog et al., 1996). Atherosclerosis also falls under this umbrella and it is thought that atherosclerotic plaques, especially in the extracranial carotid arteries, causes chronic or episodic cerebral hypoperfusion, increasing the risk of cognitive decline (Hofman et al., 1997). Stroke, atrial fibrillation and history of traumatic brain injury, have also shown to be proven risk factors of developing LOAD, and they are also thought to contribute to a very similar pathway: causing hypoperfusion of susceptible brain areas (Honig et al., 2003; Ott et al., 1997). Hypercholesterolemia is thought to alter the metabolism, cleavage or transport of Aβ, altering the development of the neurological amyloid plaques (Kuller and Lopez, 2011). Obesity, diabetes mellitus and impaired

glucose metabolism also seem to increase the susceptibility towards LOAD, thought to be through the effects of hormone compounds or the effects of glucose toxicity on amyloid metabolism (Kivipelto et al., 2005).

Further risk factors have also been suggested: vitamin D deficiency (Pogge, 2010); history of depression (Geerlings et al., 2008); high intake of saturated fat (Luchsinger and Mayeux, 2004); peripheral inflammation (Engelhart et al., 2004) as well as high serum homocysteine and fibrinogen concentrations (Bots et al., 1998; Kalmijn et al., 1999). Education level has also been suggested as modulating the degree to which AD neuropathology is expressed phenotypically. However, the evidence suggests that, although milder AD symptoms may be masked by an increased cognitive reserve in those who are more highly educated, there is no difference in cognitive function when AD neuropathology is more advanced (Koepsell et al., 2008).

In the most recent Alzheimer's Association recommendations (in conjunction with the World Dementia Council) they state that, should individuals wish to reduce their risk of cognitive decline and dementia, they should partake in regular physical activity and manage the cardiovascular risk factors associated with dementia (for example diabetes, obesity, smoking and hypertension). Along with this a healthy diet (such as the Mediterranean diet) and lifelong learning or cognitive training are also recommended in order to reduce the disease risk (Baumgart et al., 2015).

1.2 Genome Wide Association Studies

As mentioned previously, LOAD is highly heriditable but a vast proportion of its genetic risk was unknown until the advent of Genome Wide Association Studies (GWAS). GWAS aim to identify variants' associated with a certain phenotype, commonly one of a disease status. The methodology involves comparing allele frequencies of polymorphisms between groups of cases and controls. When the frequencies alter significantly, the gene that contains that polymorphism is identified as being associated to that phenotype. This is in comparison with linkage studies which are used to identify markers and traits in families, linking these markers to the trait. These kinds of studies are extremely powerful when identifying highly penetrant phenotypes, such as EOAD, linked with rare alleles, such as the *PSEN* and *APP* variations. In comparison, association studies, such as GWAS, are significantly more powerful in identifying common, low penetrance variants in these lower penetrance phenotypes, such as LOAD, provided there are enough cases and matched controls. However, GWAS so far are only estimated to have accounted for less than 50% of heriditability in a variety of complex diseases. It is thought that low allele frequency variants (ones with a minor allele frequency (MAF) below 1%) and rare variants (MAF of below 0.01%) also provide a major contribution towards these disease risks' and variants such as these cannot be picked up by association testing such as GWAS, which focuses on more common variants (Zuk et al., 2014). A large proportion of GWAS identified loci are also located in gene deserts or within introns. The functional effect of these variants is much more difficult to elucidate but as many as 80% of GWAS signals may be regulatory variants, therefore still affecting phenotype (Buil et al., 2014; Schierding et al., 2014). GWAS, therefore, can provide the basis for many future studies, for example large-scale exome sequencing projects, in order to reveal the rare causative variants in the genes or the surrounding areas they highlight.

However, for many complex traits, such as LOAD, heriditability may be distributed over thousands of variants with small effect sizes meaning identifying disease associated variants could continue indefinitely.

In the 2000's, several GWASs were undertaken in an attempt to identify genetic risk loci for LOAD. These identified Single Nucleotide Polymorphisms (SNPs) in 11 genes which may be involved in LOAD; *ABCA7, BIN1, CD2AP, CD33, CLU, CR1, EPHA1, MS4A6A, MS4A4E, MS4A4A* and *PICALM* (Harold et al., 2009; Hollingworth et al., 2011; Lambert et al., 2009; Adam C Naj et al., 2011; Seshadri et al., 2010). In 2013 a large meta-analysis of these GWAS was undertaken by IGAP, identifying a further 11 susceptibility loci (*CASS4, CELF1, FERMT2, HLA-DRB5/HLA-DRB1, INPP5D, MEF2C, NME8, PTK2B, SLC24A4/RIN3, SORL1* and *ZCWPW)* involved in LOAD as well as confirming all of the GWAS loci with the exception of *CD33* (Lambert et al., 2013). Although GWAS has played a large part in identifying common genetic variants with small effect size on disease risk, alternative methodologies are required in order to recognize low-prevalence variants with a moderate to high effect size on disease risk. Next Generation Sequencing (NGS) has gone on to identify further variants both within these GWAS genes and others. These include rare variants in *TREM2* (Guerreiro et al., 2013), *UNC5C* (Wetzel-Smith et al., 2014) *PSEN1* and *PSEN2* (also associated with EOFAD), as well as rare protective and risk variants in *APP* (Jonsson et al., 2012). All of these projects utilised similar methodologies: all performed whole-genome or whole-exome sequencing before performing quality-control, alignment and variant calling. Some projects (Guerreiro et al., 2013) then predicted pathogenicity of said variants utilising PolyPhen2 before carrying out direct genotyping, whilst others performed direct genotyping on the variants associated most strongly with disease status before examining their pathogenicity through *in vitro* assays (Jonsson et al., 2012; Wetzel-Smith et al., 2014), identifying any possible loss-of-function variants.

All of the variants associated with LOAD are shown in Table 1.2 with the ones identified through GWAS and the meta-analysis coloured in white, the loci identified through sequencing studies in blue and the *APOE* variants, the first established genetic risk factor of LOAD, shown in red.

Table 1.2
Alzheimer's Disease susceptibility genes and the most significant associations identified within these genes. Most have been identified through Genome Wide Association Studies and replicated in a meta-analysis in 2013 (white rows in table). *APOE* haplotype has been a long established genetic risk factor of LOAD, coded for by two variants (highlighted in red). Additional rare variants in several other genes have also been associated with this disease, highlighted in blue. *p* -value, Odds Ratio (OR) and Confidence Interval (CI) are shown as well as the effect of the minor allele (protective or risky for developing LOAD). Modified from Lambert *et al* (2013), Ridge *et al* (2016), Guerreiro *et al* (2013) and Wetzel-Smith *et al* (2014).

Gene	Genomic Location	SNP	*p* -value	OR (CI)	Minor Allele Effect
ABCA7	19:1063444	rs4147929	1.06E-15	1.15 (1.11-1.19)	Risk
APOE	19:44908822	rs7412	1.04E-295	2.53 (2.41-2.66)	Protective/Risk
	19:44908684	rs429358			
APP	21: 25880550 - 26171128	Multiple	-	-	Protective/Risk
BIN1	2:127135234	rs6733839	6.94E-44	1.22 (1.18-1.25)	Risk
CASS4	20:56443204	rs7274581	2.46E-08	0.88 (0.84-0.92)	Protective
CD2AP	6:47520026	rs10948363	5.20E-11	1.10 (1.07-1.13)	Risk
CD33	19:51224706	rs3865444	2.97E-08	0.94 (0.91-0.96)	Protective
CELF1	11:47536319	rs10838725	1.12E-08	1.08 (1.05-1.11)	Risk
CLU	8:27610169	rs9331896	2.77E-25	0.86 (0.84-0.89)	Protective
CR1	1:207518704	rs6656401	5.69E-24	1.18 (1.14-1.22)	Risk
DSG2	18:31508995	rs8093731	1.05E-04	0.73 (0.62-0.86)	Protective
EPHA1	7:143413669	rs11771145	1.12E-13	0.90 (0.80-0.93)	Protective
FERMT2	14:52933911	rs17125944	7.94E-09	1.14 (1.09-1.19)	Risk
HLA	6:32610753	rs9271192	2.94E-12	1.11 (1.08-1.15)	Risk
INPP5D	2:233159830	rs35349669	3.17E-08	1.08 (1.05-1.11)	Risk
MEF2C	5:88927603	rs190982	3.23E-08	0.93 (0.90-0.95)	Protective
MS4A6A	11:60156035	rs983392	6.14E-16	0.90 (0.87-0.92)	Protective
NME8	7:37801932	rs2718058	4.76E-09	0.93 (0.90-0.95)	Protective
PICALM	11:86156833	rs10792832	9.32E-26	0.87 (0.85-0.89)	Protective
PSEN1	14: 73136418 - 73223691	Multiple	-	-	Risk
PSEN2	1: 226870184 - 226896105	Multiple	-	-	Risk
PTK2B	8:27337604	rs28834970	7.37E-14	1.10 (1.08-1.13)	Risk
SLC24A4/RIN3	14:92460608	rs10498633	5.54E-09	0.91 (0.88-0.94)	Protective
SORL1	11:121564878	rs11218343	9.73E-15	0.77 (0.72-0.82)	Protective
TREM2	6:41161514	rs75932628	1.43E-07	4.59 (2.49-8.46)	Risk
UNC5C	4:95170280	rs137875858	9.50E-03	2.15 (1.21-3.84)	Risk
ZCWPW1	7:100406823	rs1476679	5.58E-10	0.91 (0.89-0.94)	Protective

In 2016, Ridge *et al* attempted to calculate the amount of genetic variance of LOAD which is explained by the known AD risk genes as shown in Table 1.2. Using data from 9,699 individuals and 8,712,879 SNPs they assessed genetic variance shown in AD case samples and what proportion of this variance is accounted for by known markers. They established that the genetic variance explained by the known GWAS AD risk SNPs, those red and white in Table 1.2, only explained 30.62% of the total genetic variance exhibited in AD cases. Of this, 25.21% was the *APOE* SNPs (red in Table 1.2) leaving only 5.41% to be accounted for by the sentinel SNPs identified through GWAS (white in Table 1.2). However, when variation in the regions surrounding all the variants listed in Table 1.2 (50 kilo base pairs (kbp) upstream and downstream) but not including the known risk variant itself was included, 28.63% of genetic variance was accounted for (Ridge et al., 2016). This highlights that it is not necessarily the SNP itself identified through GWAS that is causing the association but its surrounding region. Despite this, the majority of the genetic variance of LOAD remains unknown and novel approaches may be necessary in order to identify the additional variants and loci.

Despite this, having identified a selection of loci associated with LOAD, a pattern does begin to develop in which pathways and mechanisms are associated with the disease process. These pathways can be seen in Figure 1.5 and may provide areas to target in the search for therapeutic interventions (Beecham et al., 2014; Karch et al., 2012; Morgan, 2011). A recent pathway analysis of these GWAS loci, performed by IGAP, again highlighted the importance of the immune system in AD pathology. It specifically highlighted endocytosis regulation, protein ubiquitination and cholesterol metabolism processes (International Genomics of Alzheimer's Disease Consortium (IGAP) and International Genomics of Alzheimer's Disease Consortium IGAP, 2014). The immune system and inflammation association has also since gained ground with the theory that, instead of an immune reaction being a pathophysiological response to

AD neurological events, it may instead contribute to and exacerbate AD pathogenesis (Heppner et al., 2015). More recently, inflammatory changes have also been seen much earlier in the disease process, present even at the MCI stage of the disease. The amyloid cascade hypothesis suggests immune changes are a result of Aβ deposition. However, recent findings suggest that the immune processes may, in fact, drive AD pathology as opposed to purely being a result of it (Heppner et al., 2015).Once a definitive causative variant in each locus is identified, it may assist in contributing to clinical management of patients. This may be either through diagnosis and premature preventative treatment, guiding treatment selection (once improved therapies have been developed), or even predicting an individual's disease risk, phenotype and disease progression through a genetic panel screen (Schellenberg and Montine, 2012).

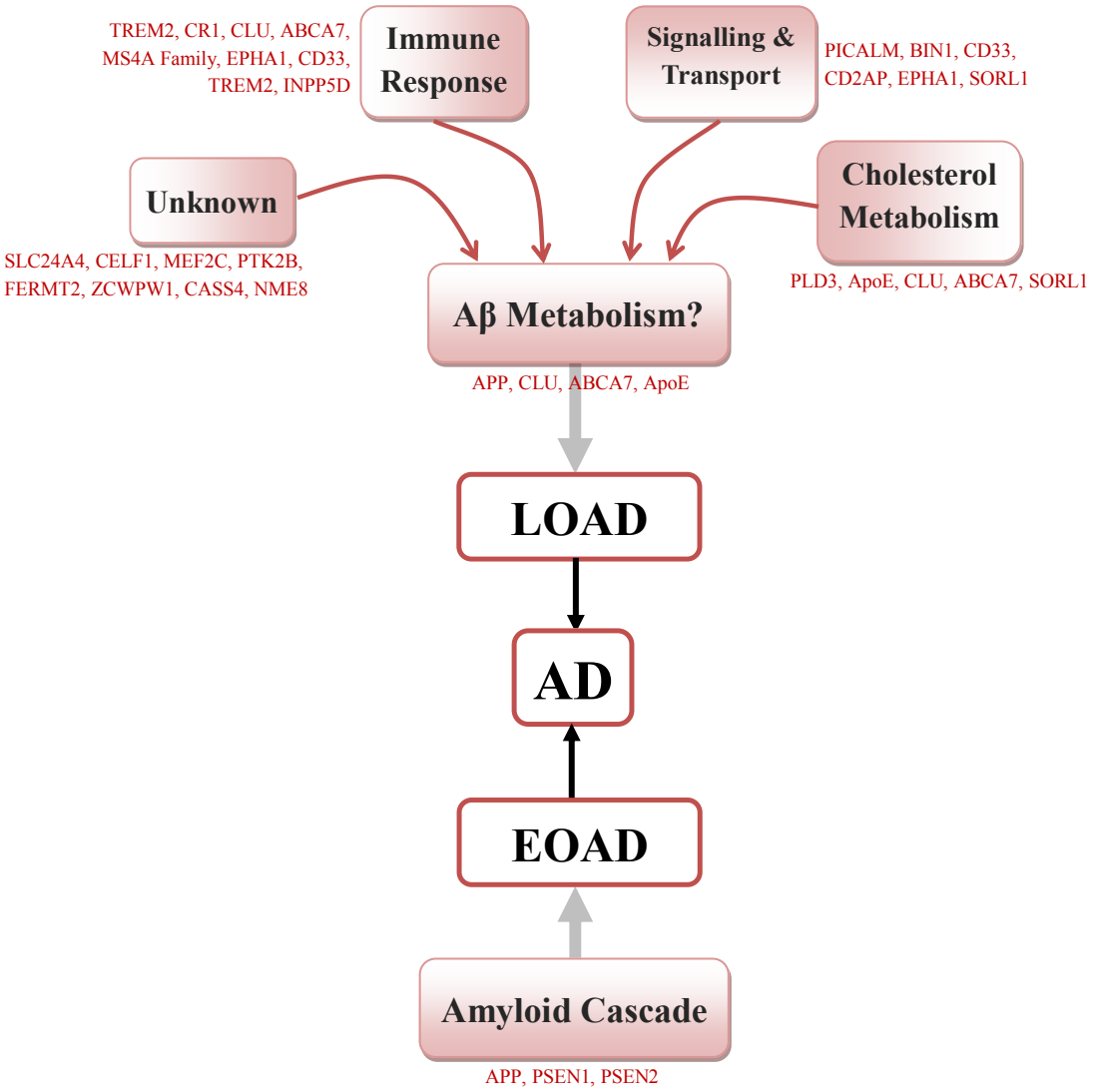


Figure 1.5
Disease pathways implicated in AD, both EOAD and LOAD, based on genes associated with the disease and their function. Adapted from Morgan (2011), Medway & Morgan (2014), Karch *et al* (2012) and International Genomics of Alzheimer's Disease Consortium (2014). The pathways highlighted (Cholesterol metabolism, immune response and signalling and transport) may feed into the amyloid processing pathway but may also provide vital therapeutic targets.

1.3 ABCA7

The ATP-Binding Cassette (ABC) transporters are a family of proteins thought to be one of the largest gene families (Iida et al., 2002; Klein et al., 1999) and are highly conserved, both evolutionary (85% conserved from bacteria all the way through to *homo sapiens)* (Higgins, 1992) and amongst the family itself (averaging 60% conserved) (Allikmets et al., 1998; Broccardo et al., 1999). They are known to transport a large variety of substrates across both intracellular and plasma membranes, switching between inward- and outward-facing conformations, against the substrates' electrochemical gradient by utilising the energy produced by ATP hydrolysis (Vasiliou et al., 2009; Zoghbi et al., 2016). In humans, there are 49 ABC proteins in total, as well as 21 pseudogenes, split into 7 subgroups labelled A-G based on phylogenetics, amino acid and protein sequence identity as well as the organization of their nucleotide binding domains (Dean and Annilo, 2005; Iida et al., 2002; Langmann et al., 2003; Vasiliou et al., 2009). This family of proteins have become a large area of research since several proteins have been linked with diseases, such as *ABCC7*, also known as *CFTR*, the cystic fibrosis gene (He et al., 2009).

The ABCA sub-family itself contains 12 members (ABCA1-13) with *ABCA11* known to be a pseudogene (Kaminski et al., 2006). The "A" family are the biggest subfamilies of the ABC transporters although they do not exhibit genetic clustering, being spread throughout the genome (Kim, 2004).

1.3.1 Identification of ABCA7

The 7th member of the A family - *ABCA7* was first identified in 2000 in human macrophages. cDNA of *ABCA7* was extended stepwise using primers designed from known sequences of other genes in the ABCA family, establishing its sequence. This same study assessed the mRNA tissue distribution, identifying *ABCA7* in myelo-lymphatic tissue, such as leukocytes, thymus, spleen and bone marrow, as well as in other peripheral blood cells. The amino acid sequence obtained contained all of the hallmark features of ABC transporters, which are described later (Kaminski et al., 2000a). In 2001, the transcriptional start sites were also predicted utilising MatInspector and 5' rapid amplification of cDNA ends (Broccardo et al., 2001).

1.3.2 The ABCA7 Gene

ABCA7 itself is located on chromosome 19p13.3 (Chr19:1,040,101-1,065,572 in human genome assembly GRCh38, as shown in Figure 1.6A), assigned by genomic probe in 2001, and has no known pseudogene (Broccardo et al., 2001; Zheng and Gerstein, 2006). Neighbouring *ABCA7* within 1.7kbp of its 3' end is the gene *HA-1*, coding for the human minor histocompatibility antigen, and 1kbp upstream of its 5' end is *CNN2*, coding for calporin, as seen in Figure 1.6B (Kaminski et al., 2000b).

ABCA7 was originally thought to consist of 46 exons (Kaminski et al., 2006, 2000b), although in 2006, a 47th exon was identified, 1019 base pairs upstream of exon one, shown in Figure 1.6C (Iwamoto, 2006). The gene size is now known to be 32kbp (a remarkably condensed genomic area) resulting in cDNA of 6.8kbp coding for 2146 amino acids of a molecular weight of 220 kilo Daltons (kDa) (Kaminski et al., 2006; Piehler et al., 2012).

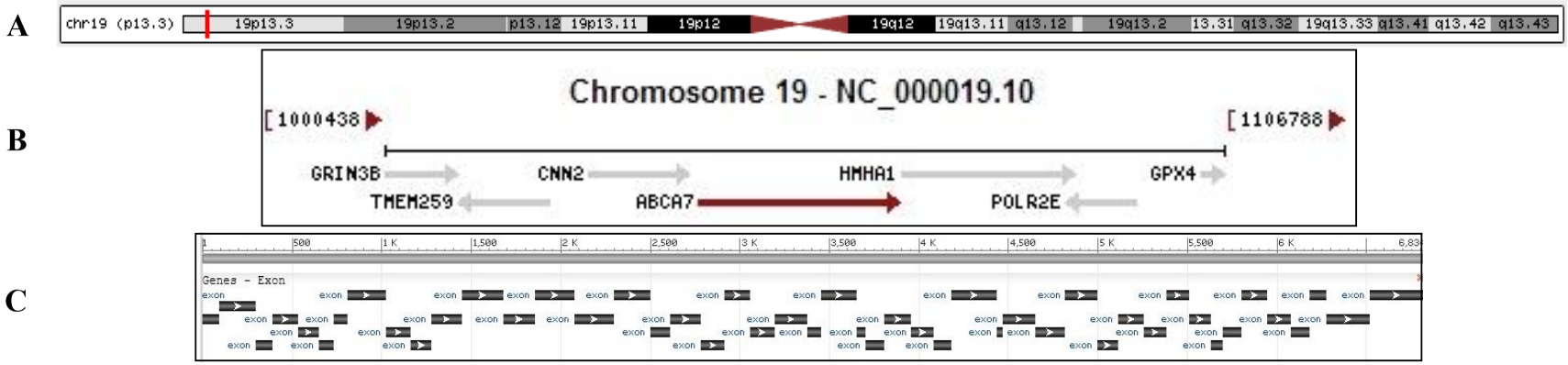


Figure 1.6

A: Location of the *ABCA7* on chromosome 19 position 13.3 (shown as a red bar). Image taken from the UCSC Human Genome Brower, RefSeq accession NM_019112.3 (http://genome-euro.ucsc.edu).

B: Location of genes neighbouring *ABCA7* on chromosome 19p13.3. Image taken from NCBI website, RefSeq accession NM_019112.3 (http:// www.ncbi.nlm.nih.gov)

C: Position of the 47 exons within *ABCA7*. Image taken from the NCBI website, RefSeq accession NM_019112.3 (http://www.ncbi.nlm.nih.gov).

1.3.3 The ABCA Proteins

All 48 ABC proteins share a similar structure (see Figure 1.7) which was originally identified in 1999 (Klein et al., 1999). They are comprised of four domains in total; two Nucleotide Binding Domains (NBDs) and two Trans-Membrane Domains (TMDs) (Dean, 2002). Half-size transporters also exist within the family, formed by one NBD and one TMD which are expressed separately before dimerising in order to form a functional unit (Barbet et al., 2012; Geillon et al., 2014). However, all of the ABCA family are "full-size" transporters meaning they are encoded in a single polypeptide chain (Kaminski et al., 2006). They are also the largest proteins within the ABC family (Piehler et al., 2012).

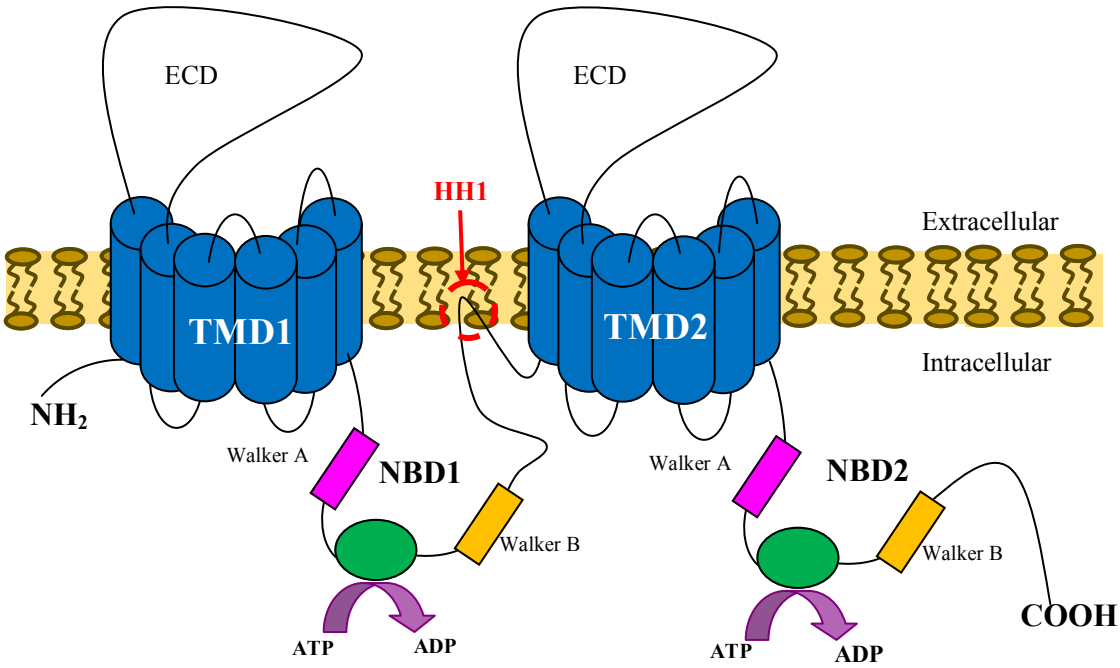


Figure 1.7
Schematic of an ABCA full-size transporter (of which ABCA7 is one) containing two Transmembrane Domains (blue), which are made up of six α– helices each and anchor the transporter in the lipid membrane, and two Nucleotide Binding Domains which contain the Walker A (pink) and Walker B (yellow) motifs as well as the ATP signature sequence, also known as Walker C (green). This is where ATP binds to and is hydrolysed to ADP to provide the energy in order to transport substrates across the membrane. The characteristic hydrophobic loop (HH1) is also shown (circled—red) as well as the characteristically large Extracellular Domains (ECD). These are both features which are specific to the ABCA proteins. Figure modified from Albrecht and Viturro *et al* (2005), Piehler *et al* (2012) and Kim *et al* (2008).

The two TMDs are hydrophilic and comprise of six α-helices each (blue in Figure 1.7), anchoring the transporter in the lipid membrane (Kim et al., 2008; Piehler et al., 2012). One thing which is unique to the ABCA sub-family are the large extracellular domains (ECD) located between the first and second transmembrane helices in each TMD (see Figure 1.7) (Bungert et al., 2001; Kim et al., 2006; Tanaka et al., 2001) as well as a hydrophobic domain (HH1) between the two TMDs (highlighted in red in Figure 1.7) (Albrecht and Viturro, 2007). The substrates are transported across membranes (either intracellular or extracellular membranes) through a pore formed by these TMDs. This pore is stabilised by metal ions, often between pore-lining cysteine residues on TMDs 1, 6 and 12 (El Hiani and Linsdell, 2014).

Each NBD contains three highly conserved motifs (Kim, 2004): the Walker A motif (pink in Figure 1.7); a Walker B motif (orange) and the ABC signature sequence or C-motif (green). The conserved sequences of these domains are shown in Table 1.3. These areas are where ATP binds to and is hydrolysed in order to transport the substrates across the membrane. The range of substrates within the ABC family is large and includes: metal ions; peptides; amino acids; sugars; xenobiotics; inorganic ions; polysaccharides; vitamins; lipids as well as a variety of hydrophobic compounds and cell metabolites (Borst et al., 2000; Borst and Elferink, 2002; Dean, 2002; Sasaki et al., 2003; Takahashi et al., 2005; Vasiliou et al., 2009). However, the ABCA family are more specialised, transporting just lipid molecules (Kim et al., 2008). As it is the NBD area to which ATP binds, any non-synonymous mutations in these regions cause a loss of functionality (Hrycyna et al., 1999).

Table 1.3
Residues of the conserved regions of ABCA proteins shown in single-letter amino acid code.
* corresponds to variable residues, residues shown underlined are conserved in all subclasses of the ABC transporter family. Modified from Broccardo *et al* 1999.

	Walker A	**Walker B**	**C Motif**
NBD1	**G**Q****L**G**HN**G**A**GK**TTT	L**D**EPT*G***D**P	L**SGG**M*RK
NBD2	**G**ECFGLL**G**VN**G**A**GK**STT	L**D**EPTTGM**D**P	Y**SGG***KRK

The Walker A, Walker B and C-Motifs are the areas which are highly conserved throughout the whole ABC family, and especially the ABCA group of proteins. The amino acid sequences of these particular motifs are given in Table 1.3 including the residues which are highly conserved throughout all subclasses of the ABC family (Broccardo et al., 1999). It is known that, in any ABC protein, if either the lysine residue in the Walker A motif (K in Table 1.3) or the aspartic acid residue in the Walker B motif (D in Table 1.3) are altered, then ATP cannot hydrolyse and therefore protein function is lost (Hrycyna et al., 1999).

1.3.4 The ABCA7 Protein

The ABCA7 protein specifically is a 220kDa protein and shares a 54% homology with its sibling protein ABCA1 (Kaminski et al., 2000a) . It was this similarity which initially suggested that the two proteins shared a degree of functional properties. ABCA7 also shares a 49% homology with ABCA4 (also known as ABCR) (Broccardo et al., 2001; Kaminski et al., 2000a).

ABCA7 Expression Patterns

When ABCA7 was originally identified, mRNA master blots were performed on multiple human tissues in order to identify its expression pattern. The majority of ABCA7 expression was found to be in myelo-lymphatic tissues including peripheral leukocytes (lymphocytes, granulocytes and monocytes), thymus, spleen and bone marrow as well as foetal tissues (Kaminski et al., 2000a). This expression pattern is remarkably different to that of *ABCA1* despite their sequence homology. In 2003, *ABCA7* RNA levels were also quantified using real-time Reverse-Transcription Polymerase Chain Reaction (RT-PCR), finding very similar results with the exception of high levels in the trachea (Langmann et al., 2003).

In 2004 Western blotting to examine the tissue expression of ABCA7 showed that the highest levels were found to be in the brain and this is now thought to be the primary site of ABCA7 expression. The exact regions of the brain where ABCA7 is predominant are, nonetheless, still to be determined. In human neuronal cells, ABCA7 was found to be expressed chiefly in microglia cells (which are principally involved in the immunological defence of the CNS) but they have also been shown to be highly expressed in CA1 neurones within the hippocampal area of the brain (Kim, 2004; Kim et al., 2006; Kreutzberg, 1995).

The expression pattern of ABCA7 within tissues are summarised in Figure 1.8 from the BioGPS resource. This resource has mapped the expression patterns of many proteins using Affymetrix Arrays and protein specific probes (219577_s_at for ABCA7) (Wu et al., 2009). Quantitative real-time PCR on cDNA prepared from RNA extracted from Human Embryonic Stem Cells (hESC) has also shown that ABCA7 is highly expressed in a variety of pluripotent stem cells, suggesting that it may also play a vital role during cellular maturation (Barbet et al., 2012).

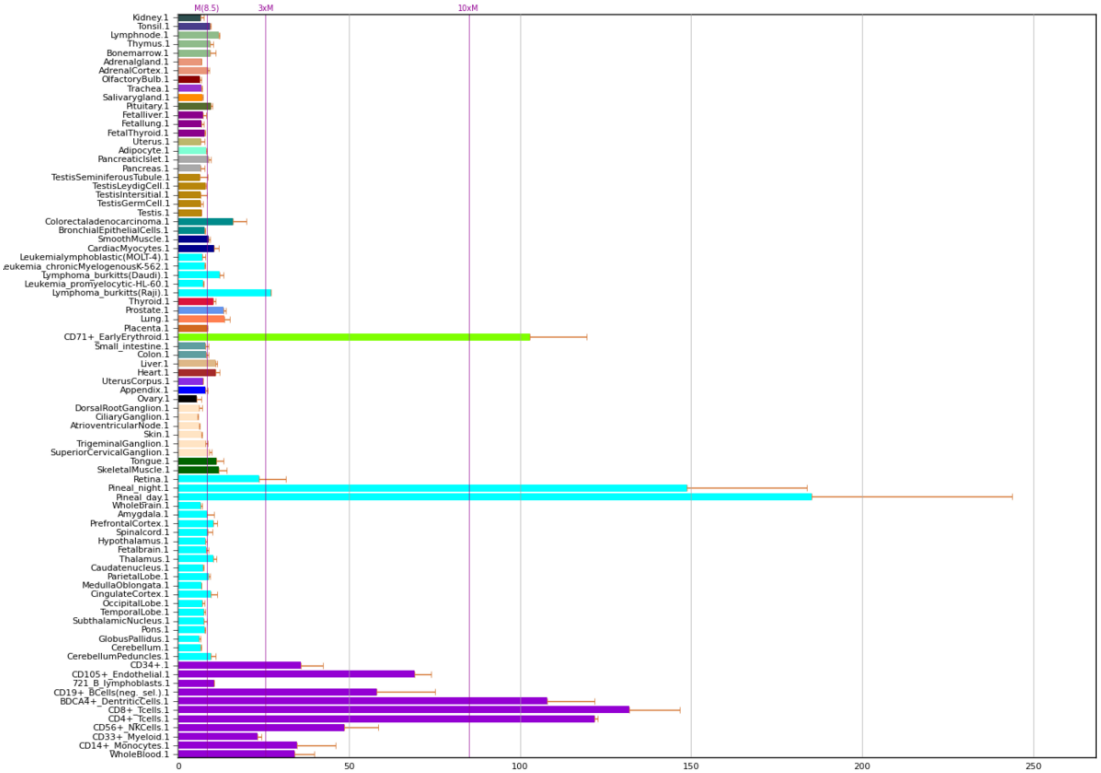


Figure 1.8
Tissue specific expression for *ABCA7* mRNA determined by the probe 219577_s_at by the GeneAtlas Affymetrix U133A array (Wu *et al.* 2009). M = median, 3xM = 3 x median, 10xM = 10 x median. The coloured bars represent the interquartile ranges of comparative enrichment of ABCA7 in the indicated tissues on a log2 scale. The greatest expression was apparent in the pineal gland (light blue) followed by cells of the immune system (purple) and precursors (light green). Image taken from the BioGPS tool (http://biogps.org/#goto=genereport&id=10347).

Cellular Location of ABCA7

The location of ABCA7 on the subcellular level is also a matter of some debate. In 2003, the protein was located on the plasma membrane of human embryonic kidney (HEK293) cells using western analysis and anti-ABCA7 antibodies (Wang, 2003) and this was verified in 2006 (Iwamoto, 2006). However, a year later, and in the same cell line, ABCA7 was found to be present on both the intracellular membranes and the extracellular plasma membrane (Abe-Dohmae et al., 2004). Despite this, upon immunofluorescent imaging of mouse macrophages, the protein was only present intracellularly, although they did locate it to the plasma membrane in renal tubule tissues from mice (Linsel-Nitschke et al., 2005), suggesting that ABCA7's subcellular location may be tissue specific.

Splicing

When ABCA7 was first identified in 2000, it was apparent that *ABCA7* has a series of relatively small introns, all consisting of multiples of three, suggesting the possibility of alternative splicing products (Kaminski et al., 2000b). Since then, at least two isoforms of ABCA7 have been identified due to alternative splicing (Ikeda et al., 2003). The Type II splice variant was shown, in 2003, to be formed by splicing out exon 5B. It was also shown that this variant is only located in the membrane of the endoplasmic reticulum of HEK293 cells whereas the Type I variant – containing exon 5B – is present in both intracellular and extracellular membranes (Ikeda et al., 2003). Exon 5B contains two termination codons with a methionine (initiation) codon following the second one, therefore, the Type II variant contains a novel N-terminus of 28 amino acids as opposed to the 166 amino acids of the Type I variant (Abe-Dohmae et al., 2006). The fact that the Type II variant is only located intracellularly may suggest that this loss is the reason the protein cannot localise to the plasma membrane (Abe-Dohmae et al., 2004). However, in both variants, the large ECD loop between the first and second transmembrane domain (known as ECD1) remains in the same orientation, exposed to the outer membrane area (Ikeda et al., 2003).

The two variants are also expressed differently in different tissues. Type I is found more abundantly in the brain and bone marrow, whereas Type II has a higher expression in lymph nodes, the spleen, thymus and trachea. Functional differences have also been identified between the two splice variants. HEK293 cells expressing the Type I ABCA7 variant demonstrated apolipoprotein A1 (apoA-I) mediated cholesterol release, whereas Type II transfected cells did not (Ikeda et al., 2003). The function of ABCA7 as a protein will be discussed in more detail later. However, this suggests the presence of a non-functional, intracellular splice variant.

In RefSeq there is also another transcript listed: NM_033308.1. This transcript is referred to as obsolete and is missing the first 138 amino acids, including the first TMD, resulting in a final protein of 2008 amino acids.

1.3.5 Transcriptional Regulation of ABCA7

High-Density Lipoprotein 3 & Low-Density Lipoprotein

It was known that ABCA1 expression increased in the presence of cholesterol (Venkateswaran et al., 2000) and, given their homologies, it was thought that ABCA7 may have a similar pattern. This was supported when *ABCA7* mRNA and protein levels where found to increase in the presence of low-density lipoprotein (LDL), and fell by as much as 70.6% in the presence of high density lipoprotein 3 (HDL3), which removes cholesterol from cells (Iwamoto, 2006). This supports the suggestion that ABCA7 is inversely regulated by cholesterol levels within the macrophage cells, similar to the reaction of ABCA1, meaning it is sterol-regulated and may function as a lipid exporter (Iida et al., 2002; Kaminski et al., 2000a).

Liver X Receptor & Retinoid X Receptor

As mentioned before, ABCA1 is known to be positively regulated by cholesterol levels within cells and this is through a pathway involving the liver X receptor (LXR) system, a system which also regulates transcription levels of apoE (Oram and Heinecke, 2005; Skerrett et al., 2015). This works in tandem with the retinoid X receptor (RXR) system, where LXR agonists are activated by binding to RXR agonists (Repa et al., 2000). In 2003, the effect of LXR agonists (with and without RXR agonists) on the effect of *Abca7* mRNA levels in mouse peritoneal macrophages were looked at and it was found that the levels were not altered at all (Abe-Dohmae et al., 2006; Wang, 2003). These results were replicated in bone-marrow macrophages

with free cholesterol, LXR and RXR agonists having no effect on ABCA7 protein levels (Kim, 2004).

Sterol Regulatory Element Binding Proteins

When the genomic organisation of *ABCA7* was investigated, a sterol regulatory element binding protein 1 (SREBP1) binding motif at its 5' end was discovered (Kaminski et al., 2000b). This family of transcription factors are frequently involved in moderating transcription of genes which regulate cholesterol metabolism (Horton and Shimomura, 1999) so it may not be surprising that it is involved in regulating the ABCA family.

It has been found that when the mouse cell line BALB/3T3 are forced to express SREBP1a and SREBP2, they both increase the expression of *ABCA7* mRNA and protein by 19% and 50% respectively (Iwamoto, 2006). However, this is the opposite of what happens with ABCA1, as when forcing Human Umbilical Vein Endothelium Cells (HUVEC) to express SREBP2, *ABCA1* mRNA expression is decreased by 50% (Zeng et al., 2004).

1.3.6 Function of ABCA7

General ABCA Function and Pathogenesis

The ABC family are generally known to transport molecules across membranes (Borst et al., 2000), using energy from ATP hydrolysis to move these substrates against their gradient. The range of molecules is vast as mentioned previously. However, ABCA transporters are known to specifically specialise in transporting lipids (Kim et al., 2008).

Many of the ABCA transporters have now been linked with pathology and disease due to the lipid transport pathway they are involved in. For example: ABCA1

transports cholesterol and contains the causative variants contributing to Tangier Disease, playing a vital role in maintaining plasma HDL levels (Assmann et al., 2001; Rust et al., 1999); ABCA3 transports the lipid rich pulmonary surfactant and is linked to neonatal respiratory failure (Shulenin et al., 2004); *ABCA4* mutations can cause abnormal lipid accumulation in the retina causing Stargadt's macular degeneration (Kielar et al., 2003; Lewis et al., 1999); ABCA12 has been connected with the rare keratinization disorder lamellar ichthyosis type 2 implying it is involved in lipid trafficking within keratinocytes in the skin (Akiyama, 2005; Lefèvre et al., 2003) and ABCA13 has been associated with schizophrenia and related diseases However the exact biological pathways behind this association are unknown, meaning the reason(s) why *ABCA13* is associated with these diseases remains a question, as it does for *ABCA7* and LOAD (Knight et al., 2009).

From these disease causing mutations, it can be implied that ABCA7 may be linked to disease pathways. However, the exact biological function of ABCA7 has not yet been characterised in detail. The following few sections will merely summarise the literature currently available on ABCA7's function.

Transport of Cholesterol and Phospholipids

As mentioned before, ABCA1 is known to be involved in cholesterol and phospholipid transport, exporting them to apolipoprotein receptors such as apoA-I or apoE to form HDL (Michael L Fitzgerald et al., 2004). As ABCA7 is 54% homologous with ABCA1 it was originally hypothesised that they may have similar functions. This idea is supported by the fact that *ABCA7* mRNA is up-regulated by cholesterol loading and down-regulated by cholesterol efflux in macrophages (Kaminski et al., 2000a), as well as the fact that it is regulated by SREBP1a, a transcription factor known to be involved in cholesterol metabolism (Iwamoto, 2006; Kaminski et al., 2000b).

The studies that have been performed in order to analyse ABCA7's transport of cholesterol and phospholipids are summarised in Table 1.4. Some agree with this original theory and some do not. However, all provide methological differences so the comparisons that can be drawn between these studies may be limited.

Table 1.4
Lipid release by ABCA7 in different studies. Where ↑ is an upregulation in release, ↓ is a downregulation in release and ND is Not Described. This table demonstrates the conflicting literature available on ABCA7's transporter function.

Reference	Cell Line	Cholesterol Release	Phospholipid Release
Hayashi *et al.* (2005)	HEK293	↑	ND
Ikeda *et al.* (2003)	HEK293	↑	↑
Wang (2003)	HEK293	↓	↑
Chan *et al.* (2008)	HEK293	↑	ND
Abe-Dohmae *et al.* (2004)	HEK293	↑	↑
Kim (2005)	Mouse macrophages	↓	↓
Linsel-Nitschke *et al.* (2005)	Human *ABCA7* in mouse macrophage	↓	↑
Quazi & Molday (2013)	Liposomes	ND	↑

In 2003, it was suggested that ABCA7 covalently binds to apoA-I when they are in proximity (Wang, 2003). However, further work went on to establish that, although ABCA7 does efflux phospholipids to apoA-I (specifically phosphatidylcholine and sphingomyelin) it does not efflux cholesterol to the same receptor (Wang, 2003). This contraindicates several other studies all of whom agree, to various degrees, that ABCA7 effluxes cholesterol out of HEK293 cells onto apoA-I (Abe-Dohmae et al., 2004; Hayashi et al., 2005; Ikeda et al., 2003). However, one study in 2012, performed on *Abca7* knock-out mice, proposing that Abca1 may compensate for Abca7 deficiency by becoming up-regulated (Meurs et al., 2012).

The idea that ABCA7 has a role in phospholipid transport is also supported by confirmation that, in liposomes, ABCA7 transports phospholipids (specifically phosphatidylserine) by flipping them from the cytoplasmic to the exocytoplasmic membrane leaflets when visualised fluorescently (Quazi and Molday, 2013).

In 2005, *Abca7* null (-/-) mice were developed (Kim, 2004) – despite reports that these mice could not be bred as this knockout was embryonically lethal (Jehle et al., 2006; Linsel-Nitschke et al., 2005). However, the macrophages taken from these mice showed no difference in cholesterol or phospholipid efflux when compared to *Abca7* wild-type mouse macrophages (Kim, 2004). Instead, they did report that the females of these *Abca7* -/- mice had up to 50% less white adipose tissue than their male counterparts or wild-type. This suggested that Abca7 may be involved in fat metabolism, a theory backed up by the fact that when 3T3L1 cells (a pre-adipose, fibroblast-like, mouse embryonic cell line) differentiate to mature adipocytes, their Abca7 expression greatly increases (Kim, 2004).

As many of these studies involved mouse Abca7, human and mouse ABCA7 functionality have been compared to see if that is where the differences in experimental results were originating. However, it was found that both human and mouse ABCA7 in HEK293 cells caused an increase of phospholipid efflux to apoA-I but neither of them transported cholesterol to the same apolipoprotein. In this same study *ABCA7* function was also knocked down using both small interfering RNA (siRNA) and heterozygote, knock-out mouse macrophages. In both of these cell lines there was no alteration in the phospholipid or cholesterol efflux to apoA-I in resting macrophages (Linsel-Nitschke et al., 2005).

All of these results (see Table 1.4) suggest that ABCA7 is involved in apolipoprotein mediated phospholipid release to lipid-free apolipoproteins such as apoA-I and apo-II (Abe-Dohmae et al., 2006, 2004; Kim, 2004; Linsel-Nitschke et al., 2005; Wang, 2003). However, it does not seem to play a significant role in plasma HDL levels, suggesting it may be involved in a more subtle role, perhaps in cellular cholesterol homeostasis in specific tissues, such as macrophages (Abe-Dohmae et al., 2006, 2004).

Immune Function and Phagocytosis

As mentioned previously, the genomic position of the *ABCA7* gene is within close proximity of the human minor histocompatibility antigen *HA-1* with its 5' end being within 1.7kbp of *ABCA7*'s 3' end (see Figure 1.6) (Kaminski et al., 2000b). HA-1 is involved in host defence and immune response, especially in response to Graft-versus-host disease (Mutis et al., 1999). It is known that genes within close proximity may share a similar function, for example, the α- and β-globin gene clusters (Lauer et al., 1980) and the Tumour Necrosis Factor gene pair on Chromosome 6 (Carroll et al., 1987), so it could be possible that ABCA7 and HA-1 also share a functional relationship although there is little homology between the two genes, suggesting there is no evolutionary relationship between them. This is also supported by the fact that they both have transcription factor binding sites for c-myb which plays a role in haematopoiesis (Gewirtz and Calabretta, 1988; Kaminski et al., 2000b) and they have similar tissue expression patterns (Kaminski et al., 2000b). The highest expression of ABCA7 is also in cells with immune function: macrophages in the body and microglia in human brain cells (Kim et al., 2006). Microglia make up 20% of the glial population within the CNS and are the resident macrophages within this system (Kreutzberg, 1995). There is therefore a high chance that ABCA7 is involved in the immune system in some way.

Both in 1996 and more recently in 2006, it has been found that *ABCA1* was highly homologous with the *CED-7* gene from the *Caenorhabditis elegans* nematode's and was thought to be its functional orthologue (Broccardo et al., 1999; Luciani and Chimini, 1996; Wu and Horvitz, 1998). This gene is part of the CED family which controls engulfment of cell corpses and phagocytosis (Ellis et al., 1991). They are structurally similar to the ABCA proteins, containing 2 NBDs and located in the plasma membrane (Wu and Horvitz, 1998). As it had been found that *CED-7* and *ABCA1* are highly homologous, in 2006, *ABCA7* was also compared with *CED-7*. It

was found that ABCA7's amino acid structure is 25% identical and 43% similar to CED-7 meaning it is the most homologous mammalian gene to *CED-7* (Jehle et al., 2006). There was, therefore, a likelihood of it being involved in human phagocytosis as it had already been identified at high levels in macrophages whose primary role is phagocytosis (Kreutzberg, 1995).

During phagocytosis, ABCA7 localises to the plasma membrane, specifically into phagocytic cups, focusing along with low-density lipoprotein receptor-related protein 1 (LRP1) which has a similar sequence and function to CED and plays a role in phagocytosing apoptotic cells (Ogden et al., 2001; Su et al., 2002).

When *ABCA7* was silenced in mouse peritoneal macrophages using siRNA (by 60-80%), it decreased phagocytosis of apoptotic cells by as much as 50-70%. However, this had no impact on FcR-mediated phagocytosis (Jehle et al., 2006). This inhibition of phagocytosis is replicated in *Abca7*+/- mice (reducing phagocytosis by 41%) and when macrophages are stimulated with apoptotic cells, Abca7 levels on the cell surface are raised (Jehle et al., 2006). Levels of SREBP1a and SREBP2, and therefore ABCA7 levels, are also increased during phagocytosis (Castoreno et al., 2005), indicating that SREBPs are a regulator of membrane biogenesis during phagocytosis due to their high phospholipid content. ABCA7 may, therefore, be involved in regulating phagocytosis (Iwamoto, 2006).

In 2010, it was demonstrated that Abca7 is stabilized by helical apolipoproteins, specifically apoA-I and apoA-II, in mouse peritoneal macrophages (Tanaka et al., 2010). They prevent Abca7 being broken down on the cell surface membrane by calpain, a calcium dependant protease (Goll et al., 2003). This increases the rate of phagocytosis, suggesting that the host defence system is enhanced by HDL components (Tanaka et al., 2010; Wang, 2003). One year later a report was also

published on the effect of HMG-CoA reductase inhibitors (statins) on phagocytosis as well as *ABCA1* and *ABCA7* mRNA levels. Statins are known to lower plasma LDL levels by up-regulating the LDL receptor through the SREBP system (Goldstein and Brown, 2009). Statins were found to increase ABCA7 expression and, through this, increase cellular phagocytic function both *in vitro*, in J774 cells, and *in vivo* in *Abca7* knockout mice. It was also found that statins lowered *ABCA1* mRNA levels (Tanaka et al., 2011). This implies that there is a link between cholesterol homeostasis and phagocytosis and that ABCA7 is involved in, and perhaps regulated by, this pathway.

From these studies it looks increasingly likely that ABCA7 is involved in phagocytosis, one of the substrates of which is apoptotic debris. There are an increasing number of autoimmune disorders which have been linked to insufficient removal of apoptotic cells as they go on to cause an inflammatory reaction (Lauber et al., 2004). This may be a clue to a pathological process in which ABCA7 is involved, especially as immune function is a known pathway associated with LOAD pathology (see Figure 1.5).

Epidermal Lipid Reorganisation

In 2003, mRNA expression profiling of all the ABC proteins during differentiation of both keratinocyte cells and Normal Human Epidermal Keratinocytes was performed. It was found that ABCA7 was highly expressed in both of these cell lines and, when they were forced to differentiate *in vivo*, *ABCA7* mRNA levels increased three-fold. Overexpression of ABCA7 in HeLa cells (a cervical cancer cell line) also increased the cellular ceramide expression levels, a lipid molecule found in high concentrations within the cell membrane involved in inhibiting cell proliferation and inducing cell death. This lipid is derived from sphingomyelin, a phospholipid which ABCA7 is known to transport (Wang, 2003). This suggests that ABCA7 may be involved in regulating cellular ceramide transport, as well as keratinocyte differentiation during

which intense lipid rearrangements occur (Abe-Dohmae et al., 2006; Kielar et al., 2003). However, no further work has been done on ABCA7's involvement in this area.

In summary, ABCA7 is known to transport phospholipids across cell membranes to extracellular apoA-I and apoE. It also has the capability of transporting cholesterol but to a much lesser extent (Abe-Dohmae et al., 2004; Chan et al., 2008; Wang, 2003). However, it also potentiates phagocytosis (Jehle et al., 2006; Kaminski et al., 2000a) which is a logical follow on from its high expression in both macrophages and microglia (Kaminski et al., 2000a; Kim et al., 2008).

1.4 The ABC Family and Disease

As mentioned previously, the ABC protein family have been linked with a variety of diseases. However, it was in 2004 when the ABCA sub-family was first linked to AD, with a conclusion that increasing expression of *ABCA2* increases the amount of APP and Aβ proportionally in neuroblastoma cells (Chen et al., 2004). Based on these results, as well as the fact that increased plasma cholesterol levels increase AD risk (Puglielli et al., 2001) and that ABCA2 is a cholesterol responsive gene (Kim et al., 2006), *ABCA2* was the sequenced in 230 EOAD cases. A synonymous SNP was identified, rs908835, as being associated with early onset AD, although the reasons behind this association are still unknown (Macé et al., 2005). Recently *ABCA5*, which has no known neurological function, was shown to be highly expressed in neurones and, to some degree, moderated Aβ production in cell systems through its function as a cholesterol transporter (Fu et al., 2015).

ABCA1 has also been linked to increased AD risk as *ABCA1* polymorphisms have been linked to altered CSF apoE and cholesterol levels. Lipid homeostasis is known to

be a large contributor to neuronal degeneration and may, therefore, increase the risk of AD as well as other neuro-degenerative conditions (Kim et al., 2006; Koldamova et al., 2005; Wollmer et al., 2003).

As mentioned previously, ABCA1's function has been well defined as a cholesterol and phospholipid transporter to HDL apolipoproteins, primarily apoA-I in plasma and apoE in the CNS (Wahrle et al., 2004). Neurologically, it is thought to prevent the accumulation of cholesterol in neurons by interacting with apoE to transport cholesterol across the blood-brain barrier to the periphery (Koldamova et al., 2003). ABCA1 deficiency leads to Tangier disease, a severe reduction of plasma HDL, leading to increased intracellular cholesterol deposits and a significantly increase risk of cardiovascular disease (Assmann et al., 2001; Bodzioch et al., 1999; Rust et al., 1999). This process has also been seen in patients infected with the HIV type I which inhibits ABCA1 function, increasing atherosclerosis in these patients (Jennelle et al., 2014). However, deficiency of neuronal Abca1 reduces apoE by about 80%, with the remaining apoE being poorly lipidated, consequently increasing the amyloid load in AD mouse models (Jiang et al., 2008; Koldamova et al., 2005). This suggests that Aβ processing and clearance is strongly influenced by the lipidation status of apoE and ABCA1 plays a role in this lipidation step. Loss-of-function mutations within the *ABCA1* gene have also been linked to low plasma apoE levels, increasing the risk of LOAD as well as other cerebrovascular diseases (Nordestgaard et al., 2015). However, further exploration of this relationship is required to provide all the details, possibly even providing therapeutic insights (Hirsch-Reinshagen et al., 2008).

As highlighted already, there is a distinct relationship between certain ABCA genes and disease. However, the genetic inheritance patterns of these associations are variable: *ABCA1* is inked to Tangier disease in an autosomal dominant manner (Rust et al., 1999). This had been identified a year previously when, through genome-wide

graphical linkage exclusion strategies, the genomic region containing *ABCA1* (chromosome 9q31) had been linked to Tangier's disease (Rust et al., 1998). Exonic sequencing of *ABCA1* subsequently identified several variants within this gene to be associated with the disease, including truncating and insertion/deletion (INDEL) variants (Rust et al., 1999). *ABCA3* has also been linked to neonatal surfactant deficiency in an homozygous loss-of-function pattern (Shulenin et al., 2004); *ABCA4* to Stargadt's disease in a recessive manner (Shulenin et al., 2004) and *ABCA12* to lamellar ichthyosis type 2 and *ABCA13* to schizophrenia in both compound heterozygous and homozygous relationships (Akiyama, 2005; Knight et al., 2009). The association between *ABCA13* and schizophrenia and other complex neurological conditions is of particular interest. Originally a chromosome abnormality was identified in one schizophrenic patient before exome sequencing was performed in 100 cases and 100 controls. This recognized multiple rare, coding variants which were then sequencing in a much larger cohort (>1600 cases and >950 controls), establishing that, in combination, the frequency of these rare variants was significantly higher in cases when compared to controls (OR = 1.93, p = 5.70e-03) (Knight et al., 2009).

Due to this variety it is therefore very difficult to predict the genetic inheritance patterns which *ABCA7* would exhibit with any disease it may be associated with. However, most of the genetic variants associated with LOAD thus far are common variants (most with a MAF of greater than 5%), behaving as risk factor variants, as opposed to causal factors, similar to the pattern seen in *ABCA13* for schizophrenia and bipolar disorder as mentioned above (Farrer, 2015; Knight et al., 2009). The likelihood is, therefore, that any disease-associated ABCA7 variants will act in a similar pattern: that of common variants altering disease risk.

1.5 ABCA7 and Disease

1.5.1 Sjögren's Syndrome

Sjögren's syndrome is an organ-specific autoimmune disease (Jacobsson et al., 1989), mainly affecting the salivary glands and tear ducts and, in its primary form, causing dry eyes and salivary glands (Sjögren, 1935; Sood et al., 2000) but, in its secondary form, can also cause connective tissue disease, most commonly, rheumatoid arthritis (Bloch et al., 1992). Sjögren's syndrome is surprisingly common with a prevalence of between 0.5% and 2.7% depending on the population (Jacobsson et al., 1989). In 1994, the autoantigen SS-N was found to be associated with Sjögren's syndrome (the "Sjögren's epitope") (Lopez-Longo et al., 1994) and in 2001, it was found that the amino acid sequence of this autoantigen is identical to residues 186-360 of ABCA7 which codes for its large first ECD (ECD-1: "ABCA-SSN") (Tanaka et al., 2001). This implies that ABCA7 is somehow involved in the pathogenesis of this autoimmune disease within the exocrine glands, the hypothesis being that the ECD-1 region of the protein can be cleaved off and circulated to act as this autoantigen (Tanaka et al., 2001). This involvement is supported by the fact that ABCA7 is expressed in the salivary glands of patients suffering from Sjögren's syndrome, detected by monoclonal antibodies against the ECD-1 region of the protein (Toda et al., 2005), an area which, normally, does not express ABCA7 at a high level (see Figure 1.8).

1.5.2 Late Onset Alzheimer's Disease

ABCA7 itself was first associated with LOAD in March 2011 when two GWAS were simultaneously published (Hollingworth et al., 2011; Adam C Naj et al., 2011). rs3764650 was identified as being associated with LOAD, achieving genome-wide significance. This is a SNP in intron 13 with a MAF of 0.10 (Hollingworth et al.,

2011). It was also found to be in Linkage Disequilibrium (LD) with rs3752246 with a D` of 0.89. This SNP is in exon 32, encoding a glycine to alanine substitution at amino acid position 1527. However, this is predicted to be a benign change (Adzhubei et al., 2010). In 2013, a further GWAS study identified another SNP, rs4147929, to be associated with LOAD (see Table 1.2). This SNP is in intron 42 with a MAF of 0.19 (Lambert et al., 2013). All of the above studies were completed in American and European populations although in 2013 the association of rs3764650 with AD was replicated in the African-American population and in 2014 it was replicated in the Han Chinese population in Taiwan as well as the Korean population (Chung et al., 2014; Liao et al., 2014; Reitz et al., 2013). All minor alleles of the variants within *ABCA7* associated with LOAD are related to an increased risk of disease.

It is known that GWAS signals identify loci associated with a phenotype, although it may not be that particular loci which is causing the signal, but also any neighbouring genes, which may cause the association. However, in 2014 association studies of collapsed (that is to say many rare variants, treated as one allele) loss-of-function *ABCA7* mutations proved that it was *ABCA7* involved in LOAD pathology as opposed to any surrounding genes (Steinberg et al., 2015). The fact that these variants were loss-of-function also indicates that it is a reduction of ABCA7 function that may contribute to AD risk. However, only rare variants were examined (that is ones with MAF of less than 0.5%) which, in combination associated with AD (Odds Ratio (OR) = 2.12, $p = 2.2 \times 10^{-13}$), but none located in combination with GWAS SNP (Steinberg et al., 2015). This study therefore does not account for the common variant association and therefore there are still hereditary gaps to be filled.

ABCA7's association with LOAD has also been validated by a study in 2014 where a GWAS was performed based on the neuropathological definition of AD as opposed to the clinical definition. Here, *ABCA7* was associated with the neuropathological

"hallmarks" of AD with an even stronger association then has been previously described (OR = 1.32, $p = 0.01$) (Beecham et al., 2014).

rs3764650 is now the GWAS tag SNP for *ABCA7* but the functional effect, if there is any, is still unknown. The minor allele (G) has been shown to decrease ABCA7 expression, perhaps explaining a pathological link (Vasquez et al., 2013). Conflictingly, it has been shown that ABCA7 expression is actually increased in AD cases and, in fact, expression increases in parallel with cognitive decline decreasing, but the authors suggest that this increase is a compensatory change after disease onset (Karch et al., 2012; Vasquez et al., 2013). This cognitive decline is also associated with increased ABCA7 expression levels in peripheral blood in rs3764650 risk allele (G) carriers (Ramirez et al., 2016). This minor allele of rs3764650, although an intronic variant, has also been linked to a later age of onset of disease as well as shorter disease duration when compared with major allele (T) carriers. Nonetheless, both of these studies are flawed with only 57 samples being used in the first and there being only marginal associations in the second, which may not survive multiple correction testing (Karch et al., 2012; Vasquez et al., 2013). However, in 2014 a larger study consisting of over 2,000 American individuals associated the minor allele of rs3764650 with increased rates of memory decline in both LOAD and MCI patients (Carrasquillo et al., 2014). This variant has also been associated with neuritic plaque burden with the burden increasing by 0.18 per minor allele (G) dosage (Shulman et al., 2013), as well as being linked to a greater cortical thinning in the frontal cortex and hippocampus (Ramirez et al., 2016). An alternative GWAS tag SNP (rs3752246) has also been linked, in the CG genotype, with a 2-fold increase with the odds of being Aβ plaque positive, similar to the risk of carrying the ε4 allele of *APOE* (Hughes et al., 2014).

The aim of GWAS is to pinpoint candidate genes which are then to be followed up, with the actual GWAS signal not necessarily having any functional significance, as in *ABCA7* with rs3764650 (Holmen et al., 2014). However, the GWAS signals themselves may not always be relevant and may merely indicate a nearby functional variant in a neighbouring gene (Sanna et al., 2011). It is therefore important to keep looking into these genes, and genes neighbouring them, in order to identify disease causing variants (Holmen et al., 2014).

1.6 ABCA7 and Alzheimer's Disease Pathways

GWAS and subsequent studies have therefore linked *ABCA7* to LOAD. However, the reasons behind this association are still unknown (Piehler et al., 2012).

It has been shown that ABCA7 plays a minor role in cholesterol and phospholipid transport (Abe-Dohmae et al., 2004; Chan et al., 2008; Wang, 2003). As the brain contains 25% of the bodies cholesterol (which itself is only 2% of body weight) it follows that ABCA7 may well have a role here in the brain if is part of the cholesterol cycle (Dietschy and Turley, 2001). apoE is the primary cholesterol carrier neurologically (Hirsch-Reinshagen et al., 2008), as well as being the primary genetic risk factor for AD (Holtzman et al., 2012), therefore they all may be linked.

ABCA7 is also known to colocalise with the LRP1 when it moves to the cell surface membrane as part of the phagocytic cup, a process which is defective in the absence of ABCA7 (Jehle et al., 2006). LRP1 is also known to localise with apoE as well as other AD associated proteins, such as CLU (which has also been associated with LOAD, see Table 1.2), in order to transport cholesterol into cells as well as uptake cellular debris (Neher et al., 2012) and APP, where it becomes internalizes in order to produce Aβ (Satoh et al., 2015). It is thought that if this process of absorbing

cholesterol is disrupted and neurones are starved of cholesterol, AD may develop (Liu et al., 2007). However, it is unknown if ABCA7 is also involved in this process.

Overexpression of ABCA7 has also been show to decrease the secretion of Aβ from APP secreting cells by retaining APP in perinuclear locations (Chan et al., 2008). ABCA7 may also have a role in increased Aβ accumulation as it has been shown that lipidated apoE normally reduces Aβ build up. However, if it is abnormally lipidated (which ABCA7 may have a small role in) then Aβ may accumulate (Wahrle et al., 2004). *ABCA7* loss-of-function, in several cell lines as well as *Abca7* knockout mice, has also been shown to significantly increase the proteolysis of APP, due to the loss of ABCA7-LRP1 interaction, therefore increasing Aβ production, perhaps providing the most direct link between ABCA7 and LOAD to date (Satoh et al., 2015). So far, these suggestions all work on the basis that ABCA7's primary role is cholesterol transport. However, as shown above, ABCA7 also plays a role in phagocytosis.

Deletion of *Abca7*, has been proven to double cerebral Aβ accumulation in AD mouse models, so far correlating with the above two theories. Although there was no alteration in APP or apoE levels, the mouse models did have significantly reduced spatial reference memory problems. They also showed that macrophages taken from these mice had reduced phagocytic activity and their ability to take up Aβ was reduced by 51% (Kim et al., 2013). This suggests that Abca7 regulates Aβ cerebral levels through its phagocytic and immunological function and could explain the link to AD.

Further studies support this. For example, ABCA7 has the highest expression neurologically in microglia (Kim et al., 2006) which contribute to phagocytic removal of apoptotic debris from the brain (Kreutzberg, 1995). It is known that microglia contribute to neuronal loss in AD mouse models as activated microglial cells surround

Aβ plaques phagocytosing viable neurones in the neighbouring area, contributing to neurodegeneration (Fuhrmann et al., 2010; Griffin et al., 1998). However, what has not been clarified is whether this abnormal immune response is an initiating event or a consequence of AD pathology.

If microglia cannot phagocytise dead or dying neurones than they may go on to release inflammatory markers contributing to AD pathology (Neher et al., 2012). Alternatively, high levels of Aβ are known to be toxic to neurones and this could initiate phagocytosis of otherwise viable neurones and synapses (Neher et al., 2012; Piehler et al., 2012). Phagocytosis may also be functioning abnormally, generating atypically folded protein plaques, leading to a variety of neurodegenerative disorders (Piehler et al., 2012).

In conclusion, there are several pathways by which ABCA7 may be linked to AD, whether this is through cholesterol and phospholipid transport or through phagocytosis although this link cannot be defined for sure. What therefore needs to be established is whether variants within *ABCA7* alter its function and contribute to AD pathology (Carroll et al., 1987).

1.7 Study Aims

This study therefore aims to:

- Create a catalogue of *ABCA7* exonic variants utilising previous sequencing projects as well as online catalogues.
- Analyse these exonic variants *in silico* to see if they are predicted to affect ABCA7 function.
- Perform genotyping assays on these variants to see if they are associated with late onset Alzheimer's disease.
- Design and carry out *in vitro* assays to investigate if variants alter ABCA7's function and how they do so.

2 Examining *ABCA7* Exonic Variants

2.1 Introduction & Background

Genome Wide Association Studies (GWAS) do not identify causative variants but merely associate genomic loci with a particular phenotype. However, these loci can theoretically be examined in further detail in order to elucidate the true causative variants of disease. These causative variants are commonly in linkage disequilibrium (LD - the lack of independence between alleles at different loci (Barnes, 2007)) with or in close proximity to the single nucleotide polymorphisms (SNPs) identified through GWAS. Indeed, it has been shown that the causative variant of a disease is less likely to be the SNP identified through GWAS, but more probably a variant in LD with it (Rosenthal et al., 2014). GWAS, therefore, links a certain LD block, or genomic location, with a phenotype rather than a specific variant (Visscher et al., 2012). Through identifying variants within these loci it also means rarer variants are also catalogued which may not have been identified through GWAS due to methodological limitations.

GWAS examines the hypothesis of "common variant - common disease," where the accumulation of a selection of common variants leads to a disease phenotype. However, rarer variants may still implicate disease risk, in the "common disease - rare variant" hypothesis. These rare variants may have a larger effect, causing disease in a small number of individuals, or a smaller effect, acting in combination with each other creating "disease haplotypes" (Gibson, 2012; Pabinger et al., 2014; Rivas et al., 2011). Examples of this "rare variant" hypothesis have been found specifically in Alzheimer's disease: carriers of a rare variant (MAF of 0.002) within the gene *TREM2*

(a protein involved in regulating microglial phagocytosis), coding for the amino acid substitution R47H, have been shown to have a significantly greater risk of developing LOAD in a small number of individuals upon whole exome sequencing of 281 LOAD patients and 504 control subjects (Guerreiro et al., 2013). The *TREM2* gene was originally linked with other, much rarer forms of dementia through pedigree studies in 2002 (Paloneva et al., 2002) before first being linked with LOAD in 2009. This was through a meta-analysis of the existing linkage studies of LOAD patients at the time. Despite the fact that there were only a very small number of included studies (n=5), meaning no regions reached genome-wide significance, the genomic region containing *TREM2* (6p21.1-q15) was one of the strongest signals ($p = 0.0169$) (Butler et al., 2009). This paved the way for rare, single variants, such as R47H, to be directly associated with LOAD through more high-throughput methodologies (Guerreiro et al., 2013). However, it cannot be said for sure that all of LOAD cases are caused by individual, rare mutations. Common variants must, therefore, also be examined, both individually and in haplotypes.

In a recent re-sequencing project involving Icelandic, European and American populations, loss-of-function *ABCA7* variants increased AD risk with a combined odds ratio (OR) of 2.03 and *p*-value of 6.8×10^{-15} (Steinberg et al., 2015). This has also been replicated in a Belgium cohort (OR = 4.03 and $p = 0.0002$) (Cuyvers et al., 2015). This implies that these loss-of-function variants may be the explanation behind *ABCA7*'s association with LOAD. As such, variants which are predicted to have a functional affect will be examined in more detail in this chapter.

Next Generation Sequencing (NGS) has been used previously in order to identify variants in risk loci identified by GWAS, which contribute to the risk of inflammatory bowel disease and type I diabetes, both complex diseases like LOAD (Nejentsev et al., 2009; Rivas et al., 2011). NGS was, therefore, utilised in this lab in 2011 (providing

the basis for previous PhD projects - Dr Jenny Lord, Dr James Turton and Dr Anne Braae) in order to catalogue the lower frequency genetic variants (ones with minor allele frequency of 1-5%), screening the eight original GWAS loci (*ABCA7*, *BIN1*, *MS4A*, *CD2AP*, *CD33*, *APOE* and *CR1* - see Table 1.2). NGS was selected as being the correct platform to perform this targeted re-sequencing due to its cost efficiency and ability to target specific loci. Several NGS platforms were researched before the Illumina HiSeq 2000 (Illumina, California, USA) was selected based on cost efficiency and its relatively low error rate of about 0.4% (Quail et al., 2012).

Ninety-six samples underwent whole genome NGS in total; all LOAD cases confirmed using the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) criteria and all collected with ethical approval and consent of all individuals. These individuals had DNA extracted from whole blood samples which were quality and quantity assessed using 1% agarose gel electrophoresis, as per lab protocol, and the Nanodrop 1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA), as per manufacture protocol. Eight pools of 12 samples each were generated, making sequencing cheaper - by up to 20% - when compared to bar-coding and creating unique libraries for individual samples (Schlötterer et al., 2014). Although this methodology is cheaper, it does come with caveats, especially in relation to low frequency variants which need to be distinguished from sequence errors (Altmann et al., 2012). However, once this bioinformatical challenge has been overcome, variant detection and allele frequency approximation is not significantly altered in NGS pooled studies when compared to using individual bar coded samples (Ingman and Gyllensten, 2009). A sample size of 96 was selected in order to provide enough power to the study to be able to provide a 95% chance of detecting variants with a MAF of 1-2% whilst still being a manageable size in order to screen and validate these pooled samples for the variants of interest which come to light when the data is analysed.

Before sequencing itself was undertaken, target enrichment was performed in order to create the libraries. The Agilent SureSelect system (Agilent, Santa Clara, USA) was chosen, utilising the in-solution hybrid capture methodology. This has been shown to be the best enrichment method for pooled samples (Day-Williams et al., 2011). RNA sequences were designed complementary to the areas of interest within the fragmented DNA. Magnetic beads were then used to pull the targeted areas out of solution. This DNA was polymerase chain reaction (PCR) amplified and clusters generated. Clusters (groups of identical DNA fragments) were used in order to increase the number of reads, increasing the likelihood that rarer variants will be identified. These clusters were amplified on a solid surface, in this case a flow cell, creating eight DNA libraries, one for each pool, in readiness for sequencing.

The Illumina HiSeq 2000 utilised Sequencing by Synthesis chemistry. This involved each cluster being sequenced base wise in real time by having all four deoxynucleotides (dNTPs) present, each with its own fluorescently labelled, reversible, terminator. These labels were detected as they became incorporated before they were cleaved off, allowing sequencing to continue. Sequencing was performed cyclically and, after a number of cycles, the clusters were regenerated and then sequenced in reverse using a second sequencing primer. This, again, increased the number of reads performed and meant data was generated from each end of the template, allowing context specific information to be captured upon the sequence being aligned through algorithms later in the process. The sequence data and quality scores were then produced in FASTQ files and sent to this lab where a bioinformatics pipeline was designed in order to analyse and annotate it.

Under this pipeline, the raw data was subjected to various bioinformatics tools in order to quality control the reads, align them, visualise, call the variants, annotate them before validating the variants, imputing them and performing association tests

upon them. The formation of this pipeline formed a large part of Dr Turton and Dr Lord's PhD projects and is summarised in Figure 2.1. As can be seen, once the sequence had been aligned and quality controlled, SNPs were called and frequencies calculated using the Comprehensive Read analysis for Identification of SNPs from Pooled sequencing (CRISP - Scripps Genomic Medicine, La Jolla, CA, USA) utility. These were quality checked again before being annotated using a variety of programmes. A number of variants called were then validated both *in silico* and *in vitro*. *In silico* validation involved utilising existing databases, the Exome project (Guerreiro et al., 2013) and the 1000Genomes browser (European Bioinformatics Institute, Cambridge, UK), comparing variants found, and their frequency, in order to corroborate the methodology used. SNPs were also validated *in vitro*, using Sanger sequencing methodology.

Before filtering, CRISP called a total of 209 variants in *ABCA7*, including 18 insertions and 17 deletions. This was reduced to 148 variants (6 insertions and 3 deletions) once they had been filtered based on the quality and depth of the reads. This number also included 24 variants which did not map to ones found in dbSNP build 137 (Bethesda, MD, USA, accessed in January 2013) or 1000Genomes and were therefore considered novel. This dataset was then imputed in order to establish if any of these rare NGS variants were associated with LOAD. In order to perform this, a control dataset (one not containing AD patients) was used as a reference (1000Genomes, Phase 1) and then association was measured using SNPtest.

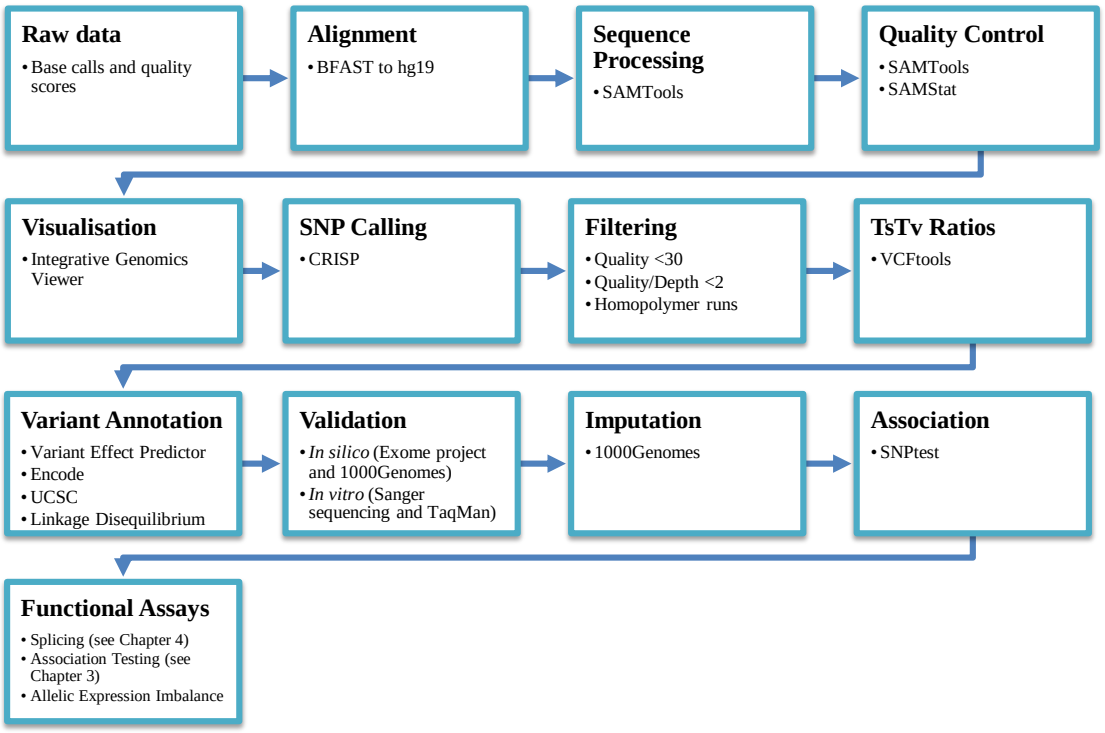


Figure 2.1
Simplified diagram of the bioinformatic pipeline which was applied to Next Generation Sequencing data in order to generate much of the data used in this current project. Adapted from Dr James Turton's and Dr Anne Braae's PhD theses.

As this chapter is focussing on the functional effect variants have on the protein, only the exonic variants from this NGS project will be examined. The first aim of this project is to create a catalogue of all exonic *ABCA7* variants and annotate them, using bioinformatics tools, to provide a selection of variants which have a suggestive effect on protein function. In combination with this NGS data, variants from the Exome Variant Server (EVS - NHLBI, Seattle, WA) will also be utilised. This browser contains exonic NGS data from 2203 African-Americans and 6503 European-Americans from a variety of phenotypes.

2.2 Methods

2.2.1 Catalogue Creation

Lists of variants from the NGS data and the online EVS tool were first achieved and co-ordinates of these variants obtained from the reference human genome build 37 (GRCh37). The EVS inventory was obtained by searching for the *ABCA7* gene name. Variants within introns were excluded from both lists by comparing their genomic location to the exonic boundaries of *ABCA7* as catalogued on the Ensembl database NM_019112.3 (EBI, Cambridge, UK). The exon number, cDNA location and consequential amino acid change caused by that variant were also recorded utilising Ensembl and EVS. rs ID numbers were obtained from the dbSNP database, build 137, if they existed. Variants which did not have a dbSNP ID were classed as "novel." MAFs for each of the variants were then obtained. As there is such a variety of MAFs depending on the database used, and the population examined, four alternate figures were collated, as well as the MAF calculated in the CRISP programme for the NGS variants. These programmes included 1000Genomes, EVS (the average MAF from all populations were used for both of these databases), dbSNP and HapMap where the European population MAFs were used. All MAFs were presented in a percentage format as opposed to a decimal format.

The effect these variants have on protein functionality was predicted using three different online tools: SeattleSeq (National Heart, Lung, and Blood Institute, Bethesda, MD, USA); Polyphen-2 (Harvard, MA, USA) and SIFT (La Jolla, CA, USA). These three programmes all make predictions on the effect these variants may have on the protein function based on the amino acid substitution they cause, methodology also utilised in previous studies (C. Sassi et al., 2014). SeattleSeq generates a Grantham Score based on the chemical properties and class differences (if

any) of the amino acid substitution. This score is then classified: “conservative” (a score from 0 to 50); “moderately conservative” (a score from 51 to 100); “moderately radical” (a score 101 to 150) or “radical” (a score of 151 or more) (Li et al., 1984). PolyPhen-2 (Polymorphism Phenotyping v2) annotates variants based on multiple sequence alignment of the surrounding region to homologous proteins and whether the site of variation is in a CpG region or a hypermutable site. This results in a score from 0 to 1 which is then classified: a “probably damaging” variant is one with a score above 0.85; “possibly damaging” is above 0.15 and “benign” for the remainder (Adzhubei et al., 2010). SIFT assigns scores based on conservation of the sequence surrounding the variation and gives a score ranging from 0 to 1. Any that score below 0.05 are then classified as “damaging,” and all others are “tolerated” (Kumar et al., 2009). For this purpose the SIFT Human Protein DB programme was used, utilising Ensembl build 63 with the ABCA7 protein ID of ENST00000263094. For the SeattleSeq programme the input involves the chromosome number, genomic location, as per hg37, and the alleles involved. All of this data was obtained from information already collated. For PolyPhen-2, the protein sequence in FASTA format was obtained from Ensembl (NP_061985.2), inputted along with the amino acid position and substitution effected, again, from data already collected.

These data were all compiled between 6th February and 29th March 2014.

2.2.2 Catalogue Refinement

The catalogue created was subsequently refined. However, before this initial refinement, in order to prevent potentially novel variants from being eliminated due to a lack of data (especially ones identified through the NGS), a separate list was created of only novel variants, with them being defined as ones without a dbSNP ID (rs number). Any of these novel variants which appeared in the NGS dataset and had a predicted damaging effect on protein function were included in the final list of

variants to be examined further. Only variants which had appeared in the NGS data were included as this meant they had been seen in our samples, increasing the likelihood of us being able to investigate them further, due to them being present in at least one sample in our DNA bank.

The first refinement performed on the whole dataset was executed based on the MAF of the variants. Power calculations were carried out in order to calculate the minimum allele frequency that could be genotyped using the samples held in the lab (approximately 1500 cases and 1500 controls). These calculations were performed in QUANTO v1.2.4 (USC Biostats, California, USA) using the gene-environment interaction model (as LOAD has both environmental and genetic components) and a study design of matched case-controls. Default settings were utilised except for the following which required manual configuration:

- The allele frequency was inputted at 0.005 to 0.05 (0.5% ~ 5%) in increments of 0.005 (0.5%) in order to view the power that could be achieved at a variety of MAFs
- The inheritance model was set as log additive
- Population prevalence was set to 0.24 (representing an approximation of the environmental component of LOAD being equal to 24%)
- Population risk was inputted to be 0.12 (12%) as an estimated risk of LOAD between the ages of 65 and 100
- The genetic effect size was estimated to have an odds ratio of 2
- The environmental effect size was estimated to have an odds ratio of 1.2
- The sample size was inputted as 1500 cases and 1500 controls
- Type I error rate was set to 0.05 (the maximum tolerated type I error rate)

- The programme was instructed to perform a 2df test as this enables the calculation of power whilst taking into account the effects from both genetic and environmental factors

The catalogue was then refined based on the lowest MAF which provided a power of at least 70%. As the MAFs between different programmes sometimes differed largely, as long as one of the databases displayed a MAF of equal to or over this figure, the variant was included.

As this project is focused on variants which have a functional effect on the ABCA7 protein, the next refinement was to exclude all synonymous variants. It is acknowledged that these synonymous SNPs, which were excluded at this step, may have functional effects such as on splicing or protein regulation (Edwards et al., 2012). However, at this stage, only SNPs affecting the protein function itself were selected. With this in mind, the next refinement was performed by also eliminating variants which were not expected to be damaging by at least one of the prediction programmes used. A damaging prediction was defined as a "Moderately Radical" or "Radical" Grantham Score, a "Possibly Damaging" or "Probably Damaging" PolyPhen-2 score or a "Damaging" score from SIFT.

2.2.3 Annotation of Variants

The final lists of variants, with the exception of any novel ones, were annotated with their LD scores in relation to the GWAS tag SNP – rs3764650. Novel variants were unable to be interrogated in this manner due their absence from the 1000Genomes database which was used to calculate these LD scores. This annotation was completed in a Linux environment where the 1000Genomes data for *ABCA7* was extracted, utilising Tabix (a tool which indexes genomic position sorted files in TAB-delimitated formats (Li, 2011)) with the following command:

```
$tabix -hf ftp://1000genomes.ebi.ac.uk/vol1/ftp/release/20110521/ALL.chr19.phase1_release_v3.20101123.snps_indels_svs.genotypes.vcf.gz 19:1040103-1065569 > ABCA7raw1kGSNPs.vcf
```

This produced a Variant Call File (VCF) containing all of the SNPs and insert/deletions (indels) within the base pairs 1,040,103 to 1,065,569 on chromosome 19, the co-ordinates of *ABCA7*. The European gene frequencies were then extracted using a Perl script, written by Dr Christopher Medway, and comparing it to a list of samples in the GBR and CEU populations:

```
$perl vcf-subset -c Phase1_Samples_GBR_CEU ABCA7raw1kGSNPs.vcf > ABCA7_CEU_GBR.vcf
```

This also produced a VCF file which was then converted to a PLINK file using the vcftools software (Danecek et al., 2011):

```
$vcftools --vcf ABCA7_CEU_GBR.vcf -plink-tped --out ABCA7_CEU_GBR
```

PLINK was then used to calculate the LD between the SNP of interest ("rs1") and the GWAS SNP - rs3764650. PLINK is another command-line programme frequently used to analyse large genetic data sets (Purcell et al., 2007). The following command line was used:

```
$plink -tfile ABCA7_CEU_GBR -ld rs3764650 rs1
```

This command was repeated for all the variants of interest. Using this method, PLINK provides the D` score (the measurement of evidence of recombination between loci calculated from haplotype frequencies), r^2 score (covariance in allelic value between the loci), the estimated haplotype frequency compared to the frequency expected under linkage equilibrium as well as the haplotypes which are in phase (Balding et al., 2007).

All variants, including any novel ones, were then visualised on the tertiary structure of ABCA7 utilising the TOPO2 online programme. The FASTA amino acid sequence of

ABCA7 was uploaded (NP_061985) along with transmembrane residue numbering acquired from UniProt (UniProt Consortium 2012, ID Q8IZY2). TOPO2 provides an image of the transmembrane protein. The amino acids of the variants of interest were inputted into the TOPO2 user interface to visualise where they are located on the protein. The locations of the three identifying motifs (Walker A, Walker B and ATP signature sequence or C motif) were also examined using the amino acids sequences as identified in Table 1.3 (Broccardo et al., 1999). Once the locations of these three motifs had been established, the full list of variants was re-examined to see if there were any variants of interest within these functionally vital areas.

The effect these variants have on the structure of the transmembrane domains was scrutinised in order to analyse if they alter the transmembrane structure of the ABCA7 protein. This was performed using the online programme TMHH Server version 2.0 (Centre for Biological Sequence Analysis, Technical University of Denmark, Lyngby, Denmark), which predicts transmembrane regions utilising the hidden Markov model with the N-best algorithm (Krogh et al., 2001). This programme has been shown to be the most accurate transmembrane prediction programme (Möller et al., 2001). The wild type amino acid sequence of ABCA7 (NP_061985) was inputted into the programme as well as the sequence with the mutated amino acids for any variants which had been shown to be located within one of the transmembrane domains as demonstrated in the TOPO2 output. This resulted in a list of the transmembrane domains in numerical format as well as a graph of probabilities of these being correct.

2.3 Results

The total number of *ABCA7* exonic variants obtained from EVS was 238 and 38 from the NGS dataset. Both of these data groups contained the only exonic GWAS SNP, rs3752246 (19:1056492). When all of these were collated, a list of 240 variants was created, available in Appendix A. A breakdown of the whole catalogue can be seen in Figure 2.2, showing the refinement process and the number of variants excluded at each stage. In total, 204 variants were excluded due to their low MAF, 21 due to being synonymous variants, 11 due to being benign changes, leaving five variants as being potentially damaging. The total list contained a total of 11 "novel" variants, available in Table 2.1. Of these, only two variants were present in NGS dataset although only one was predicted to be damaging by both PolyPhen2 and SIFT (present at 19:1056958) and will therefore be included in the final list of variants to be annotated further.

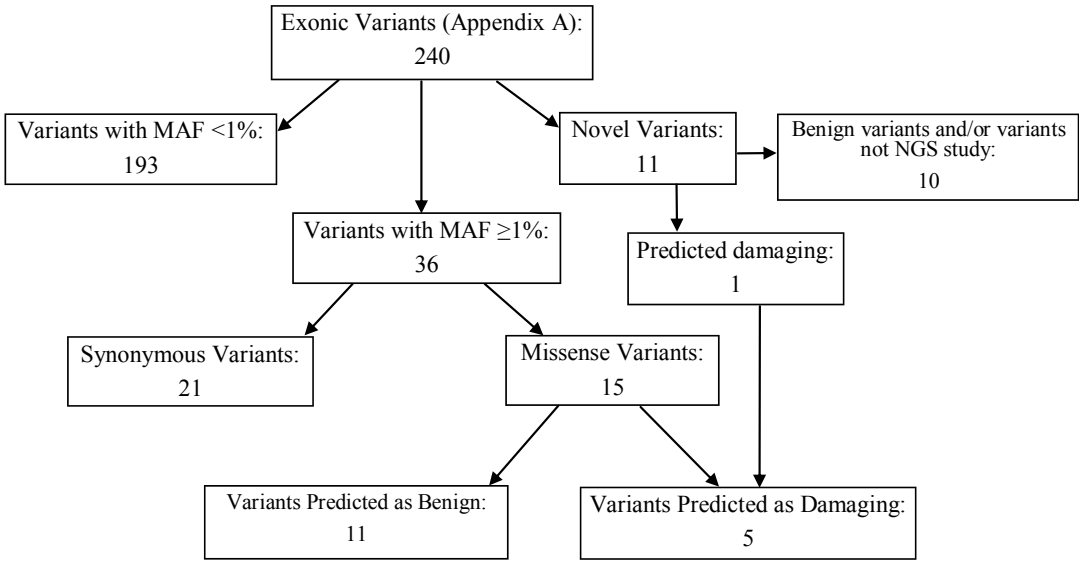


Figure 2.2
Breakdown of the variants catalogued. The final five variants will be analysed in more detail through genotyping assays. The complete list of variants, as well as all their details, can be seen in Appendix A.

Table 2.1

All of the novel variants within the full database (available in Appendix A) created from the NGS data and the EVS database. Novel was classed as variants without rs IDs, according to dbSNP build 137. All cDNA changes were annotated using reference sequence NM019112.3. This table was created in order to not miss any possibly pathogenic variants, especially ones present in the NGS dataset. The most promising of these variants is 19:1056958 which is predicted to be damaging by two of the three programmes used and is present in the NGS data. It will therefore be included in the list of variants for further analysis. All MAFs are presented in percentage format.

						MAF						Affect Prediction					
	CHR	GENOMIC POSITION	dbSNP ID	GENE LOCATION	cDNA	HapMap (European) (%)	EVS (All) (%)	dbSNP (%)	1000Genomes (All) (%)	CRISP (%)	PROTEIN	Grantham Score	Polyphen2	SIFT	NGS	GWAS	EVS
1	19	1047253	-	Exon 15	c.1944_1946del	NA	0.1039	NA	NA	NA	A648_F649delinsA			Damaging (0.04)			•
2	19	1047507	-	Exon 16	c.2124_2130del	NA	0.2825	NA	NA	NA	E709Afs*86	FRAMESHIFT					•
3	19	1048961	-	Exon 17	c.2337C>T	NA	NA	NA	NA	0.775	P779P			Tolerated (0.636)	•		•
4	19	1051493	-	Exon 21	c.2871del	NA	0.0122	NA	NA	NA	S958Pfs*34	FRAMESHIFT					•
5	19	1055907	-	Exon 31	c.4208del	NA	0.0639	NA	NA	NA	L1403Rfs*7	FRAMESHIFT					•
6	19	1056958	-	Exon 34	c.4639T>C	NA	NA	NA	NA	0.663	S1547P	Moderately Conservative (74)	Probably Damaging (1)	Deleterious (0)	•		
7	19	1057946	-	Exon 36	c.4914_4916del	NA	0.1519	NA	NA	NA	S1638_F1639delinsS			Damaging (0.00)			•
8	19	1058026	-	Exon 36	c.4994del	NA	0.04	NA	NA	NA	T1666Pfs*15	FRAMESHIFT					•
9	19	1058640	-	Exon 38	c.5173_5174insA	NA	0.008	NA	NA	NA	L1725Hfs*55	FRAMESHIFT					•
10	19	1061813	-	Exon 41	c.5497del	NA	1.9652	NA	NA	NA	K1833Rfs*	FRAMESHIFT					•
11	19	1065320	-	Exon 47	c.6338_6340del	NA	0.3116	NA	NA	NA	Q2113_K2114delinsQ	DELETION/INSERTION					•

The output of the power calculations in QUANTO can be seen in Figure 2.3. When this output was analysed, it is apparent that, with approximately 1500 cases and 1500 controls, 76% power can be achieved when analysing a variant with a 1% MAF. A MAF of 1% or more in any of the databases analysed was therefore used in order to perform the first refinement. This shortened the list considerably to 36 variants, again including the GWAS SNP at a MAF of around 18% (see Table 2.2). Upon further refinement to only missense variants, this list was again shortened to 15, still containing the GWAS SNP which codes for a glycine to alanine change at amino acid position 1527 (see Table 2.3). Four of these variants were predicted to have a damaging effect on the protein and therefore made the final refinement (see Table 2.4). Here the GWAS SNP was eliminated as it was not predicted by any of the programmes to be damaging. This provided a final list of four variants: rs3752233 (R463H); rs59851484 (A676T); rs3752239 (N718T) and rs114782266 (R1812H), as well as one novel mutation from Table 2.1 (19:1056958, S1547P)), which was also predicted to be damaging.

File Parameters Wizards View Help

Outcome:		Disease	
Design:		Matched case-control	
Hypothesis:		Gene-environment interaction	
Sample size:		1500 case-control pairs	
Significance:		0.050000, 2-sided	
Gene			
Mode of inheritance:		Log-additive	
Allele frequency:		0.0050 to 0.0500 by 0.0050	
Binary environmental factor			
Prevalence:		0.2400	
Disease model		Summary parameters	
$^*P_0$	0.114312	k_P	0.120000
R_G:	2.0000	$^*\bar{R}_G$:	1.9988
R_E:	1.2000	$^*\bar{R}_E$:	1.1998
R_{GE}:	1.0000	(*indicates calculated value)	

Parameter	Null	Full	Reduced	df
Interaction	$\beta_{GE}=0$	$\beta_{GE},\beta_G,\beta_E$	β_G,β_E	1
Gene	$\beta_G=0$	β_G	——	1
Environment	$\beta_E=0$	β_E	——	1
Interaction, Gene	$\beta_{GE}=0,\beta_G=0$	$\beta_{GE},\beta_G,\beta_E$	β_E	2

	Power						
Frequency	Interaction	Gene	Environment	Interaction, Gene	P_0	$\bar{R}_G$	$\bar{R}_E$
0.005000	0.0500	0.5641	0.5817	0.4599	0.114312	1.9988	1.1998
0.010000	0.0500	0.8476	0.5810	0.7661	0.113415	1.9988	1.1996
0.015000	0.0500	0.9536	0.5803	0.9131	0.112526	1.9988	1.1995
0.020000	0.0500	0.9869	0.5796	0.9707	0.111647	1.9988	1.1993
0.025000	0.0500	0.9965	0.5789	0.9907	0.110775	1.9988	1.1991
0.030000	0.0500	0.9991	0.5783	0.9972	0.109912	1.9988	1.1989
0.035000	0.0500	0.9998	0.5777	0.9992	0.109057	1.9988	1.1988
0.040000	0.0500	0.9999	0.5770	0.9998	0.108211	1.9988	1.1986
0.045000	0.0500	0.9999	0.5764	0.9999	0.107372	1.9988	1.1985
0.050000	0.0500	0.9999	0.5758	0.9999	0.106542	1.9988	1.1983

Figure 2.3
Output from QUANTO upon doing the power calculations. The Interaction, Gene column is used when analysing the power as LOAD has both an environmental and genetic component which interact. This shows that, with an allele with Minor Allele Frequency of 1% (0.01), a power of 76.61% (0.7661) can be achieved with 1500 case samples matched with 1500 controls (case-control pairs). This 1% cut-off was therefore used.

Table 2.2

All of the variants from Appendix A with a Minor Allele Frequency of 1% or above in at least one of the databases. This produces a list of 36 variants which was then further refined. All cDNA changes were annotated using reference sequence NM019112.3. All MAFs are presented in percentage format.

						MAF						Affect Prediction					
	CHR	GENOMIC POSITION	dbSNP ID	GENE LOCATION	cDNA	HapMap (European) (%)	EVS (All) (%)	dbSNP (%)	1000Genomes (All) (%)	CRISP (%)	PROTEIN	Grantham Score	Polyphen2	SIFT	NGS	GWAS	EVS
1	19	1041347	rs182233998	Exon 2	c.-14T>C	NA	1.1456	31.80	1.00	0.61	-	-	-	-	•		•
2	19	1041352	rs3752229	Exon 2	c.-9A>G	6.70	3.4676	12.00	12.00	7.2	-	-	-	-	•		•
3	19	1041852	rs3764644	Exon 4	c.183G>T	NA	6.7426	7.10	7.00	1.963	L61L			Tolerated (1)	•		•
4	19	1042809	rs3764645	Exon 7	c.563A>G	NA	38.8667	41.28	41.00	41.025	E188G	Moderately Conservative (98)	Benign (0.244)	Tolerated (0.647)	•		•
5	19	1043103	rs72973581	Exon 8	c.643G>A	NA	4.6301	2.20	2.00	4.238	G215S	Moderately Conservative (56)	Benign (0.029)	Tolerated (0.625)	•		•
6	19	1043748	rs3752232	Exon 10	c.955A>G	4.50	11.3025	10.10	10.00	1.788	T319A	Moderately Conservative (58)	Benign (0)	Tolerated (0.882)	•		•
7	19	1044712	rs3764647	Exon 11	c.1184A>G	4.50	10.7896	9.70	10.00	1.788	H395R	Conservative (29)	Benign (0)	Tolerated (1)	•		•
8	19	1045026	rs10405305	Exon 12	c.1241C>G	NA	3.1471	2.70	3.00	NA	A414G	Moderately Conservative (60)	Benign (0.055)	Tolerated (0.35)			•
9	19	1045173	rs3752233	Exon 12	c.1388G>A/C	0.00	2.9837	5.90	6.00	2.263	R463H	Conservative (29)	Probably Damaging (0.997)	Tolerated (1)	•		•
10	19	1047002	rs3752234	Exon 14	c.1824A>G	NA	46.6346	40.10	40.00	58.075	A608A			Tolerated (1)	•		•
11	19	1047161	rs3752237	Exon 15	c.1851A>G	38.30	37.5905	31.80	32.00	61.988	G617G			Tolerated (1)	•		•
12	19	1047336	rs59851484	Exon 15	c.2026G>A	NA	3.8354	3.00	3.00	NA	A676T	Moderately Conservative (58)	Probably Damaging (1.000)	Tolerated (0.55)			•
13	19	1047537	rs3752239	Exon 16	c.2153A>C	NA	2.7002	5.80	6.00	2.063	N718T	Moderately Conservative (65)	Possibly Damaging (0.529)	Tolerated (0.239)	•		•
14	19	1048982	rs9282560	Exon 17	c.2358C>T	NA	0.1618	0.60	1.00	NA	C786C			Tolerated (0.07)			•
15	19	1049269	rs4147914	Exon 18	c.2385G>A	NA	15.2863	21.90	22.00	17.875	L795L			Tolerated (1)	•		•
16	19	1049305	rs4147915	Exon 18	c.2421C>A	NA	13.1846	21.70	22.00	9.013	V807V			Tolerated (1)	•		•
17	19	1050996	rs74176364	Exon 19	c.2629G>A	NA	0.5167	1.50	2.00	NA	A877T	Moderately Conserved (58)		Tolerated (0.38)			•
18	19	1051214	rs3752240	Exon 20	c.2745A>G	36.90	36.5462	30.60	31.00	37.175	V915V			Tolerated (0.11)	•		•
19	19	1052005	rs3764652	Exon 22	c.3027C>T	44.60	40.1544	38.50	38.00	40.963	A1009A			Tolerated (1)	•		•
20	19	1052086	rs61576791	Exon 22	c.3108G>A	NA	5.9452	6.00	6.00	NA	T1036T			Tolerated (1.00)			•
21	19	1053524	rs3752241	Exon 24	c.3417C>G	21.20	16.3879	16.40	16.00	16.45	L1139L			Tolerated (1)	•		•
22	19	1054060	rs3752243	Exon 26	c.3528A>G	42.70	47.3551	49.30	49.00	34.888	L1176L			Tolerated (0.352)	•		•
23	19	1055191	rs3745842	Exon 30	c.4046G>A	NA	40.6467	39.50	39.00	38.375	R1349Q	Conserved (43)	Benign (0.002)	Tolerated (0.546)	•		•
24	19	1056065	rs881768	Exon 32	c.4239A>G	45.00	45.9372	43.80	44.00	24.75	R1413R			Tolerated (1)	•		•
25	19	1056227	rs111322482	Exon 32	c.4401T>C	NA	1.8761	1.80	2.00	NA	A1467A			Tolerated (0.66)			•
26	19	1056421	rs113711363	Exon 33	c.4509G>A	NA	1.9837	1.80	2.00	NA	P1503P			Tolerated (0.32)			•
27	19	1056426	rs113269196	Exon 33	c.4514G>A	NA	0.7919	0.60	1.00	NA	R1505H	Conserved (29)	Benign (0.016)	Tolerated (0.18)			•
28	19	1056492	rs3752246	Exon 33	c.4580G>C	18.30	12.7884	17.60	18.00	77.2	G1527A	Moderately Conservative (60)	Benign (0)	Tolerated (1)	•	•	•
29	19	1058176	rs4147918	Exon 37	c.5057A>G	4.50	3.014	5.60	6.00	1.638	Q1686R	Conservative (43)	Benign (0.001)	Tolerated (0.234)	•		•
30	19	1059056	rs114782266	Exon 40	c.5435G>A	NA	2.6838	2.70	3.00	2.963	R1812H	Conservative (29)	Benign (0.108)	Deleterious (0.017)	•		•
31	19	1061804	rs78320196	Exon 41	c.5487T>C	NA	3.9142	4.30	4.00	2.2	N1829N			Tolerated (1)	•		•
32	19	1062192	rs4147921	Exon 42	c.5592T>C	NA	2.9755	5.60	6.00	1.6	A1864A			Tolerated (0.501)	•		•
33	19	1064193	rs4147930	Exon 45	c.5985G>A	0.00	29.5659	36.20	36.00	73.6	L1995L			Tolerated (1)	•		•
34	19	1065018	rs4147934	Exon 46	c.6133G>T	NA	25.0256	36.80	37.00	77.338	A2045S	Moderately Conservative (99)	Benign (0.057)	Tolerated (0.962)	•		•
35	19	1065044	rs4147935	Exon 46	c.6159C>T	NA	26.1408	23.30	23.00	44.725	G2053G			Tolerated (0.133)	•		•
36	19	1065563	rs2242437	Exon 47	c.6580G>C	29.30	NA	36.40	36.00	75.863	-	-	-	-	•		•

Table 2.3

All of the variants from Appendix A with a MAF of 1% or above in at least one of the databases as well as coding for a missense variation and therefore more likely to have a functional effect on the ABCA7 protein. All cDNA changes were annotated using reference sequence NM019112.3. All MAFs are presented in a percentage format.

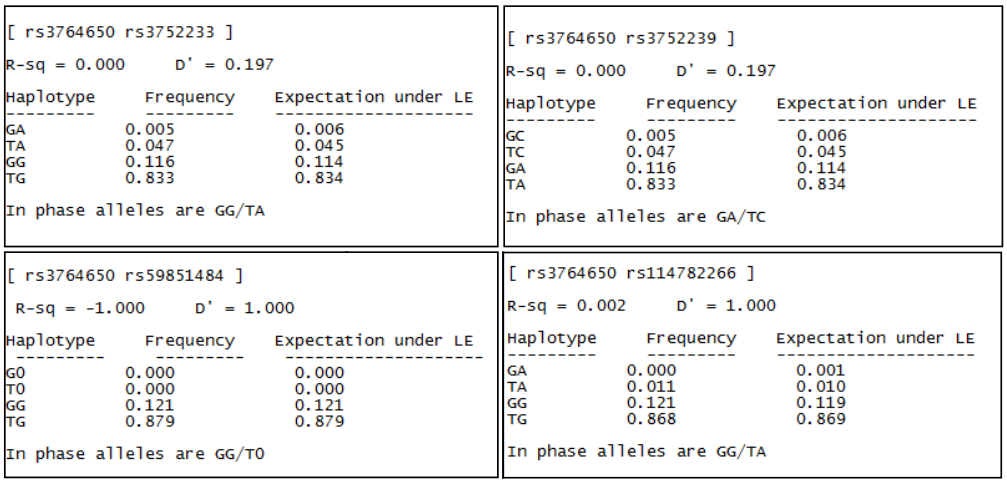
						MAF						Affect Prediction					
	CHR	GENOMIC POSITION	dbSNP ID	GENE LOCATION	cDNA	HapMap (European) (%)	EVS (All) (%)	dbSNP (%)	1000Genomes (All) (%)	CRISP (%)	PROTEIN	Grantham Score	Polyphen2	SIFT	NGS	GWAS	EVS
1	19	1042809	rs3764645	Exon 7	c.563A>G	NA	38.8667	41.28	41.00	41.025	E188G	Moderately Conservative (98)	Benign (0.244)	Tolerated (0.647)	•		•
2	19	1043103	rs72973581	Exon 8	c.643G>A	NA	4.6301	2.20	2.00	4.238	G215S	Moderately Conservative (56)	Benign (0.029)	Tolerated (0.625)	•		•
3	19	1043748	rs3752232	Exon 10	c.955A>G	4.50	11.3025	10.10	10.00	1.788	T319A	Moderately Conservative (58)	Benign (0)	Tolerated (0.882)	•		•
4	19	1044712	rs3764647	Exon 11	c.1184A>G	4.50	10.7896	9.70	10.00	1.788	H395R	Conservative (29)	Benign (0)	Tolerated (1)	•		•
5	19	1045026	rs10405305	Exon 12	c.1241C>G	NA	3.1471	2.70	3.00	NA	A414G	Moderately Conservative (60)	Benign (0.055)	Tolerated (0.35)			•
6	19	1045173	rs3752233	Exon 12	c.1388G>A/C	0.00	2.9837	5.90	6.00	2.263	R463H	Conservative (29)	Probably Damaging (0.997)	Tolerated (1)	•		•
7	19	1047336	rs59851484	Exon 15	c.2026G>A	NA	3.8354	3.00	3.00	NA	A676T	Moderately Conservative (58)	Probably Damaging (1.000)	Tolerated (0.55)			•
8	19	1047537	rs3752239	Exon 16	c.2153A>C	NA	2.7002	5.80	6.00	2.063	N718T	Moderately Conservative (65)	Possibly Damaging (0.529)	Tolerated (0.239)	•		•
9	19	1050996	rs74176364	Exon 19	c.2629G>A	NA	0.5167	1.50	2.00	NA	A877T	Moderately Conserved (58)		Tolerated (0.38)			•
10	19	1055191	rs3745842	Exon 30	c.4046G>A	NA	40.6467	39.50	39.00	38.375	R1349Q	Conserved (43)	Benign (0.002)	Tolerated (0.546)	•		•
11	19	1056426	rs113269196	Exon 33	c.4514G>A	NA	0.7919	0.60	1.00	NA	R1505H	Conserved (29)	Benign (0.016)	Tolerated (0.18)			•
12	19	1056492	rs3752246	Exon 33	c.4580G>C	18.30	12.7884	17.60	18.00	77.2	G1527A	Moderately Conservative (60)	Benign (0)	Tolerated (1)	•	•	•
13	19	1058176	rs4147918	Exon 37	c.5057A>G	4.50	3.014	5.60	6.00	1.638	Q1686R	Conservative (43)	Benign (0.001)	Tolerated (0.234)	•		•
14	19	1059056	rs114782266	Exon 40	c.5435G>A	NA	2.6838	2.70	3.00	2.963	R1812H	Conservative (29)	Benign (0.108)	Deleterious (0.017)	•		•
15	19	1065018	rs4147934	Exon 46	c.6133G>T	NA	25.0256	36.80	37.00	77.338	A2045S	Moderately Conservative (99)	Benign (0.057)	Tolerated (0.962)	•		•

Table 2.4

All of the variants from Appendix A with a MAF of 1% or above in at least one of the databases and coding for a predicted damaging missense mutation. These variants provide the final list of possibly functionally damaging variants and will therefore be investigated further along with 19:1056958 from Table 2.1. All cDNA changes were annotated using reference sequence NM019112.3. All MAFs are presented in percentage format.

						MAF						Affect Prediction					
	CHR	GENOMIC POSITION	dbSNP ID	GENE LOCATION	cDNA	HapMap (European) (%)	EVS (All) (%)	dbSNP (%)	1000Genomes (All) (%)	CRISP (%)	PROTEIN	Grantham Score	Polyphen2	SIFT	NGS	GWAS	EVS
1	19	1045173	rs3752233	Exon 12	c.1388G>A/C	0.00	2.9837	5.90	6.00	2.263	R463H	Conservative (29)	Probably Damaging (0.997)	Tolerated (1)	•		•
2	19	1047336	rs59851484	Exon 15	c.2026G>A	NA	3.8354	3.00	3.00	NA	A676T	Moderately Conservative (58)	Probably Damaging (1.000)	Tolerated (0.55)			•
3	19	1047537	rs3752239	Exon 16	c.2153A>C	NA	2.7002	5.80	6.00	2.063	N718T	Moderately Conservative (65)	Possibly Damaging (0.529)	Tolerated (0.239)	•		•
4	19	1059056	rs114782266	Exon 40	c.5435 G>A	NA	2.6838	2.70	3.00	2.963	R1812H	Conservative (29)	Benign (0.108)	Deleterious (0.017)	•		•

LD calculations were performed on the final five variants and the output from PLINK can be seen in Figure 2.4. However, the novel variant was not able to be included in the LD calculations at this stage due to it not being present in the 1000Genomes dataset as mentioned previously. Previous analysis has shown that D` is the most accurate way of describing LD as it does not take the allele frequencies of either loci into account (Devlin and Risch, 1995), therefore, this is the measure that will be analysed. As can be seen in Figure 2.4, two variants (rs59851484 and rs114782266) have a D` of 1, implying these loci are in complete correlation with the GWAS SNP rs3764650. However, the calculations for rs59851484 may not be totally accurate as the minor allele (A) does not appear to have been seen in the 1000Genomes dataset used to calculate these LDs (see Figure 2.4). Unfortunately, genotypes of rs3764650 were not available for these samples so the true LD was unable to be ascertained: however, if more time had been permitted, this GWAS tag SNP would also have been genotyped in order to establish the true LD.



[rs3764650 rs3752233]

R-sq = 0.000 D' = 0.197

Haplotype	Frequency	Expectation under LE
GA	0.005	0.006
TA	0.047	0.045
GG	0.116	0.114
TG	0.833	0.834

In phase alleles are GG/TA

[rs3764650 rs3752239]

R-sq = 0.000 D' = 0.197

Haplotype	Frequency	Expectation under LE
GC	0.005	0.006
TC	0.047	0.045
GA	0.116	0.114
TA	0.833	0.834

In phase alleles are GA/TC

[rs3764650 rs59851484]

R-sq = -1.000 D' = 1.000

Haplotype	Frequency	Expectation under LE
G0	0.000	0.000
T0	0.000	0.000
GG	0.121	0.121
TG	0.879	0.879

In phase alleles are GG/T0

[rs3764650 rs114782266]

R-sq = 0.002 D' = 1.000

Haplotype	Frequency	Expectation under LE
GA	0.000	0.001
TA	0.011	0.010
GG	0.121	0.119
TG	0.868	0.869

In phase alleles are GG/TA

Figure 2.4
Results of the Linkage Disequilibrium calculations preformed between the variants of interest (rs3752233, rs3752239, rs59851484 and rs114782266) and the ABCA7 intronic GWAS tag SNP rs3764650. This GWAS SNP codes for a G to T change. These present the r^2 and D` scores for each allele as well as the haplotypes observed, their frequency in the dataset used (1000Genomes Phase I) against the frequency expected under linkage equilibrium (if they were totally dependant on each other and, therefore, have identical minor allele frequencies). rs59851484's result may not be entirely representative, however, as the minor allele (A) is not present in the 1000Genomes dataset. This may because it is rarer then predicted. However, rs114782266 is predicted to be in LD with the GWAS tag SNP with a D` score of 1. The frequency of the variants presented here are in decimal format.

The location of these five variants, as well as the ABC motif's within ABCA7, can be seen in Figure 2.5, visualised on the tertiary structure of ABCA7 from TOPO2. As can be seen, there is no commonality in the location of these variants within the protein, being spaced throughout the structure. However, there are two variants within one of the extracellular domains (ECD) between transmembrane domains five and six. This may mean that this particular ECD is important in ABCA7's function. There is also one variant, rs3752233, within the first ECD which is known to be functionally important.

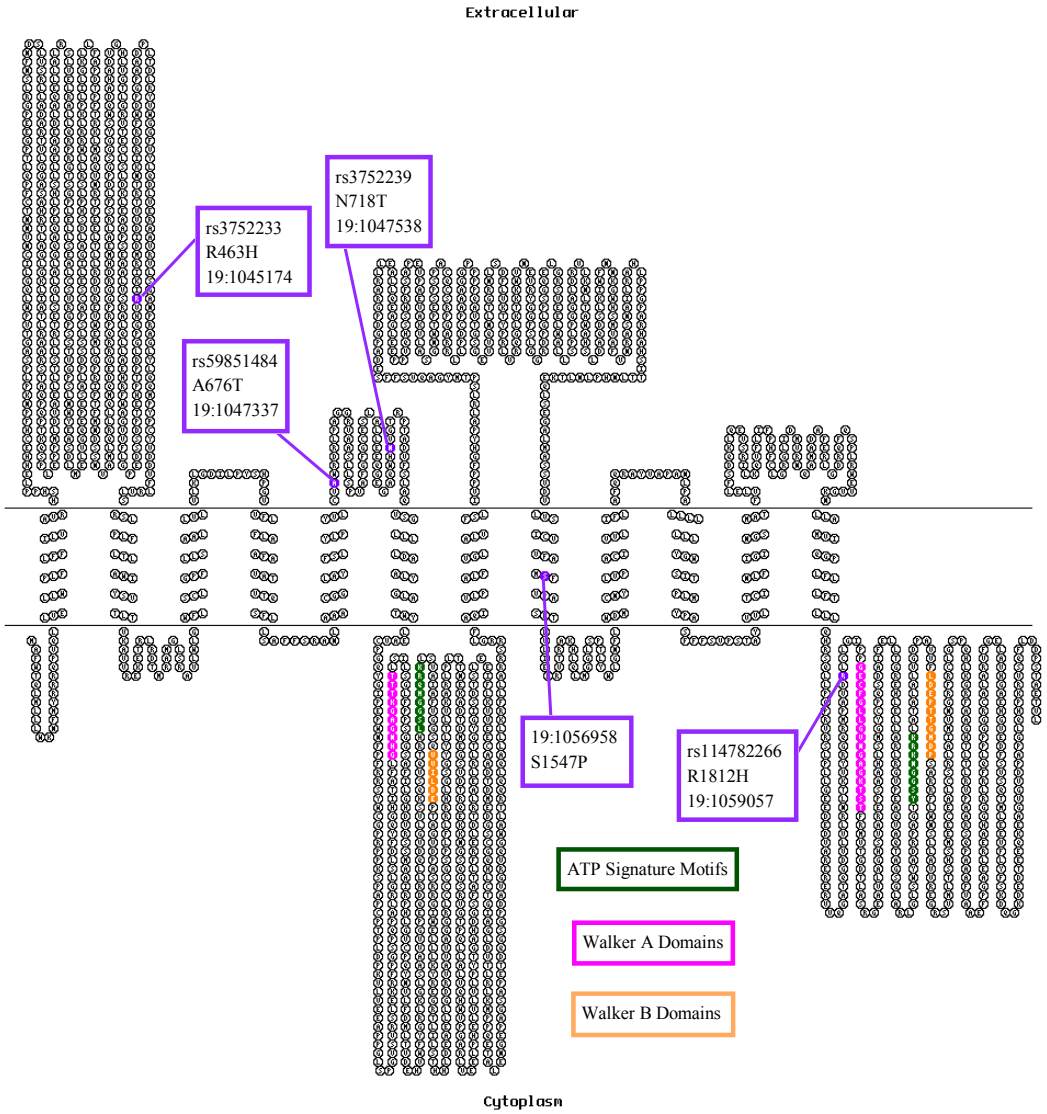


Figure 2.5

2D representation of ABCA7's structure within the cell membrane, created using TOPO2 (www.sacs.ucsf.edu/TOPO2/). The variants which will be examined in more detail (rs3752233, rs5981484, rs3752239. 19:1056958 and rs117482266) are highlighted along with the amino acid they affect. The three motifs vital to ABCA function are also highlighted which are located within the Nucleotide Binding Domains (NBDs). These areas are vital to any ABCA protein's function, primarily because they are where ATP binds to and becomes hydrolysed in order to provide energy to transfer molecules across the membrane through a pore created by the transmembrane domains (TMDs).

As can be seen in Figure 2.5, the novel variant 19:1056958 is located within transmembrane domain eight and was therefore analysed to see if it alters the transmembrane domains through analysis on the TMHMM server. The results from these are available in Figure 2.6 and suggest that this variant does not alter the overall structure of the transmembrane domains.

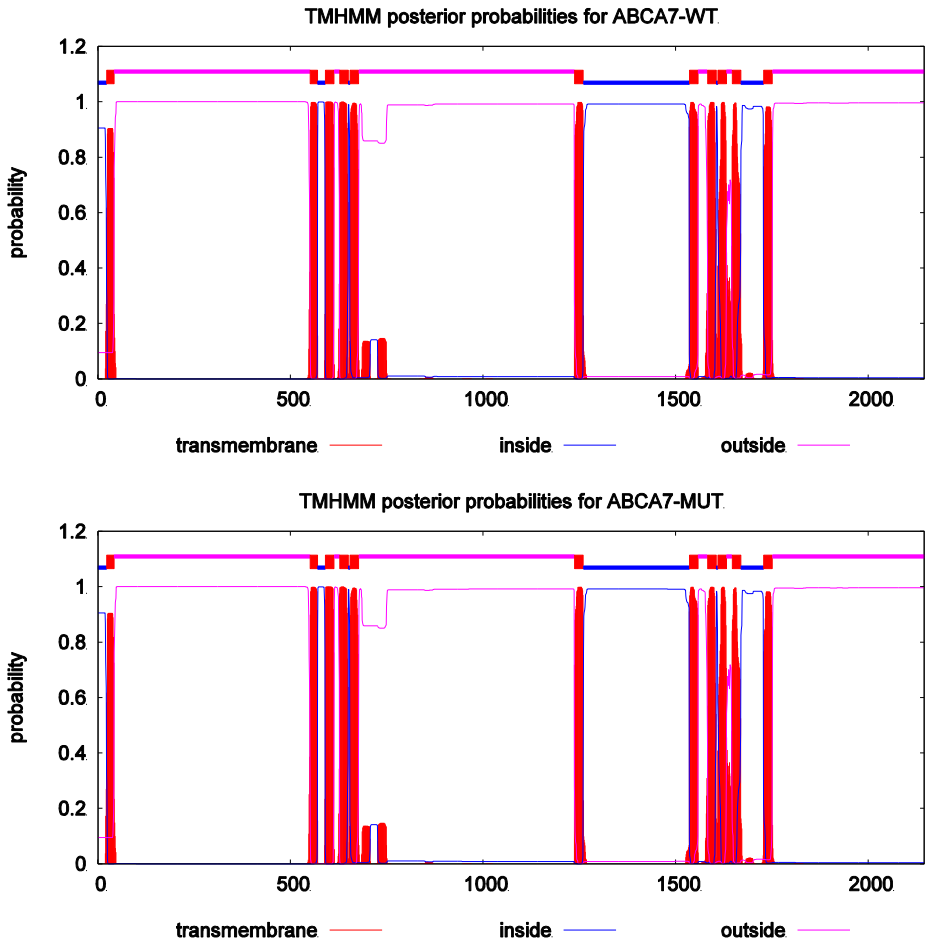


Figure 2.6
These images contain the output from the TMHH server version 2.0 (http://www.cbs.dtu.dk/services/TMHMM/), predicting the transmembrane domains in the wild type amino acid sequence of ABCA7 (ABCA7-WT) as well as the amino acid sequence where position 1547 is altered from Serine to Proline (ABCA7-MUT). This change is caused by a mutation at genomic position 19:1056958 with a T to C change. These graphs show the probability of certain regions being transmembrane domains. There is no difference between the two meaning that this variant does not appear to alter the transmembrane domain overall structure, however, it may still alter the pore structure within the cell membrane and may, therefore, still alter ABCA7's transporter function.

As the locations of the functionally important domains in ABCA7 were mapped to its amino acid sequence in TOPO2, a list of variants in these areas was also created. This included ten variants in all six domains and can be seen in Table 2.5.

Table 2.5

Variants within functionally important areas of *ABCA7* . These ten variants are spread between all six of the motifs found in all ABCA proteins. These motifs (Walker A, Walker B and the ATP motif or ABC signature sequence) are vital to all the ABCA proteins' functions as they bind and hydrolyse ATP to ADP, providing energy for the transporter to transport substrates across the cell membrane. If any of these variants alter this process, they will cause loss of function to the protein. As can be seen, eight of the ten presented variants are predicted to be damaging to the protein function by at least one of the three prediction programmes used. All cDNA changes were annotated using reference sequence NM019112.3. All MAFs are presented in percentage format.

							MAF						Affect Prediction					
		CHR	GENOMIC POSITION	dbSNP ID	GENE LOCATION	cDNA	HapMap (European) (%)	EVS (All) (%)	dbSNP (%)	1000Genomes (All) (%)	CRISP (%)	PROTEIN	Grantham Score	Polyphen2	SIFT	NGS	GWAS	EVS (E/A)
1	Walker A	19	1049410	rs368179689	Exon 18	c.2526C>T	NA	0.0077	NA	NA	NA	H842H			Tolerated (1.00)			•
2		19	1049414	rs371982554	Exon 18	c.2530G>C	NA	0.0077	NA	NA	NA	G844R	Moderately Radical (125)	Probably Damaging (0.995)	Damaging (0.00)			•
3	ATP Motif	19	1051293	rs200140277	Exon 20	c.2824G>C	NA	0.0154	NA	NA	NA	G942R	Moderately Radical (125)	Probably Damaging (1.000)	Damaging (0.00)			•
4	Walker B	19	1051505	rs148962540	Exon 21	c.2882T>C	NA	0.0077	NA	NA	NA	V961A	Moderately Conserved (64)	Probably Damaging (1.000)	Damaging (0.00)			•
5	Walker A	19	1059079	rs143614132	Exon 40	c.5458G>A	NA	0.0384	NA	NA	NA	G1820S	Moderately Conserved (56)	Probably Damaging (0.994)	Tolerated (0.07)			•
6		19	1061804	rs78320196	Exon 41	c.5487T>C	NA	3.9142	4.30	4.00	2.2	N1829N			Tolerated (1)	•		•
7		19	1061813	-	Exon 41	c.5497del	NA	1.9652	NA	NA	NA	K1833Rfs*	FRAMESHIFT					•
8	ATP Motif	19	1063612	rs148635111	Exon 43	c.5782G>A	NA	0.0077	NA	NA	NA	G1928R	Moderately Radical (125)	Probably Damaging (1.000)	Damaging (0.00)			•
9		19	1063624	rs114787084	Exon 43	c.5794C>T	NA	0.0617	0.09	<1.00	NA	R1932C	Radical (180)	Probably Damaging (1.000)	Damaging (0.00)			•
10	Walker B	19	1063775	rs373283316	Exon 44	c.5864G>T	NA	0.008	NA	NA	NA	G1955V	Moderately Radical (109)	Probably Damaging (1.000)	Damaging (0.00)			•

2.4 Discussion

From this bioinformatical interrogation of data acquired from our own NGS and from data freely available on the EVS, a list of 240 exonic variants within *ABCA7* was reduced to a final list of five variants. Four of these are predicted to be damaging to the protein and have a minor allele frequency of at least 1%. The final variant on this list is a novel variant arising from the NGS data. Its MAF is therefore unknown (CRISP approximates it at 0.663%) but it is predicted to be functionally damaging by both the PolyPhen2 and SIFT programmes.

The methodology behind this approach was validated by a colleague - Dr Christopher Medway in the Mayo Clinic, Jacksonville. He had also, independently, interrogated the EVS database, examining it for variants likely to be associated with Late Onset Alzheimer's Disease in *ABCA7*. The approach he used, which involves examining haplotypes of different variants and therefore not just focussing on the variants predicted to be damaging to the protein function, produced a list of 27 variants. All of the variants in Table 2.4 are included in this list. The remaining 23 variants are all included in Appendix A, with the exception of two variants in the 5' untranslated region (UTR) area, one in the 3' UTR and two in intronic regions. Of the remainder, four have a MAF of less than 1%, nine are predicted to not be damaging to the protein structure or function and one is a synonymous amino acid substitution. Four in total are novel variants, all seen in Table 2.1. However, none are seen in our NGS data. Nevertheless, the fact that the variants in Table 2.4 are also included in Dr Medway's list corroborates the approaches used here.

When the catalogue was initially refined, eliminating variants with a MAF of less than 1%, this reduced it drastically, removing 204 variants. The vast majority of these were present on EVS but not in any of the other databases analysed, such as 1000Genomes

or dbSNP. The variants within EVS have not all been validated, including all of the insertion/deletion (INDEL) calls reported. This implies that, when they do not also appear in any of the other databases, they may not be true variations, including the novel variants only seen in EVS, such as the majority of the variants in Table 2.1. All of these novel variants were not seen in HapMap, dbSNP or 1000Genomes and only two of the 11 were seen in our NGS data. As it is unlikely for novel variants only present in the EVS database to also be present in the samples held in this lab, only ones present in the NGS dataset were examined in more detail. Of these, two possible variants existed although only 19:1056958 is predicted to be damaging to the protein and therefore, this one has been interrogated further. The removal of such a high number of rare variants increases the chances that any singleton coding variants which may contribute to risk would be missed. Methods of examining these rare variants exist, for example, burden analysis of rare variant load programmes. However, in order to carry out these tests, full genetic data of each individual is required in order to "collapse" all the rare variants into one score and test association between this score and the disease trait (S. Lee et al., 2014). This data was not available for any of the study subjects included in this analysis at the time this project was done and, therefore, this option could not be explored.

When looking purely at the NGS dataset, a 1% cut off for MAF seems extremely low, especially as only 96 individuals were sequenced. However, when these samples are supplemented with the data from EVS, 1000Genomes and dbSNP, the results may be more reliable, increasing the power. Indeed, some of the variants present in Appendix A with a MAF of less than 1% have been shown to be real and in association with disease. However, both of these studies were performed in much bigger datasets, massively increasing the power and, therefore, the ability to reliably establish the association between these rare variants and disease (Dykxhoorn, 2016; Xia, 2016).

This project is specifically looking at variants which affect the structure and function of the ABCA7 protein. However, intronic variants - as well as ones in the UTR of the gene - cannot be ignored. Variants in these areas have been linked to many things affecting the protein, mainly splicing and regulation of mRNA, and therefore protein levels (Pagani and Baralle, 2004). Annotation of these variants has proven to be very difficult, especially due to the extremely high numbers of these variants (the NGS project produced a list of 96 intronic SNPs tentatively associated with LOAD, a great deal more than the 38 exonic ones). Synonymous variants have also been excluded from analysis in this process. However, again, they can similarly alter the protein. This can also be through affecting splicing but may again alter transcription, mRNA transport and translation, rendering these so-called "silent mutations" far from silent (Goymer, 2007). Frequently, more data than just genomic data is required to annotate these variants, both intronic and synonymous. For example, transcriptomic and proteomic data is needed in order to analyse the functional significance of these variants on protein levels and alternate isoforms, far beyond the scope of this project (Ritchie et al., 2015). However, this type of analysis is going to become more and more vital as the amount of genomic data available increases. As many of the phenotypes under investigation are complex, and therefore arise from interactions among multiple factors, the complete biological model is only going to become apparent if all of these factors are considered within the analysis. The automatic elimination of intronic and synonymous variants is, therefore, a drawback of this analysis.

As mentioned previously, commonly the GWAS tag SNP is not the disease-causing variants, but frequently it is a variant in LD with it (Rosenthal et al., 2014; Schaub et al., 2012). In Figure 2.4, it can be seen that two of the variants under investigation, rs114782266 and possibly rs59851484, are in strong LD with the GWAS tag SNP rs3764650. rs59851484 is only possibly associated as its' minor allele is not seen in

the dataset, therefore the calculations may not be representative. This might imply that this variant may actually be rarer than first thought and is over-represented in the datasets used to obtain its MAF. This is supported by the fact that this is the only potentially pathogenic variant that does not appear in the NGS data. The other two variants examined, rs3752233 and rs3752239, both have a D` of 0.197 implying that these loci have been separated from rs3764650 through recombination. However, D` values of less than 1 are more difficult to interpret and therefore no conclusions as to the levels of recombination can be made from this data (Barnes, 2007). Unfortunately, the genotype of rs3764650 is not known for the majority of the samples that will be used in the next chapter so the accuracy of these *in silico* LD calculations could not be assessed. However, a future project is planned to acquire these genotypes, as well as the other GWAS tag SNPs listed in Table 1.2 through genotyping studies, in order to evaluate the true LD scores.

The location of all the variants in Table 2.4, as well as the novel variant 19:1056958, can be seen in Figure 2.5, as well as the location of the functionally important areas mentioned in the ABCA7 protein. The location of variant 19:1056958 is of particular interest as it is located in one of the transmembrane domains of the ABCA7 protein, which are highly conserved both within the *Homo Sapien* ABCA family and between species, forming the pore the protein transports substrates through. This variant was therefore examined in further detail in order to see if is predicted to alter the physical structure of the transmembrane domains as there is a possibility a substitution in this region could affect the formation of the α-helix necessary to form the transmembrane domain (Kobayashi et al., 1990). From Figure 2.6 it can be seen that an alteration of amino acid number 1547 from Serine to Proline does not grossly affect the structure of the transmembrane domains in ABCA7. The programme predicts 11 transmembrane domains, one fewer than the ABCA proteins are known to have, eliminating transmembrane domain six. However this is consistent in both the wild

type and mutant versions of the protein. The fact that this variant introduces a serine to proline substitution is of particular interest as this particular kind of change is known to produce kinks in alpha-helical structures (such as the ones which make up the TMDs in the ABC proteins) (Barnes, 2007). Consequently, even if this variant does not alter the physical transmembrane structure, it may still alter the way substrates are transported through it by altering the pore conformation. An example of this is mutations in TMD 10 of the ABCC1 protein (a multidrug resistance protein) altering substrate specificity (Zhang et al., 2006a). The most accurate way of assessing this in the ABCA7 protein would be by performing functional assays, evaluating ABCA7's transport function in both wild type and mutant (containing the S1547P substitution) versions of the protein.

The location of variant rs3752233 is also of interest as it is within the first ECD. Variants within this region in other ABCA proteins, such as *ABCA1*, are known to cause disease, such as Tangier's disease, and therefore mutations in the same area in *ABCA7* may cause similar disease causing changes (M. L. Fitzgerald et al., 2004).

Only two of the variants in Table 2.4 have been reported in the literature previously and, as mentioned earlier, the variant at position 19:1056958 has not been reported before. rs3752233 and rs3752239 were both reported in the same study examining common coding variants within LOAD GWAS loci, as this chapter does specifically for *ABCA7* (Holton et al., 2013). However, although this report identifies these two variants as having been investigated within this study, the results for these individual variants are not presented and, therefore, the findings are not obvious (Holton et al., 2013). When examining the location of all of these variants in the ABCAM database (a mutation knowledge base mining all of the ABC protein mutations from available literature and aligning them between the members of the family (Gyimesi et al.,

2012)), it becomes apparent that there are no reported variants in any of these locations in other ABCA protein in the available literature.

When specifically looking at variants which are in known functionally important areas of the ABCA7 protein, ten become apparent, available in Table 2.5. The majority of these are predicted to be damaging to the protein. However, all apart from one of these are only present in the EVS database and may, therefore, not be real. Only one variant in this list, rs78320196 a synonymous mutation, is present in the NGS dataset. One variant - the novel one at co-ordinates 19:1061813 - is also included in Dr Medway's list of potentially damaging variants. All of these variants have the potential to be extremely damaging to ABCA7's function as a transporter if they are, in fact, genuine. Variants within these regions are known to cause disease in other ABC genes. For example, variants in the first Nucleotide Binding Domain (containing the first Walker A, B and ATP Signature Motifs) in the *ABCA4* gene are known to cause Stargardt's eye disease and in the *CFTR* gene (also known as *ABCC7*), Cystic Fibrosis (Sabirzhanova et al., 2015). The problem with the majority of these variants is that they are very rare with only two have a MAF of above 1%, implying that further analysis of these variants may be problematic.

In relation to further functionally significant areas, recently a loss-of-function mutation in the *ABCA1* gene was linked to low plasma levels of apoE and an increased risk of Alzheimer's disease (Nordestgaard et al., 2015). The location of this mutation in the ABCA1 protein (N1800H) can be seen in Figure 2.7 (created following the same methodology as in Section 2.2.3) and, when compared to the location of variants shown in Figure 2.5, it can be seen that none of the variants of interest map to a similar region (the intracellular domain between transmembrane domains 10 and 11). However, the *ABCA1* variant is very rare (seen in 1 in every 500 individuals) and, when the full catalogue of variants (in Appendix A) is examined,

there are three variants within this intracellular domain which may, or may not, have the same loss-of-function affect as the N1800H *ABCA1* variant. These three variants (19:1057946, rs141237099 and rs371036349) are also all very rare (all with a MAF of below 0.5%) and are all predicted to be damaging to the ABCA7 protein function. Therefore, these variants may prove to be potentially rare, loss-of-function mutations within *ABCA7*. However, sufficient power (i.e. increased patient samples) would be required in order to substantiate this hypothesis, something which is not achievable currently.

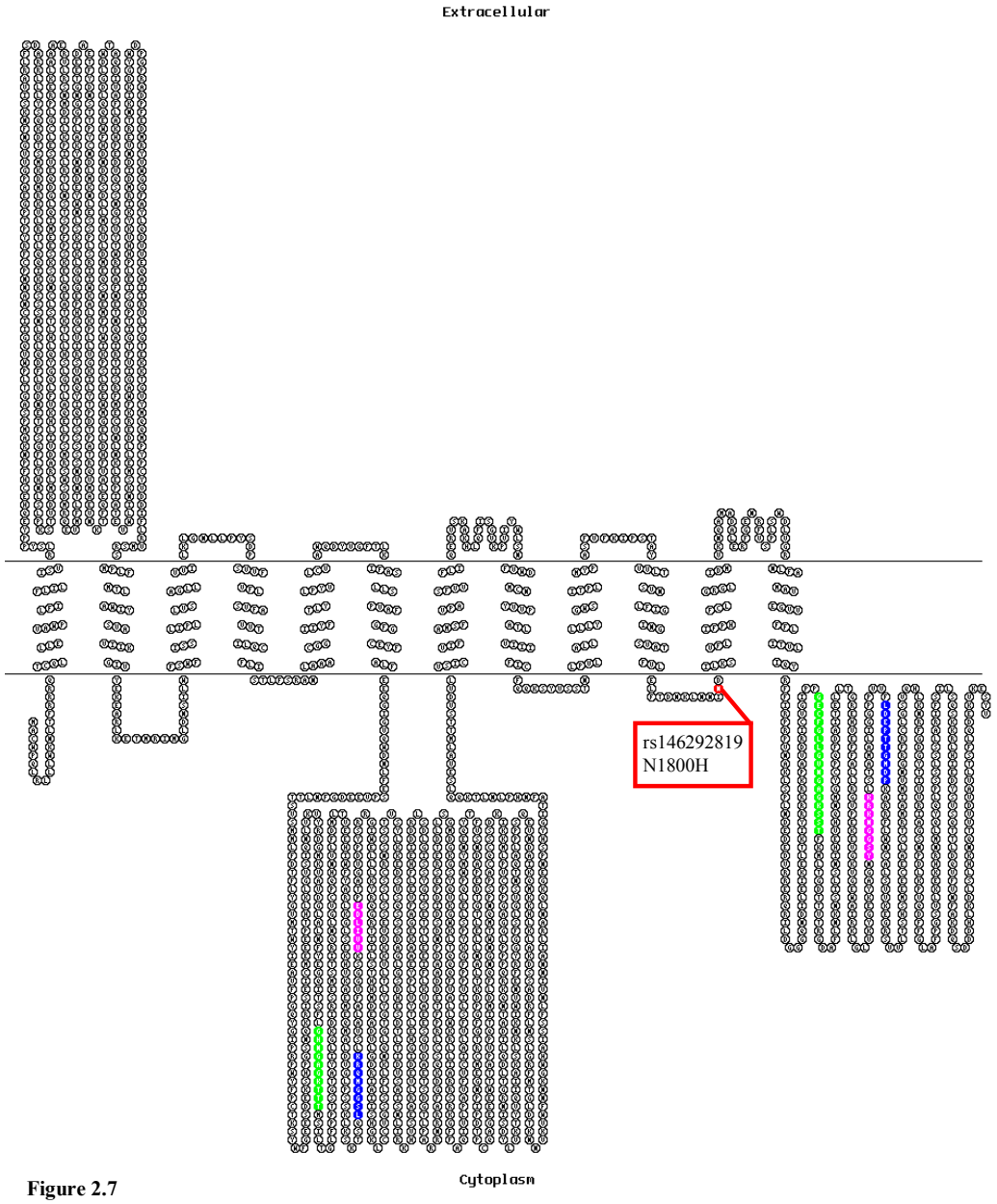


Figure 2.7

2D representation of ABCA1's structure within the cell membrane, created using TOPO2 (www.sacs.ucsf.edu/TOPO2/) and the transmembrane domain information retrieved from UniProt (http://www.uniprot.org/ - ID O95477) and the three motifs as in Figure 2.5. The variant highlighted in red can be seen in the intracellular region between the transmembrane domains 10 and 11. This is reported to be a loss-of-function variant, increasing the risk for Alzheimer's disease (Nordestgaard *et al*, 2015). However, none of the variants which have been highlighted in Figure 2.5 are in the equivalent region in the ABCA7 protein.

In order to examine whether the final variants selected have an effect on the ABCA7 protein, and therefore alter an individual's risk of Late Onset Alzheimer's disease, they must be first scrutinized to see if they have a putative association with disease. This will initially be done for the four variants in Table 2.4 (rs3752233, rs3752239, rs114782266 and rs59851484) as well as the novel variant 19:1056958 as these are all predicted to be damaging to the ABCA7 protein. If this genotyping does show an association of LOAD, functional experiments will then be designed in order to see the precise affect these variants do have.

The variants in Table 2.5 and those present in the same intracellular domain as the loss-of-function *ABCA1* variant would be investigated in an ideal world. However, with their frequency being as low as it is, a much bigger DNA bank then the one currently available would be necessary in order to perform any association tests with meaningful power. Furthermore, only one of the variants (rs78320196) has been seen in the NGS dataset and this is one of the few variants in this sub-catalogue which is not predicted to be damaging to protein function. This makes it increasingly unlikely (especially given their MAF) that any of these variants will be seen in the samples available to us.

2.5 Conclusions

In conclusion, the variants enumerated in Table 2.4 as well as 19: 1056958 from Table 2.1 will be genotyped in both Late Onset Alzheimer's disease cases and control samples in order to establish any potential association with disease.

3 Case-Control Genotyping of Putative Damaging *ABCA7* Variants

3.1 Introduction

As shown in Chapter 2, there are five variants which, when analysed employing bioinformatic tools, have been shown to be potentially damaging to the function of the ABCA7 protein. Four of these variants (rs114782266, rs3792233, rs3792239 and rs59851484) are relatively common with all of them having a minor allele frequency (MAF) of 2% or more in the full contingency of databases examined, as well as being anticipated to be harmful to the protein function by at least one of the prediction tools used. The fact that these variants are all common increases the likelihood of power being achieved to detect any association with LOAD in the genotyping assays which will be carried out in this chapter.

An additional variant is also predicted to be damaging to the ABCA7 protein: present at genomic position 19:1056958, this variant was highlighted in the Next Generation Sequencing (NGS) project performed in 2011. It is assumed to be much rarer than the previous variants mentioned as, not only is it novel and therefore not been catalogued previously, but it was also present only in one of the 96 samples which underwent NGS and even then in a heterozygous fashion. This would provide it with an estimated MAF of 0.663% (as calculated by the CRISP programme, employed to analyse the NGS data). However, it is also in a functionally important area of ABCA7 – one of the transmembrane domains – and will therefore still be investigated further.

In order to ascertain their association with disease, all of these variant will therefore be genotyped; both in cases (confirmed LOAD patients) and controls. This will inform as to whether the variants are more frequent in one of these two groups and may, therefore, cause a change in the ABCA7 protein which is associated with disease (if more frequent in the disease group) or protective against LOAD (if more frequent in the control group). The samples used are part of the Alzheimer's Research UK (ARUK) DNA Bank collected from a variety of centres (Belfast, Ireland; Bonn, Germany; Bristol; Leeds; Manchester; Nottingham; Southampton and Oxford, UK), all collected with ethical approval and consent of all individuals. All variants will be genotyped in 1984 samples (1013 case samples and 971 control samples). However, further samples are held of which stocks are low of (a total of 624 cases and 462 controls from Bristol; Oxford and Southampton, UK) and will therefore only be used to genotype the rarer variant (19:1056958). For all samples, the following demographics are held: age of AD diagnosis (age of sampling is used for controls); sex; *APOE* allele status and the centre from which they were collected. These samples, having been extracted from brain and/or blood and stored at -80°C, are stored in 96-well plates at 20ng per well in a 2µl volume for use in genotyping assays.

In order to carry out the genotyping, Kompetitive Allele Specific Polymerase chain reaction (KASP™, LGC Genomics, Middlesex, UK) assays were designed for each of the five variants. KASP™ assays were utilised, as opposed to the more commonly used TaqMan® assays, as a result of comparison studies performed in this laboratory: contrasting the genotyping results achieved for both TaqMan® and KASP™ assays for the same variant (Braae et al., 2014). The KASP™ assay was shown, not only to be significantly cheaper to purchase and run, but also generated improved clustering of the genotype groups with fewer samples failing. This has also been repeated with other assays, including one of the assays discussed here: rs59851484. This was initially run as a TaqMan® assay by a student (Ahmed Alahmad). However, it failed

to distinguish the positive heterozygous and homozygous mutant controls as separate clusters. When a KASP™ assay was designed instead, the clusters became much clearer and more distinguishable.

The KASP™ Genotyping system comprises of the KASP™ Assay Mix and KASP™ Master Mix. Within the assay mix are two competitive allele specific forward primers and one common reverse primer. Incorporated into the two forward primers is one of two florescent resonance energy transfer (FRET) cassettes specific to either the FAM or HEX dye depending on which allele the primer is specific for. Within the KASP ™ master mix are the two FRET cassettes (FAM and HEX combined with a quencher molecule) as well as the reference dye (ROX), Taq polymerase, free nucleotides and a commercially optimised concentration of $MgCl_2$ (0.2mM). During the thermocycle, the DNA under investigation is denatured and the forward primer relevant for its allele anneals. Upon the next denaturing step, the appropriate FRET cassette can then anneal to the DNA on the complementary sequence to the forward primer, causing the cassette to fluoresce at a specific frequency to the dye in question, allowing the genotype of each individual sample to be distinguished when read on a qPCR system. This methodology is summarised in Figure 3.1.

All five assays were performed; four by students: rs3752239 and rs114782266 were carried out by Akili Mata, rs59851484 by Ahmed Alahmad and 19:1056958 by Sara Garin as part of their Masters projects. Their results will be presented as part of this chapter.

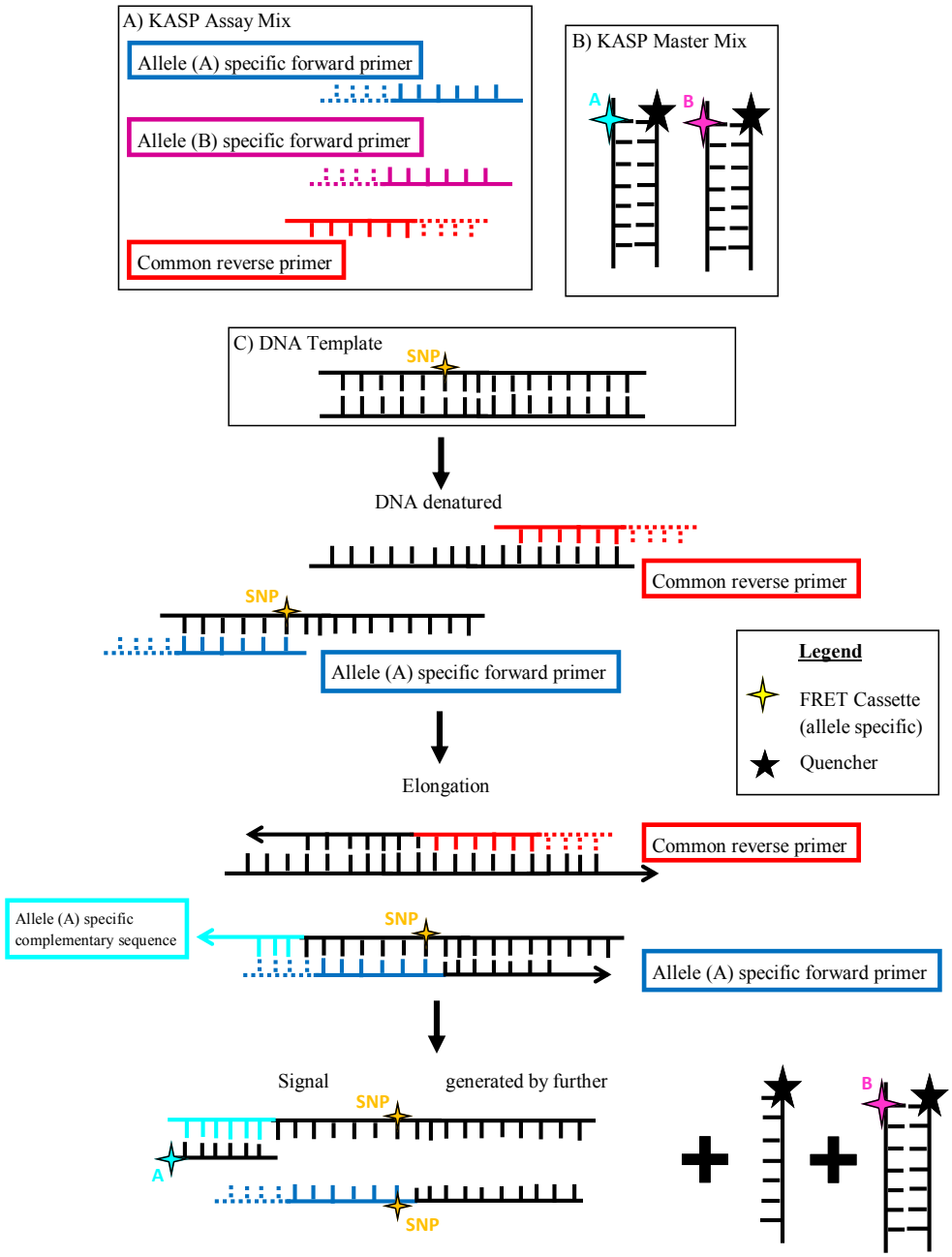


Figure 3.1

Methodology of the KASP™ genotyping assays. This is based on competitive allele-specific PCR. Components include the KASP™ assay mix, containing two allele specific forward primers and one common reverse primer, and the KASP™ master mix containing universal FRET cassettes, one labelled with the FAM dye and one with the HEX dye. During PCR thermal cycling, the relevant allele-specific primer binds to the DNA template and elongates. The primer also contains a tail sequence which corresponds to the FRET cassette which, upon subsequent thermocycles, then becomes incorporated into the DNA sequence. As these are amplified, the FRET cassette can bind to the sequence complimentary to this tail sequence, separating the dye from the quencher so it can therefore fluoresce. In this example, allele B is not present so this florescent signal remains quenched.

3.2 Materials & Methods

3.2.1 Identification of Positive Controls

Positive controls for all of the variants were required in order to ensure the genotyping assays were calling the genotypes correctly. Alongside No Template Controls (NTCs) these were used to clarify that the clusters within the genotyping assays were being formed accurately. A colleague (Dr Christopher Medway) developed a Perl script which imputed 1000Genomes and GWAS data of samples held by the lab, in order to identify samples which carried the heterozygous genotype for each of the variants, using the Genomes Phase 1 integrated variant set (v3) as a reference panel. This was performed for all four of the known variants. For the novel variant, the sample which was identified to carry the variant during the NGS study was used as the positive control. All of these samples were held as part of the ARUK DNA Bank with the exception of the samples for the variant rs59851484. Samples could not be identified within the ARUK DNA Bank which carried this variant so DNA was obtained from the Mayo Clinic, Jacksonville which had also genotyped and identified samples carrying this variation. The samples used, as well as the genotypes for the relevant variations and all demographics held for the samples, are shown in Table 3.1. All control samples (those with no clinical disease) were population controls that donated blood samples with a wide age range – from 38 to 100.

Table 3.1
Sample demographics of the samples used as controls in genotyping assays. The variants the samples were used for are shown, as well as the major (Ref) and minor (Alt) alleles for each of these variants. The genotype assigned to each sample through Sanger sequencing is shown as well as all demographics held within the ARUK DNA database. The positive control samples used for the rs59851484 were acquired from the Mayo DNA bank (Jacksonville, USA), due to no positive controls being found in the ARUK DNA database. NA = data not available.

Variant	Ref./Alt.	Sample ID	Genotype	Sex	Centre of Origin	Age at Death	Age at Onset	Disease Status	ApoE Status
19:1056958	T/C	M642	T C	Female	Manchester	66	NA	Confirmed AD	3 3
rs3752239	A/C	AD295	A C	Female	Belfast	NA	NA	Confirmed AD	3 3
		M544	A A	Female	Manchester	64	58	Confirmed EOAD	3 4
rs59851484	G/A	3356	A A	Female	Mayo, USA	NA	86	Confirmed AD	3 2
		11219	G A	Female	Mayo, USA	NA	88	Confirmed AD	3 3
		M134	G G	Male	Manchester	NA	57	Probable EOAD	3 3
rs3752233	G/A	AD111	G G	Male	Nottingham	88	NA	Confirmed AD	3 3
		M233	G A	Male	Manchester	NA	58	Probable EOAD	2 3
rs114782266	G/A	AD202	G G	Female	Nottingham	71	NA	Confirmed AD	3 4
		BN0636	G A	Male	Bonn, Germany	NA	NA	Control	3 3

3.2.2 Polymerase Chain Reaction Primer Design

The positive controls identified above were Polymerase Chain Reaction (PCR) amplified and sequenced in order to ensure they did indeed carry the single nucleotide polymorphism (SNP) of interest. Primers were designed to include approximately 200 base pairs (bp) both upstream and downstream of all the variants, resulting in an amplicon of at least 400bp for each assay. The sequences were obtained from UCSC Genome Browser (UCSC Genome Informatics Group, Santa Cruz, California, USA) and verified on the Ensembl Genome Browser (EBI, Cambridge, UK), release 74. These sequences were inputted into Primer3 (University of Tartu, Tartu, Estonia) to provide primer sequences. The UCSC *in-silico* PCR program (UCSC Genome Informatics Group) was used to ensure the specificity of the selected primers. The SNPcheck software (NGRL, Manchester, UK) was then used to check there were no SNPs in the primer binding sites. These were all performed in July 2014 and are shown in Table 3.2 along with the amplicon size achieved and the annealing temperature used during PCR. These primers will also be used to validate the genotypes called in the KASP™ Assays.

Table 3.2
Table showing the primer sequences, amplicon size and annealing temperatures used for the PCR amplification of all variants.

Primer	Sequence	Amplicon Size (base pairs)	Annealing Temperature (°C)
ABCA7_novel_Forward	CATCCTCATCCCACCAACCT	494	58
ABCA7_novel_Reverse	TCAAGACTTTACCCAGGCCA		
ABCA7_rs3752239_Forward	GGCCTACTTCTCCCTCTACC	425	61
ABCA7_rs3752239_Reverse	CACCCCTAGCTCAGCCTC		
ABCA7_rs59581484_Forward	GGTGACCCAGAGCTTCCTG	421	65
ABCA7_rs59581484_Reverse	ACACAGCTTCCAGGTACCAG		
ABCA7_rs3752233_Forward	CAGTGCCTGTCCTTGGACA	370	62
ABCA7_rs3752233_Reverse	TGGCTACCATACCACTGTTGA		
ABCA7_rs114782266_Forward	GGGGATGTGTTGGTGCTG	299	62
ABCA7_rs114782266_Reverse	AGTAGGTGCCCAATAAAAGGTG		

3.2.3 Polymerase Chain Reaction

The amplification reactions for rs114782266, rs3792233 and 19: 1056958 consisted of 1x Roche PCR Buffer (consists of 100mM Tris-HCl, 15mM $MgCl_2$, 500mM KCl at a pH of 8.3 - Roche Diagnostics, Basel, Switzerland), 0.2mM dNTPs, 0.5pmol/µl of the relevant forward and reverse primers (designed previously), 1 unit (U) of Taq DNA Polymerase (Roche Diagnostics) and 10ng of the DNA before being made up to 30µl with dH_2O. For the PCR involving the assay rs3752239, 1x of Q solution was also added (a DNA additive which modifies the melting behaviours of DNA in order to facilitate the amplification of problematic DNA regions, such as GC-rich areas, QIAGEN, Limburg, Netherlands). For rs59851484, a hot start Taq was instead used as well as a different buffer (containing ammonium sulphate) in order to fluidise the DNA, increasing primer binding: 1x Fermentas Taq Buffer with $(NH_4)_2SO_4$ (750mM Tris-HCl, 200mM $(NH_4)_2SO_4$, pH 8.8 - Thermo Scientific, Pittsburgh, PA, USA), 2mM $MgCl_2$, 0.2mM dNTPs, 0.5pmol/µl of each primer, 1x Q solution and 2.5 units of Hot Start Taq (QIAGEN). The thermal cycles performed for each of the assays can be seen in Table 3.3. All were performed in a Veriti 96 Well Thermal Cycler (Applied Biosystems, Carlsberg, CA, USA). PCR products were examined by electrophoresis at 80 volts (V) for 35 minutes in a 1% agarose gel (30ml 1x TAE buffer (40mM Tris, 20mM acetic acid, and 1mM EDTA), 0.3g of peqGOLD Universal Agarose (PEQLAB, Wilmington, DE, USA) and 5µl ethidium bromide) in 1x TAE buffer alongside a 1kbp plus DNA ladder (Thermo Scientific). The electrophoresis gel was photographed under ultraviolet (UV) light to ensure only one product was present of the correct size.

Table 3.3

Polymerase Chain Reaction (PCR) thermocycle parameters used to amplify and confirm the positive controls for the KASP™ genotyping assays. A proportion of samples genotyped were also amplified and Sanger sequenced utilising these conditions in order to validate that the assays were clustering the haplotypes correctly.

	19:1056958			**rs3752239**			**rs59851484**		
	Temperature (°C)	Time		Temperature (°C)	Time		Temperature (°C)	Time	
Initial	94	2 minutes		94	2 minutes		95	15 minutes	
Denaturation	94	30 seconds	} x30	94	30 seconds	} x30	94	30 seconds	} x30
Annealing	58	30 seconds		61	30 seconds		65	45 seconds	
Extension	72	1 minute		72	1 minute		72	1 minute	
Final Extension	72	7 minutes		72	7 minutes		72	10 minutes	
Hold	10	∞		10	∞		10	∞	

	rs3752233			**rs114782266**		
	Temperature (°C)	Time		Temperature (°C)	Time	
Initial	94	2 minutes		94	2 minutes	
Denaturation	94	30 seconds	} x30	94	30 seconds	} x30
Annealing	62	30 seconds		62	30 seconds	
Extension	72	1 minute		72	1 minute	
Final Extension	72	7 minutes		72	7 minutes	
Hold	10	∞		10	∞	

3.2.4 Sanger Sequencing

The PCR products were first purified using ExoSAP-IT (Affimetrix, Santa Clara, California, USA); Exonuclease I (ExoI) to remove primers and Shrimp Alkaline Phosphatase (SAP) to dephosphorylate excess dNTPs. This was done in a final volume of 7µl, containing 5µl of PCR product and 2µl of ExoSAP-IT. This reaction mix was incubated in the thermal cycler at 37°C for 15 minutes followed by another 15 minutes at 80°C to inactivate the enzymes.

The sequencing reaction took place in a 10µl reaction volume consisting of 5µl of the ExoSAP product, 1x BigDye Terminator Sequencing Buffer (Applied Biosystems), 2µl BigDye (contains DNA polymerase, deoxynuleotides and fluorescently labelled dNTPs), and 0.5pmol/µl forward primer. A thermal cycle was performed consisting of 25 cycles of a denaturation step at 96°C for 30 seconds, an annealing step at 50°C for 15 seconds and an extension step at 60°C for 4 minutes. This was performed in a Veriti 96 Well Thermal Cycler.

The sequencing products were purified following the manufacturer's protocol for the Performa DTR Gel Filtration Cartridges (EdgeBio, Gaithersburg, MD, USA): centrifuging the cartridges at 800 x g for 3 minutes before adding the Sanger sequencing products. The cartridges were again centrifuged at 800 x g for 3 minutes, filtering the products through the gel in order to remove unincorporated BigDye, dNTPs and salts. The elute was dried out at 90°C before being read by capillary electrophoresis on a 3130 Genetic Analyser (Applied Biosystems) provided by the Molecular Diagnostics Lab.

The sequences were received as chromatogram files and analysed using the Chromas Lite software (Technelysium, South Brisbane, Australia), aligning the sequences with the reference sequence (obtained from UCSC when the primers were designed) using

the Clustal Omega online tool (EMBL-EBI, Wellcome Trust, Cambridge, UK). This aligned sequence was examined in order to elucidate their genotype for the variants of interest.

3.2.5 KASP™ Assay Design

The surrounding sequences for each of the variants (obtained from UCSC Genome Browser in July 2014) were submitted to LGC Genomics (Hoddesdon, UK) who designed and validated the allele specific primers. These primers, as well as version 4.0 of the low ROX KASP™ Master Mix, were then received from them. The sequences designed are shown in Table 3.4.

Table 3.4

Primer sequences designed for use in the KASP™ genotyping assays. The FAM primers are specific to the major allele and contain the FRET cassette for the FAM dye. All assays also contain a common reverse primer in order to allow product amplification.

ID	Primer_AlleleFAM	Primer_AlleleHEX	Primer_Common	AlleleFAM	AlleleHEX
ABCA7_Novel	CATCTGTGTGGTCTTTGCCATGT	CATCTGTGTGGTCTTTGCCATGC	GAGTGAAGCTGGCCGGGACAAA	T	C
ABCA7_rs3752239	AGGGCGCGCAGTGGCACAA	GGGCGCGCAGTGGCACAC	AGAGACCTGGGCCAGGCTGAA	A	C
ABCA7_rs59851484	GCCCTACGTGCTGTGTGTGG	GCCCTACGTGCTGTGTGTGA	GGGCAGCCGGTCCCGCCAAG	G	A
ABCA7_rs3752233	CAATGTCCATGCGGATTTTGATGC	GTCAATGTCCATGCGGATTTTGATGT	CCAGACCTGGGCCCCGGC	G	A
ABCA7_rs114782266	GAGGATGCCAGCTGTTGACCG	AGAGGATGCCAGCTGTTGACCA	AGGGGGAATCCCCAGGCACAA	G	A

3.2.6 KASP™ Assays

Each assay was performed initially with only 12 samples (including the previously identified positive controls and two NTCs). Each reaction contained 1x KASP™ Master Mix, 0.11µl KASP™ Primer mix and 20ng of each DNA sample. These were made up to 8µl with dH_2O and then the following thermocycle was performed on a Veriti 96 Well Thermal Cycler: 15 minutes at 94°C for an initial incubation followed by 10 cycles of 94°C for 20 seconds, 65-55°C for 60 seconds (dropping 0.6°C every cycle) then 26 cycles of 94°C for 20 seconds and 55°C for 60 seconds before being held at 10°C. Each run could then be re-cycled (at 94°C for 10 seconds then 57°C for 60 seconds) depending on how the genotype clusters appear when they are visualised.

In order to visualise the clusters, the samples were read on a qPCR machine (Mx3000P, Stratagene, San Diego, California, USA) using the MxPro Software. An allele discrimination program was set up with all samples being set as "unknown genotype" with the exception of the NTCs which were set as such. The fluorescent dyes were set as HEX (minor allele) and FAM (major allele) for bi-allelic discrimination and ROX as the internal control reference dye. The samples were read and then analysed in the Rn-post setting, generating a dual colour scatter plot. The allele clusters were set manually with the FAM axis displaying the major allele genotype and HEX the minor allele.

Once the 12 initial samples had been run, they were also PCR amplified and Sanger sequenced, as described in Sections 3.2.3 and 3.2.4, in order to ensure the KASP™ assay distinguished the genotypes correctly. Once this was confirmed, 96-well assay plates were run. All assays were run on at least 1984 samples, as previously mentioned, as well as a subsequent 1086 samples being run for the variant at position 19:1056958. All case samples had been diagnosed as probable or possible AD following the NINCDS-ADRDA or the CERAD guidelines.

3.2.7 Validation of Genotyping Results

On average, approximately 3% of the samples genotyped were also PCR amplified and Sanger sequenced in order to validate the genotypes generated. This was performed utilising the conditions described previously in Sections 3.2.3 and 3.2.4. Generally samples which clustered closer to the positive control were sequenced (in order to validate their carrying a minor allele) as well as a "control" sample from the homozygous wild-type cluster.

3.2.8 Statistical Analysis of Genotype Results

Results from each genotyping assay were subsequently analysed using the command line program PLINK (Purcell et al., 2007), performing allelic association tests with both Fisher's exact test and logistic regression analysis. The latter also allows for covariate correction (in this case age, sex, centre of sample origin and *APOE* status). In order to analyse large datasets, PLINK accepts data in tab delimited file in the format of .map, .ped and .covar files. The .map files contain the details of all SNPs being analysed – a unique identifier (for example an rs number), the chromosome and base pair co-ordinates of the variant and the genetic distance specified in Morgans. The latter is set to zero in all cases described here as population separation was not being analysed. The .ped file contains the family ID and individual ID for all samples (these are identical in the population of samples utilised here), paternal and maternal ID (again, set to zero in all samples here as no family analysis was performed), sex (1 was set to male, 2 was female), phenotype (1 was control samples, 2 was case samples) and the genotype of all variants listed in the corresponding .map file (where 2 is the major allele and 1 is the minor allele so a 2 1 genotype would be heterozygote and 1 1 would be homozygous for the minor allele). For all of these fields an entry of 0 or -9 meant the information was missing. The .covar file again contains the family and individual IDs, in order to map the samples to the ones contained in the .ped file,

as well as including the phenotype and sex using the same coding as previously described. Also included in this file were the centres the samples originate from (for coding used here see Table 3.5), age of AD diagnosis (in cases) or age of sampling (in controls) and number of ApoE4 alleles carried (0, 1 or 2 – established through two TaqMan® assays for variants rs429358 and rs7412, performed when the samples were first incorporated into the ARUK DNA Bank). Examples of all of these files can be seen in Figure 3.2.

Code	Centre
1	Bonn, Germany
2	Leeds, UK
3	Nottingham, UK
4	Manchester, UK
5	Belfast, Ireland
6	Oxford, UK
7	Southampton, UK
8	Mayo Clinic, USA
9	Bristol, UK

Table 3.5
Table showing the codes used for the geographical origin of samples used in this genotyping project. This data is part of the .covar file (see Figure 3.2) in order to correct for sample origin.

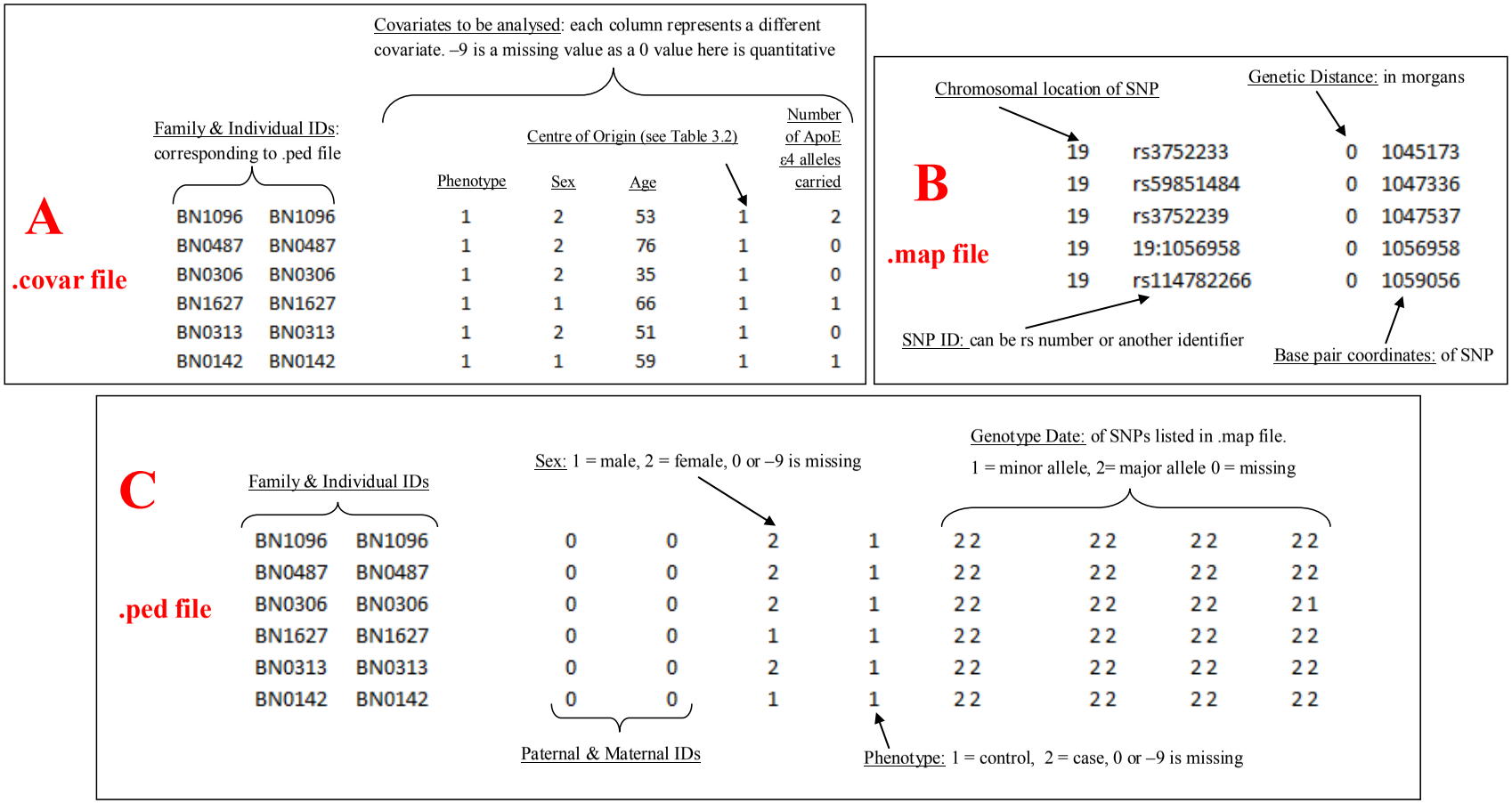


Figure 3.2

Examples of the three files used in order to statistically analyse the genotyping results in PLINK. The .covar file (A) contains data on the covariates corrected for in the Logistic regression test, the .map file (B) contains the details of the variants being tested and the .ped file (C) contains the genotypic data for all variants for each individual sample.

.map, .ped and .covar files were therefore created for all of the variants before Fisher's exact test and logistic regression analysis (both with and without the covariant correction) were performed on all data. The following commands were inputted into the command prompt in order to perform these respective tests:

- → plink --map *mapfilename* --ped *pedfilename* --fisher --allow-no-sex --ci 0.95
- → plink --map *mapfilename* --ped *pedfilename* --logistic --allow-no-sex --ci 0.95
- → plink --map *mapfilename* --ped *pedfilename* --covar *covarfilename* --logistic --allow-no-sex --ci 0.95

The --allow-no-sex flag was also included, allowing the program to include in the analysis the individuals who were missing gender data. As this was a large number of individuals, this increased the power of the association tests. A command was also included (--ci) in order to include confidence intervals of 95% in the output.

For the Fisher's exact test a file entitled plink.assoc.fisher was generated as the output and for the logistic regression test, the output was a file entitled plink.assoc.logistic. These were then examined as text files in order to establish the association of each of the individual variants with LOAD.

In order to increase power further (as the variants under investigation may not be as common as stated in databases) a further .ped file was obtained from colleagues in the Mayo Clinic, Jacksonville (Dr Christopher Medway and Professor Steven Younkin) who, as well as corroborating the bioinformatics approaches used here (see Chapter 2) also genotyped all of the variants (with the exception of 19:1056958) in their much larger cohort. This provided the genotypes of a further 7972 samples (4246 controls and 3726 cases). Statistical analysis was performed on just the Nottingham samples as well as a merged dataset using the command line prompts previously stated.

3.3 Results

Positive controls for four out of the five assays were identified. These were confirmed through PCR amplification and Sanger sequencing. However, for one of the assays (rs59851484) a positive control within the DNA samples help in this lab could not be identified. Fortunately this variant had also been genotyped in the Mayo Clinic, Jacksonville by colleagues (Dr Christopher Medway and Professor Steven Younkin). Heterozygous and homozygous mutant positive controls were therefore obtained from them and confirmed again through Sanger sequencing.

As the KASP™ assays were run; genotype plots were created by the MxPro Software for each 96 well plate. Examples of these plots, for the rs3752239 and 19:1056958 variations, can be seen in Figures 3.3 and 3.4 respectively. The homozygous wildtype samples are highlighted on the FAM axis (blue circles), homozygous mutant samples on the HEX axis (red diamonds), heterozygous samples on the scale between the two (green squares) and no template controls or failed samples at the meeting of the two axis (yellow triangles). As can be seen, all assays provide clear, distinct clusters for the three genotypes. Samples which fail to amplify appear as a red cross (for example the samples clearly distinct from any genotype cluster in Figure 3.3) or within the NTC cluster. The total number of samples successfully genotyped for each assay is shown in Table 3.6. This also presents the success rate of each of the assays, showing that no more than a maximum 5.2% of samples failed in any of the assays. Most of these failures are presumed to be due to poor quality DNA as the majority of these samples also failed to amplify in PCR as shown in Figure 3.5 which demonstrates one of the PCRs performed to confirm the genotypes called by all of the five KASP™ assays.

Table 3.6

Table showing the number of Nottingham samples successfully genotyped utilising the KASP™ assays (N). Success rate when compared to overall number of samples genotyped is also presented. As can be seen, all assay returned an impressive success rate, all being over 95% with the exception of 19:1056958. However, this is believed to be because this assay was run on extra plates which were known to contain poorer quality DNA.

Assay	N	Success Rate (%)
rs3752233	1933	97.53
rs3752239	1919	96.00
rs114782266	1937	97.20
rs59851484	1934	97.58
19:1056958	2912	94.82

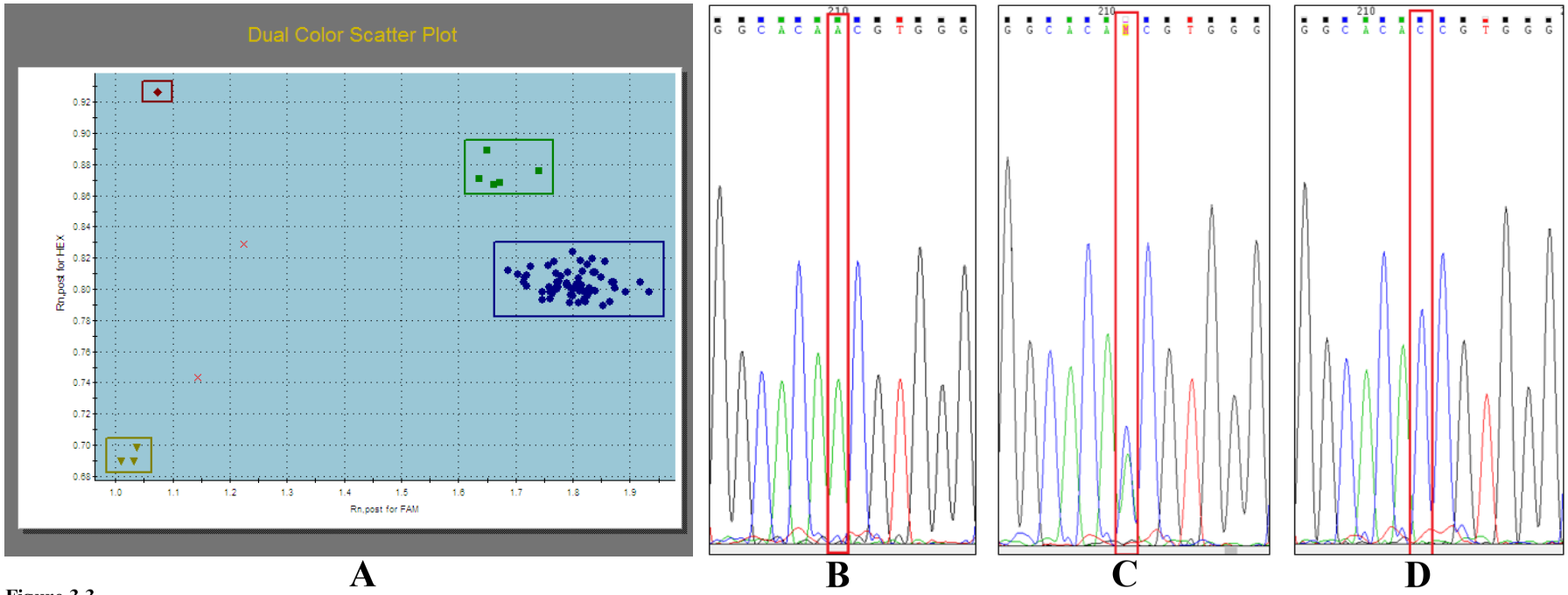


Figure 3.3

(A) KASP^{TM} dual colour scatter plot showing the distinct genotype clusters for the variant rs3752239. The homozygous wild type samples return a strong FAM signal (blue) and the sequence of one of these samples can be seen in (B) with the location of the variant highlighted in red. This sample is homozygous for the major allele (A). Heterozygous samples return a combination of FAM and HEX signal and can be seen in green in (A). One of the samples in this cloud was run as a positive control and the Sanger sequence of one of the other samples can bee seen in (C) where an A and C signal is seen at the location of the variant. Homozygous mutant samples are shown in the red cloud in (A) and the confirmation electropherogram in (D), shows this sample is homozygous for the minor allele (C). Also seen in (A) is the group of No Template Controls returning no florescence signal. This group (as well as the red crosses seen as outliers) also contained some DNA samples which failed to amplify.

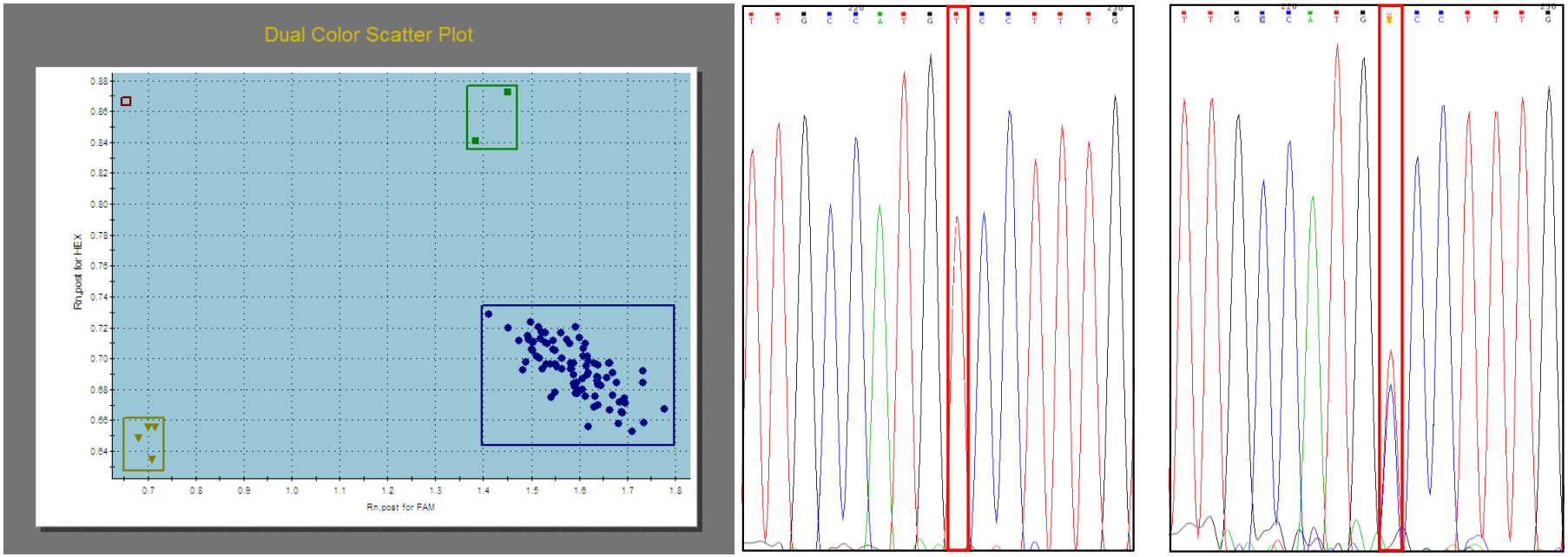


Figure 3.4

(A) KASP^{TM} dual colour scatter plot showing the distinct genotype clusters for the novel variant at genomic position 19:1056958. No homozygous mutant samples were identified for this variant. However, on this plate one heterozygous sample was identified (green—the other is a positive control). The Sanger sequence confirming this can be seen in (C) with the electropherogram of one of the samples in the wild type cluster (blue in (A)) can be seen in (B). Again, the group of samples which failed to amplify, as well as the No Template Controls, can be seen in the bottom left of the scatter plot.

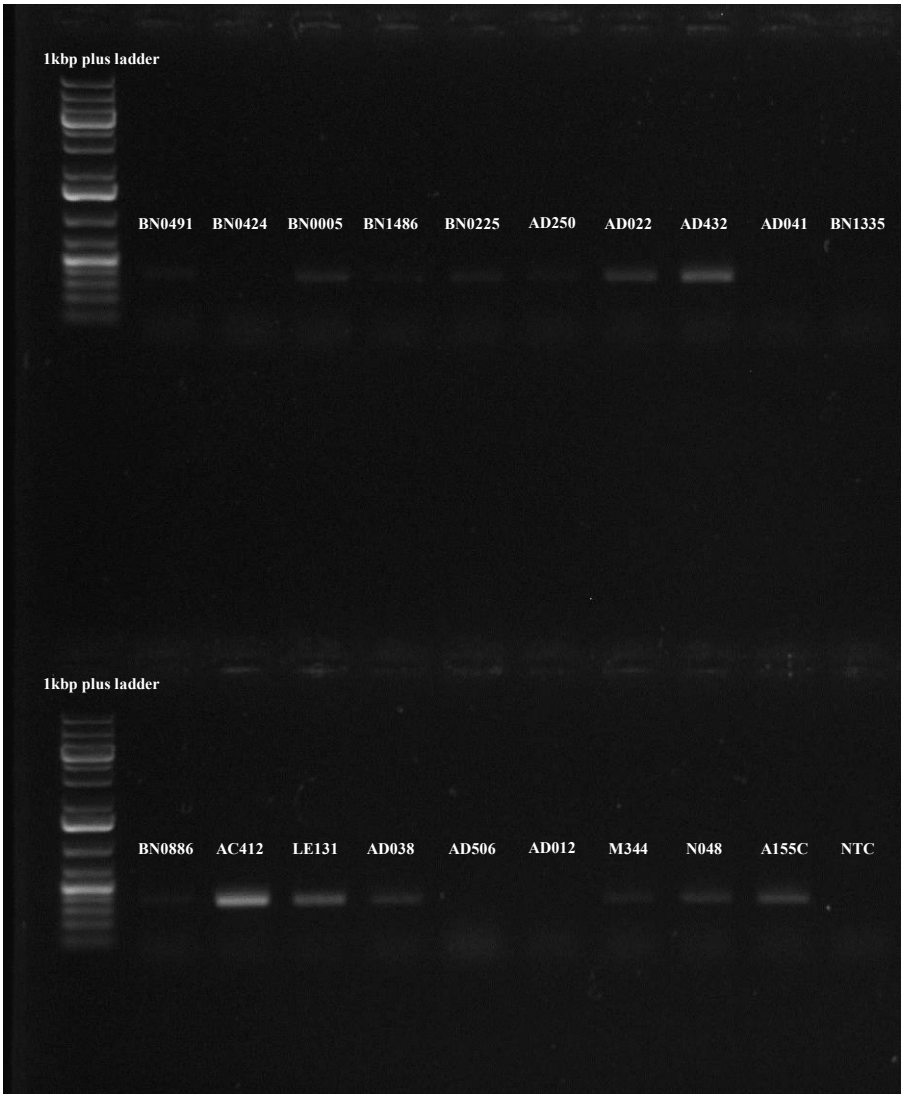


Figure 3.5

1% ethidium bromide stained agarose gel run at 80V for 45 minutes. Sample IDs for each lane are shown as well as the No Template Control (NTC) and the 1kbp ladder they were run against. These samples were genotyped for the variant rs3752233 and were subsequently PCR amplified and Sanger sequenced in order to validate the genotypes called in the KASP™ assays. Some of the samples (BN0424, AD041, BN1335, AD506 and AD012) failed to amplify in these assays and were therefore PCR amplified to attempt to acquire a genotype. However, as can be seen here, they also failed to PCR, suggesting that the DNA for these samples was degraded. The remaining samples all produced a band of the correct size (370bp), despite some staining faintly on the gel, and were Sanger sequenced to confirm the genotypes called in the KASP™ assays, demonstrating the high accuracy of this assay. This was repeated for the other four variants also genotyped.

For rs3752239, 1919 samples were successfully genotyped with four homozygous mutant and 134 heterozygous samples being identified. The success rate of this genotyping assay was 96% and around 3% of samples were sequenced to verify the results, providing an accuracy of 97.7%. For rs114782266 1937 samples were genotyped with a 97.2% success rate, identifying one homozygous mutant sample and 20 heterozygous samples. 2.7% of samples were sequenced in order to validate the results providing an accuracy of 92.3% for this genotyping assay. For rs3752233 1933 samples were successfully genotyped giving a success rate of 97.53%. From these, 144 heterozygous and five homozygous mutant samples were identified. For rs59851484 1934 samples were successfully genotyped but no minor alleles were found in either case or control samples. For the variant at genomic position 19:1056958, 2912 samples were successfully genotyped. However there was a higher failure rate of 94.8% due to the poorer quality of the DNA of the extra samples genotyped for this variant. In total, five heterozygous samples were identified for this variation.

The results outputted by PLINK can be seen in Tables 3.7-3.11. As can be seen, when the statistical tests are run only with "Nottingham" samples, none of the tests return significant results. However, when the American samples are merged with the Nottingham samples, a few of the statistical tests do return a *p*-value of less than 0.05. For rs3752239, both the Fisher's exact and logistic regression test produce *p*-values < 0.05 - highlighted in red in Table 3.8. The odds ratios (OR) and confidence intervals (CIs) for these tests also reinforce these *p*-values (both OR = 0.86 (0.74-0.99)). In Table 3.10 it can be seen that, although no minor alleles for rs59851484 were found in the Nottingham samples whatsoever (hence the statistical tests on these samples returning a "NA" result - see Table 3.10), when this data is combined with the American samples, this variant appears to be associated with the phenotype. Both Fisher's exact test and the logistic regression analysis result in *p*-values of less than

0.02 and 0.03 respectively. The odds ratios and confidence intervals also support this (both OR = 0.28 (0.09-0.83)). However, when these tests take into account covariates, these tentatively associated results do not withstand. Correction for multiple testing should also be considered. As five variants are being examined, a *p*-value of 0.01 should be used to signify a significant result, therefore, none of these results withstand this correction. Also of interest in Table 3.11 is the difference of the minor allele frequency between cases (F_U) and controls (F_U) for the novel variant 19:1056958. Although this test is not significant there is a large difference in the MAFs between these two groups (MAF of 0.13% in cases and 0.04% in controls).

Table 3.7

Statistical results from the rests run on rs3752233 at genomic position 19:1045174. N = number of non-missing individuals used in analysis (for example, some may be removed from analysis due to incomplete covariant data), F_A = frequency of the minor allele in cases, F_U = frequency of minor allele in controls, STAT = coefficient t-statistic (in Logistic Regression test), P = exact *p* -value or asymptomatic *p* -value for t-statistic (in logistic regression tests only), OR = Odds Ratio, L95 = Lower 95% Confidence Interval, U95 = Upper 95% Confidence Interval. As can be seen, none of these tests returned a *p* -value of < 0.05.

Statistical Test	Samples Used	Covariates	N(cases)	N(controls)	F_A	F_U	STAT	P	OR	L95	U95
Fisher's Exact	Nottingham	-	1630	1403	0.03691	0.04184	-	0.4567	0.8775	0.6342	1.214
Fisher's Exact	Nottingham & Mayo	-	5356	5649	0.03717	0.04221	-	0.07718	0.876	0.7577	1.013
Logistic Regression	Nottingham	-	1630	1403	-	-	-0.7783	0.4364	0.8805	0.6392	1.213
Logistic Regression	Nottingham	ALL	1484	1338	-	-	3.50E-17	1	1	0	inf
		Phenotype	1484	1338	-	-	0.03356	0.9732	2.61E+18	0	inf
		Sex	1484	1338	-	-	3.52E-16	1	1	0	inf
		Age	1484	1338	-	-	9.79E-16	1	1	3.74E-50	2.68E+49
		Centre	1484	1338	-	-	5.16E-17	1	1	1.87E-307	5.34E+306
		ApoE4 Dosage	1484	1338	-	-	6.36E-17	1	1	0	inf
Logistic Regression	Nottingham & Mayo	-	5356	5649	-	-	-1.79	0.07352	0.8758	0.7575	1.013
Logistic Regression	Nottingham & Mayo	ALL	5208	5580	-	-	-0.05631	0.9551	0.0001466	5.14E-138	4.19E+129
		Phenotype	5208	5580	-	-	0.1221	0.9028	7.62E+29	0	inf
		Sex	5208	5580	-	-	0.04539	0.9638	1.70E+05	2.48E-221	1.17E+231
		Age	5208	5580	-	-	-0.6962	0.4863	0.9137	0.7088	1.178
		Centre	5208	5580	-	-	0.06337	0.9495	59.29	8.59E-54	4.09E+56
		ApoE4 Dosage	5208	5580	-	-	1.364	0.1725	7.458	0.4157	133.8

Table 3.8

Statistical results from the rests run on rs3752239 at genomic position 19:1047538. N = number of non-missing individuals used in analysis (for example, some may be removed from analysis due to not having all of the covariant data present), F_A = frequency of the minor allele in cases, F_U = frequency of minor allele in controls, STAT = coefficient t-statistic (in Logistic Regression test), P = exact *p* -value or asymptomatic *p* -value for t-statistic (in logistic regression tests only), OR = Odds Ratio, L95 = lower 95% Confidence Interval, U95 = upper 95% Confidence Interval. The tests returning a *p* -value of 0.05 or less are highlighted in red although these results do not withstand correction for covariants or multiple testing.

Statistical Test	Samples Used	Covariates	N(cases)	N(controls)	F_A	F_U	STAT	P	OR	L95	U95
Fisher's Exact	Nottingham	-	1630	1403	0.03347	0.04126	-	0.2325	0.8045	0.5755	1.125
Fisher's Exact	Nottingham & Mayo	-	5356	5649	0.03716	0.04302	-	0.04022	0.8584	0.7429	0.9919
Logistic Regression	Nottingham	-	1630	1403	-	-	-1.269	0.2045	0.8053	0.5764	1.125
Logistic Regression	Nottingham	ALL	1484	1338	-	-	3.86E-15	1	1	0	inf
		Phenotype	1484	1338	-	-	0.03348	0.9733	2.61E+18	0	inf
		Sex	1484	1338	-	-	-8.36E-17	1	1	0	inf
		Age	1484	1338	-	-	-1.19E-16	1	1	3.39E-50	2.95E+49
		Centre	1484	1338	-	-	1.37E-16	1	1	3.00E-308	3.33E+307
		ApoE4 Dosage	1484	1338	-	-	-2.19E-16	1	1	0	inf
Logistic Regression	Nottingham & Mayo	-	5356	5649	-	-	-2.069	0.03852	0.8585	0.743	0.992
Logistic Regression	Nottingham & Mayo	ALL	5208	5580	-	-	-0.04154	0.9669	8.49E-05	6.60E-197	1.09E+188
		Phenotype	5208	5580	-	-	0.1203	0.9042	2.48E+29	0	inf
		Sex	5208	5580	-	-	0.04533	0.9638	1.70E+05	1.22E-221	2.37E+231
		Age	5208	5580	-	-	-0.6873	0.4919	0.9151	0.7106	1.178
		Centre	5208	5580	-	-	0.06487	0.9483	52.87	4.59E-51	6.09E+53
		ApoE4 Dosage	5208	5580	-	-	1.356	0.1751	7.449	0.4087	135.8

Table 3.9

Statistical results from the rests run on rs114782266 at genomic position 19:1059057. N = number of non-missing individuals used in analysis (for example, some may be removed from analysis due to not having all of the covariant data present), F_A = frequency of the minor allele in cases, F_U = frequency of minor allele in controls, STAT = coefficient t-statistic (in Logistic Regression test), P = exact *p* -value or asymptomatic *p* -value for t-statistic (in logistic regression tests only), OR = Odds Ratio, L95 = lower 95% Confidence Interval, U95 = upper 95% Confidence Interval. As can be seen, none of these tests returned a *p* -value of <0.05.

Statistical Test	Samples Used	Covariates	N(cases)	N(controls)	F_A	F_U	STAT	P	OR	L95	U95
Fisher's Exact	Nottingham	-	1630	1403	0.004555	0.007376	-	0.2984	0.6157	0.2659	1.426
Fisher's Exact	Nottingham & Mayo	-	5356	5649	0.006786	0.007062	-	0.8623	0.9608	0.6833	1.351
Logistic Regression	Nottingham	-	1630	1403	-	-	-1.135	0.2563	0.614	0.2645	1.425
Logistic Regression	Nottingham	ALL	1484	1338	-	-	-7.74E-17	1	1	0	inf
		Phenotype	1484	1338	-	-	0.03354	0.9732	2.61E+18	0	inf
		Sex	1484	1338	-	-	-2.61E-16	1	1	0	inf
		Age	1484	1338	-	-	-6.63E-16	1	1	1.94E-50	5.16E+49
		Centre	1484	1338	-	-	-2.08E-17	1	1	3.38E-307	2.96E+306
		ApoE4 Dosage	1484	1338	-	-	-5.13E-17	1	1	0	inf
Logistic Regression	Nottingham & Mayo	-	5356	5649	-	-	-0.2311	0.8172	0.9605	0.6823	1.352
Logistic Regression	Nottingham & Mayo	ALL	5208	5580	-	-	-0.03499	0.9721	0.0007856	9.75E-178	6.33E+170
		Phenotype	5208	5580	-	-	0.115	0.9085	2.88E+28	0	inf
		Sex	5208	5580	-	-	0.04463	0.9644	1.71E+05	2.61E-225	1.12E+235
		Age	5208	5580	-	-	-0.6822	0.4951	0.9158	0.7113	1.179
		Centre	5208	5580	-	-	0.06482	0.9483	48.43	5.42E-50	4.33E+52
		ApoE4 Dosage	5208	5580	-	-	1.352	0.1764	7.453	0.4052	137.1

Table 3.10

Statistical results from the rests run on rs59851484 at genomic position 19:1047337. N = number of non-missing individuals used in analysis (for example, some may be removed from analysis due to not having all of the covariant data present), F_A = frequency of the minor allele in cases, F_U = frequency of minor allele in controls, STAT = coefficient t-statistic (in Logistic Regression test), P = exact *p* -value or asymptomatic *p* -value for t-statistic (in logistic regression tests only), OR = Odds Ratio, L95 = lower 95% Confidence Interval, U95 = upper 95% Confidence Interval. As can be seen, no minor alleles for this variant were found in the Nottingham samples therefore all results return a "NA" result, although some of the combined data sets return a *p* -value of less then 0.05 and these are highlighted in red. However, as shown, these results do not withstand correction for covariants or multiple testing.

Statistical Test	Samples Used	Covariates	N(cases)	N(controls)	F_A	F_U	STAT	P	OR	L95	U95
Fisher's Exact	Nottingham		1630	1403	0	0	-	NA	NA	NA	NA
Fisher's Exact	Nottingham & Mayo		5356	5649	0.0004284	0.001547	-	0.01394	0.2767	0.09247	0.8279
Logistic Regression	Nottingham	-	1630	1403	-	-	NA	NA	NA	NA	NA
Logistic Regression	Nottingham	ALL	1484	1338	-	-	NA	NA	NA	NA	NA
		Phenotype	1484	1338	-	-	NA	NA	NA	NA	NA
		Sex	1484	1338	-	-	NA	NA	NA	NA	NA
		Age	1484	1338	-	-	NA	NA	NA	NA	NA
		Centre	1484	1338	-	-	NA	NA	NA	NA	NA
		ApoE4 Dosage	1484	1338	-	-	NA	NA	NA	NA	NA
Logistic Regression	Nottingham & Mayo	-	5356	5649	-	-	-2.299	0.02151	0.2764	0.09233	0.8273
Logistic Regression	Nottingham & Mayo	ALL	5208	5580	-	-	-0.001635	0.9987	1.18E-06	0	inf
		Phenotype	5208	5580	-	-	0.1145	0.9088	9.01E+26	0	inf
		Sex	5208	5580	-	-	0.045	0.9641	1.70E+05	2.95E-223	9.74E+232
		Age	5208	5580	-	-	-0.6856	0.493	0.9153	0.7106	1.179
		Centre	5208	5580	-	-	0.06457	0.9485	47.52	6.03E-50	3.74E+52
		ApoE4 Dosage	5208	5580	-	-	1.357	0.1749	7.546	0.4071	139.9

Table 3.11

Statistical results from the tests run on the variant at genomic position 19:1056958. N = number of non-missing individuals used in analysis (for example, some may be removed from analysis due to not having all of the covariant data present, F_A = frequency of the minor allele in cases, F_U = frequency of minor allele in controls, STAT = coefficient t-statistic (in Logistic Regression test), P = exact *p* -value or asymptomatic *p* -value for t-statistic (in logistic regression tests only), OR = Odds Ratio, L95 = lower 95% Confidence Interval, U95 = upper 95% Confidence Interval. For this variant there was no American cohort with which to merge the Nottingham genotyping data and, as can be seen, this did not return results with *p* -values of <0.05, likely due to rarity of the variant and the relatively small sample size.

Statistical Test	Samples Used	Covariates	N(cases)	N(controls)	F_A	F_U	STAT	P	OR	L95	U95
Fisher's Exact	Nottingham	-	1630	1403	0.001293	0.000366	-	0.3797	3.533	0.3946	31.63
Logistic Regression	Nottingham	-	1630	1403	1	-	1.129	0.2589	3.536	0.3947	31.67
Logistic Regression	Nottingham	ALL	1484	1338	1	-	4.99E-14	1	1	0	inf
		Phenotype	1484	1338	1	-	0.0416	0.9668	2.61E+18	0	inf
		Sex	1484	1338	1	-	-8.88E-16	1	1	0	inf
		Age	1484	1338	1	-	3.13E-16	1	1	1.71E-41	5.86E+40
		Centre	1484	1338	1	-	-1.57E-15	1	1	1.64E-164	6.12E+163
		ApoE4 Dosage	1484	1338	1	-	1.52E-15	1	1	0	inf

3.4 Discussion

This chapter has examined the relationship between five variants, predicted to be damaging to the function of the ABCA7 protein, and the late onset Alzheimer's disease phenotype. Despite three of the variants (rs3752239, rs59851484 and 19:1056958) returning tentative results, none of the statistical tests performed survived correction for covariates or multiple testing. There was an extremely high success rate of genotyping these variants (all above 94%) and the failures among these are presumed to be due to poor quality DNA as the majority failed to amplify through PCR as well. The genotyping assays were also confirmed through Sanger sequencing a selection of these samples, confirming highly accurate genotyping assays. Sanger sequencing was utilised as it is still considered the gold standard of sequencing and therefore ascertaining genotypes (Wetterstrand, 2014). One caveat to this study was that all control samples used were population controls and therefore they may have gone onto developing AD after donation of their DNA.

In order to ascertain the associations, two statistical tests were performed: Fishers exact test and Logistic Regression analysis. Fisher's exact test is a straight forward association test whereas Logistic Regression analyses the dosage of the minor allele, not just the presence of it taking into account as to whether the individual is heterozygous or homozygous for this minor allele. The Logistic Regression test can also be corrected for covariates which, in this case were sex, age of diagnosis (in the case samples) or age of sampling (in the control samples), centre of origin (see Table 3.5 for a list of these) and, most importantly in this phenotypic analysis, ApoE4 allele dosage. In terms of inheritance patterns, the majority of minor alleles seen were carried in as compound heterozygotes. Only rs37552233 and rs3752239 were seen in a homozygous manner, to be expected as these are the two commonest variants when looking at the MAFs in dbSNP and 1000Genomes. However, even then, only 15 and

16 samples respectively are seen carrying these variants in a homozygous pattern out of a total of 11,005 each.

For rs3752233, there were no significant tests for the Nottingham cohort alone but when these samples were combined with the American cohort, this subsidised dataset did improve the *p*-value, dropping from approximately 0.46 to 0.08 in both the Fisher's exact test and the Logistic Regression analysis. The Fisher's exact test also returns minor allele frequencies (MAFs) in both cases and controls and, for this variant, there was virtually no difference between the two; 3.69% in cases and 4.18% in controls. This marginally higher frequency in controls returns an odds ratio of below 1 (0.88 in both statistical tests performed), indicating a protective action of the minor allele of this variant if it did have an effect on phenotype, which seems unlikely given these insignificant results.

The association tests performed on rs3752239, on the other hand, do return some results with *p*-values less than 0.05. Unsurprisingly none of these are from data sets only containing the Nottingham samples due to limited numbers, but both the Fisher's exact and Logistic Regression tests return *p*-values of 0.04 when the American cohort of samples are also included in the analysis. Despite this significant test, there is not a large difference in MAFs between cases and controls, with Fisher's exact test returning a MAF of 4.3% in controls and 3.7% in cases. The ORs for both statistical tests are 0.86, again implying a protective action of the minor allele (C). This protective result further increases doubt as to whether this result is meaningful due to *ABCA7* returning risky ORs (>1) in all GWAS studies (see Table 1.2). However, the Fisher's result does not withstand correction for multiple testing (due to five variants analysed, the *p*-value needs to be less than 0.01 to be considered significant, as per the Bonferroni correction). Likewise, the logistic regression effect disappears when the data is adjusted for known covariates.

When examining the results for rs114782266 from just the Nottingham samples, although not significant, there is a relatively large difference of MAFs between case and controls: 0.46% compared to 0.74%. These are both quite different to any MAF recorded in any of the databases examined in Chapter 2: an average of 2.8% across the four databases. However, when the MAFs in different populations are examined in the 1000Genomes database (as opposed to the combined "All" option used in Chapter 2), it can be seen that the frequency in European samples is a lot lower at only 1%. This is also apparent when the American samples are merged with the Nottingham samples, increasing the power but not the significance (the *p*-value goes from 0.30 to 0.86) as well as increasing the OR from 0.62 to 0.96. The difference in MAFs between the two groups is also almost eliminated - 0.68% in cases and 0.71% in controls - in the merged dataset. This contradiction could be explained by examining the differences in population MAFs. The MAF in African populations, according to the 1000Genomes database, is 10% which, given the nature of American ancestry, may explain the skewing of these results. This effectively "nullifies" the apparent protective effect seen in analysis of just the British samples. However, studies looking at the effect of rare variants (<5%) in asthma have suggested that the association of these rare variants with complex disease are ethnic specific and therefore do not contribute to the "missing heritability" of complex diseases such as asthma and LOAD (Igartua et al., 2015). Studies such as these, therefore, need to be more ethnicity focused and, in this particular case, the British sample size needs to be increased. Alternatively, if the data were available, the results could be adjusted for the population stratification to control for this potential confounder.

The effect of ethnic dependant population substructure is also seen in rs59851484. The first obvious finding is that there are no minor alleles found at all in the British samples. This correlates with the findings relating to this variant shown in Figure 2.4 where the minor allele was not found in any of the 1000Genome Phase 1 samples (and

therefore an LD score could not be calculated) and the fact that no positive control could be identified in our DNA bank. This suggested that this variant may be rarer than first thought, at least in certain populations. Again, looking at the MAF in different populations in the 1000Genomes Phase 3 data, even though the "All" MAF is 3%, the MAF in European, East and South Asian is all 0% (European/ American MAF is seen as 0.02% in EVS). It is the American and African MAFs which pulls the overall MAF up, with ethnic specific MAFs of 1% and 12% respectively. Despite the fact that the samples from the Mayo Clinic are labelled as "American," there is likely to be cryptic African ancestry (including the controls as well as the cases) and a wide range of ancestry in American samples is known to skew genetic association tests (Campbell et al., 2005). This has likely artificially increased the MAFs in both groups in this test (0.04% in cases and 0.16% in controls). When analysing the merged dataset for rs59851484, there is a low OR utilising both Fisher's exact and the Logistic Regression tests. In both the OR is around 0.276 and the CI is 0.09-0.83 which, although quite a wide confidence interval, suggests a potential protective influence of the minor allele of this variant (A). However, data correction (for both multiple testing and covariates) suggests that in both tests the results are not significant.

The analysis for rs59851484 has emphasised some of the limitations of this study. The first limitation has already been highlighted in that examining the cumulative MAF in databases such as 1000Genomes is not the most effective methodology, especially when there is such a major difference between populations. Such a large difference, as exhibited with rs59851484, may require different approaches when it comes to genetic testing. For example, this variant requires treating as a rare variant in British and European samples (where the MAF is close to 0%) but as a common variant in Africans (where the MAF is 12%). Not only was the small sample size of the British samples a problem but it highlighted how population stratification can create false

positive results within genetic association tests. In order to reduce the chances of this happening, family based studies need to be performed, matching samples based on ancestry as well as increasing sample sizes as a whole (Cardon and Palmer, 2003). Principal Component Analysis (PCA) can also be used in order to correct for this population substructure. PCA analyses the proportion of genetic variance and reduces it down to a single principal component. If this is large, then one can obtain a value as to the presence of population substructure that is present. This can then be used to correct for the population substructure in any disease association studies performed. However, whole genome data is required for any samples in order to perform PCA, something which was not available for these samples, both the Nottingham and the Mayo cohorts.

Sample size was also an issue when testing the novel variant at genomic position 19:1056958 although this was expected to be a problem due to its known low frequency. This variant was validated in the NGS data but still only had a MAF of 0.66% in this cohort of samples. This, as well as the fact that there was no additional data from the Mayo Clinic with which to supplement our data, drastically reduced the power for this association study. Instead, approximately an extra 1000 samples were run in house in order to try and improve the power of this test. However, this was not successful in achieving a statistically significant test, categorizing this as purely an exploratory test. Although there is no imputation, association or LD data available for this variant due to its novelty, it is predicted to be functionally important; it is in a conserved region (the area from 19:1056903 to 19:1057085 is known to be highly conserved, encompassing this variant (Turton, 2014)) and functionally vital area (TMD 10) and is annotated as being probably damaging or deleterious. The fact that this variant codes for a Serine to Proline amino acid substitution also means it is likely to alter function as this specific deviation is known to produce kinks in alpha-helical structures, such as TMDs. This suggests they are not able to adopt the main-chain

confirmations seen in other amino acid chains, potentially altering the TMD this variant is located in and this modifying the pore ABCA7 forms in cellular membranes (Barnes, 2007).

Despite the lack of power in this assay (a power of only 57.7% was achieved), the results are suggestive of an effect. The MAF in cases is three-fold higher than in controls (0.13% as opposed to 0.04%) and the Fisher's exact test yields an OR of 3.53, tentatively suggesting that this variant notably increases the risk of LOAD. However, the CI is extremely wide, ranging from 0.40 to 31.63, emphasising that this assay needs to be repeated in many more samples (thousands) in order to determine if this observation is accurate. At the beginning of this project, this variant was classed as "novel," having been seen in our NGS data but none of the online datasets. However, it has now been identified in other studies and assigned an rs ID number: rs778244634. This is reported in the dbSNP as being part of the ExAC dataset and reports a MAF of 0.0008%, being seen in only one individual out of 60,706. However, it has not been validated as part of this study so any further information is not available.

Due to the rarity of this variant, 2281 case-control matched pairs would be required in order to receive 75% power, 3360 in order to achieve 90% power. Single variant association tests, such as the ones performed in this chapter, frequently fail to detect rare variants purely due to their lower MAF. In order to make these kind of tests meaningful and worthwhile, variants of higher MAFs need to be focused on or DNA datasets need to be combined in order to obtain the power as a scientific community (Lord et al., 2014). However, rare variants still need to be investigated in some shape or form as they are known to contribute to the missing heritability of LOAD, mainly due to them being missed by high throughput genotyping techniques such as GWAS. For example, the *TREM2* variant identified in 2013 through exome sequencing is an

example of this; a rare variant which contributes to this missing heritability (Guerreiro et al., 2013). Another way of obtaining the power required for association testing of rare variants is by "collapsing" them or aggregating several rare variants (for example, each with a MAF of ~0.01%) in order to boost their association power. This panel of variants can then be genotyped in order to investigate their cumulative effect (S. Lee et al., 2014; Steinberg et al., 2015). The variants aggregated can be linked due to proximity (for example all within TMD 10) or functionality (for example within both TMDs 10 and 11 as these are the domains which identify substrates) (Bansal et al., 2010; Zhang et al., 2006b). Although these methodologies are still difficult to implement due to limited samples available and complex bioinformatics required in order to interpret the results, they are going to be required increasingly in the future in order to complete the gaps currently present in the knowledge of the genetics of LOAD.

3.5 Conclusions

In conclusion, despite the measures taken in Chapter 2, limiting the variants investigated to only ones of a minor allele frequency of above 1%, they still were not common enough to achieve power when genotyped in the test subjects for which DNA is held in this lab. Although the majority of the variants tested here had also been genotyped in an American cohort, merging the two datasets in order to increase the power often caused more complications, introducing population substructure into the samples, making statistical analysis of them problematic.

Despite some of the data looking promising, that of rs59851484, rs3752239 and 19:1056958 T>C specifically, none of the test results survived being corrected. They all need to be tested in larger datasets, specifically ones of European origin, to determine if there is an association with the LOAD phenotype before time is invested in investigating them functionally.

4 Minigene Splicing Assays

4.1 Introduction

Splicing is a post-translational process whereby introns are removed from precursor messenger RNA (pre-mRNA) in order to form mature messenger RNA (mRNA) containing only the coding segments (exons). At first glance, having such a large proportion of the genome not coding for proteins (it is estimated that 95% of nuclear RNA is spliced out) appears remarkably wasteful and inefficient. However, introns do play a vital role in eukaryotic gene expression, enhancing their efficient expression by containing transcriptional enhancer elements, recruiting the machinery required for gene expression and, as discussed in further detail later, increasing the variety of the transcriptome through alternate splicing as well as offering "mutation buffering" (Elliott, 2010; Scotti and Swanson, 2015). In order to facilitate the elimination of these introns, protein complexes bind to conserved *cis*-sequences defining the intron-exon boundaries in the pre-mRNA sequence, forming the spliceosome complex. These *cis*-sequences comprise of the donor site (5' end - GU in mRNA) and the acceptor site (3' end - AG in mRNA), shown in detail in Figure 4.1, creating what is thought to be approximately 50% of the information defining splice sites (Wu and Hurst, 2015). There is also a branch site (typically 18 to 35 base pairs upstream from the 3' site) characteristically consisting of a polypyrimidine tract leading to the branch point containing an adenine, again shown in Figure 4.1 (Taggart et al., 2012). During splicing, five small nuclear RNAs (snRNAs), complementary to the *cis*-sequences (U1, U2, U4, U5 and U6 - so called due to the high levels of the nucleotide uridine within their sequences), as well as between 50 and 100 other polypeptides, are recruited to the donor and acceptor sites.

The 5’ and 3’ ends of the intron are subsequently brought together into close proximity, forming the spliceosome (Cartegni et al., 2003). The 5’ sense strand is first cleaved when the -OH group of the branch point adenosine attacks the phosphodiester bond between the upstream exon and the intron. This forms a new bond between the 5’ end of the intron and the branch point, creating a loop of mRNA or the “lariat intermediate.” The 3’ -OH group of the upstream exon then attacks the phosphodiester bond between the downstream exon and the intron, releasing the lariat loop (which goes on to be rapidly degraded) and ligating the two exons together, forming the mature RNA, summarised in Figure 4.2 (Clancy, 2008; Elliott, 2010; Taggart et al., 2012).

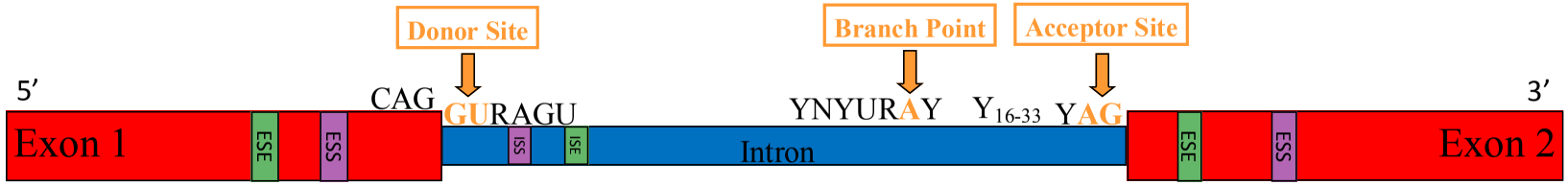


Figure 4.1
cis-sequences vital to successful splicing. Y = pyrimide, R = purine and N = any nucleotide. The donor, acceptor and branch sites are highlighted (GU, AG and A respectively) which is where the snRNA's act (arrows). U1 acts on the donor site (GU), subunits of U2 act on the polypyrimide tract point ($U2AF^{65}$) and the acceptor site (AG—$U2AF^{35}$). SF1/BBP binds to the branch point (A). Other snRNAs, especially U5 which acts on both the donor and acceptor sites, form a bridge between the 5' and 3' splice sites as well as interacting with U4/6 in order to form the mature spliceosome. The A in the branch point (arrow) donates the 2'-OH group used in the first catalytic step of splicing (see Figure 4.2). Also shown here are the Exonic Splicing Enhancer's (ESE), Exonic Splicing Silencer (ESS), Intronic Splicing Silencer (ISS) and Intronic Splicing Enhancer (ISE) elements which can be located within the exons and introns respectively. The enhancers can promote splicing through binding of SR proteins and the silencers can block splicing by the binding of heterogeneous ribonucleoprotein particles. Adapted from Elliot (2010) and Vuong *et al* (2016).

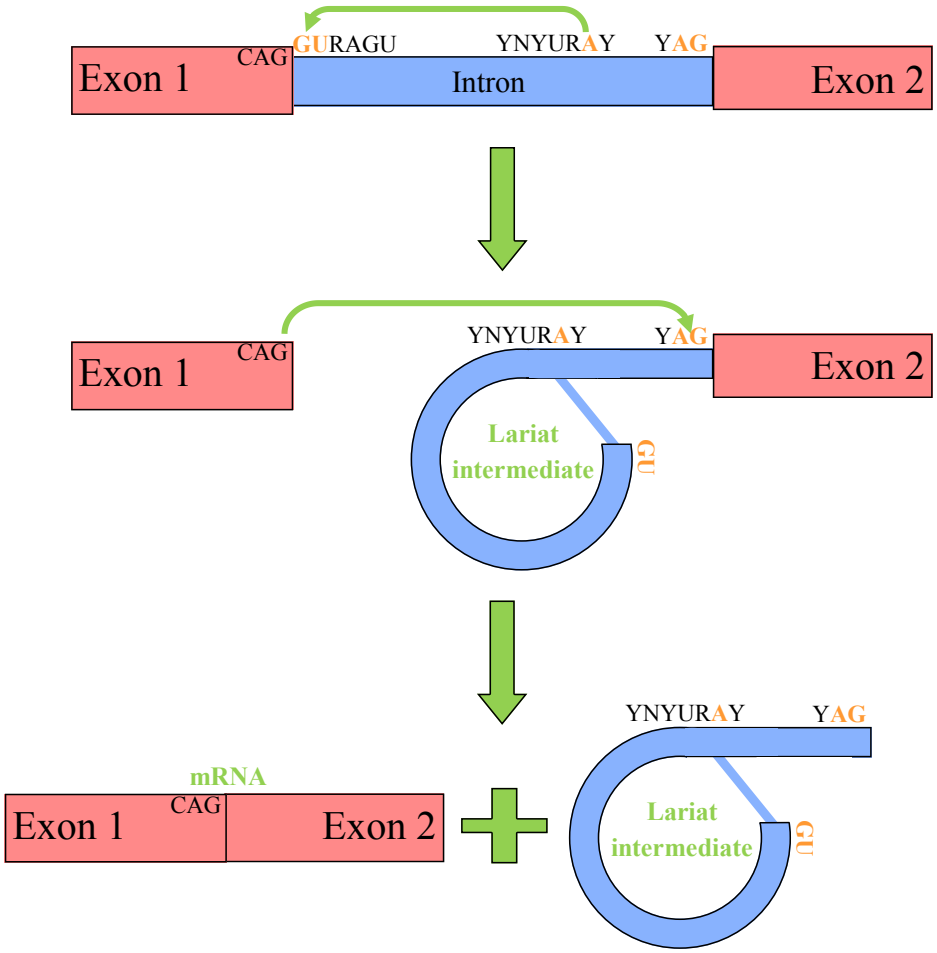


Figure 4.2
The three main steps involved in eukaryotic mRNA splicing. Y = pyrimide, R = purine and N = any nucleotide In the first step, the –OH of the branch point adenosine (A) attacks the phosphodiester bond between the upstream exon (Exon 1) and the intron. A new bond is therefore formed between the 5' end of the intron and the branch point, forming the lariat intermediate. The 3' -OH groups of the upstream exon then attacks the phosphodiester bond between the downstream exon (Exon 2), releasing the lariat loop and ligating the two exons together.

Alternative splicing can be achieved by exon skipping, alternating the exon size or retaining introns into the mature mRNA product, creating different transcripts or altering the reading frames and, consequently, alternative protein isoforms (Elliott, 2010). It is estimated that 90% of the protein coding genes in the human genome undergo alternative splicing, immensely increasing the transcriptome diversity through a variety of consequences, as shown in Figure 4.3 (Pan et al., 2008; Scotti and Swanson, 2015). This variety is especially important in complex tissues, for example the brain, where many alternative spliced isoforms are exhibited, both between areas of the brain as well as at different time points during development or along the neurological cell lineage (Kang et al., 2011; Vuong et al., 2016). Alternative splicing is regulated by *cis*- and *trans*-factors, silencing or activating the splice sites (Matlin et al., 2005). These modulation *cis*-acting enhancer regions occur in both exons and introns (ESEs and ISEs respectively) as do the silencer regions (ESSs and ISSs respectively - all demonstrated in Figure 4.1). Enhancer motifs are identified through the binding of serine arginine (SR) proteins, and snRNPs interact with the silencer sequences (Clancy, 2008; Pagani and Baralle, 2004). It is predominantly the concentration of these proteins which determine whether or not the exon is skipped, or included, in the final protein transcript by altering the spliceosome at these splice sites (Vuong et al., 2016). However, other factors may also play a role in alternative splicing; chromatin architecture, histone modifications and dephosphorylation of splicing factors are all examples of these (Matlin et al., 2005; Tilgner et al., 2012).

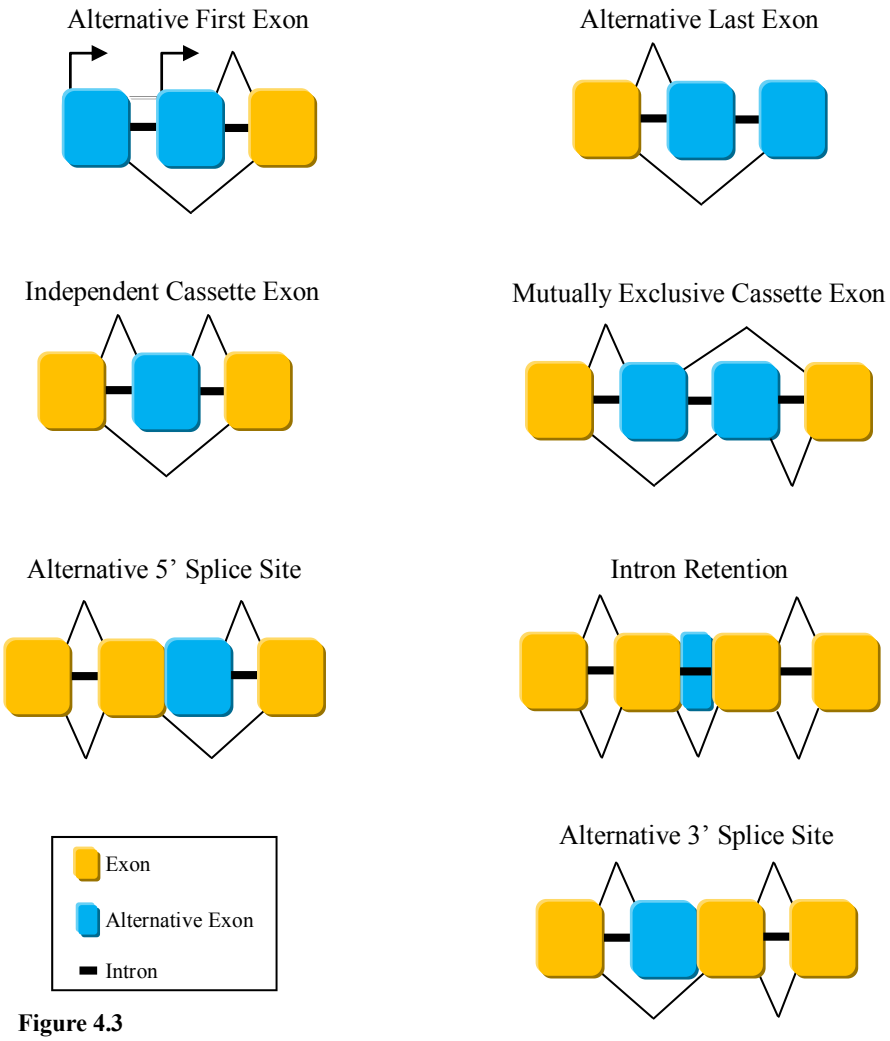


Figure 4.3

The consequences of the complex machinery involved in alternative splicing. Alternative first and last exons can be included, introns retained, a cassette exon included (either independently or in a mutually exclusive manner) as well as alternative 5' and 3' splice sites being present. Both cassette exon skipping and intron retention are frequent in humans and mammals. These patterns can be demonstrated through RNA-sequencing (RNA-seq). Adapted from Scotti and Swanson 2015 and Vuong *et al* 2016.

It has been estimated through computational predictions, comparative genomics and transcriptome profiling of both normal and diseased tissue that a surprisingly high fraction of pathogenic mutations are due to their effect on splicing (Singh and Cooper, 2012). As such, approximately one in three mutations in the human genome are believed to go on to cause aberrant splicing, both somatic and hereditary, and, therefore, tight regulation is imperative. These mutations can alter transcription in several ways: they can cause loss-of-function by introducing a premature stop codon; cause an exon to be spliced out or an intron to be included; reduce the specificity of a splice site, varying its location by inserting or deleting amino acids or altering the reading frame; previously existing pseudo-splice sites can also be activated, altering the balance of splice isoforms or they can displace the splice site, creating longer or shorter exons than are coded for originally (Baralle and Baralle, 2005; Singh and Cooper, 2012). In addition, variants can also alter the secondary structure of the RNA, preventing the enhancer or silencer proteins from gaining proximity to their complementary sequences (Pagani and Baralle, 2004). Variants effecting splicing are commonly *cis*-acting, located in the core consensus sequences of the donor and acceptor sites, the branch point or the ESEs, ESSs, ISEs and ISSs. Any dysregulation of pre-mRNA splicing has the risk of causing pathogenesis although this can be used to a biological advantage. The introduction of a premature stop codon, for example, can act as an on-off switch for certain genes, such as the *male-specific lethal 2* gene in female Drosophila (Smith and Valcárcel, 2000).

Over 90% of synonymous variants known to be associated with disease are thought to be associated due to their impact on splicing (Wu and Hurst, 2015) and that splice site altering mutations, such as these, cause 15% of all human genetic diseases (Cooper and Mattox, 1997). Several variants within genes associated with LOAD have already been linked to alternative splicing. For example, *MAPT* (the gene coding for tau), *PSEN1*, *PSEN2*, *PICALM* and *CD33* all have documented splice site variants which

have been associated with LOAD risk, all confirmed through RT-PCR of alternative splice isoforms from brain tissue samples, with the exception of the *PICALM* variant (Cruts and Van Broeckhoven, 1998; Donahue et al., 2006; Malik et al., 2013; Celeste Sassi et al., 2014; Sato et al., 1999; Schnetz-Boutaud et al., 2012). *APP* is also documented to have several different isoforms, the ratios of which can alter pathological risk (O'Brien and Wong, 2011). Members of the ABC family have also had alternative splicing implicated in disease: two *ABCA1* variants, both within intronic regions, result in the skipping of exon 7 and exon 32 and cause hyperalphalipoproteinaemia and premature coronary heart disease (Rhyne et al., 2009). *ABCA7* itself also has a splice site variant (c.5570+5 G>C, a loss-of-function variant) which significantly increases an individual's risk of LOAD (OR = 1.97, $p = 5.3x10^{-10}$) when genotyped in 3,419 case and 151,805 control samples (Steinberg et al., 2015).

All of these examples highlight how vital correct splicing is in order to prevent pathogenesis, including that of LOAD. In order to address this, during the NGS data analysis previously discussed in Chapter 2, the variants identified within the *ABCA7* gene were annotated by the Variant Effect Predictor (VEP) in order to infer if any of these variants may affect splicing. This program highlighted six variants which have the potential to alter splicing of the ABCA7 protein. However, the original annotation of the NGS data using VEP does have drawbacks. In its methodology, it examines the variants to see if they are within the first three bases of an exon or within the first eight bases of an intron. As seen previously, this does not include the entirety of the splicing machinery; spliceosome signals, the branch site, and possibly the enhancer and silencer regions are therefore not examined. However, high-throughput analysis of possible splicing variants is problematic with only 84% of splice site defects detected *in silico* and, of the ones which are detected *in silico*, only 50% of these are accurate due to splice site sequences being extremely short and degenerate (Houdayer

et al., 2008; Smith and Valcárcel, 2000). In an ideal situation, these predicted splice site variants would be examined in an *in vivo* environment with RNA being extracted from targeted brain regions and the whole spliceosome examined through RNA sequencing (RNA-seq) methodology in correlation with individual genomic background (Di Resta et al., 2014). However, this is unfeasible for a small number of variants, such as the ones presented here, due to the labour and cost. A commonly utilised alternative method for analysing the functional effect of these variants is a side-by-side analysis of the two alleles within a minigene assay. This involves cloning an insert containing the *cis*-element(s) of interest including the exon(s) and flanking intronic sequences into the Exontrap pET01 vector, specifically designed to examine the splicing machinery in eukaryotic cells, as it contains the conserved splice site sequences as seen in the human transcriptome. This cloned vector is then transfected in specific cell lines where it undergoes translation and the resulting RNA can be extracted and examined. This methodology is summarised in Figure 4.4. Minigene assays in theory allow for extremely long transcripts (thousands of nucleotides can be included in the genomic segment (Cooper, 2005)) to be cloned into cells and are relatively quick. In previous studies it has been shown to exhibit a 100% concordance with splice site altering variants previously classified using RT-PCR analysis on RNA from patient samples within the *BRCA1* gene, demonstrating it to be an accurate *in vitro* representation (Steffensen et al., 2014).

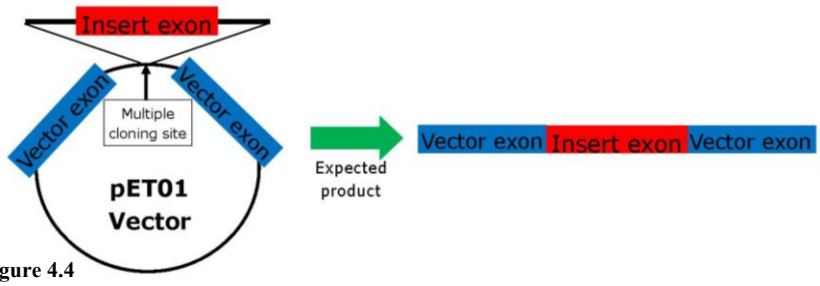


Figure 4.4
Methodology behind the minigene assay. The exon(s) of interest are inserted into the vector along with intronic flanking sequence. Once cloned into the Exontrap pET01 vector, the insert and vector exons are flanked by 5' and 3' eukaryotic exonic sequences as well as a polyA tail resulting in the intronic sequences being removed through splicing. This vector also contains the ampicillin resistance gene and bacterial eukaryotic enhancer sequences. The expected product, after undergoing transcription within the cells it is transfected into, contains both vector and the inserted exons.

The six variants presented in Table 4.1 are those predicted by VEP as having an effect on the splicing of the ABCA7 protein. Further annotation of these variants is also presented; annotation by PolyPhen-2 has previously been discussed in Chapter 2 and predicts the functional effect of the minor allele; the level of conservation of the region, vital in the *cis*-sequences involved in splicing, is annotated by PhastCons; ESEfinder examines the binding sites for SR-proteins, and thus the locations of ESEs (Cartegni et al., 2003); the Berkeley Drosophila Genome Project (BDGP) annotates predicted donor and acceptor sites by examining the degree of base pair complementarity between the splice sites and snRNAs, forming the basis of "strong" and "weak" splice sites (Reese et al., 1997) and Human Splicing Finder examines several different aspects of the splicing machinery including the acceptor and donor sites, branch points, ESEs and ESSs (Desmet et al., 2009). All of these programs were accessed in March 2015. Only two variants are predicted to be damaging to the splicing process by all three of these prediction programs: 19:1054696 (G>C) and rs881768 (A>G). The remainder of these putative splicing SNPs will not be examined any further due to the predictions in at least one of the three programs (ESEFinder, BDGP and Human Splicing Finder) determining no difference between the major and minor alleles; a threshold which has been suggested for *in silico* tools such as these as part of this work (Clement et al., 2016).

Table 4.1

In silico investigation of splicing variants in the *ABCA7* gene as identified in the NGS study. The reference and alternative allele as called in CRISP are shown (Ref/Alt) as well as the minor allele frequency (MAF) as predicted by CRISP and the flanking sequence for bioinformatics analysis. PolyPhen-2 predicts damaging scores of the variant on protein function and PhastCons highlights the level of conservation within the variant region with a score of 0 to 1. SRSF = Serine/Arginine-Rich Splicing Factors, ESE = Exonic Splicing Enhancer and ESS = Exonic Splicing Suppressor.

Genomic Location - GRCh38	rsID	Ref/Alt	CRISP MAF	Flanking sequence for bioinformatics analysis	Consequence (PolyPhen-2)	Conservation (PhastCons)	ESEfinder	BDGP	Human Splicing Finder
19:1046225	-	G/T	0.01	TGCCCCTCTC**K**CAGGTTTTGG	Intron variant, splice region	0.21	SRSF1 & SRSF5 sites lost	-	-
19:1049013	rs4147912	A/C	0.77	AGGGTGAGGC**M**CTACGAGGCT	Intron variant, splice region	0.00	SRSF1 site gained	-	An intronic ESE site is created.
19:1054696	-	G/C	0.01	TTCTTCAGGT**S**GGTGCAGAAG	Intron variant, splice region	0.97	SRSf2 site gained	Known donor site with a score of 0.91 reduced to a score of 0.38	Alteration of an intronic ESS site.
19:1056066	rs881768	A/G	0.25	GGCCCCACAG**R**TACGGAGGCT	Synonymous	0.00	SRSF1 site gained SRSF6 site lost	Novel donor site created with a score of 0.98	A cryptic donor site is activated. An ESE site is created and an ESS site is altered.
19:1061893	rs200538373	G/C	0.01	CCACAGGTGA**S**GGGTGCCAGG	Intron variant, splice region	0.28	SRSF5 gained	-	An intronic ESE site is created.
19:1062165	rs4147920	T/C	0.02	GGCGCCCCCA**Y**CCCCAGCGTG	Intron variant, splice region	0.15	SRSf1 & SRSF5 sites gained SRSF2 site lost	-	One intronic ESS site is altered, another created

Due to the variant at position 19:1054696 having already been investigated by a colleague (Dr James Turton, presented in his thesis (Turton, 2014)), this chapter will focus on the A>G variant rs881768 (MAF = 0.25). The *in silico* prediction for this variant is shown in Table 4.1 and pictorially in Figure 4.5. It suggests that the introduction of the minor allele G generates a new donor site at the beginning of exon 32, as well strengthening the existing acceptor site (already weak with a score of 0.56 in the major allele, increasing to 0.76 in the minor allele transcript). The novel donor site introduced is stronger than exon 32's natural donor site, if only slightly (0.98 compared to 0.93). Although this is not a large difference, it may mean that this novel donor site, much further upstream, is utilised instead. This implies that the presence of this minor allele may result in the removal of all or the majority of exon 32, similar to one of the pathogenic variants reported in *ABCA1*, mentioned earlier (Rhyne et al., 2009). A gene fragment, containing exon 32 and its surrounding introns and splicing machinery, will therefore be examined utilising the minigene system as previously described. In order to attempt to examine the transcripts expressed in the more natural neurological environment, RNA will also be extracted from brain tissue samples from individuals who are homozygous for both the major allele (A) and the minor allele (G), total cDNA will then be synthesised and PCR amplified in order to analysis the *ABCA7* mRNA present neurologically. This will provide insights as to whether the transcripts expressed by both of these genotypes differ *in vivo* or not.

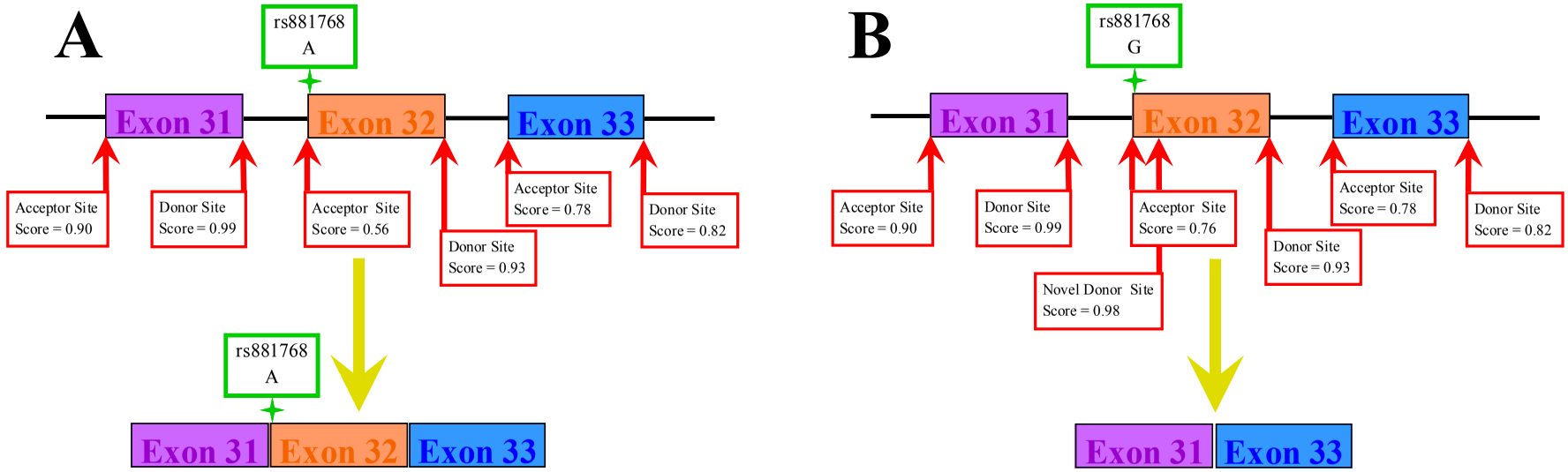


Figure 4.5
The *in silico* predictions of the acceptor and donor *cis*-sequences in exons 31, 32 and 33 of the *ABCA7* gene according to the BDGP prediction programme. (A) presents the predictions with the major allele (A) of the variant rs881768 and (B) presents the predictions with the minor allele (G) of this variant. The minor allele (G) introduces a novel donor site 126 base pairs downstream of the existing acceptor site suggesting that this allele could activate a cryptic splice site or cause a change in protein isoform ratios. However, it also strengthens the score of the existing acceptor site from 0.56 to 0.76 making *in silico* prediction of this variant very difficult to interpret. Other *in silico* tools also predict this variant to effect splicing: ESEfinder predicts it to introduce a SRSF1 site and loose a SRSF6 site—both of which are exon splicing enhancer (ESE) proteins. Both of these predictions are corroborated by the Human Splicing Finder programme (see Table 4.1).

4.2 Materials & Methods - Minigene Assay

4.2.1 Primer Design

Primers were designed to include exon 32 of *ABCA7*, and its flanking introns, from genomic DNA as previously described in Section 3.2.1. The primers were designed so that the sense primer contained the digestion site for the restriction enzyme *SalI* and the antisense primer contained the digestion site for *XbaI*. This was to ensure the insert could be digested out of and ligated into vectors. Webcutter 2.0 (Yale University, New Haven, CT, USA) was used to confirm that the digestion sites for the two enzymes only occurred within the primers themselves within the inserts. The manufacturer's protocols for the vectors to be used were also examined to ensure there were no digestion sites within the vectors themselves, other than those present in the multiple cloning site.

4.2.2 Polymerase Chain Reaction

A selection of heterozygous samples for rs881768, as well as samples containing the major and minor alleles were identified through a Perl script written by a college (Dr Christopher Medway, personal communication), imputing the genotype of this variant from known genotypes in these samples. These samples were held as part of the ARUK DNA bank and all samples had been collected with ethical approval and with consent of all individuals. The relevant samples were PCR amplified and sequenced in order to validate their sequence as well as ensuring that they did not contain any other SNPs in the region.

A high fidelity Taq DNA polymerase was used in order to avoid introducing nucleotide substitutions. Amplification was performed on predicted heterozygous samples in a 30µl reaction volume and consisted of 1x Fermentas Taq Buffer with

$(NH_4)_2SO_4$ (750mM Tris-HCl, 200mM $(NH_4)_2SO_4$, 0.1% Tween 20 and 20mM $MgCl_2$ - Thermo Scientific, Pittsburgh, PA, USA), 2mM $MgCl_2$, 0.2mM dNTPs, 0.25pmol/µl of the sense and antisense primers designed as above (see Table 4.2), 2.5U Expand High Fidelity Taq (Roche, Burgess Hill, UK) and 50ng DNA template (the samples identified as above). PCR was performed using a Veriti 96 Well Thermal Cycler (Applied Biosystems, Carlsberg, CA, USA). The thermal cycle was programmed for 2 minutes at 94°C for initial denaturation, followed by 10 cycles of 15 seconds at 94°C for denaturation, 30 seconds at 62°C for annealing and 40 seconds at 72°C for extension. A further 20 cycles were performed, identical to the previous cycles, but extending the extension step by 5 seconds every cycle before a final extension step was then performed at 72°C for 7 minutes. PCR products were examined by electrophoresis at 80V for 35 minutes on a 1% agarose gel (30ml 1x TAE buffer, 0.3g of peqGOLD Universal Agarose (PEQLAB, Wilmington, DE, USA) and 5µl ethidium bromide) in 1x TAE buffer alongside a 1kbp DNA ladder (Thermo Scientific). The electrophoresis gel was photographed under UV light.

Table 4.2
Sequences of all primers used in this splicing assay. For primer rs881768_Forward the digestion site of restriction enzyme *SalI* is shown in bold and for the primer rs881768_Reverse the digestion site for enzyme *XbaI* is shown in bold.

Primer	Sequence
rs881768_Forward	TGGGGT***gtcgac*** CCCCAGTTCCACTCCCATG
rs881768_Reverse	GCCACC***tctaga*** TGCCAGCCTTTGTTGTTGAA
M13_Forward	GTAAAACGACGGCCAG
M13_Reverse	CAGGAAACAGCTATGAC
pET01_Forward	GATCGATCCGCTTCCTG
pET01_Reverse	CACTGGAGGTGGCCCG
Exon32_Forward	GGCTCTGGGGAAGTGGTTC
Exon32_Reverse	GGTTGAGTGTGGTGATGCTG

4.2.3 Sanger Sequencing

The PCR products were first purified using ExoSAP-IT (Affymetrix; Exonuclease I (ExoI) to remove excess primers and Shrimp Alkaline Phosphatase (SAP) to dephosphorylate excess dNTPs) This was performed at a ratio of 5µl of PCR product to 2µl of ExoSAP-IT. The reaction mix was incubated in the thermal cycler at 37°C for 15 minutes followed by another 15 minutes at 80°C to inactivate the enzymes.

The Sanger sequencing reaction took place in a 10µl reaction volume as described in Section 3.2.4 with the exception of the rs881768_Forward primer being used at a concentration of 0.5pmol/µl. The sequencing products were purified following the manufacturer's protocol of the Performa DTR Gel Filtration Cartridges as per Section 3.2.4 (EdgeBio, Gaithersburg, MD, USA) before being read by capillary electrophoresis on a 3130 Genetic Analyser (Applied Biosystems) provided by the Molecular Diagnostics Lab.

4.2.4 Sequence Alignment

The sequences were received as chromatogram files and analysed using the Chromas Lite software (Technelysium, South Brisbane, Australia). The sequence was converted into FASTA format and then inputted into the ClustalW2 online program (EMBL-EBI, Cambridge, UK) along with the reference sequence obtained from Ensembl, transcript ENST000000263094 (obtained in January 2014) in order to align the two sequences. One sample (AD115) was selected which was heterozygote for rs881768 whilst matching the remainder of the reference sequence completely. Demographics available for this sample (as part of the ARUK DNA database) are shown in Table 4.3.

Table 4.3
Sample demographics of the sample used to create the constructs for this minigene assay. This sample is heterozygous for the variant rs881768 (genotype A/G). The individual this DNA was obtained from was female and 94 when she died and AD was confirmed as part of a post mortem. The age of onset of the disease is unknown (NA), however, it is known that she carried one ε2 and one ε4 allele of the ApoE gene.

Sample ID	Sex	Centre of Origin	Age at Death	Age at Onset	Disease Status	ApoE Status
AD115	Female	Nottingham	94	NA	Confirmed AD	2 4

4.2.5 Cloning of the PCR Products

The selected sample was then inserted into the pCR©2.1-TOPO vector following the protocol of the TOPO TA Cloning kit, chemical transformation pathway (Invitrogen, Carlsbad, CA, USA). 0.5µl of fresh PCR product from Section 4.2.2 was combined with 1µl of the salt solution provided and 1µl of the TOPO vector before being made up to 5µl with dH_2O. The products were transformed into the cells immediately and the remainder stored at -20°C.

4.2.6 Transformation

The cloned products were transformed into One Shot© TOP10 *E. coli* cells again following the TOPO TA Cloning kit protocol. One vial of the TOP10 cells was defrosted on ice before 2µl of the TA clone was added. This was incubated on ice for 15 minutes before being heat shocked at 42°C for 30 seconds. The cells were incubated on ice for 2 minutes before 250µl of warmed SOC media (Super Optimal Broth with Catabolite repression - Invitrogen) was added and the mixture shaken horizontally (250rpm) at 37°C for one hour.

10µl and 50µl of the cells were incubated on plates containing Circlegrow® bacterial growth media (4% – MP Biomedicals, Santa Ana, California, USA), agarose (1.5%), ampicillin (0.1%) and X-gal (1.6mg per plate – Fisher Scientific, Waltham, MA, USA), made up to 100µl with SOC media.

4.2.7 Plasmid DNA Extraction

Positive colonies, ones containing the cloned vector, were coloured white. This was due to the inserted genomic region interrupting the β-galactosidase gene in the vector, inhibiting its production. When this protein is produced, it hydrolyses the X-gal to form a blue pigment, highlighting the colonies containing the vector without the insert. The positive, white colonies were, therefore, incubated in 5ml of liquid media (identical to the media used in Section 4.2.6 with the exception of agarose being omitted) for a further 18 hours at 37°C, shaking horizontally at 250rpm before their plasmid DNA was extracted following the PureYield Plasmid Miniprep System protocol (Promega, Southampton, UK). 3ml of this liquid media was centrifuged at maximum speed in order to pellet the bacterial cells. This cell pellet was resuspended in 600µl of dH_2O before 100µl of Cell Lysis Buffer was added. The tubes were mixed well to allow complete lysis of the cells before being deactivated with 350µl of chilled Neutralization Solution and mixed well. Samples were then centrifuged for 3 minutes at maximum speed in order to pellet the cell debris. The supernatant was transferred to a PureYield Minicolumn and placed in a PureYield Collection tube. This was centrifuged at maximum speed for 15 seconds to anneal the plasmid DNA to the column, with the flow through being discarded. The column was then washed with 200µl of Endotoxin Removal Wash and 400µl of Column Wash, being centrifuged under the same conditions as previously for both wash steps. The elute was discarded and the columns transferred to fresh 1.5ml tubes. 30µl of the Elution Buffer was added to the column and incubated at room temperature for 1 minute before being centrifuged at maximum speed for 30 seconds to elute the plasmid DNA. The plasmid DNA concentration was then determined using the Nanodrop 1000 Spectrophotometer (Thermo Scientific) as per manufacturer's protocol.

4.2.8 Sanger Sequencing

The plasmid DNA was sequenced following the protocol for the BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems) using the M13_Forward primer (see Table 4.2 for sequence – binds to the vector region ahead of the insert). The sequencing reaction took place as previously described in Sections 3.2.4 and 4.2.3, utilising 150-300ng of plasmid DNA and 0.5pmol/µl M13_Forward primer. A thermal cycle was performed and the sequencing product purified as mentioned above. The sequence was then aligned in order to identify colonies containing the major and minor alleles but no other variations from the reference sequence.

4.2.9 Restriction Enzyme Digest

Digestion of the two plasmid DNA samples was performed in a 20µl reaction volume consisting of 1000ng of the plasmid DNA, 1x FastDigest Green Buffer (Fermentas, Vilnius, Lithuania), 0.05U/µl of *SalI* and 0.05U/µl of *XbaI*.

Another digestion reaction was set up with the pET01 vector, a vector which is specifically designed to assess splicing activity. This reaction took place in a 50µl volume containing 5000ng of the vector, 1x FastDigest Green Buffer, 0.03U/µl of *SalI*, 0.03U/µl of *XbaI* and 0.04U/µl of FastAP Alkaline Phosphatase in order to remove the 5' phosphate groups at the ends of the digested vector to prevent it from self-ligating (Thermo Scientific).

A thermal cycle was performed on a TRIO-Thermoblock (Biometra, Göttingen, Germany). The thermal cycle programmed consisted of 37°C for 30 minutes to digest the product and 65°C for 20 minutes to inactivate the enzymes. The products – as well as undigested plasmids as a comparator - were then examined by electrophoresis at 80V for 25 minutes on a 1% agarose gel, containing 4µl SYBR Safe (Life

Technologies, Carlsbad, CA, USA) alongside a 1kbp DNA ladder (Thermo Fisher Scientific, Massachusetts, USA). This was examined under a DarkReader Transilluminator (Clare Chemical Research, Dolores, CO, USA).

4.2.10 Gel Extraction

The two inserts and the linearized, dephosphorylated vector were extracted from the SYBR Safe gel following the protocol of the QIAquick Gel Extraction Kit (QIAGEN, Hilden, Germany). All bands of the correct size were excised from the gel using a sharp scalpel under the DarkReader and placed in 1.5ml tubes. The weight of the agarose slices were calculated before three volumes of QG buffer was added per gel volume. The gel was melted by incubating the tubes at 50°C, vortexing every few minutes, until the gel was completely dissolved. The pH was corrected by the addition of 3M sodium acetate, if indicated by the pH indicator within the buffer, before 1 gel volume of isopropanol was added and mixed. This was then added to a QIAquick column and centrifuged at maximum speed for 1 minute to anneal the DNA to the column, before a further 0.5ml of QG buffer was added and centrifuged again, removing the flow through, in order to remove all traces of agarose. 0.75ml of the PE wash buffer was added and again centrifuged for a further 2 minutes. The columns were then transferred to fresh 1.5ml tubes and 30µl of EB elution buffer was added, incubating for 1 minute at room temperature to disassociate the DNA from the column before centrifuging at maximum speed for 1 minute to elute the DNA. The concentration of the products was determined using the Nanodrop 1000 Spectrophotometer (Thermo Scientific) before being stored at -20°C.

4.2.11 Ligation

The inserts and the vector were ligated together following the protocol of the T4 DNA Ligase Kit (Invitrogen). A vector ligated to itself was also created as a control for the

ligation reaction. 100ng of vector was ligated to each of the inserts. The following equation was used to calculate the amount of insert needed:

$$\frac{100ng\ vector\ X\ 665bp\ insert\ size}{4461bp\ vector\ size}\ X\ 3 = 44.7ng\ of\ insert$$

The reaction itself took place in a 20µl reaction volume consisting of 100ng of vector, 44.7ng of insert (not required in the control reaction), 1x T4 DNA Ligase Buffer and 0.05U/µl of T4 DNA ligase. This was incubated at room temperature for 1 hour.

4.2.12 Transformation

The ligation products (including the vector ligated to itself) were transformed separately into the NEB High Efficiency 5-alpha chemically competent *E. coli* cells C2987I (New England BioLabs, Ipswich, MA, USA) by adding 2µl of each of the ligation products to 50µl of these cells. They were incubated on ice for 30 minutes before being heat shocked at 42°C for 30 seconds. They were then incubated in 950µl of SOC media at 37°C for 1 hour, horizontally shaking at 250rpm.

Two plates were created for each transformation product; one containing 50µl (made up to 100µl with SOC media) and one containing 100µl of the cells and all were incubated at 37°C overnight. The plates were identical to the ones used previously (see Section 4.2.6) with the exception of X-gal which was omitted. The plates containing the cells transformed with the vector ligated to itself act as a control. These plates produced considerably fewer colonies than those containing the insert, showing that the transformation and ligation reactions had been performed successfully.

Colonies from each of the plates containing the inserts (one containing the major and one the minor allele) were then selected and incubated in 25ml of media (consisting of

4% Circlegrow® and 0.1% of ampicillin) at 37°C for 18 hours, horizontally shacking at 250rpm.

4.2.13 Endotoxin Free Plasmid DNA Extraction

The plasmid DNA was extracted from these colonies following the protocol for the NucleoBond Xtra Midi Plus EF Kit (Fisher Scientific). The bacteria were pelleted by centrifuging the full 25ml of liquid culture at 6000 x g for 15 minutes before resuspending the pellet in 4ml of the Resuspension Buffer. 4ml of the Lysis Buffer was then added and the tubes inverted before incubating for 5 minutes at room temperature in order to lyse the cells completely. 4ml of the Neutralization Buffer was added and, again, the tubes inverted in order to terminate the lysing. The mixture was transferred to the NucleobondXtra Midi Columns which, prior to, had been hydrated with 15ml of Equilibration Buffer. The cell mixture was filtered through these, discarding the flow-through, before washing the columns with three different wash buffers (FIL-EU (5ml), ENDO-EF (35ml) and WASH-EF (15ml)) in turn. The columns were then transferred to fresh 50ml tubes and 5ml of Elution Buffer was added and filtered through the columns in order to elute the DNA. The DNA was precipitated by the addition of 3.5ml of room temperature isopropanol, vortexing well, before centrifuging at 4°C, 6000 x g for 90 minutes. The supernatant was removed and the pellet washed with the addition of 3ml of endotoxin-free 70% ethanol, centrifuging again at 4°C, 6000 x g for 20 minutes. The pellet was air-dried for 10 minutes before being resuspended in 100µl of endotoxin-free TE buffer (10mM Tris, 1mM EDTA) overnight at 4°C. The concentrations were then determined using the Nanodrop spectrophotometer. The samples were sequenced to ensure consistency following the BigDye Terminator v3.1 Cycle Sequencing Kit protocol as before (see Section 3.2.4) except the primer pET01_Forward was used (see Table 4.2).

4.2.14 Glycerol Stocks

All colonies confirmed to contain the vectors expected, in both Sections 4.2.8 and 4.2.13, had glycerol stocks made of them in order to prevent having to retransform the plasmid if more DNA was necessary. This was done by combining 200µl of the liquid culture with 800µl of glycerol in a 1.5ml tube. This was labelled well and stored at -80°C long term.

4.2.15 Culturing Cells

COS-7 cells are monkey African green kidney cells and were obtained from the European Collection of Cell Cultures. These cells are known to represent accurately the eukaryotic splicing environment, containing a complex spliceosome, as is seen *in vivo*. COS-7 were cultured in Dulbecco's Modified Eagle Medium (DMEM) with 10% foetal bovine solution (FBS), 2mM L-Glutamine, 100U/ml penicillin-streptomycin (Sigma, St Louis, Missouri, USA) and 100U/ml fungizone (Gibco, Walton, MA, USA) to prevent bacterial and fungal contamination.

BE(2)-C cells are neuroblastoma cells and were therefore cultured to examine the splicing products under a neurological specific spliceosome, linking these results to the location of LOAD pathology and to demonstrate reproducibility. These were cultured in a 1:1 ratio of Eagle's Minimal Essential Medium (EMEM) to Ham's F12 with 1% non-essential amino acids, 2mM Glutamine, 15% FBS, 100U/ml penicillin-streptomycin (all Sigma) and 100U/ml fungizone (Gibco).

All cell work (including transfections and RNA extractions) was performed in a category II biological cabinet in order to reduce the risk of contamination (Labcard Class II Biological Safety Cabinet, Triple Red, Long Crendon, UK). All media and additives used were pre-warmed to 37°C in a water bath before use, in order to

prevent shocking the cells, and all were filtered through 0.22µm sterile filters prior to use. Cells were maintained in 75cm^2 Corning Tissue Culture flasks (Corning, New York, USA) in 13ml of their respective media in an incubator set at 37°C with a humidified atmosphere of 5% CO_2 in air (ICNFlow Automatic CO_2 Incubator Model 106, Thermo Scientific). Once they reached ~90% confluence they were passaged by removing the old media, washing the adhered cells with 10ml of Phosphate-buffered Saline (PBS - Sigma) and incubating for 5-10 minutes in the incubator with 3ml of trypsin (Sigma) in order to digest the adherent proteins which attach the cells to the base of the flask. Once the cells were microscopically free from the flask base, they were re-pipetted with 3ml of their respective media in order to deactivate the trypsin enzyme and split between two flasks.

In order to create stocks of all cells, once they were trypsinised and re-pipetted, they were transferred to a 15ml tube and centrifuged at 300 x g for 15 minutes. These cells were then re-suspended in 6ml of 90% FBS and 10% dimethylsulfoxide (DMSO - a cyroprotective agent) and 1ml was transferred to 2ml screw topped tubes. These were stored at -80°C for at least 24 hours in a CoolCell© LX (BioCision, Larkspur, CA, USA) to delay the cooling process (cooled by 1°C per minute) in order to reduce the amount of ice crystals formed. They were then removed from the CoolCell© and stored in liquid nitrogen indefinitely. Upon removal of the cells from the liquid nitrogen store, they were kept on dry ice before being defrosted in a water bath warmed to 37°C. The cells were then transferred to a 25cm^2 Corning Tissue Culture flask containing 5ml of the respective media, warmed to 37°C. This flask was incubated until the cells were ~90% confluent upon which they were transferred to a 75cm^2 flask using 3.5ml of PBS and 1.5ml of trypsin, following the methodology previously described. They were again incubated until ~90% confluent before being passaged as above, replacing the media if necessary.

4.2.16 Transfection

COS-7 cells were transferred to 60mm plates at a concentration of 3.5 x 10^5 cells per plate and BE(2)-C at a concentration of 6 x 10^5 cells per plate and incubated for 24 hours before the two DNA samples were transfected into both cell lines following the TransFact protocol (Promega). TransFact is a lipid reagent which creates lipid vesicles around the DNA, delivering it into eukaryotic cells.

9µl of TransFact (9nl of TransFact per 1ng of DNA), 1000ng of plasmid DNA and 2ml of serum-free media were incubated at room temperature for 15 minutes. The media was then removed from the plates of cells and the TransFact mixture was added. These were incubated at 37°C, 5% CO_2 for 1 hour before a further 4ml of complete media (either DMEM or EMEM:F12 depending on the cell line) was added to the plates. These were then incubated for a further 24 hours.

4.2.17 RNA Extraction

Total RNA was extracted from the COS-7 and BE(2)-C cells following the RNeasy Mini Kit protocol (QIAGEN). The cells were first trypsinised as previously described, washing them with warmed PBS before releasing them from the plates with 1.5ml of trypsin per dish. The cells were transferred to 15ml tubes and pelleted by centrifuging at 300 x g for 5 minutes. They were resuspended in 600µl of RLT Buffer and 6µl of β-Mercaptoethanol in order to lyse the cells as well as reduce RNase activity by reducing its' disulphide bonds. The cells were homogenised by passing them at least 10 times through a 20-guage needle fitted to a sterile 10ml syringe before the addition of 600µl of 70% ethanol (in order to remove salts), mixed and transferred to the RNeasy Mini Column. The columns were centrifuged at 10,000 x g for 30 seconds and the flow through discarded, in order to anneal the total RNA to the column. The column was washed with 350µl of the RW1 buffer before an additional DNase step

was performed in order to ensure complete removal of genomic DNA. This was achieved by the addition of 1U of RNase free DNase I diluted in 70µl of Buffer RDD (QIAGEN) to the columns and incubating at room temperature for 15 minutes. The columns were washed again with 350µl of RW1 buffer, centrifuged, and the flow through discarded. An additional wash step was performed with 500µl of RPE buffer before the columns were transferred to fresh 1.5ml tubes where 30µl of RNase-free dH_2O was added and incubated at room temperature for 1 minute before being centrifuged at 10,000 x g for 1 minute in order to elute the RNA. The total RNA was quantified on the Nanodrop spectrophotometer and stored at -80°C.

4.2.18 cDNA Synthesis

Total cDNA was synthesised from the total RNA extracted following the protocol for the AffinityScript Multiple Temperature cDNA Synthesis kit (Stratagene, La Jolla, CA, USA). In this synthesis step, two different primers were used: oligo(dT) and random primers. Control reactions were also performed where the enzymes were absent to ensure that the entirety of the genomic DNA was removed in the previous step.

In all of the reactions, 2µg of RNA was required in a reaction volume of 15.7µl. To this, 500ng of oligo(dT) primer or 300ng of random primer were added. These were incubated at 65°C for 5 minutes and then cooled to room temperature for 10 minutes to allow the primers to anneal to the RNA and to prevent the enzymes about to be added from being heat inactivated.

The reaction volume was made up to 20µl with 1x AffinityScript RT Buffer and 1.6mM of dNTP mix. To the positive reactions, 1U/µl of RNase Block Ribonuclease Inhibitor and 1U of AffinityScript Multiple Temperature Reverse Transcriptase were also added. A thermal cycle was then performed on a TRIO-Thermoblock (Biometra)

consisting of 25°C for 10 minutes to extend the primers, 42°C for 60 minutes for the cDNA synthesis and 70°C for 15 minutes to inactivate the reaction. The resulting total cDNA was then stored at -20°C.

4.2.19 cDNA Polymerase Chain Reaction

Amplification of the target cDNA, i.e. the inserted genetic region, was performed in a 30µl reaction volume and consisted of 1x LongAmp Taq Reaction Buffer (60mM Tris-SO_4, 20mM $(NH_4)_2SO_4$, 2mM $MgSO_4$ at pH 9.1 - New England BioLabs), 0.2mM dNTPs, 1ng/µl of the sense and antisense primers (designed for the pET01 vector - see Table 4.2), 0.5U LongAmp Taq DNA Polymerase (New England BioLabs) and 1µl cDNA template. PCR was performed using a Veriti 96 Well Thermal Cycler (Applied Biosystems). The thermal cycle was programmed for 30 seconds at 94°C for initial denaturation, followed by 30 cycles of 15 seconds at 94°C for denaturation, 30 seconds at 59°C for annealing and 50 seconds at 65°C for extension and 10 minutes at 65°C for the final extension. PCR products were examined by agarose gel electrophoresis as described above (see Section 4.2.2).

4.2.20 Sanger Sequencing

The PCR products were sequenced to allow comparison of the two sequences. The PCR products were first treated with ExoSAP-IT as described previously in Section 3.2.4 and were then sequenced utilising the BigDye methodology as previously mentioned, using the pET01_Forward primer (see Table 4.2).

The sequencing product was purified following the manufacturer's protocol for the EdgeBio Performa DTR Gel Filtration Cartridges. The products were then read by capillary electrophoresis and analysed using the ChromasLite and ClustalW2 software.

4.3 Materials & Methods - RNA Extraction from Brain

4.3.1 RNA Extraction from Brain

RNA was extracted from cortex brain tissue samples containing both the major and minor alleles of this variant in order to examine the *ABCA7* isoforms present, endorsing the results obtained in the minigene assays. Samples which we held brain tissue for and with the genotype of homozygous major or minor alleles for this variant were identified in Sections 4.2.2 to 4.2.4 and their demographics are available in Table 4.4. Unfortunately, no data is available regarding the age of these individuals at death.

Table 4.4
Sample demographics of the samples used to examine the RNA of *ABCA7* isoforms in brain tissue, related to the rs881768 genotype. Sample M647 was identified as carrying the major allele (AA) and M648 the minor allele (GG) of this variant. These samples were chosen as brain tissue was available as part of the ARUK DNA Brain Bank. The sex of the individuals, the centre of origin, disease state (both confirmed Alzheimer's disease (AD) at post-mortem for these samples) and the two ApoE alleles carried are presented. For both of these samples, their age at death and the age of disease onset are unknown (NA).

Sample ID	Sex	Centre of Origin	Age at Death	Age at Onset	Disease Status	ApoE Status
M647	Female	Manchester	NA	NA	Confirmed AD	3 4
M648	Female	Manchester	NA	NA	Confirmed AD	3 4

All equipment during this protocol was cleaned with 2% trigene and 70% ethanol and was cooled on dry ice prior to use in order to keep the brain tissue as cold as possible for as long as possible in an attempt to reduce any further RNA degradation. The RNeasyPlus Universal Midi Kit (QIAGEN) protocol was followed. 250mg of tissue for each sample were first homogenised by covering them in liquid nitrogen before crushing with a pestle and mortal to a fine powder. This was transferred to a pre-cooled 50ml tube before 5ml QIAzol Lysis Reagent was added and passed through a 19-guage needle and 1ml syringe until the tissue powder was homogenised as much as possible. This was then incubated at room temperature for 5 minutes, in order to promote the disassociation of nucleoprotein complexes, before 500µl of the genomic

DNA Elimination Solution was added and mixed vigorously for 15 seconds. 1ml of chloroform was then added, again mixed vigorously for 15 seconds, before being incubated at room temperature for 3 minutes in order to promote phase separation. The samples were then centrifuged at 5000 x g for 17 minutes at 4°C to separate the solution into three phases: the upper, colourless, aqueous phase containing the RNA, the middle, white phase as well as a lower, red, organic phase. The upper, aqueous phase was transferred to a new tube, 1 equal volume of 70% ethanol was added and mixed well before being transferred to the RNeasy Midi Spin Column. This was centrifuged at 5000 x g for 5 minutes, to anneal the RNA to the column, and the flow-through discarded. 4ml of Buffer RWT was added and an identical centrifuge performed in order to wash the membrane. An additional wash step was performed by adding 2.5ml of Buffer RPE and again centrifuging. This was repeated, creating a total of three wash steps, before the column was transferred to a fresh tube and 200µl of RNase-free dH_2O added and incubated for 1 minute at room temperature in order to disassociate the RNA from the membrane. This was centrifuged at 5000 x g for 3 minutes in order to elute the RNA. The sample concentrations were then determined utilising the Nanodrop 1000 Spectrophotometer (Thermo Scientific) to ascertain the success of the extractions.

4.3.2 cDNA Synthesis

This was performed as in Section 4.2.18 following the protocol for the AffinityScript Multiple Temperature cDNA Synthesis kit. Both oligo(dT) and random primers were used.

4.3.3 cDNA Polymerase Chain Reaction

Primers within the exons of *ABCA7* were designed as in Section 4.2.1. Primers were designed to be within exons 30 and 33 in order to ascertain whether exon 32 was

included in the transcripts or not. Exon 30 was selected as opposed to exon 31 due to exon 31 being so short (33 base pairs). The primer sequences obtained are shown in Table 4.2 (Exon32_Forward and Exon32_Reverse).

All of the products from the cDNA synthesis were PCR amplified in a 30µl volume consisting of 1x PCR Buffer with $MgCl_2$ (750mM Tris-HCl, 200mM $(NH_4)_2SO_4$, pH 8.8 - Roche), 0.2mM dNTPs, 1pmol/µl of the primers designed (Exon32 primers as seen in Table 4.2) and 1U of Taq DNA Polymerase (Roche). The following thermocycle was executed: initial denaturation was performed at 94°C for 2 minutes, and then a further 30 seconds at 94°C, 58°C for 30 seconds and 72°C for 1 minute repeated 30 times. A final extension step was set at 72°C for 7 minutes. The products were examined by 1% agarose gel electrophoresis before being Sanger sequenced utilising the Exon32_Forward primer located within exon 30, as previously described.

4.4 Results

All primers used, including the ones designed specifically for this assay, are listed in Table 4.2. The selection of samples PCR amplified and Sanger sequenced in order to verify their genotype can be seen in Figure 4.6. These samples contained individuals who carried the major and minor alleles in a homozygous fashion as well as those who were heterozygous for variant rs881768 and were utilised in both the minigene assays and to identify brain tissue samples in order to extract total RNA. Demographics for the samples selected for these assays can be seen in Tables 4.3 and 4.4.

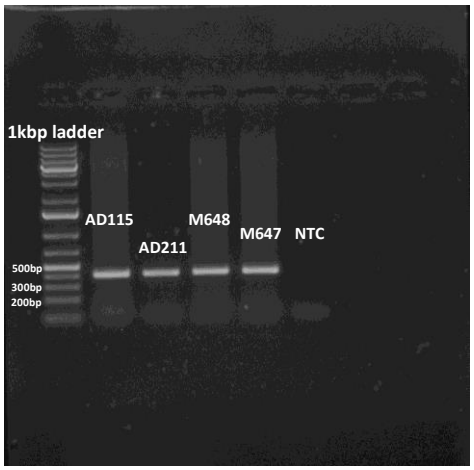


Figure 4.6
1% ethidium bromide stained agarose gel showing the samples amplified, utilising the rs881768 primers presented in Table 4.2, in order to ascertain their genotype for this variant by Sanger sequencing these PCR products. Sample AD115 was shown to be heterozygote and was therefore used to construct the minigene assays (see Table 4.3), sample M648 carried the minor allele only and sample M647 the major allele only. Brain tissue was therefore acquired for these two samples from which RNA was extracted to examine the *ABCA7* isoforms present (see Table 4.4). Sample AD211 was also shown to be heterozygous but was not used in these assays. Also shown is a 1kbp ladder, to compare the product sizes to, and a no template control (NTC) to show no contamination was present.

Once constructs had been created for both the major and minor allele within the pCR®2.1-TOPO vector, restriction enzyme digests were performed in order to ensure the DNA region had been inserted into the vector. An example of these digests can be seen in Figure 4.7, where one major allele construct and one minor allele construct can be identified.

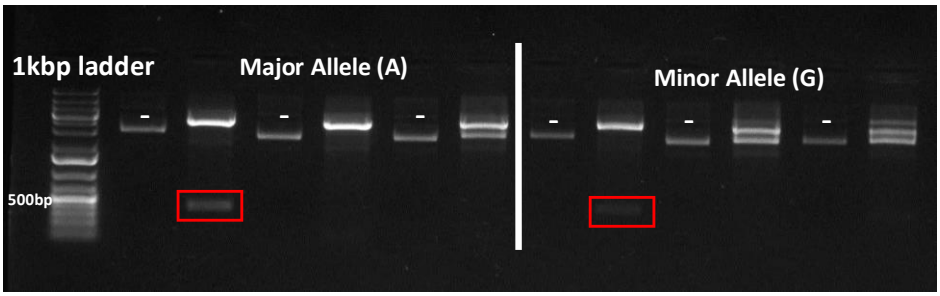


Figure 4.7
Plasmid DNA containing both the major (A) and minor (G) allele of the variant rs881768 in the pCR©2.1-TOPO vector in order to conduct minigene assays for this variant. Every second lane shows the vector which has been digested by the restriction digest enzymes *SalI* and *XbaI*, run alongside a vector which has not been digested for comparison (lanes marked with a -). The red boxes show vectors containing the insert as this has been digested out of the vector to produce a band of 386bp. These inserts were excised from the gel before being ligated into the pET01 vector and transfected into cells in order to examine the RNA isoforms produced.

Transfections into both COS-7 and BE(2)-C cells were replicated multiple times (four in COS-7 cells and three times in BE(2)-C cells) and all cDNA created from these transfections, upon PCR amplification, can be seen in Figure 4.8 (from COS-7 cells) and Figure 4.9 (from BE(2)-C cells). In both Figures 4.8 and 4.9 it can be seen that the samples containing the major allele all contain only one band when the synthesised cDNA is PCR amplified. This band is of the correct size to be only vector exons and Sanger sequencing of this sized product (after gel extraction of the band, as in Section 4.2.10) confirms this (see Figure 4.10). Also in Figures 4.8 and 4.9 it is apparent that the samples carrying the minor allele show two bands - one of identical size to the band present in major allele samples (Sanger sequencing confirms the bands are of identical composition - see Figure 4.10) but a larger band is also present. Upon Sanger sequencing of this product (see Figure 4.11), it was established that (although it does not sequence entirely cleanly) it contains both the vector exons and the inserted exon 32 as expected. Upon visualising the gel under the black light and excising the section of gel where this band would be present for all samples, containing both the major and minor alleles, this larger band was present in all minor allele samples and none of the major alleles. All of these bands were Sanger sequenced in order to verify that all of the samples were consistent and they were, as presented in Figures 4.10 and 4.11. This implies that exon 32 is spliced out in samples containing the major allele (G) but included some of the time with the presence of the minor allele (A) of this variant.

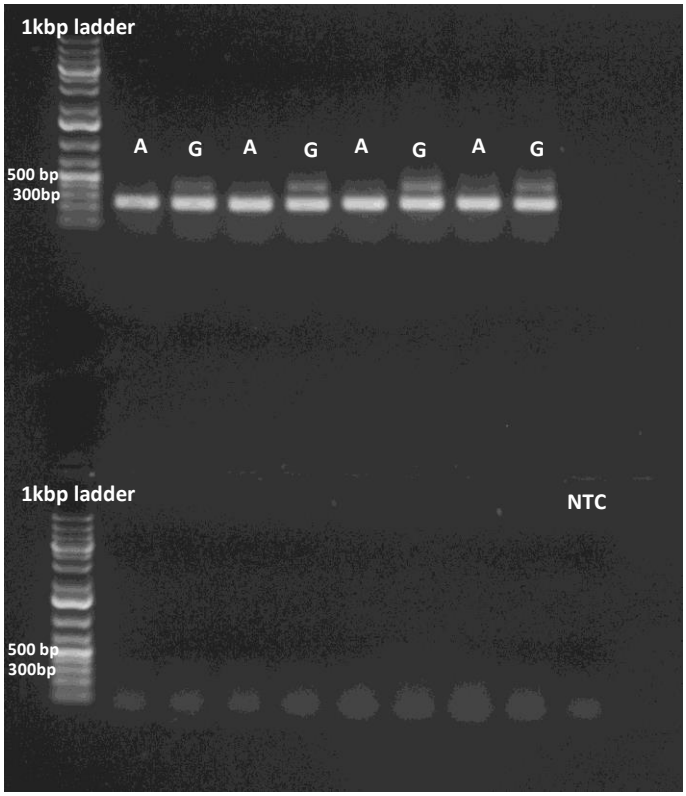


Figure 4.8

cDNA synthesised from four separate transfections into COS-7 cells in reference to a 1kbp ladder. A = vector transfected into the cells containing an insert containing the major allele (A), G = vector transfected with an insert containing the minor allele (G). All samples contain the smaller band of approximately 240 base pairs but only the minor allele samples contain the larger band of approximately 420 base pairs. The sequence and alignment of both of these bands can be seen in Figures 4.10 & 4.11 respectively.

The bottom row of wells show the amplification of the "enzyme negative" samples, showing no genomic DNA is present, only RNA was amplified. All sequences of the bands were consistent between transfections. There is also a blank No Template Control lane (NTC) showing no contamination is present.

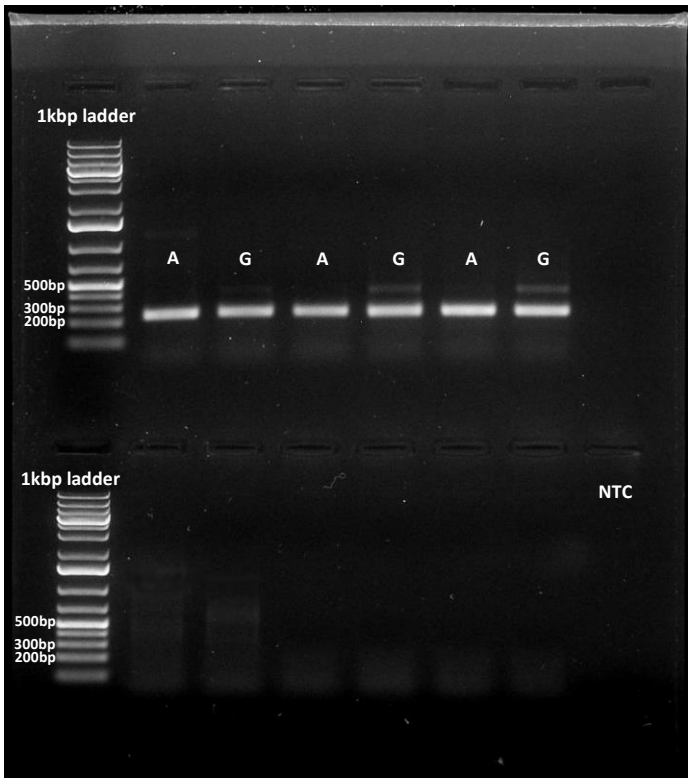


Figure 4.9
cDNA synthesised from three separate transfections into BE(2)-C cells in reference to a 1kbp ladder. A = vector transfected into the cells containing an insert containing the major allele (A), G = vector transfected with an insert containing the minor allele (G). All samples contain the smaller band of approximately 240 base pairs but only the minor allele samples contain the larger band of approximately 420 base pairs, correlating with what is seen in COS-7 cells. The sequence and alignment of both of these bands can be seen in Figures 4.10 & 4.11 respectively. The bottom row of wells show the amplification of the "enzyme negative" samples, showing no genomic DNA is present, only RNA was amplified with the exceptions of the first transfection where genomic DNA was present. The top band in these lanes was extracted and sequenced and shown to be genomic DNA by the presence of intronic DNA. These genomic DNA bands are also present in the "enzyme positive" lane of the major allele vector of this transfections and the sequence also contained intronic DNA. The minor allele of this transfection showed a weaker larger band although DNA was still extracted and sequenced from this region. All sequences of the bands were consistent between transfections. There is also a blank No Template Control lane (NTC) showing no contamination is present.

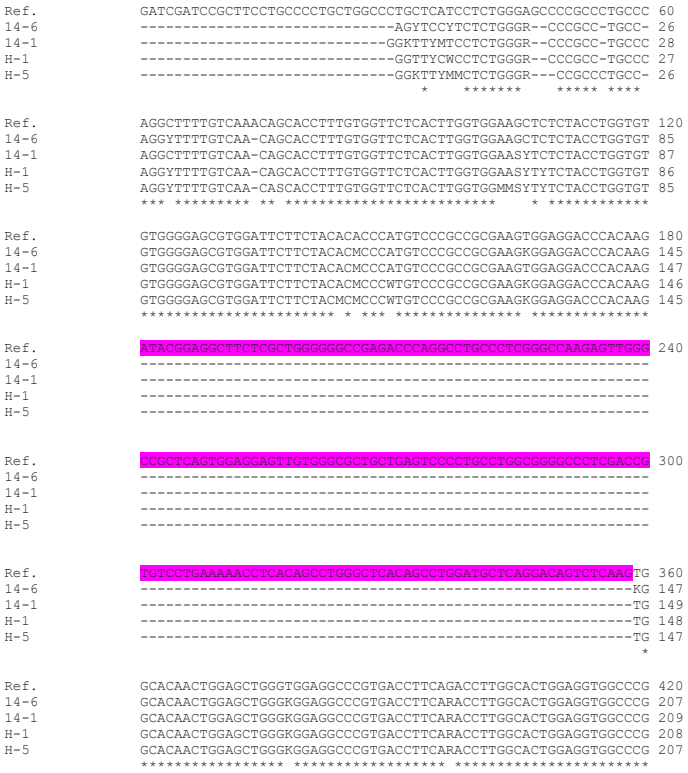


Figure 4.10
Alignment of the band of size 240 base pairs, present in all samples, with the reference sequence (Ref.). The reference sequence contains only exonic sequence: the vector exons with the insert exon (exon 32 - shown here highlighted in pink) in-between them. This shows that the band which is the major product in both major allele (14-6 and 14-1) and minor allele samples (H-1 and H-5) does not contain the insert exon, indicating it has been spliced out.

Figure 4.11
Alignment of the band of size 420 base pairs, present in only the samples containing the minor allele, with the reference sequence (Ref.). Although these samples have not sequenced very cleanly (despite cloning the samples to attempt a clean sequence), they have sequence well enough to establish all expected exons are present. This suggests that exon 32 (highlighted in pink) is only included in the mRNA when the minor allele (G) of variant rs881768 is present.

Figure 4.12(A) shows the cDNA amplified from RNA extraction from brain tissue samples M647 and M648. It can be seen that there is a no difference between cDNA harbouring the major allele (A) and minor allele (G), although there is genomic DNA present, shown as a faint band at approximately 1000bp. This was proven to be genomic DNA as it is also present weakly in the enzyme negative reactions. This genomic DNA is of the size expected if introns are included in this PCR amplification. The smaller band of around 400bp was also extracted from the gel and Sanger sequenced. Upon alignment with the exonic reference sequence (Figure 4.12(B)), it can be seen that all *ABCA7* exons are present in both samples.

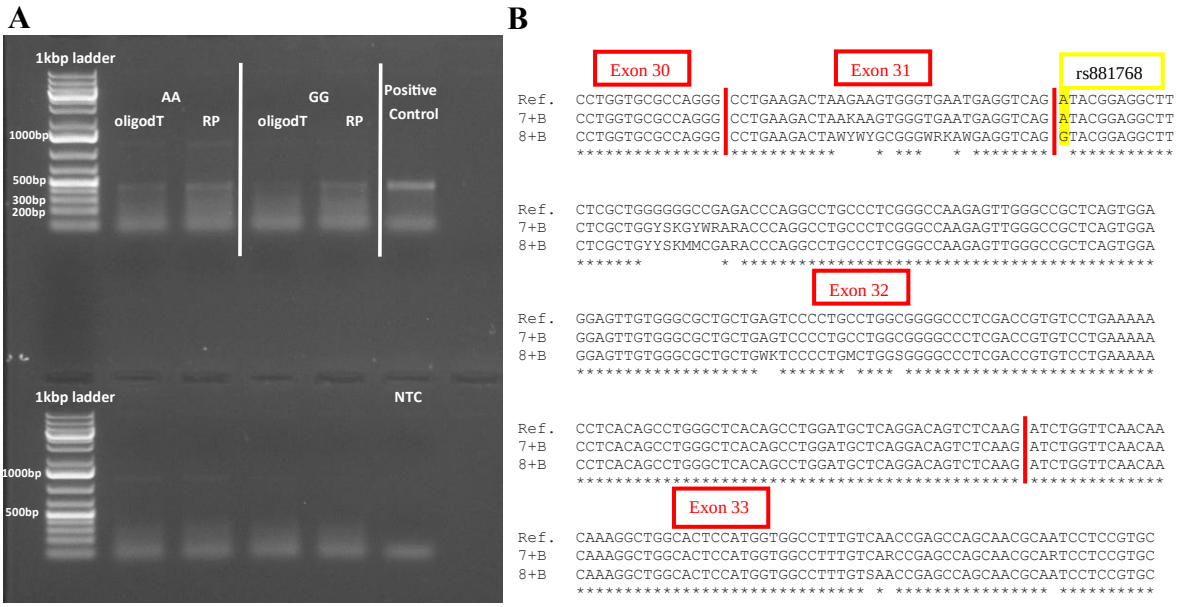


Figure 4.12

(A) cDNA PCR amplified from RNA extracted from brain tissue carrying both the major (A) and the minor (G) alleles for the splicing variant rs881768. Both samples had total cDNA synthesised utilising both the oligodT and random primers (RP) and both were PCR amplified. A positive control cDNA sample (made from RNA extracted from cells) was also PCR amplified in order to ensure the PCR was working. As can be seen, the products between the two brain tissue samples were identical with the major product being around 400 base pairs, the size expected when all *ABCA7* exons are splicing into the cDNA product. This band was extracted from the gel and Sanger sequenced, confirming it did contain all of the exons expected. There was also a faint band of size 1000 base pairs, the correct size of genomic DNA (i.e. if introns were included) for this region. This was also present in the enzyme negative reactions (bottom row), proving that it was, in fact, genomic DNA.

(B) Alignment of the band of size of around 400 base pairs, present in both samples carrying the major allele (7+B) and the minor allele (8+B) . This shows that exon 32 was present in both.

4.5 Discussion

This work aimed to identify if the minor allele of a variant, rs881768-G, present at the very beginning of exon 32, affected the splicing of the *ABCA7* gene. It had been highlighted by annotation programs during analysis of NGS data and further scrutiny by a variety of programs supported the fact that it may result in alternative splicing of ABCA7 (see Table 4.1). Minigene assays in two different cell lines (COS-7 and BE(2)-C) showed that only the minor allele forced inclusion of exon 32 while RNA extracted from brain tissue showed that exon 32 was included in all transcripts of *ABCA7*, irrespective of rs881768 genotype.

Unfortunately current *in silico* tools aiming to identify damaging functional variants (whether this functional effect be through splicing or other means) are still imperfect, therefore further *in vitro* validation methods are required. The methodology utilised - the minigene splicing assay - has previously been identified as accurately representing the pathogenicity of possible splicing variants (Steffensen et al., 2014) although some argue that the absence of the flanking sequences in the minigene construct may alter the spliceosome formed (Cooper, 2005). However, in this study we have also examined total RNA extracted from brain tissue, where the full length pre-mRNA is present and splicing takes place in its physiological context.

The cell lines utilised here were also chosen carefully. The COS-7 cell line is known for its high transfection efficiency as well as accurately representing the complex eukaryotic splicing environment seen *in vivo.* The BE(2)-C cell line was selected to test the tissue-specificity of splicing, being a neuroblastoma cell line, as different cell lines may exhibit different ratios of alternatively spliced products, therefore being unrepresentative of the regulation occurring in the tissue of origin (Cooper, 2005).

Utilising a neuronal cell line, therefore, will more accurately represent the effect this variant may have on AD pathology.

The synonymous variant rs881768 is located at the first base of exon 32 in *ABCA7* (transcript ID NM_019112). Although it is not thought that the first base of an exon is directly involved in the mechanisms of splicing (see Figure 4.1), *in silico* predictions suggest that this A to G allele substitution creates a new donor site with a similar score to the original donor site, located only 126bp away (a score of 0.98 compared with 0.99) and changes the binding of two ESE proteins (Table 4.1). These predictions suggest the G allele could activate a cryptic splice site or cause a change in protein isoform ratios. Additionally, score predictions for the acceptor site shows that the A allele has a lower score than the G allele (0.56 compared to 0.76) which could lead to exon skipping. This makes *in silico* functional prediction and interpretation of this variant very difficult. Population data from 1000Genomes shows that the G allele has an allele frequency of 0.44 (combined across all populations although when this variant was analysed in the NGS dataset utilising the CRISP tool (see Section 2.1), the minor allele was identified as having a frequency of 0.25. The discrepancy in these allele frequencies could suggest that this variant may also have some kind of pathological effect, perhaps protective of disease due to the lower frequency in the disease population.

ABCA7 has an interesting pattern of alternative splicing, with many introns consisting of multiples of three, potentially allowing in-frame addition and deletion of introns, although no such transcripts have been identified (Kaminski et al., 2000b). However, elimination of exon 32 does results in a frame-shift change creating a termination codon and causing early truncation of the protein at amino acid position 1457, located within the subsequent exon (exon 33). Exon 32 codes for part of the extracellular loop between TMDs 7 and 8, shortening it. It is not known whether the deletion of exon 32

affects the binding areas, and therefore the function, of this extracellular loop. However, this truncated protein is non-functional as it lacks the last five transmembrane domains and the second nucleotide binding domain. A similar splicing variant has been identified in the ABCA1 protein upon analysis of patient obtained cDNA from a familial sufferer of low serum HDL and premature coronary heart disease. This variant is also located at the exon/intron junction of exon 32 and causes skipping of exon 32, introducing eight new amino acids and a premature stop codon. This produces a protein of 1496 amino acids in length compared to the normal ABCA1 protein of 2261 amino acids (Rhyne et al., 2009). rs881768 truncates ABCA7 in a similar manner - shortening the protein from the typical 2146 amino acids to 1457 when exon 32 is eliminated.

This shortening of the protein (quite dramatically due to ABCA7 having 47 exons and 2146 amino acids in total) may suggest, therefore, that a surveillance pathway, such as nonsense-mediated decay (NMD) may eliminate the truncated mRNA, causing sole expression of the complete isoform *in vivo*. NMD is a surveillance pathway present in all eukaryotes and identifies mRNA transcripts containing premature stop codons which may produce damaging isoforms. It goes on to eliminate these isoforms, limiting the translation of abnormal proteins. This would explain why, when total RNA from brain tissue is examined, only the full isoform is present, as these surveillance pathways, or secondary structures affecting transcription, are present in this environment (Wadkins, 2000). An alternative explanation may be that the deletion of exon 32 results in the protein breaks down naturally, with no premature stop codon introduced, therefore making NMD unnecessary.

The primary limitation of this study is that the minigene methodology only examines the RNA produced in cell lines, with only one exon present, making it a very artificial system. In order to examine what occurs to exon 32 *in vivo*, RNA from brain sections

of carriers of both the major and minor alleles was examined, mapping it to the location of LOAD pathology. However, studies involving RNA from post-mortem tissue must always be viewed with caution. The handling and treatment of the tissue prior to it being stored in our laboratory at -80°C was unknown and, therefore, RNA might have degraded at source and consequently be un-representative of what occurs *in vivo*. Further studies could also be performed by creating clones containing more than just exon 32, increasing the amount of the natural splicing machinery in the minigene assays.

Further issues, especially with this particular assay, may be the accuracy of the RT-PCR. This may mean that the ratios of the two PCR amplicons are not representative of the different mRNA splice variants or that low levels of transcripts are not even shown upon electrophoresis examination. This caveat can be addressed by removing aliquots of the PCR reaction following a range of cycle numbers and quantifying the bands. If they both increase linearly then the PCR is representative of the correct ratios.

Since the start of this project, new expression databases have become available, for example, an expressed sequence tag dataset known as GTEx (GTEx Consortium, 2015). When *ABCA7* isoforms are examined within this dataset, it can be seen that exon 32 is incorporated in all *ABCA7* transcripts, expressed at the same level (if not higher) than the surrounding exons (Version 6p, accessed October 2016). This was demonstrated in all neurological tissue this database incorporates (pituitary, cerebellum and cerebral hemisphere) as well as whole blood and spleen which had relatively high expression levels of *ABCA7* when compared to other tissues. Examining the effect of rs881768 on transcript data in GTEx shows the variant has no effect on splicing, corroborated this chapter's findings in brain tissue where exon 32 of *ABCA7* is incorporated regardless of the rs881768 genotype.

From this data it can be seen that, *in vitro*, variant rs881768-G, stabilises the splice site, increasing the inclusion of exon 32 in the mRNA. However, in the more natural environment, in the total RNA extracted from brain tissue and in the GTEx database, where the entirety of the gene is present, exon 32 is included in samples containing both the major and minor alleles. This suggests that, in the *in vivo* environment, something may stabilise the splice site, such as secondary structure interactions throughout the pre-messenger RNA during splicing (Wadkins, 2000).

Upon analysis of the protein when exon 32 is deleted, it is also apparent it is a frame shift deletion, truncating the protein within the subsequent exon. This is a remarkably shortened protein considering *ABCA7* contains 47 exons and may, therefore, suggest that a surveillance pathway, such as nonsense-mediated decay, may eliminate the truncated mRNA, causing only the full protein, including exon 32, to be expressed. Therefore, assays were performed in lymphoblastoid cell lines, containing the entire *ABCA7* genomic region, in order to inhibit nonsense-mediated decay to see if this is indeed the case.

4.6 Conclusions

Upon analysis of cDNA extracted from cells transfected with minigene vectors containing both the major and minor alleles of the rs881768 variant, it is apparent that only the minor allele forces the inclusion of exon 32 into the transcript. However, this is not the case when RNA is extracted and RT-PCR performed from brain tissue samples containing these two alleles. This suggests that there may be a surveillance pathway in place, controlling which isoform of ABCA7 is expressed. Further assays will therefore be performed in order to examin the possible mechanisms of these surveillance systems.

5 Nonsense Mediated Decay

Nonsense mediated decay (NMD) is a post-transcriptional mechanism, active in all eukaryotic organisms, in order to control the quality of mRNA translated. It acts as a surveillance pathway in order to eliminate any mRNA present which would result in abnormal transcripts, the most common situation being the presence of a premature stop codon (PTC) in the mRNA (Elliott, 2010). This would result in a truncated protein, due to premature termination of translation, which can cause deleterious gain-of-function or dominant-negative functions (Sonenberg, 2000). Not all PTCs have negative functions; they can encourage molecular diversity: cellular adaptability and improve viability (Elliott, 2010). They may be introduced through incomplete or inaccurate splicing, chromosomal translocations, deletions, insertions and point mutations. As mentioned previously, approximately 55% of human pre-mRNA undergoes alternative splicing, providing a dramatic increase in genomic diversity although this can lead to introduction of these PTCs, usually due to a frameshift effect (Maquat, 2004). Point mutations can also initiate alternative splicing, occasionally through altering a *cis*-acting splicing element but it is thought that the principal pathological action of splicing affecting nonsense mutations are due to them introducing a PTC. To emphasise this it is thought that around one third of inherited diseases are due to the premature termination of proteins (Elliott, 2010; Maquat, 2004; Sonenberg, 2000).

Two things are required in order to initiate NMD post-translation; a ribosome halted at a stop codon and a downstream *cis*-sequence (Elliott, 2010). In mammals, this *cis*-sequence is the exon-junction complex, occurring around 20 base pairs upstream of the exon-exon junction and at least 50 base pairs downstream of the termination

codon. These aspects of NMD and their locations are illustrated in Figure 5.1. This implies that, if the premature termination codon is present in the last exon of a gene (where there is no downstream exon-exon junction), NMD will not be initiated but as long as there is a "spliceable" intron downstream of the PTC, NMD will take place (Elliott, 2010; Maquat, 2004).

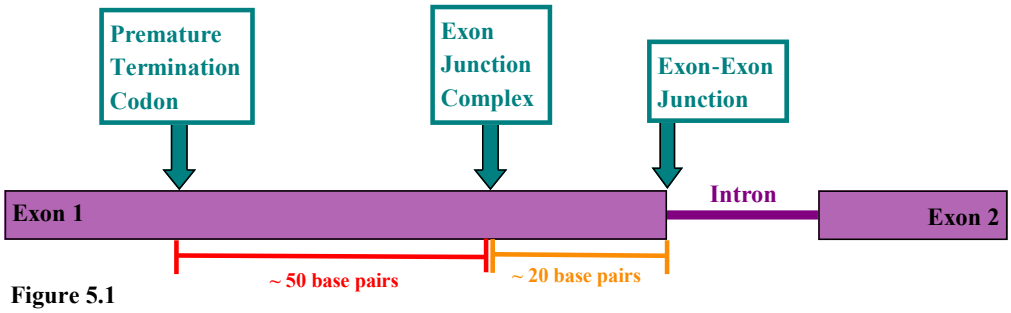


Figure 5.1
Pictorial representation of the *cis*-sequences required in order to initiate nonsense mediated decay. The premature termination codon exists approximately 50 base pairs upstream of the exon junction complex which itself is approximately 20 base pairs upstream of the exon-exon junction between this exon and the downstream intron. The premature termination codon stalls the ribososome, causing translation release factors to be liberated, recruiting protein complexes to the exon junction complex and stimulating nonsense mediated decay.

The exon-junction complex compromises of a protein core, anchoring it to the complementary mRNA sequence, and plays a part, not just in NMD, but also in mRNA export to the cytoplasm, translation and localization. This protein complex involves the Regulator of Nonsense Transcript (UPF) proteins, which are the main contributors to NMD, recruiting the mRNA degradation machinery. UPF1 is initially recruited upon translation release factors being liberated when the ribosome encounters a stop codon. It then interacts with UPF2 and UPF3, consequently phosphorylating UPF1 and forming the activated surveillance complex. This complex then recruits factors which initiate both 5' to 3' and 3' to 5' decay, either through removal of the 5' cap and then degradation by the exonuclease 1 enzyme or deadenylation from the 3' end. The importance of NMD in controlling and surveying correct gene expression is confirmed by the fact that the UPF proteins are conserved

throughout all eukaryotic organisms. Additionally mice embryos with NMD inactivated are fatally reabsorbed by their mothers (Elliott, 2010; Maquat, 2004; Sonenberg, 2000).

As shown previously, the variant rs881768 appears to alter the inclusion of the exon 32 in ABCA7. The G allele appears to be the only allele which causes the inclusion of this exon in the artificial minigene environment. However, when RNA from brain tissue which harbours both genotypes of this variant is examined, exon 32 appears to always be included. It is therefore postulated that a surveillance pathway, such as NMD, is degrading the mRNA not containing exon 32. This is thought to be as the deletion of exon 32, caused by the A allele of rs881768, is a frameshift mutation and causes a subsequent termination codon in the next exon, truncating the protein by 13 exons, including vital functional domains such as the final five transmembrane domains and the second nucleotide binding domain.

It is known that aminoglycoside antibiotics bind to ribosomes, affecting their fidelity and allowing stop codons to be read as amino acids, inhibiting NMD (as do a selection of other ribosomal binding drugs and molecules). Using these antibiotics, inhibition of NMD will therefore be attempted in cell lines containing known human genetic material in order to identify if NMD is controlling the expression of this truncated ABCA7 isoform. Epstein-Barr transformed B-lymphoblastoid cell lines (LCLs) containing genetic material from two 1000Genomes individuals identified as carrying the major allele (HG00255 - AA) and the minor allele (HG00137 - GG) were obtained after being identified using the 1000Genomes browser (accessed April 2014).

Puromycin is an aminoglycoside antibiotic which will inhibit NMD. It does this through inactivating the 60S ribosomal subunit, preventing the reading of the mRNA and, therefore, protein synthesis. Transcripts which undergo NMD are targeted by the

60s ribosome and, as such, treatment with puromycin will interrupt this process and prevent the transcript from being targeted by NMD (Carter et al., 1995). Following a literature search (performed in January 2015) it was confirmed that there is no commonality in the concentration and incubation period for puromycin to definitively impede NMD in cell lines. Several papers examined (Andreutti-Zaugg et al., 1997; Castellsagué et al., 2010; Lamba et al., 2003; Nguyen-Dumont et al., 2011; Noensie and Dietz, 2001; Ware et al., 2005) as well as personal communication with a colleague (Dr Anne Braae) showed that the puromycin concentrations used to successfully inhibit NMD ranged from 10µg/ml to 1000µg/ml. The incubation times also varied widely; from two to six hours. It was therefore decided that a range of both concentrations and time points would be used in order to ascertain the potential effect of NMD inhibition on this truncated form of ABCA7. These concentrations and time points reflect the conditions that have shown to be successful in the literature. There is a risk that, if the concentration or incubation time is too high, all of protein synthesis could be inhibited, killing the cells due to no proteins being made. Due to this, a positive control was also identified in order to ensure that NMD had been inhibited in these assays, consisting of a four base pair deletion in the *CD33* gene. Total RNA was extracted from both of the cell lines (HG00255 of the AA genotype and HG00137 of the GG genotype) and cDNA synthesised which was then PCR amplified, both for ABCA7 and for the positive control, and Sanger sequenced in order to establish the action of NMD.

5.1 Materials & Methods

5.1.1 Culturing Cells

The two LCLs (HG00255 and HG00137) were obtained from the Coriell Cell Repositories from the National Human Genome Research Institute Sample repository for Genetic Research. They were cultured in Roswell Park Memorial Institute (RPMI) 1640 media supplemented with 2mM L-glutamine, 15% FBS, 1% Fungizone and 1% Penicillin-Streptomycin (all Sigma). Cells were maintained in suspension in $25cm^2$ Corning Tissue Culture flasks in 15ml of media in a 37°C, 5% CO_2, humidified environment at a cell density of between 200,000 and 500,000 cells/ml. In order to transfer, passage and plate out the cells, the cell culture media was transferred to a 50ml tube before being centrifuged at 300 x g for 5 minutes. The spent media was removed and the cells were re-suspended in fresh media. All other basic cell culture methodology described previously was followed, as in Section 4.2.15.

5.1.2 Treatment of Cells with Puromycin

The cells were plated out into 12 well plates, at a concentration of 2×10^5 cells/ml with 2ml of cells per well. Four wells were seeded for each cell line in order to accommodate the different concentrations of puromycin being used (50µg/ml, 100µg/ml, 200µg/ml and 1000µg/ml) as shown in Figure 5.2. A plate was created for each of the four time points being examined - 0.5 hours, 2 hours, 6 hours and 10 hours. An extra plate was also created in order to extract control RNA to see what ABCA7 isoforms were produced in cells with no puromycin treatment. The cells were plated out and 48 hours later, once the cells were well established, the appropriate concentration of puromycin (Sigma) was added to each well. The plates were incubated for the appropriate period of time at 37°C, 5% CO_2 with humidity before

their RNA was extracted. Each plate for each time point was repeated in order to show the results were replicable.

Figure 5.2
Representation of the 12 well plates (denoted by the dark blue lines) created in order to examine the effect of nonsense mediated decay inhibition on the isoform of *ABCA7* transcribed in two lymphoblastoid cell lines - HG00137 and HG00255 - which contain different alleles for the rs881768 variant. Due to only two cell lines used, the third row of the plate was not used. Different concentrations of puromycin are utilised to inhibit NMD due to conflicting literature reports. Four of these plates are created in order to examine four time points of incubation (0.5 hours, 2 hours, 6 hours and 10 hours). These were all replicated twice.

5.1.3 RNA Extraction

Total RNA was extracted from all cell cultures following the RNeasy Mini Kit protocol (QIAGEN), as described in Section 4.2.17, before being quantified.

5.1.4 cDNA Synthesis

The AffinityScript Multiple Temperature cDNA Synthesis kit (Stratagene) was used to synthesis total cDNA as in Section 4.2.18. Due to the low concentrations of RNA obtained, only 500ng of each sample was used to synthesis cDNA and only the random primers were utilised to synthesis the cDNA. These primers were selected (as opposed to the oligo(dT) primers) due to them synthesising better quality cDNA from total RNA from brain (see Section 4.3.2 and Figure 4.12(A)). Negative synthesis reactions (with the enzymes eliminated) were also performed to determine if genomic DNA was present.

5.1.5 cDNA Polymerase Chain Reaction

PCR was performed as in Section 4.3.3 utilising primers designed to be within exons 30 and 33 of *ABCA7*, with the reverse primer being located prior to the PTC created within exon 33 upon elimination of exon 32 (see Table 4.2). All cDNA synthesised, including the negative reactions, was PCR amplified. The PCR conditions in Section 4.3.3 were followed and the results examined by gel electrophoresis (1% agarose gel run for 55 minutes at 80V). A selection of samples were also Sanger sequenced (at least two samples per time point were sequenced) in order to confirm whether exon 32 was included or not. This sequencing (utilising the forward primer located in exon 30), and analysis of the sequencing results, was performed as previously described.

5.1.6 cDNA Polymerase Chain Reaction - Positive Control

A colleague (Braae, 2016), had previously used these cell lines to examine a variant in the *CD33* gene (rs201074739) which codes for a four base pair deletion (CCGG/-), creating a PTC and therefore creating a target for NMD. This project involved synthesising cDNA from these cell lines treated with 200µg/ml of puromycin for six hours and demonstrated that NMD does, in fact, regulate the form of *CD33* expressed, eliminating the isoform containing the four base pair deletion due to it creating a PTC. The primers used in this study (Forward: ACAGGCCCAAAATCCTCATC and Reverse: CTGTAACACCAGCTCCTCCA) were therefore used to amplify the cDNA synthesised in Section 5.1.4 in order to see if the puromycin treatments in Section 5.1.2 had, in fact, inhibited NMD. The cDNA from both cell lines not treated with puromycin, as well as from cells treated with 200µg/ml of puromycin for 6 hours were PCR amplified in a 30µl reaction volume consisting of 1x Roche PCR Buffer (750mM Tris-HCl, 200mM $(NH_4)_2SO_4$, pH 8.8), 0.2mM dNTPs, 1pmol/µl of the primers stated above and 1U of Taq DNA Polymerase (Roche). The following thermocycle was executed: initial denaturation was performed at 94°C for 2 minutes,

and then a further 30 seconds at 94°C, 59°C for 30 seconds and 72°C for 1 minute repeated 30 times. A final extension step was set at 72°C for 7 minutes. The products were examined under electrophoresis before being Sanger sequenced as previously described utilising the forward primer as above, at a concentration of 0.5pmol/µl.

5.2 Results

Figure 5.3 depicts the *ABCA7* cDNA PCR amplified from the half hour incubations of the LCL cells with puromycin. This gel is identical to ones produced for all of the remaining time points (two hours, six hours and ten hours) with the exception that the bands are slightly fainter as the time course increased due to a lower concentration of RNA being extracted from the cells. This is thought to be due to the fact that the puromycin, when exposed to the cells for longer periods of time, especially at higher concentrations, begins to kill the cells due to it inhibiting ribosomes. It is also representative of cells not treated with puromycin. Figure 5.3 and the other gels acquired for the remaining time points show that only one PCR product is produced for all concentrations of puromycin. This PCR product is the correct size to contain exon 32 and this was confirmed upon Sanger sequencing.

Two samples per time point were sequenced (usually the ones which had PCR amplified best upon electrophoresis examination) and this sequence was aligned to the reference sequence. All sequences aligned to the sequence for exon 32, showing it to be included in all samples. This sequence was also very clear, with minimal background noise, suggesting that an isoform not containing exon 32 was not even present at low levels, as demonstrated in Figure 5.4. This was also apparent in cDNA extracted from cells not treated with puromycin; indicating exon 32 is present in the ABCA7 isoform produced naturally by these cells, irrespective of the rs881768

genotype. All of these results were replicated in two puromycin treatments performed for each time point and each concentration.

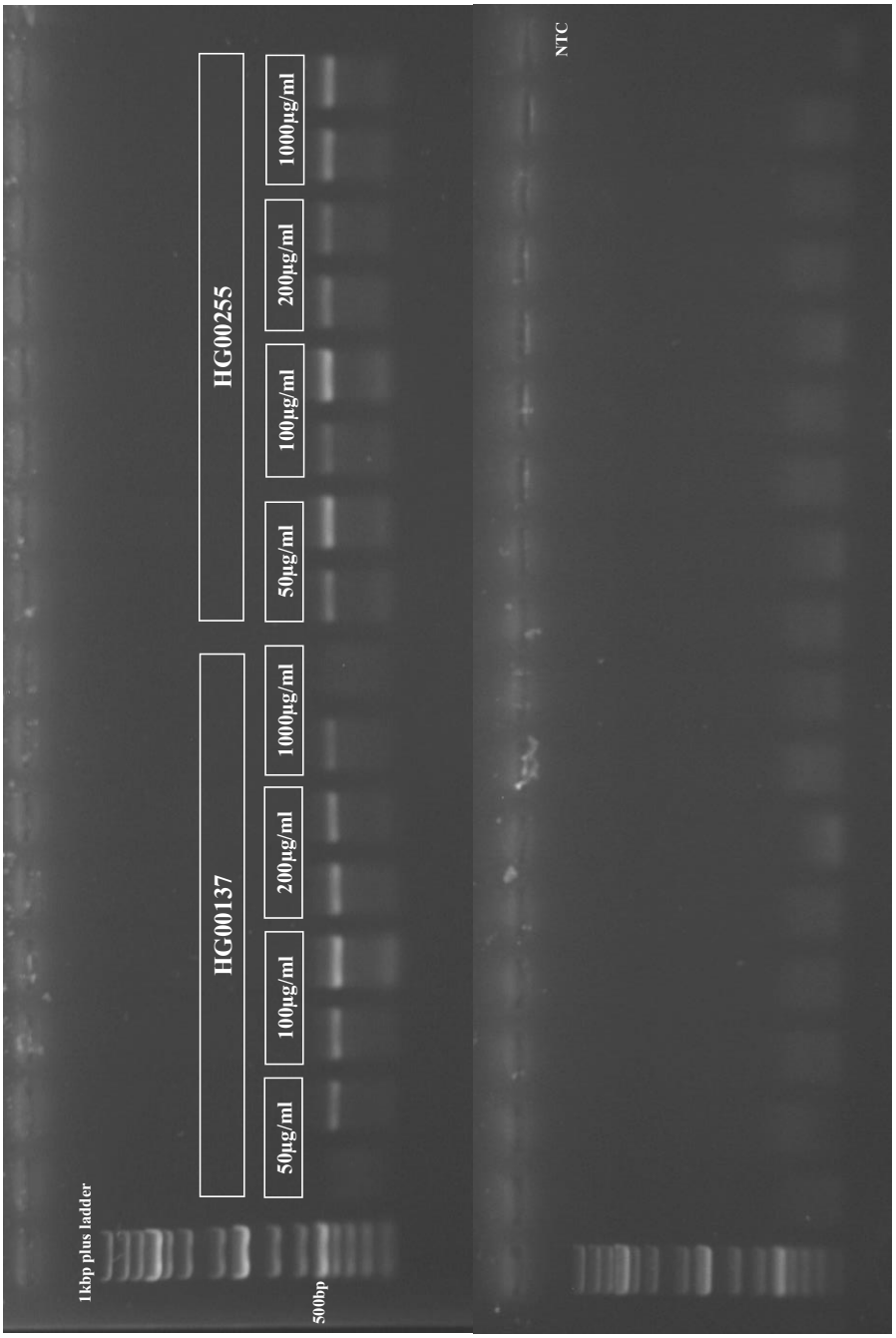


Figure 5.3
Electrophoresis gel presenting the cDNA synthesised from total RNA extracted from LCL cells incubated with puromycin for half an hour. Each puromycin concentration was repeated twice in both cell lines (with HG00137 containing the minor allele for the rs881768 variant). The bottom gel shows the enzyme negative reactions performed for each of the samples shown on the top gel, as well as a no template control (NTC), demonstrating no genomic DNA or contaminants are present. Only one product is visible at all concentrations, in both the replications, attesting that the truncated isoform of *ABCA7*, with exon 32 absent, is not expressed in these cells even with nonsense mediated decay inhibited. This gel is identical to ones produced when cells are incubated for two hours, six hours and ten hours, as well as when no puromycin treatment is performed.

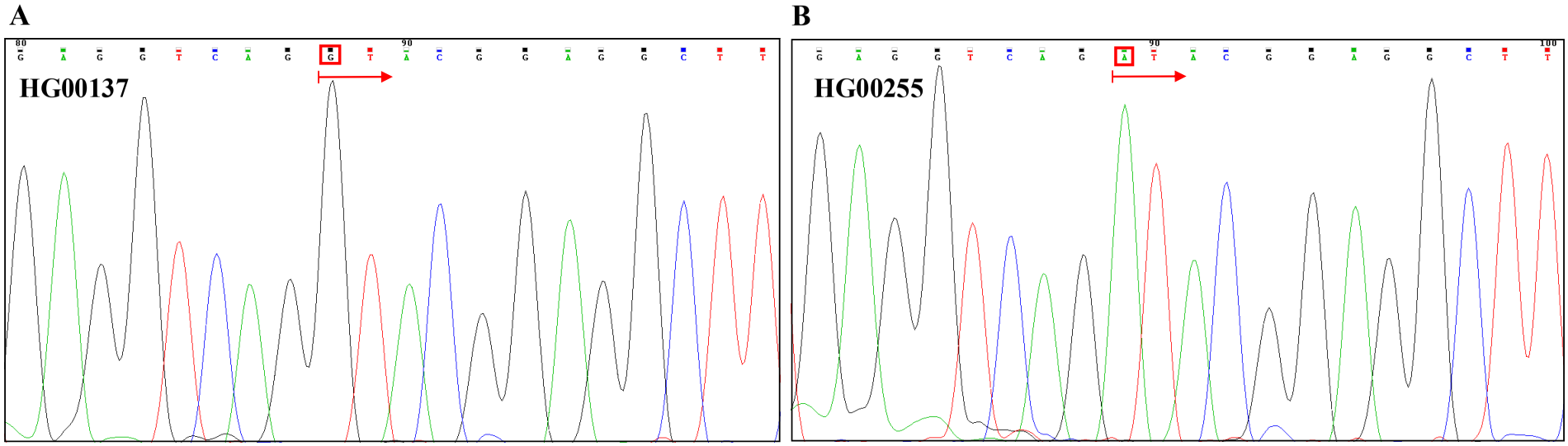


Figure 5.4
Sequence of the PCR amplified cDNA from cells treated with 1000µg/ml of puromycin for ten hours. Figure (A) shows the sequence from cell line HG00137 containing the minor allele for this variant (G) and figure (B) shows the sequence from cell line HG00255 which contains the major allele (A) for variant rs881768 (also highlighted in red). The start of exon 32 is denoted by the arrow in both sequences, showing that neither contain an isoform of ABCA7 where exon 32 is spliced out, even at low levels, as the sequence is very clean.

The electropherogram in Figure 5.5 shows the cDNA amplified with the positive control primers in both cell lines when they were not treated with puromycin. Here it is apparent that the four base pair deletion was not present. However, in cells treated with 200µg/ml of puromycin for 6 hours, the deletion becomes apparent in the HG00137 cell line which is heterozygous for this variant (see Figure 5.6). This validated the experimental approach performed in Section 5.1.2, demonstrating that NMD has been inhibited in these assays, suggesting NMD is unlikely to be responsible for eliminating the truncated form of ABCA7 with the 32nd exon spliced out.

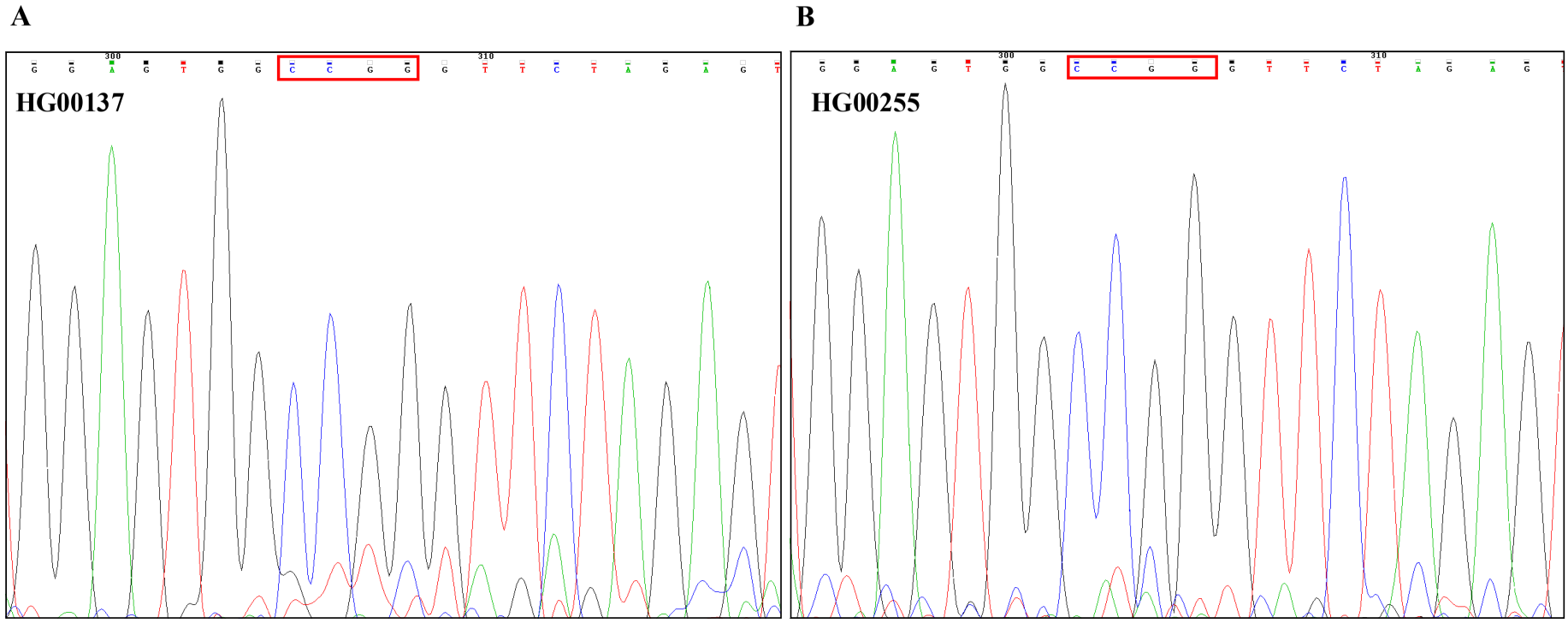


Figure 5.5
Electropherogram of both the HG00137 (A) and HG00255 (B) cell lines with no puromycin treatments. The HG00137 cell line is heterozygous for the rs201074739 variant in the *CD33* gene. This variant is a deletion of the CCGG fragment highlighted in both images, however, only the transcript containing these four bases is expressed, suggesting it has been degraded through NMD. The heterozygous deletion is present when genomic DNA from the HG00137 cell line is examined (data not shown).

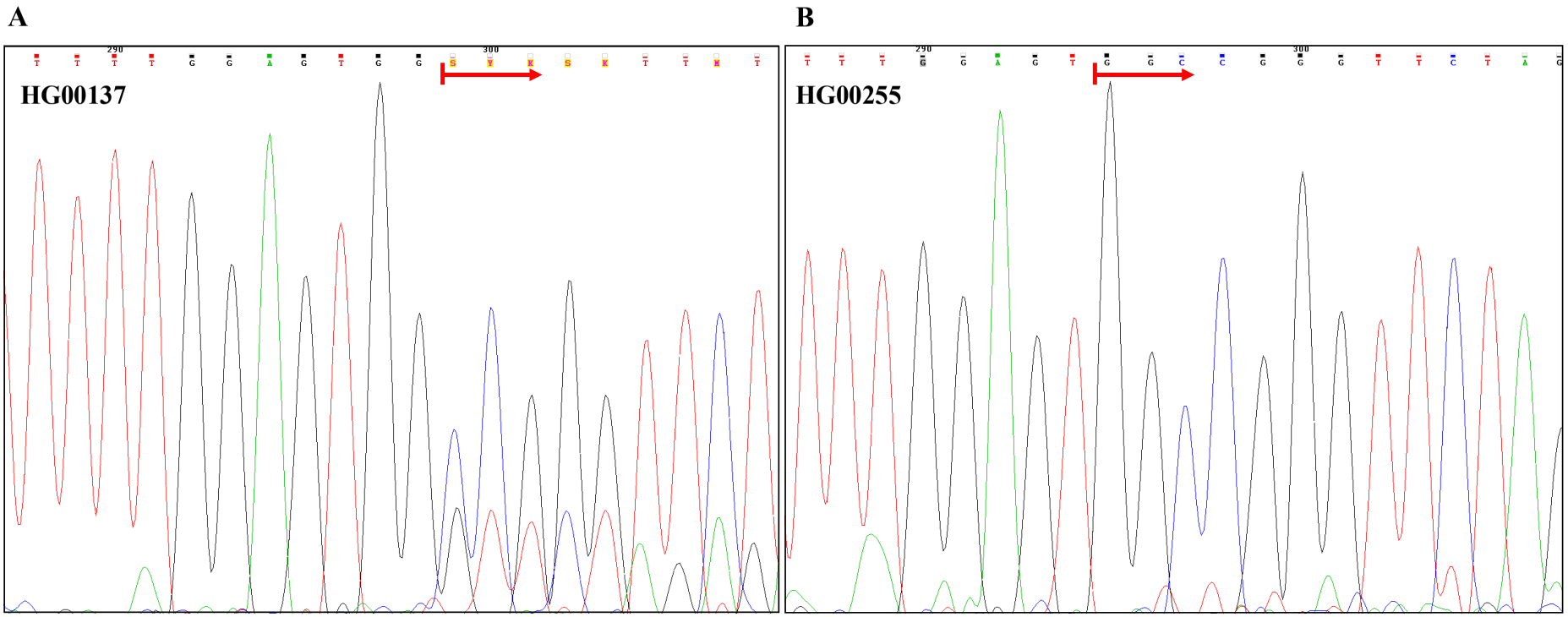


Figure 5.6
Electropherogram of both the HG00137 (A) and HG00255 (B) cell lines after treatment with 200µg/ml of puromycin for 6 hours. This shows that the cDNA extracted from the HG00137 cell line, containing the CCGG deletion in a heterozygous fashion, now exhibits two isoforms: one containing the deletion and one not, shown by the double sequence from the point of the mutation onwards (highlighted by the red arrow in (A)). The HG00255 cell line however, does not show this double sequence and only exhibits one transcript as shown in (B). This shows that the puromycin treatments done previously in this chapter have successfully inhibited NMD.

5.3 Discussion

Upon examination of cDNA extracted from lymphoblastoid cell lines containing both the minor allele (G) and major allele (A) of the variant rs881768, exon 32 of ABCA7 is included in all transcripts, contradicting the results obtained in Section 4.4 where this exon was only included at low levels in the presence of the G allele. These cells were therefore treated with puromycin at a variety of concentrations and incubation periods in order to see if the surveillance pathway nonsense mediated decay controls the isoform expressed in the *in vivo* environment. As can be seen in Figures 5.3 and 5.4, puromycin treatment had no apparent effect on the isoform of ABCA7 expressed, indicating that either NMD is not the surveillance pathway controlling this expression pattern or that the rs881768 variant does not regulate this isoform ratio. This was despite NMD being shown to be inhibited in these assays by PCR amplifying an area known to be controlled by NMD and showing a difference between cells treated with puromycin and those which had not been.

Alternative treatments of the cells could be attempted to see if different compounds, for example cycloheximide, would perhaps act in a slightly different mechanism in order to inhibit NMDs action on this isoform (Cuyvers et al., 2015). However, from this data, it would appear that NMD may not be regulating gene expression in this instance, at least not in LCLs. There are several alternative surveillance pathways, for example, non-stop decay and no-go decay. However, these are unlikely to act in this situation as they are known to be activated by a ribosome stalled at the 3' end of the mRNA (due to the absence of stop codons) or stalled at non-functional ribosomes respectively (Doma and Parker, 2006; Vasudevan et al., 2002). It appears, therefore, that the splicing out of exon 32, seen in Figures 4.8 and 4.9, may be artefact of the minigene assay and the presence of the whole *ABCA7* genomic region, as seen Figures 4.12 and 5.3, create some form of secondary structure interaction, forcing the

inclusion of exon 32 in the ABCA7 isoform. This could be examined further by designing alternative minigene constructs including more of the *ABCA7* gene, for example including the exons either side of exon 32.

Aberrant splicing may still play a part in LOAD pathology; therefore, the splicing of GWAS genes does still require investigation. However, the methodology used in this chapter is clearly able to generate false positive results and therefore alternatives need to be found. Exon arrays or RNA-seq are examples of this, both high throughput technologies examining alternative isoforms in the tissue of choice, in this case the brain.

5.4 Conclusions

Within the minigene assay system, the minor allele of rs881768 (G) is the only genotype which incorporates exon 32 into the *ABCA7* isoform due to it strengthening the acceptor site at the start of this exon. However, when the total RNA is examined from brain tissue samples and transformed cell lines exhibiting both genotypes, exon 32 is always incorporated. This is apparent even when the surveillance pathway nonsense mediated decay is inhibited, suggesting that this truncated isoform is not expressed at all *in vivo*. The truncated isoform seen in the minigene assays may, therefore, be an artefact of this particular methodology. This suggests that the *in silico* prediction programs used to predict splice site variants may not be entirely reliable and more dependable methodologies are required in order to identify potential pathological splice site variants.

6 Dual Luciferase Assays

6.1 Introduction & Background

As previously described, the primary aim of a Genome Wide Association Study (GWAS) is to identify genetic variations which are associated with a particular phenotype, for example, LOAD. However, as many as half of the variants identified are not located in, or near, any coding genes. So the question remains as to how do they actually impact on pathology (Schierding et al., 2014)? The matter is further convoluted by complex linkage disequilibrium (LD) plots between the GWAS tag SNPs and any actual causative variants, possibly genomically distant from these tag SNPs. However, it has been noted that a large proportion of loci identified through GWAS for complex diseases are, in fact, expression quantitative trait loci (eQTLs) implying that gene regulation may play a role in the risk of these complex disorders (Albert and Kruglyak, 2015; Li et al., 2016; Zhu et al., 2016). The fact that so many of these variants are within "gene deserts" supports this theory as, despite the fact that they may be distant linearly from any gene, DNA folding and interactions with proteins, mean they can act to regulate multiple distant genes (Schierding et al., 2014). In fact, recent studies have substantiated the idea that the causal genes behind GWAS signals are frequently not the nearest gene to that signal (Zhu et al., 2016). How, therefore, do these variants regulate gene expression?

Virtually any step of gene expression can be controlled, with the process of promoter activity to transcription to mRNA expression to translation to protein expression being a sequential cascade with all steps being regulated (Li et al., 2016). However, the first step - that of transcription of DNA to RNA - is modulated through a wide variety of

actions: transcription factors binding to specific DNA sequences; chromatin accessibility; DNA methylation; alternative splicing; small or long, intronic, non-coding RNA; RNA editing as well as mRNA degradation. Transcription factor (TF) proteins bind to consensus DNA sequences in order to initiate mRNA transcription and are therefore vital in controlling many cellular processes. They contain a sequence-specific DNA-binding domain as well as other domains in order to interact with other proteins, for example, recruiting RNA polymerase enzymes or other TFs and co-factors, in order to initiate transcription (Rice and Correll, 2008). Commonly it is the promoter region of DNA which is responsible for initiating transcription of its gene, located on the same strand and upstream of the gene, usually no more than 1000bp away. There are also enhancer elements, generally *cis*-acting, either upstream or downstream. Repressor and silencer regions also exist with the repressor site located close to or even overlapping the promoter region in order to block transcription when regulated by certain transcription factors. Silencer sequences mimic this but can be located elsewhere, although it is usually upstream of the gene in question (Albert and Kruglyak, 2015). RNA processing and transport is also highly regulated. Despite the fact that 40% of the human genome is transcribed to RNA, less than 2% goes on to be translated to proteins, highlighting the importance of these regulatory pathways (Rice and Correll, 2008). The flexibility of this gene regulation means that gene expression can be highly specific, as well as highly diverse, altering between cell types, even altering during differentiation of related cell types, and acting over large distances. It is also highly complex, with the distal regulatory components for just one cell equating to tens of thousands of elements, something especially vital in complex tissues such as the brain (Kitchen et al., 2014).

Many eQTLs are SNPs which affect the binding of TFs or chromatin function, altering the DNA secondary structure and therefore the accessibility of the TF binding sites (TFBS), within these enhancer or promoter regions (Albert and Kruglyak, 2015).

These eQTLs may, therefore, alter protein levels even though they are often physically or genomically distant (even up to 2Mbp away in the human genome) from the gene in question (Albert and Kruglyak, 2015). In fact, only 25% of eQTLs are thought to act on the gene closest to them and only 50% on sites within 50kbp of their location (Schierding et al., 2014; Stamatoyannopoulos, 2016). It is thought that the majority of variants affecting gene regulation do so through their effect on TF binding, more commonly insertion or deletion variations as opposed to single nucleotide polymorphisms, or their effect within promoter or enhancer sites (Albert and Kruglyak, 2015).

Due to this degeneracy, it is remarkably problematic to attempt to identify variants which may alter gene regulation and, subsequently, phenotype. There have been an increasing number of studies attempting to piece GWAS data together with expression data in an attempt to fill this gap (Gusev et al., 2016; Zhu et al., 2016). A systemic genomics approach, involving genetic, transcriptome, proteomic and phenotypic data would give a much more thorough and complete biological picture (Ritchie et al., 2015). Identifying eQTLs through studies such as eGWAS (GWAS complimented with expression data) will also make identifying the disease associated gene much easier due to the mRNA levels being from a single gene as opposed to a GWAS loci highlighting one LD block (Zou et al., 2012). Examining gene expression also provides an "intermediate phenotype" between genetic variation and the higher phenotype - particularly useful in conditions such as LOAD where diagnosis cannot be confirmed until post mortem, making any results much more biologically interpretable and clinically relevant (Gamazon et al., 2015; Gusev et al., 2016). However, mapping these eQTLs in order to analyse them in combination with genetic studies is no easy feat, the primary issue being that eQTLs are very much tissue (and even cell type) specific. They therefore need to be mapped in all biologically relevant tissues in order to achieve this.

Most expression studies mentioned here, and in the further literature, all focus on mRNA levels as a measure of gene expression. However, protein levels also need to be considered as they could be more clinically accurate and can themselves be independently regulated, as mentioned earlier. Unfortunately specimen availability and costs of these expression studies make projects like this rare, although costs are decreasing as well as technology improving, possibly making these studies much more accessible in the future (Ritchie et al., 2015; Zou et al., 2012). Despite these problems, the importance of performing such studies has become apparent. As previously mentioned, there is a significant enrichment of regulatory variants within disease associated SNPs. In a recent study, approximately 80% of the chip-based heritability of disease risk for 11 different complex diseases are present in deoxyribonuclease (DNase) I hypersensitivity sites which play a role in chromatin accessibility and, therefore, transcription, highlighting the importance of gene regulation in disease processes (Gamazon et al., 2015; Zou et al., 2012). This is despite the fact that, when expression GWAS are carried out on a variety of brain tissue samples, it has been estimated that only a maximum of 18% of expression variance is due to the "best" eQTLs. (Zou et al., 2012). The remaining variation in gene expression is predominantly caused by environmental components or due to alteration by the trait itself in a reverse causal effect (Below, 2016; Gamazon et al., 2015; Zou et al., 2012). However, this 18% is particularly vital in neurological disorders as regulation of gene expression is essential in brain function due to its complex structures, sub-structures and cell type varieties (Kitchen et al., 2014).

Gene regulation has already been shown to play a role in LOAD pathogenesis. MicroRNAs (miRNAs - noncoding and involved in post-translational gene regulation by interacting with mRNA and silencing genes) have been shown to differ in expression between AD cases and controls and, consequently, affect the expression of some AD related proteins such as BACE1 (Shewale, 2012). There has also been a

huge range of genes that have been shown to be differentially expressed between AD brains and controls (Bai et al., 2013; Golde, 2016; Nho, 2016).

Regulation of the ABCA protein family has also been linked with disease: an eQTL shown to alter *ABCA1* expression levels has been associated with coronary atherosclerosis, a minor form of Tangier disease which *ABCA1* loss-of-function mutations are known to cause (Kyriakou et al., 2004). Regulation of *ABCA7* itself has also been linked to LOAD. Previous studies have suggested that *ABCA7* mRNA levels are elevated during AD and cognitive decline (Karch et al., 2012) and the GWAS tag SNP rs3764650 is itself negatively correlated with *ABCA7* mRNA levels in human brain tissue (Vasquez et al., 2013). Other SNPs within the *ABCA7* locus have shown to significantly increase *ABCA7* mRNA levels in human cerebellum and temporal cortex samples (Allen et al., 2012). In fact, in a large parallel re-sequencing project, a low frequency intronic SNP within *ABCA7* (rs78117248) was the most significant associated variant (Cuyvers et al., 2015). This SNP is in LD with all three of the GWAS SNPs (rs3764650, rs4147929 and rs3752246) and seems to account for all of the disease risk presented by these tag SNPs due to its association with LOAD remaining after these three minor alleles were corrected for in a conditional logistic regression test (Cuyvers et al., 2015). What is even more interesting is that this SNP is located in an area of DNase I hypersensitivity and in a transcription factor biding region. When analysed using RegulomeDB (a database utilising features from the ENCODE project to score SNPs falling in regulatory elements (Boyle et al., 2012)), a score of 4 was assigned, indicating this variant has slight regulatory potential. Could, therefore, dysregulation of messenger levels of *ABCA7* be the mode of action for pathogenic mutations within this locus?

In order to identify these putative, possibly pathogenic, regulatory variants, colleagues in the Mayo Clinic, Jackson, Florida (including Dr Christopher Medway, Professor

Steven Younkin, Dr Mariet Allen, Dr Nilüfer Ertekin-Taner and Dr Minerva Carrasquillo), analysed data from RegulomeDB in order to highlight the most likely variations. One of the most promising results was that of rs2020000, scoring 1f in RegulomeDB indicating it is likely to affect binding of TFs and therefore be linked to expression of gene targets. This is mirrored in its annotation by HaploReg which, again, utilises chromatin state and protein binding annotations from ENCODE and the Roadmap Epigenomics projects (Ward and Kellis, 2011). In HaploReg rs2020000 shows an enormous amount of data for regulatory potential in brain regions, including the hippocampus (Ward and Kellis, 2011). However, it does not appear to be the causative variant of *ABCA7*'s association with LOAD due to it not being in linkage disequilibrium (LD) with the GWAS tag SNP (D` = 0.829 and r^2 = 0.013, see Figure 6.1).

Genomically rs2020000 is located upstream of *ABCA7* (see Figure 6.1), within the 3' UTR of the gene directly upstream of *ABCA7*: *CNN2*. Although it is located almost 1Kbp upstream of *ABCA7*'s defined promoter region (see Figure 6.1 and (Iwamoto, 2006)), the evidence for its regulatory potential is compelling. In Figure 6.2, taken from the UCSC Genome Browser (Kent et al., 2002), it can be seen that it is not only in an area of transcription factor binding (based on ChIP-seq data), it is also in a known area of DNase hypersensitivity and is linked to histone modification and chromatin binding in many cell lines, including neurological cell lines. This variant may be surprisingly far upstream of the gene of effect (see Figure 6.1). However, as previously discussed, eQTLs can be *cis*-acting and, therefore, up to 1Mbp away from their target gene. In fact, regulatory variants even further upstream of *ABCA7* have been computationally identified, for example, the variant rs3087680 at genomic position 1038290 on chromosome 19, a further 67 base pairs upstream of rs2020000 (19:1038223), as shown in Figure 6.1 (Nho, 2016).

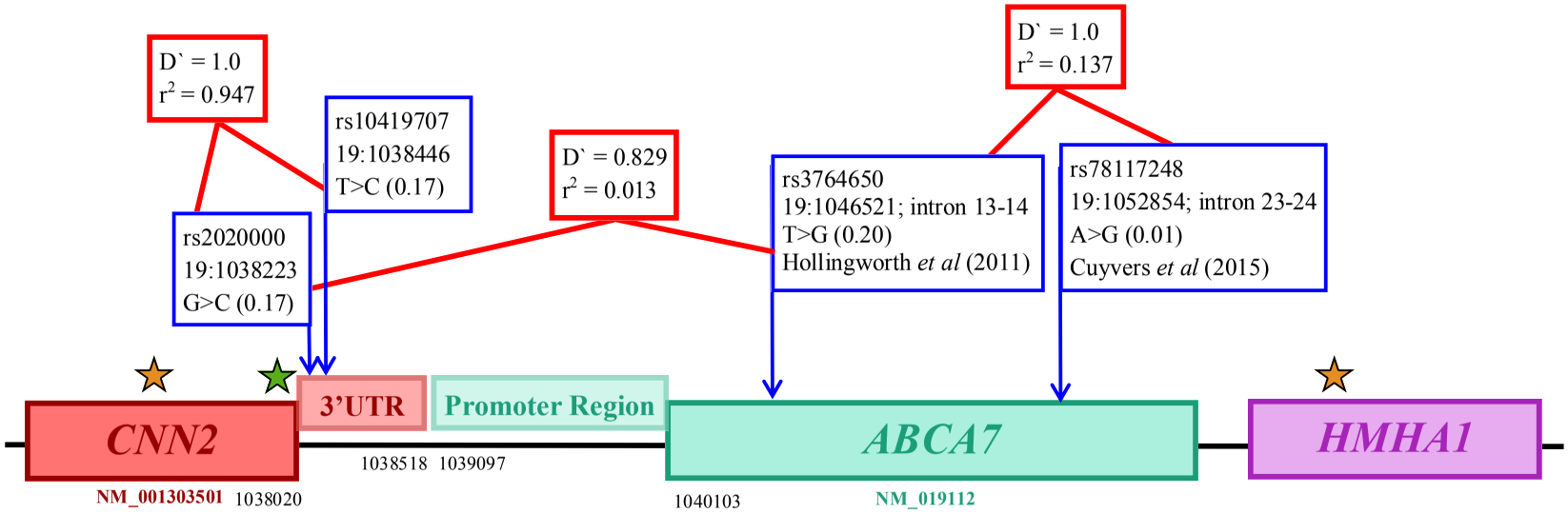


Figure 6.1

Representation of the genomic location of variant rs2020000 (position 1038223 on chromosome 19) which is a G to C change with a minor allele frequency of 0.17. This variant is predicted to alter the levels of *ABCA7* transcript but not levels of the *CNN2* transcript despite its location in that gene's 3' untranslated region (UTR). The documented promoter region of *ABCA7* is presented (Iwamoto *et al* 2006). This variant is in strong LD with the variant rs10419707 with a D` of 1.0 and r^2 of 0.947 (see Section 5.3) but not with the GWAS tag SNP rs3764650 (reported in Hollingworth *et al* 2011). Also shown is the relationship between rs78117248 and the GWAS tag SNP. This variant was reported in Cuyvers *et al* 2015 as being in strong LD with not only this GWAS SNP but two others in *ABCA7* (rs4147929 and rs3752246 - data not shown). It is reportedly in an area of regulatory elements indicating *ABCA7* regulation may be the mode of action for it's pathogenic association with LOAD. Several other putative regulatory elements have also been identified within *ABCA7*. The orange stars show the location of two computationally identified in Allen *at al* (2012) and the green star in Nho *et al* (2016). Genomic locations are relative to hg19.

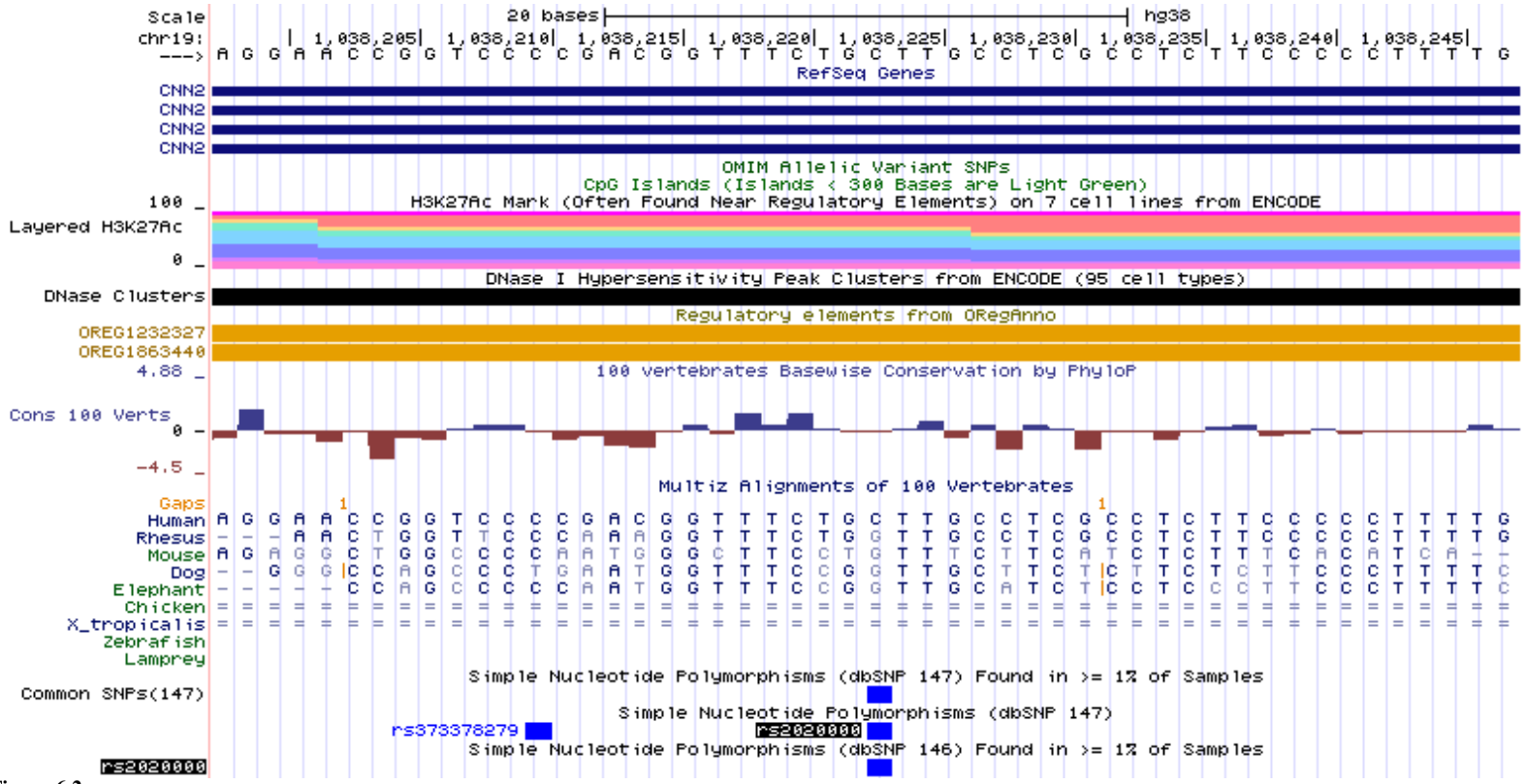


Figure 6.2

Genomic region where rs2020000 is located (highlighted on the bottom track with a blue box in the centre of the image). From this is can be seen that the variant is located in the *CNN2* gene and that it is in an area of H3K27Ac histone marks in all seven cell lines examined in the ENCODE project, including neurological cell lines (Layered H3K27Ac track). These marks are commonly found near regulatory elements and they demonstrate how accessible the chromatin is to transcription. The DNase clusters track (black) shows that this is an area of DNase I hypersensitivity, again commonly found in regulatory regions, especially promoters. The orange OREG1232327 and OREG1863440 tracks indicate that there are also transcription factor binding sites in this region. The multiple alignment shows that this genomic location (19:1038223) is conserved in all mammals present in this database, especially as the ancestral allele for this variant is G as opposed to C as presented here. Image taken from the UCSC Genome Browser (Kent *et al* 2002).

Variant rs2020000 was therefore genotyped in 271 cerebellum samples and 290 temporal cortex samples (as part of the Mayo eGWAS dataset, presented by Allen et al., 2012 with the same sample demographics, gene expression measurements and the quality control of raw probe levels as presented in this study). Genotype and mRNA levels were then correlated through the Whole Genome DASL assay (Illumina, San Diego, CA, USA), utilising a linear regression model in order to correct for both disease-related and technical covariates including diagnosis, age-at-death, sex, *APOE* ε4 allele dose, plate number and RNA integrity. This confirmed that there is a highly significant association between the rs2020000-C minor allele and elevated levels of *ABCA7* mRNA, confirmed by both *ABCA7* probes (ILMN_1743205 and ILMN_2259319) as shown in Table 6.1. For both probes the most striking effect was seen in cerebellum samples but the direction of the effect was comparable in temporal cortex tissue. Due to the genomic location of this variant being within the 3' UTR of the *CNN2* gene, association between rs2020000 allele dosage and *CNN2* mRNA levels were also examined. However, these results were not significant due to the *CNN2* probe on the WG-DASL microarray (ILMN_1770290) harbouring a polymorphism, this association did not pass quality control so no definitive comparison could be made (Zou et al., 2012).

Table 6.1
Fold increase of mRNA per allele dosage of the minor allele (C) of rs2020000. *ABCA7* mRNA levels were assayed utilising the probes ILMN_1743205 and ILMN_2259319 in cerebellum and temporal cortex samples in the Whole Genome DASL assay as part of the Mayo eGWAS dataset (previously described in Allen et al 2012). It can be seen that in all brain regions, both probes presented a significant association between this minor allele and an increase in *ABCA7* mRNA expression.

	Cerebellum	**Temporal Cortex**
ILMN_1743205	$p = 1.19e^{-10}$ Fold = 1.16 - 1.31	$p = 2.36e^{-03}$ Fold = 1.04 - 1.17
ILMN_2259319	$p = 8.39e^{-04}$ Fold = 1.06 - 1.23	$p = 4.39e^{-02}$ Fold = 1.00 - 1.21

Reporter gene assays were therefore designed in order to ascertain the true regulatory effect of this variant. These assays are based on the measurement of a reporter gene protein, in this case the bioluminescent firefly luciferase protein, following transfection of vectors coding for this protein as well as containing the DNA fragment under investigation. This reporter gene is not produced by mammalian cells and therefore should give a representative measurement of how this DNA sequence alters gene expression. However, inter-experimental variations are still likely to affect expression levels, for example cell number, pipetting error, transfection efficiency, cell lysis efficiency and assay efficiency. This can be minimised by utilising a second reporter vector, in this case the *Renilla* luciferase gene within a second vector, in order to normalize the activity of the experimental reporter protein with respect to experimental errors as well as adjusting for well-to-well variability and efficiency.

The dual luciferase reaction assay system (Promega) was therefore utilised in order to compare promoter activity between the major allele (G) and the minor allele (C) promoter-luciferase constructs of this variant, in order to establish its regulatory nature *in vitro*.

6.2 Materials & Methods

6.2.1 Identification of Positive Controls

Utilising the Perl script written by Christopher Medway (as described in Section 3.2.1), samples were obtained from the ARUK DNA Bank which contained the two alleles in homozygous fashions for the variant rs2020000. Several of each genotype were selected to be PCR amplified and sequenced in order to ensure they only differed at the location of the variant. All samples within this DNA bank were collected with ethical approval and consent of all individuals.

6.2.2 Primer Design

Primers were designed in order to amplify approximately 250bp either side of the variant, creating an insert of around 500bp. The methodology described in Section 3.2.2 was followed with the exception of once the primer sequences were finalised, they were altered in order to incorporate *att*B sites into both the forward and reverse primers. This allowed the insert to be introduced into Gateway converted vectors containing the *att*R sites. The primers were modified as in Figure 6.3.

6.2.3 Polymerase Chain Reaction and Sanger Sequencing

The samples identified in Section 5.2.1 were PCR amplified and sequenced in order to validate their genotype. The reactions consisted of 1x Roche Expand High Fidelity PCR Buffer (Roche Diagnostics), 2mM $MgCl_2$, 0.2mM dNTPs, 1pmol/µl of the primers designed as in Figure 6.3, 1U of High Fidelity Taq polymerase (Roche Diagnostics) and approximately 10ng of DNA in a volume of 30µl. They were thermocycled as follows: initially denatured at 94°C for 2 minutes then a further 15 seconds at 94°C, annealed at 60°C for 30 seconds and extended at 72°C for 45 seconds. This was repeated 10 times before the program was altered slightly in order to obtain a higher yield of the amplification products: the reactions were denatured again at 94°C for 15 seconds, annealed at 60°C for 30 seconds but the extension time was dropped to 40 seconds with an additional 5 seconds added every cycle, still at 72°C. This was repeated for 20 cycles before a final extension step of 72°C was performed for 7 minutes.

***att*B1 Forward Primer:** 5’ - GGGG - ACA - AGT - TTG - TAC - AAA - AAA - GCA - GGC - GGC - TNN - (17bp template specific sequence) - 3’

*att*B1 site

***att*B2 Reverse Primer:** 5’ - GGGG - AC - CAC - TTT - GTA - CAA - GAA - AGC - TGG - GTN - (19bp template specific sequence) - 3’

*att*B2 site

Figure 6.3
Additions to primer designs in order to utilise the PCR product in the Gateway cloning system (Invitrogen, Calsbad, California, USA). This system allows rapid cloning of products into vectors containing the *att*P and *att*R sites, as shown in Figure 6.4, using enzyme mixes as opposed to separate restriction enzyme digests and ligation reactions, increasing throughput.

All PCR products were examined by 1% agarose gel electrophoresis as described in Section 3.2.3 and sequenced as in Section 3.2.4 utilising the forward primer designed above at a concentration of 0.5pmol/µl. Demographics of the samples identified to be carrying the two alleles for variant rs2020000 upon analysis of this sequence, as performed in Section 3.2.4, are shown in Table 6.2.

Table 6.2
Sample demographics of the samples used to create the constructs for these regulatory assays. The sample AD445 carries the major (C) allele for the variant rs2020000. This individual was a male with confirmed AD and carrying two ε3 alleles for ApoE, however, the age he died or the age the disease was diagnosed is unknown (NA). Sample BRI138 carried the minor (G) allele for the variant and was confirmed early onset AD (EOAD) with symptoms presenting at the age of 58. She was known to carry two ε4 alleles of the ApoE gene.

Sample ID	Sex	Centre of Origin	Age at Death	Age at Onset	Disease Status	ApoE Status
AD445	Male	Nottingham	NA	NA	Confirmed AD	3 3
BRI138	Female	Bristol	NA	58	Confirmed EOAD	4 4

6.2.4 Polymerase Chain Reaction Product Purification

In order to ensure the PCR products were pure and clean (for example all primer-dimers, excess primers and dNTPs were removed), the QIAquick PCR Purification kit (QIAGEN) was utilised. 5 volumes of the binding Buffer PBI were added per volume of PCR product to allow binding of any single- or double-stranded PCR products in the solution. The pH was corrected with 10µl of 3M sodium acetate, pH 5.0 if indicated by the pH indicator within the buffer, before the mixture was transferred to a QIAquick spin column. This was centrifuged at 17,900 x g for 1 minute in order to bind the DNA to the column. 0.75ml of the wash Buffer PE was then added and the columns centrifuged, again at 17,900 x g for 1 minute. The flow through was removed and the centrifuge step repeated in order to remove all ethanol present from the addition of the wash buffer. The columns were then transferred to fresh tubes and 30µl of the elution Buffer EB was added and incubated at room temperature for 1 minute to uncouple the DNA from the column. The DNA was then eluted by

centrifuging the columns at 17,900 x g for 1 minute before the products were quantified as described in Section 4.2.7.

6.2.5 Gateway BP Recombination Reaction

All of the Gateway protocols were performed following those from the manufacturer (Invitrogen). This recombination reaction was performed in order to introduce the *att*B-PCR products into the pDONR221 vector (containing the *att*P sites) to create an entry clone and a by-product as described in Figure 6.4 and allows much more rapid and efficient cloning when compared to utilising restriction enzyme digests as in Chapter 4. One reaction was performed for both the major and minor allele samples by combining 15-150ng of the *att*B-PCR product with 150ng of pDONR221 and 1x of the BP Clonase™ II Enzyme mix, made up to 8µl with TE Buffer. This was then incubated at 25°C for one hour before terminating the reaction by the addition of 2µg of Proteinase K and incubating at 37°C for 10 minutes. This creates an entry clone containing the insert of interest flanked by the *att*L sites in preparation for the next step.

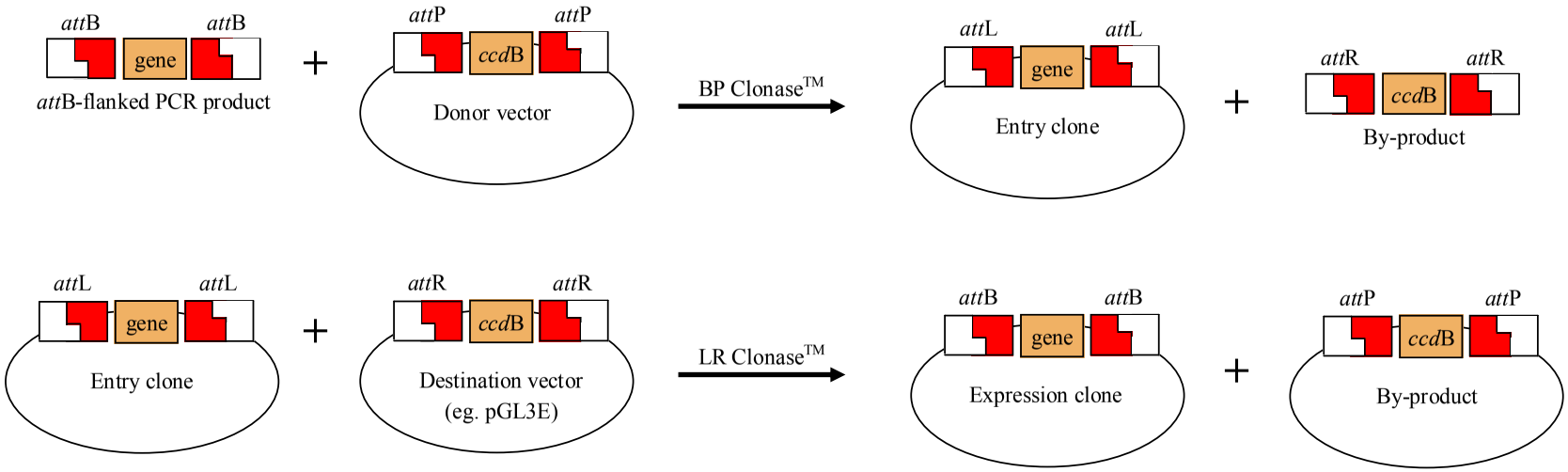


Figure 6.4

Gateway® technology utilises lambda recombination systems in order to transfer and introduce DNA sequences (flanked by modified *att* sites) between vectors. Two recombination reactions are involved in the Gateway® process: the BP reaction and LR reaction. The BP reaction transfers an *att*B substrate (in this case a *att*B-PCR product due to the introduction of *att*B sites into the PCR primers) into an *att*P substrate (in this case the pDONR221 vector (Invitrogen)) to create an entry clone containing the *att*L site. This is catalysed by the BP Clonase™ enzyme mix. The LR reaction facilitates recombination of the *att*L substrate (the entry clone created in the BP reaction) with an *att*R substrate (the destination vector—in this case the pGL3 vectors). This creates the *att*B-containing expression clones, catalysed by the LR Clonase™ enzyme mix.

6.2.6 Transformation

Both major and minor allele entry clone constructs were transformed into OneShot OmniMax 2T1 Phage-Resistant Cells (Invitrogen). 1µl of each BP reaction product was incubated with 50µl of these cells on ice for 30 minutes before being heat shocked in a 42°C water bath for 30 seconds. 250µl of SOC Media was then added to each vial of cells and these mixtures were shaken horizontally (250rpm) at 37°C for one hour.

The cultures were each spread on plates consisting of 1.5% agarose, 4% Circlegrow® and 0.05mg/µl kanamycin (as this is the antibiotic resistance gene present in the pDONR221 vector). All transformed cells were plated out at 20µl and 100µl of cells per plate. 80µl of SOC media was also added to the 20µl plates. These were incubated at 37°C for 16 to 18 hours.

Four of the grown colonies were picked for each genotype the next day and grown up in 5ml of liquid media consisting of 4% Circlegrow® and 0.05mg/µl kanamycin. These were incubated at 37°C, shaking horizontally (250rpm) for 16-18 hours before their plasmid DNA was extracted.

6.2.7 Plasmid DNA Extraction

The plasmid DNA for the colonies grown above was extracted following the PureYield Plasmid Miniprep System protocol as per Section 4.2.7 (Promega). The plasmid DNA was quantified to determine its concentration using the Nanodrop 1000 Spectrophotometer (Thermo Scientific) following the manufacturer's protocol.

All were then sequenced to validate the transformation had worked as previously described in Section 3.2.4. The primer used for this sequencing was the M13 Forward

primer (GTAAAACGACGGCCAGT) at a concentration of 0.5pmol/µl, the site of which is present upstream of the *att*P site in the pDONR221 vector.

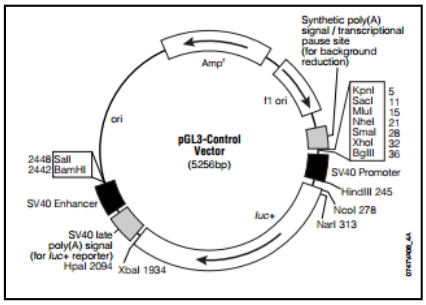
6.2.8 Gateway LR Recombination Reaction

Once constructs of each genotype were identified, the LR Clonase™ reaction was performed in order to introduce them into the destination vectors which would then be used as the reporter constructs. The destination vectors used in this instance were the pGL3 vectors (three in total - Basic, Promoter and Enhancer - Promega) which can be seen in Figure 6.5 B-D. All three of these had been previously modified (by a colleague - Dr Sally Chappell) in order to introduce the *att*R sites, as part of the Gateway Cassettes, by ligating a blunt-ended cassette containing the *att*R sites into either the 5' or 3' multiple cloning sites of theses vectors (see Figures 6.4 and 6.5). All vectors have both 5' and 3' cloning sites (in relation to the *luciferase* gene) and versions of all three vectors with the Gateway cassettes in both cloning sites were created. Due to the nature of the variant under examination and owing to its location in relation to the *ABCA7* gene (nearly 1Kbp upstream of it), constructs of all three vectors were created with the insert in the 5' cloning site. The pGL3-Basic vector was used to initially ascertain whether the 500bp region containing the variant had, as a whole, any promoter activity when compared to the identical vector not containing this insert. The pGL3-Promoter and pGL3-Enhancer vectors were then used in order to amplify the differences, if any, between the major and minor constructs. The differences in signals between the two constructs, if small, would be more apparent in these two vectors due to the presence of the Simian vacuolating virus 40 (SV40) promoter and enhancer regions in these vectors respectively, as seen in Figure 6.5 C & D. The pGL3-Promoter vector was used, in this instance, to examine whether this region was acting as an enhancer in combination with the SV40 Promoter region (as it

would with the *ABCA7* promoter region) and the pGL3-Enhancer vector was used to examine whether this region had promoter activity in isolation.

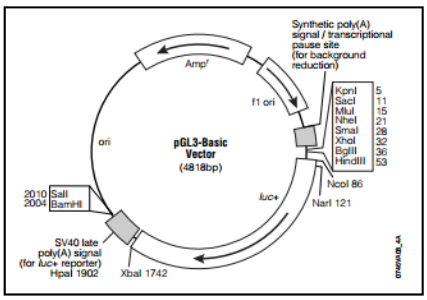
In order to introduce the inserts into these destination vectors, 150ng of the products from the plasmid DNA extractions in Section 5.2.7 were incubated with 150ng of each of the destination vector and 1x of the LR Clonase^{TM} II enzyme mix, made up to 8µl with TE Buffer. This was incubated at 25°C for one hour before terminating the reaction by the addition of 2µg of Proteinase K and incubating at 37°C for 10 minutes.

This created a destination vector (containing either the major or minor allele inserts in each of the three vectors) and a by-product as shown in Figure 6.4.



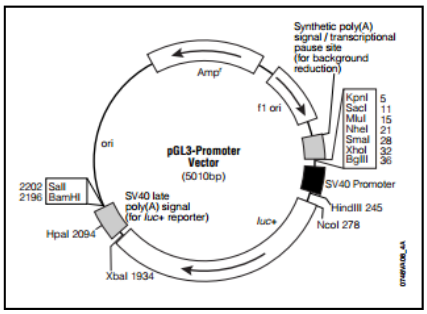
A

The pGL3-Control vector contains both the SV40 (eukaryotic) promotor and enhancer sequences. This results in a strong expression of the luciferase protein and is therefore used to monitor transfection efficiency.



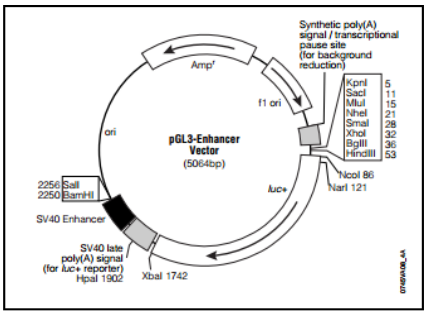
B

The pGL3-Basic vector contains no promotor or enhancer elements, therefore luciferase activity is solely dependant on the insertion of a promotor sequence upstream of *luc*+. Enhancer elements may also be inserted downstream of *luc*+.



C

The pGL3-Promoter contains an SV40 promoter upstream of the *luc*+ gene. Putative enhancer regions can be inserted upstream or downstream of the SV40/*luc*+ unit.



D

The pGL3-Enhancer vector contains a SV40 enhancer downstream of the *luc*+ reporter gene. Putative promotor sequences can be inserted upstream or downstream of the *luc*+/SV40 unit. The presence of the enhancer unit will enhancer any action of the cloned promoter regions.

Figure 6.5

Circle maps of the pGL3 Vectors. All contain the ampicillin resistance gene (*AMP*^{r}) guaranteeing ampicillin resistance in *E. coli*, cDNA coding for the modified firefly luciferase (*luc*+), f1 ori (origin of replication derived from filamentous phage) and ori (origin of plasmid replication in *E. coli*). Arrows in the *AMP*^{r} and *luc*+ genes indicate direction of transcription and the arrow in f1 ori indicates direction of strand synthesis. Multiple cloning sites are also shown with specific restriction enzyme sites listed. This is the location the Gateway Cassettes (A and B—both containing the *att*R site) were ligated.

6.2.9 Transformation

All of these vectors are transformed again, following the protocol as in Section 5.2.6. The only variation from this protocol was kanamycin was substituted with ampicillin (as this was the resistance gene the pGL3 vectors contain - see circle maps in Figure 6.5) at a concentration of 0.1%.

After an overnight incubation on agar plates, colonies were picked (two colonies per vector transformed) and incubated at 37°C in 25ml of liquid media (consisting of 4% Circlegrow® and 0.1% of ampicillin) shaking horizontally at 250rpm.

6.2.10 Endotoxin Free Plasmid DNA Extraction

These colonies had their total plasmid DNA extracted in an endotoxin free manner in order to reduce the chances of them contaminating the cells upon transfection. The NucleoBond Xtra Midi Plus EF Kit was utilised, following the protocol as in Section 4.2.13, with samples then being quantified (as previously described) and sequenced following the BigDye Terminator v3.1 Cycle Sequencing Kit protocol as before (see Section 3.2.4). Samples were sequenced in both the forward and reverse direction in order to substantiate that the sequences of the inserts were correct. The primers used to sequence were the RV3 primer (CTAGCAAAATAGGCTGTCC), the site of which is located upstream of the 5' multiple cloning site in the pGL3 vectors and the GL2 primer (CTTTATGTTTTTGGCGTCTTCCA), the site of which is located downstream of the 5' multiple cloning site in the pGL3 vectors.

All vectors were examined by 1% agarose gel electrophoresis as in Section 3.2.3 in order to ensure they were intact.

6.2.11 Growing "Empty" Vectors

In order to reduce variability and to allow the detection of small expression changes, alongside the experimental vectors, "empty" vectors were also transfected, containing no insert, as an inter-experiment comparison. This was alongside the pGL3-Control vector (see Figure 6.5A) which was also transfected alongside the above in order to provide a known level of luciferase activity i.e. a positive control. The pRL vector, containing the *Renilla* luciferase gene was also co-transfected with the experimental vectors, the "empty" vectors and pGL3-Control in order to normalize the firefly luciferase expression levels.

In order to provide enough of all of these vectors to perform repeat transfections, more were grown in XL1-Blue Competent Cells (Stratagene, San Diego, California, USA). 100µl of the cells were thawed on ice and, in pre-chilled 1.5ml tubes, combined with 1.7µl β-mercaptoethanol. These are gently mixed before being incubated on ice for 10 minutes. 50ng of the appropriate vector was then added to the tube, incubated on ice for 30 minutes before being heat shocked in a water bath heated to 42°C for 45 seconds. They were chilled on ice for a further 2 minutes before being incubated with 0.95ml of SOC media at 37°C for 1 hour being shaken horizontally at 250rpm.

These cells were plated out on agar plates consisting of 4% Circlegrow®, 1.5% agar and 0.1% ampicillin. 80µl and 20µl of cells are both plated out, both being made up to 100µl with SOC media. These were then incubated overnight at 37°C.

Two colonies were selected per transformed vector, cultured overnight and extracted in an endotoxin-free manner before being sequenced as described above in Section 5.2.10. The RV3 and GL2 primers were used in order to ensure the vectors transformed correctly. The reference sequences for the pGL3-Control, pGL3-Basic, pGL3-Promoter, pGL3-Enhancer vectors were obtained from the Promega website in

September 2015. The pRL vector did not contain any of these primer sites so, in order to ensure the transformation efficiency, this vector was restriction enzyme digested utilising the enzymes *BamHI* and *HindIII*, following the protocol as described in Section 4.2.9. The products were run on a 1% agarose gel, as in Section 3.2.3, to ensure two fragments were achieved for each digestion reaction - one of 1463bp and one of 224bp, confirming the presence of the pRL vector.

6.2.12 Glycerol Stocks

All colonies confirmed to contain the vectors expected in Sections 5.2.10 and 5.2.11 had glycerol stocks made of them in order to prevent having to retransform the plasmid if more preparations were necessary. This was done by combining 200µl of the liquid culture with 800µl of glycerol in a 1.5ml tube. This was labelled well and stored at -80°C long term.

6.2.13 Cell Culture

All transfections were performed in BE(2)-C cells which were cultured and maintained as in Section 4.2.15.

6.2.14 Transfection

In preparation for transfection, BE(2)-C cells were plated into 12 well plates at a concentration of 3 x 10^5 cells per well in 1ml of complete, filtered EMEM:F12 media. These were incubated in the plates at 37°C with a humidified atmosphere of 5% CO_2 in air for 24 hours.

All vectors to be transfected were quantified as previously described and then diluted 1µl in 9µl of dH_2O in order to increase pipetting accuracy. The pRL *Renilla* Luciferase vector was also included here in order to co-transfect it into all of the cells

to provide an internal control value to which expression of the experimental firefly luciferase could be normalized.

For each well 400µl of serum-free EMEM:F12 media, 20ng of pRL and 200ng of the appropriate vector were combined along with 9nl of TransFact^{TM} (Promega) per nanogram of vector DNA in order to facilitate the transport of the DNA into the cells through liposomes. Each vector was transfected in triplicate. These mixtures were vortexed well and incubated at room temperature for 15 minutes. The media was then removed from the cells plated out the day before and the 400µl of the prepared mixture added to each well. The plates were incubated at 37°C, 5% CO_2 for 1 hour before a further 800µl of complete, filtered EMEM:F12 media was added to each well in order to enable cell growth. The plates were incubated at 37°C, 5% CO_2 for 24 hours in order to allow the firefly luciferase and *Renilla* luciferase proteins to be translated.

6.2.15 Dual Luciferase Reporter Assay

All reagents mentioned in this section are provided as part of the Dual Luciferase Reporter Assay System protocol (Promega).

Once the 24 hour incubation was complete, the media was removed from the cells and they were washed in warmed PBS twice in order to remove any traces of the media. 200µl of 1x Passive Lysis Buffer was then added to each well in order to rapidly lyse the cells. This was incubated for 20-40 minutes on a plate tipper in order to ensure all cells were lysed and their proteins released. In order to measure the activity of both firefly luciferase and *Renilla* luciferase two different reagents were prepared. Luciferase assay reagent II (LAR II) was prepared beforehand by resuspending the lyophilized Luciferase Assay Substrate in 10ml of the supplied Luciferase Assay Buffer II. This was aliquoted out in 1ml aliquots and stored at -80°C

before being defrosted on ice prior to use. This reagent activates the firefly luciferase when combined with the products of the cell lysis. The Stop&Glo® reagent was prepared just prior to use. 1 volume of 50X Stop&Glo® substrate was added to 50 volumes of Stop&Glo® Buffer. This reagent quenches the firefly luminescence whilst simultaneously initiating the *Renilla* luciferase reaction.

All measurements were performed on a TD 20/20 Luminometer (Turner Design's, San Jose, California, USA) by pre-loading 50µl of the LARII into an adequate number of 2ml disposable cuvettes. 20µl of the cell lysis was then added to each cuvette, mixed and loaded onto the machine. Upon being loaded, the machine delayed for 2 seconds before taking a 10 second luminescence reading. This was repeated with 50µl of the Stop&Glo® for the same sample before that tube was discarded and the rest of the samples were measured. The luminometer then calculated a ratio of firefly luciferase:*Renilla* luciferase for each sample and an average ratio for each vector as well as standard deviation (SD) and a percentage coefficient of variation (CV) for these triplicate measurements. Readings were only accepted when the CV was below 21% within each experimental triplicate as any higher than this implied that there was too much intra-experimental variability present.

6.2.16 Statistical Analysis

All statistical analysis was performed on SPSS Statistics version 23 (IBM Corporation, New York, USA). After elimination of any readings with a CV of 21% or greater, as well as any transfections where the luciferase levels in the pGL3-Control vectors were low (less than 10x the values in the other "empty" pGL3 vectors), the remaining values were checked for their distribution by plotting histograms containing a normal distribution curve. None of the vectors presenting as containing normally distributed luciferase to *Renilla* ratios, also corroborated by Kolmogoroz-Smirnov tests performed within SPSS, therefore dictating which statistical tests were

performed. The luciferase levels for the experimental constructs were normalised to the luciferase levels for the pGL3 vectors not containing an insert. Due to the skewed distribution of these datasets and the fact that the data was paired (as the readings were from the same transfections), a Wilcoxon signed rank test was performed between the normalised major allele and the normalised minor allele constructs in order to establish the effect of the minor allele on firefly luciferase expression. Fold differences between this set of vectors as well as between the experimental vectors containing the major allele and the "empty" pGL3 vectors were also calculated within SPSS, utilising the "transform variable" function. Medians and interquartile ranges (IQR) for these fold changes were also calculated in order to obtain a numerical value for the effect of both the variant in question, and the 500bp region as a whole, upon expression.

6.3 Results

One sample was identified to be homozygous for the G allele and one for the C allele of the rs2020000 variant within the ARUK DNA Bank. These samples were PCR amplified and sequenced utilising the primers presented in Table 6.3, incorporating the *att*B sites into the clones and validating their genotypes. The electrophoresis gel of the samples can be seen in Figure 6.6, showing amplicons of the correct size (498bp).

Table 6.3
Primers designed in order to PCR amplify a region containing the variant rs2020000 to clone into reporter vectors. These primers contain the *att* B sites in order to introduce this clone into entry clones containing the *att* L sites. The product size achieved by these primers as well as the annealing temperature optimised for them are also shown (Tm).

	Primer Sequence	Product Size	Tm (°C)
Sense	GGGGACAAGTTTGTACAAAAAAGCAGGCTGTGAGTGTCAGCGTGGGAT	498	60
Antisense	GGGGACCACTTTGTACAAGAAAGCTGGGTTTTCACTTCCTTGTACGGCC		

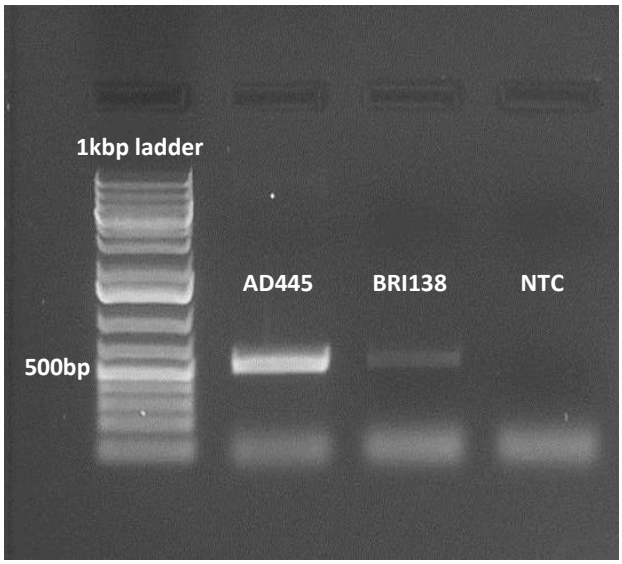


Figure 6.6
Ethidium bromide stained 1% agarose gel run at 80V for 25 minutes to amplify the genomic region for the variant rs2020000 (expected product 498bp). The samples identified here (AD445 and BRI138) were subsequently sequenced and confirmed to carry the major (C) and minor (G) alleles for this variant respectively. Also shown is a No Template Control (NTC) demonstrating there was no contamination present. As can be seen, there was a large amount of primer-dimer present so these PCR products were purified before sequencing (utilising ExoSAP-IT) as well as before they were used to create the clones (employing the QIAquick PCR Purification Kit).

The two samples (AD445 and BRI138) do not differ in their genetic sequence except at the location of the rs2020000 variant (position 19:1038223) and at a location 223bp downstream of this, 19:1038446, with the sample carrying the minor allele of rs2020000 (C), carrying a C allele at this genomic location (reference allele T). This was examined closer and the variant rs10419707 is present at this location, a T to C SNP with a MAF of 0.17. As this MAF is identical to that of rs2020000, the LD between these two variants was examined utilising Haploview (Barrett et al., 2005). As can be seen in Figure 6.7 these two variants are in complete LD, with a D` of 1.00 and an r^2 value of 0.947, implying that there is very little, if any, recombination between these two variants. Therefore, a sample containing the minor allele of both these variants was used as the minor allele construct.

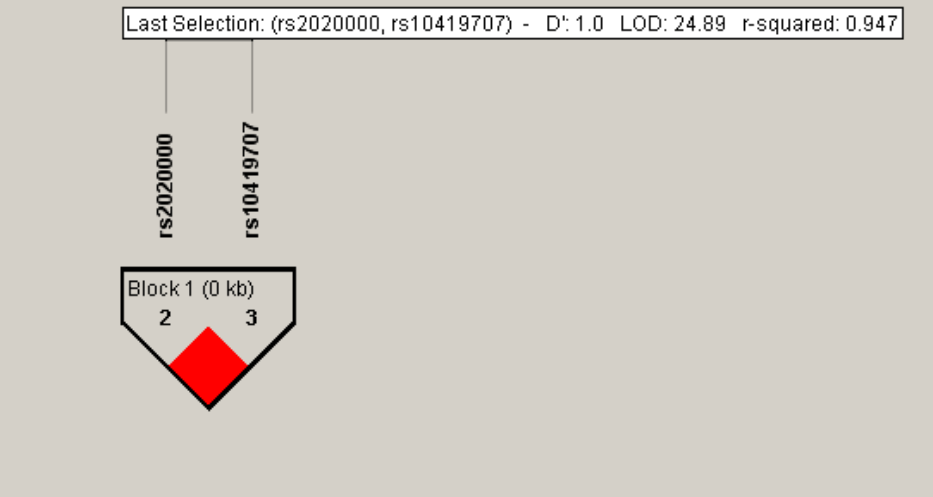


Figure 6.7
Linkage Disequilibrium (LD) plot between the variants rs2020000 and rs10419707 which have very similar minor allele frequencies (MAFs) of 0.1675 and 0.1661 respectively. A D` of 1 indicates there is an absence of recombination between the two loci. The $r^2$ value differs from 1 slightly due to the slight difference in their MAFs. The log of the likelihood odds ratio (LOD) is also presented which is a measure of confidence in the value of D`. A score of >2 obtains a dark red colour on the heat map, identifying that this score of 25 indicates extremely high levels of confidence in this D` score.

Clones were therefore created utilising these samples. Examples of the pGL3-Enhancer clones created can be seen in Figure 6.8(B). Also shown in Figure 6.8 are the pRL vectors grown for these transfections, digested by the restriction enzymes *BamHI* and *HindIII* in order to check the transformation worked efficiently. Two bands are shown, one of size 2242bp and one of 1463bp, indicating that this plasmid DNA is indeed the pRL vector. Also shown in Figure 6.8 is the pGL3-Control vector of approximately 5256bp, where the majority of the DNA is supercoiled, indicating it is suitable for use in transfections. The identity of this plasmid DNA was confirmed through sequencing using the RV3 and GL2 primers.

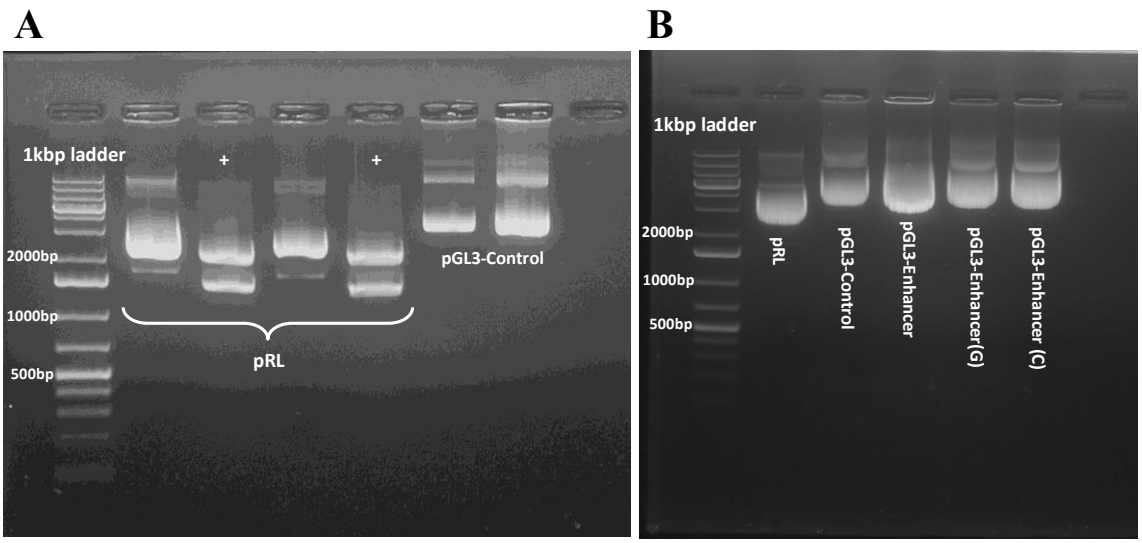


Figure 6.8
(A) shows vectors run on a 1% agrose gel, stained with ethidium bromide, alongside a 1kbp ladder. The lanes with a "+" show the *Renilla* luciferase (pRL) vectors which have been digested with the *BamHI* and *HindIII* restriction enzymes, producing bands of 1463bp and 2242bp, in order to demonstrate that this transformation was successful. Alongside these digested plasmids, the uncut plasmids are also shown as a comparison. Here it can be seen that the majority of the plasmid DNA is supercoiled. The pGL3-Control vector is also shown, this vector was Sanger sequencing in order to validate the transformation.
(B) shows the vectors used in a transfection with the pGL3-Enhancer clones. All vectors transfected are shown including the pRL vector again, the pGL3-Control vector, the pGL3-Enhancer vector without the insert and the pGL3-Enhancer vector with the inserts containing both the major (G) and minor (C) alleles. These vectors appear slightly higher on the gel due to their heavier molecular weight as they contain the insert of 498bp. All of these vectors are suitable for transfection due to the majority of the plasmid DNA being supercoiled.

The median luciferase readings, corrected to the *Renilla* readings, for all of the vectors transfected (pGL3-Basic, pGL3-Enhancer and pGL3-Promoter all with no insert and with inserts containing both the G and the C alleles) are presented in Table 6.4. From these values it can be seen that, in all three vectors, the reporter protein activity is greatly increased in the constructs containing the G allele when compared to the constructs with no insert but this activity is reduced in the constructs containing the minor allele C. These values for each vector, as well as the IQRs and any outlier readings are presented graphically in Figures 6.9-11.

Table 6.4

Median values for the luciferase levels corrected to the Renilla levels in all three reporter vectors: pGL3-Basic, pGL3-Enhancer (containing the SV40 Enhancer region) and the pGL3-Promoter (containing the SV40 Promoter region). The values in the presence of no insert (-), the insert containing the major allele (G) and the insert containing the minor allele (C) are all presented. The 25% and 75% interquartile ranges (IQR) are also displayed as well as the number (*n*) of transfections performed in order to obtain these values. Each transfection was performed in triplicate.

Vector	Allele	Median	IQR	*n*
pGL3-Basic	-	0.037	0.024 - 0.050	9
pGL3-Basic	G	0.166	0.111 - 0.226	8
pGL3-Basic	C	0.095	0.067 - 0.156	10
pGL3-Enhancer	-	0.024	0.014 - 0.069	7
pGL3-Enhancer	G	0.082	0.064 - 0.118	7
pGL3-Enhancer	C	0.062	0.022 - 0.081	7
pGL3-Promoter	-	0.869	0.434 - 1.185	8
pGL3-Promoter	G	2.865	2.196 - 5.610	8
pGL3-Promoter	C	2.785	2.146 - 4.990	8

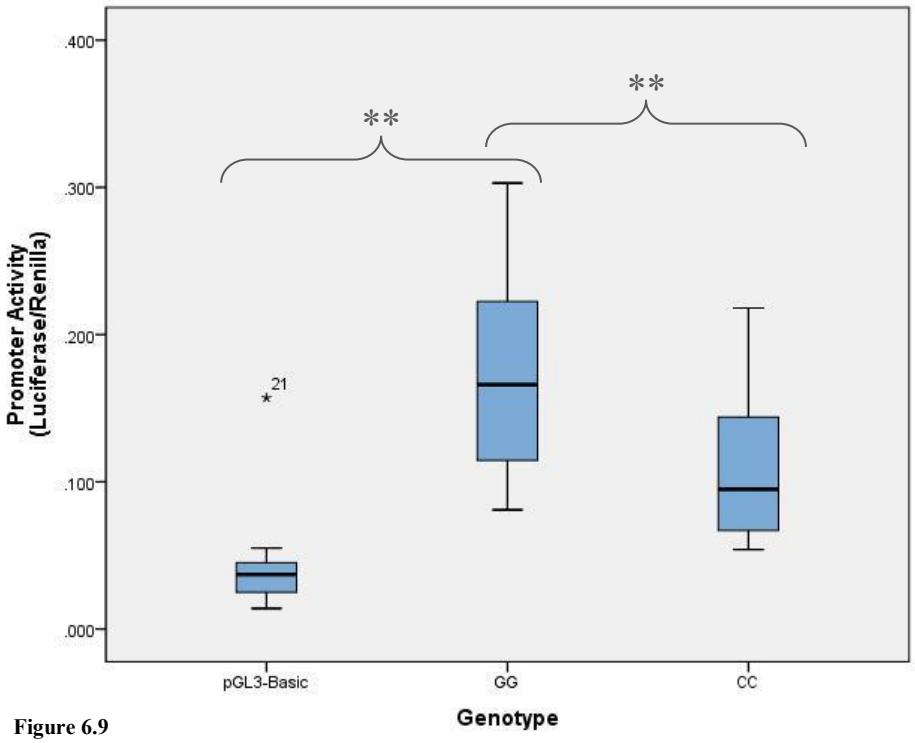


Figure 6.9

Median values for the pGL3-Basic vector containing no insert (pGL3-Basic), the insert containing the major allele (GG) and the insert containing the minor allele (CC), all denoted by the thick black lines. The edges of the blue boxes indicate the 25% and 75% interquartile ranges, the whiskers of the boxes mark the maximum and minimum values and the asterisk (*) shows an outlier value. These values are also shown numerically in Table 6.4. *n* for pGL3-Basic is 9, 8 for GG and 10 for CC, all performed in triplicate. ** indicates the statistical test for the difference between the two vectors indicated returns a *p*-value of <0.02.

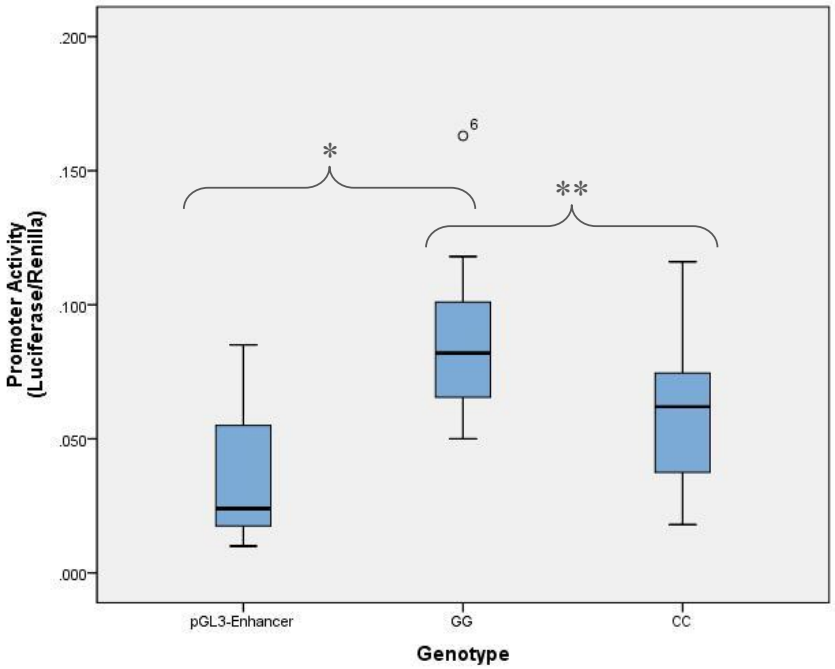


Figure 6.10
Median values for the pGL3-Enhancer vector containing no insert (pGL3-Enhancer), the insert containing the major allele (GG) and the insert containing the minor allele (CC), all denoted by the thick black lines. The edges of the blue boxes indicate the 25% and 75% interquartile ranges, the whiskers of the boxes marks the maximum and minimum values and the circle (°) shows an outlier value. These values are also shown numerically in Table 6.4. *n* for all of these vectors is 7, all performed in triplicate. * indicates the statistical test for the difference between the two vectors indicated returns a *p*-value of <0.05 while ** indicates the *p*-value returned is <0.02.

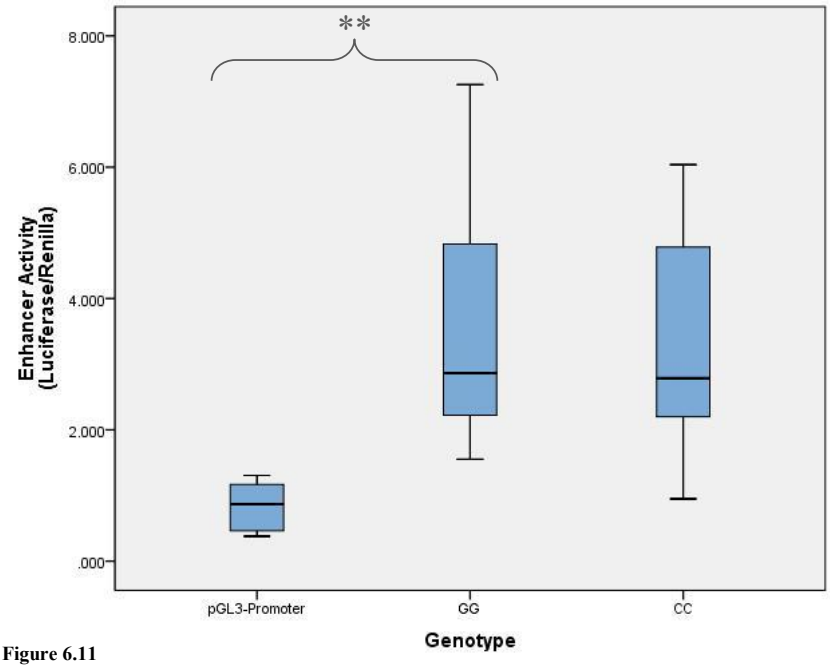


Figure 6.11
Median values for the pGL3-Promoter vector containing no insert (pGL3-Promoter), the insert containing the major allele (GG) and the insert containing the minor allele (CC), all denoted by the thick black lines. The edges of the boxes indicate the 25% and 75% interquartile ranges and the whiskers of the boxes indicate the maximum and minimum values. These values are also shown numerically in Table 6.4. *n* for all of these vectors is 8, all performed in triplicate. ** indicates the statistical test for the difference between the two vectors indicated returns a *p*-value of <0.02. If no markings are shown between the two vectors, this indicates that the *p*-value obtained was >0.05.

The median fold differences between the major and minor allele constructs as well as between the major allele constructs and the vector with no insert are presented in Table 6.5 alongside the IQRs and *p*-values as determined by the Wilcoxon signed rank tests performed. All of these fold changes indicate that the cloned 500bp region as a whole acts as an extremely strong promoter and enhancer region with an average fold increase of 5.1, 2.7 and 5.5 in the pGL3-Basic, pGL3-Enhancer and pGL3-Promoter vectors respectively with all of these tests returning a *p*-value of less than 0.02 and all of the 25% interquartile ranges being above one. The three tests between the major and minor allele also show similar patterns although only the pGL3-Basic and pGL3-Enhancer tests produce *p*-values below 0.05. All three fold differences indicate that the minor allele (C) has a weaker promoter activity than the major allele (G), decreasing expression by about 30% in the two significant tests (see Table 6.5).

Table 6.5

Fold differences between two different vectors containing the major allele insert (G) and the minor allele insert (C) or no insert (-). A fold increase of < 1 indicates that Allele 1 decreases reporter gene expression while a fold increase of > 1 signals that Allele 1 increases reporter gene activity in the constructs. The 25% and 75% interquartile ranges (IQR) are also presented as well as the *p*-values for these different tests. Tests showing a *p*-value of less than 0.05 are highlighted in red.

Vector	Allele 1	Allele 2	Median Fold Difference	IQR	*p*
pGL3-Basic	C	G	0.694	0.604 - 0.952	**0.018**
pGL3-Basic	G	-	5.089	3.681 - 8.640	**0.018**
pGL3-Enhancer	C	G	0.738	0.360 - 0.983	**0.043**
pGL3-Enhancer	G	-	2.667	1.918 - 5.619	**0.018**
pGL3-Promoter	C	G	0.832	0.735 - 1.050	0.263
pGL3-Promoter	G	-	5.484	2.409 - 6.203	**0.012**

6.4 Discussion

rs2020000 is a common variant (MAF = 0.17) located upstream of the *ABCA7* gene, within a sequence that is predicted to be regulatory based on compelling epigenetic and *in silico* data from the ENCODE Project and the Epigenetics Roadmap Project. In the Mayo brain eGWAS cohort, the minor allele of this SNP (C) was found to be associated with increased messenger levels of *ABCA7* within the cerebellum ($p<0.005$) and was also the SNP with the strongest regulome score within RegulomeDB (1f), prompting us to further investigate its functional potential. The accuracy of this computational data was therefore assessed by *in vitro* assays. Promoter constructs identified that the region inserted into these constructs acted as a strong promoter region as a whole, creating a fold increase in reporter protein expression of 5.09 ($p<0.02$) within the pGL3-Basic vector. This is comparable in the pGL3-Enhancer vector but not to the same degree (fold increase of 2.67, $p<0.02$). However, this may be due to the high variation present in the measurements for the pGL3-Enhancer vector with no insert (see Figure 6.10). It appears that the fold increase would be higher without this high variation.

It is also apparent that the region has some enhancer activity, with a fold increase of 5.48 in the pGL3-Promoter vector with a $p<0.02$. This implies that this region acts as a strong enhancer in combination with a promoter region further downstream of it, shown by the high increase in enhancer activity in the pGL3-Promoter vectors, containing the SV40 promoter site downstream of the multiple cloning region where our clone was introduced. This increases the likelihood that this region acts as an enhancer to the *ABCA7* gene due to its genomic location upstream of the documented *ABCA7* promoter region (see Figure 6.1).

The eGWAS data for this variant predicted that the minor allele (C) was associated with increased messenger levels of the *ABCA7* gene (see Table 6.1). Contradicting this, these promoter constructs showed that this allele decreased expression of the luciferase protein. In both the pGL3-Basic and pGL3-Enhancer the minor allele decreases expression by approximately 30% (p<0.05 in both). This pattern is also shown in the pGL3-Promoter construct but not to the same extent, decreasing expression by approximately 25% but returning a p-value of >0.05.

When examining this variant in GTEx, an expression database recently available, this also shows the minor allele (C) increasing *ABCA7* expression, as shown in Figure 6.12 (effect size of 1.35, $p = 2.1e^{-8}$). This supports the results from the microarray results presented in Section 6.1, even to approximately the same degree with GTEx estimating the C allele to increase expression by 35% and the eGWAS 20%.

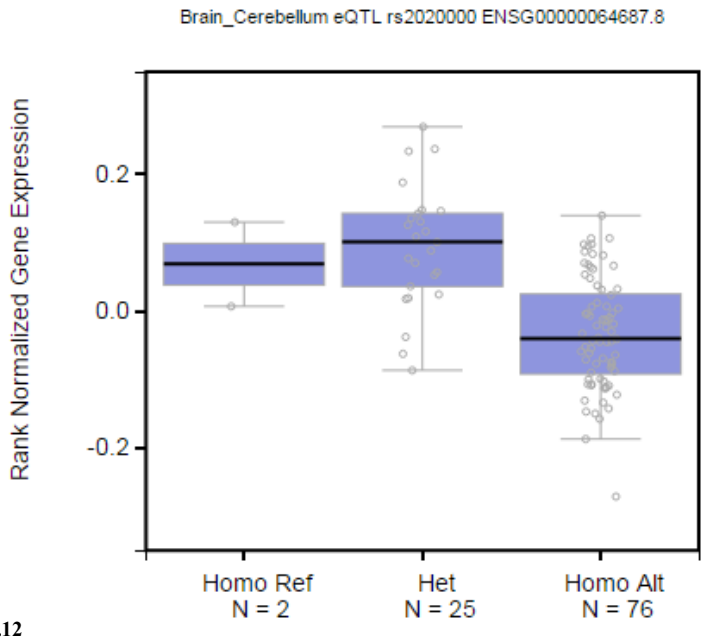


Figure 6.12
Single tissue expression data for rs2020000 transcript ENSG00000064687.8 (*ABCA7*), measured in brain cerebellum tissue. These measurements are taken from volunteer samples and, as such, are unrelated to clinical status. This demonstrates that the minor allele (Homo Ref) allele of rs2020000 (C) up regulates *ABCA7* expression levels when carried in a heterozygous and homozygous fashion (effect size of 0.35, $p = 2.1e^{-8}$). The confusion between the alleles in this figure is explained by the fact that the minor allele (C) is labelled as the reference allele in hg19/GRCh37 as shown in Figure 6.2. The fewer numbers of alleles present in the "Homo Ref" and "Het" datasets (n=25 and n=2 respectively) endorses this. This RNA-seq study therefore confirms the microarray data presented in Section 6.1, contradicting the dual luciferase assays performed in Chapter 6. This image was taken from gtexportal.org (accessed October 2016).

This inconsistency may be due to the eGWAS and GTEx examining mRNA levels whereas the luciferase assays examine protein levels, would proteomics work in brain tissue be more descriptive despite it being much more challenging (Golde, 2016)? The fact that the eGWAS data was acquired from AD brains and this disease state itself may affect the messenger expression levels must also be taken into account. An alteration in gene expression levels can be a consequence of disease, in a reverse causal effect, as opposed to the causal event (Albert and Kruglyak, 2015). Of course, all data acquired *in vitro* does not necessarily represent what happens *in vivo*. Despite many previous studies finding that dual-luciferase assays are extremely sensitive and reproducible (Brasier and Ron, 1992), *in vitro* inaccuracies may occur due to temporal regulation, tissue specificity or the remaining genetic background (Cirulli and Goldstein, 2007). It has also been shown that regions function differently depending on whether they are integrated into the cellular chromosome or acting as an episome, as the transfected vectors are in these assays (Inoue et al., 2016). This is thought to be due to the altered binding of chromatin and that chromosomally integrated sequences have a more replicable and reproducible activity then those of identical sequences in episomes. Action of the integrated sequence are also more foreseeable by ENCODE annotations (Inoue et al., 2016). However, in 2003, Hoogendoorn *et al* showed that if putative promoter constructs met three criteria, then the reproducibility of these promoters was extremely high. These three criteria were: statistically significant differences in expression between the constructs with $p<0.05$; replication with independent construct preparations and ≥ 1.5-fold difference in expression between haplotypes (Hoogendoorn et al., 2003). The first two of these criteria are met in all tests highlighted as red in Table 6.5, although, the fold difference between the two alleles is not big enough to be defined as an enhancer haplotype as per this criteria. However, this study did not include any constructs which down-regulated expression, therefore leaving a gap in these proposed standards (Hoogendoorn et al., 2003).

The contradiction seen between the microarray data and these promoter constructs may also be due to several other factors. For example, the haplotype this variant occurs in should also be considered. It may be that the minor allele (C) of rs2020000 is in strong LD with another, stronger, eQTL which increases expression of ABCA7. This may counteract the decrease in expression rs2020000-C provides, explaining both the microarray and cellular assay results presented here. The cell line used in these assays should also be considered. BE(2)-C cells are a transformed cancer cell line and may, therefore, not be entirely representative of what occurs *in vivo*. Consequently, an alternative cell line should be used to replicate these assays, for example, primary neuronal cells which would obviously be the most representative of the *in vivo* environment. However, these cells are difficult to acquire, as well as maintain, therefore an alternative may be stem cells which have been differentiated into neuronal cells. The difficulty of transfecting into these cells is also unknown, so future work would involve looking into this.

As mentioned previously, a link between levels of *ABCA7* mRNA and the risk of LOAD has already been established. In 2013 it was noted that the minor allele of the GWAS tag SNP rs3764650 (G - associated with increased risk of LOAD) was also associated with a dose-dependent decrease in *ABCA7* mRNA levels (Vasquez et al., 2013). However, increased expression of ABCA7 has been associated with more advanced cognitive decline, as seen in LOAD pathogenesis (Karch et al., 2012). The authors explain these links through ABCA7's postulated function in phagocytosis (Vasquez et al., 2013): the knockdown of *ABCA7* leads to a significant reduction of phagocytosis of apoptotic cells and increased expression of ABCA7 increases the phagocytosis of multiple substrates, decreasing Aβ cellular uptake in mouse models (Jehle et al., 2006; Kim et al., 2013). Immune response, including phagocytosis, is a pathway that has been linked to AD neuropathogenesis previously (see Figure 1.5), suggesting that *ABCA7* may reduce AD risk through activating phagocytosis,

therefore increasing the removal of apoptotic debris as well as the uptake of Aβ (Morgan, 2011). However, this does not explain the fact that increased *ABCA7* mRNA levels have been seen in AD individuals. This may be due to compensatory changes: long-term over-expression of ABCA7, as exhibited by the rs3764650 major (T) allele, reduces an individual's risk of AD and, therefore, the increased expression of ABCA7 seen in AD cases, may be an inadequate, late, compensatory change (Vasquez et al., 2013). The under-expression of *ABCA7*, as seen in the minor allele of this SNP in the promoter constructs, may therefore be a risk factor for developing LOAD but is over-compensated for further down the disease process by alternative regulatory mechanisms, increasing ABCA7 levels again, perhaps explaining the association seen in AD cases in the eGWAS cohort.

However, the association of this variant with disease is unclear and is something which does need to be established before this variant can be classed as possibly disease-causing (MacArthur et al., 2014). As mentioned previously, this variant is unlikely to be the causative, pathogenic variant within *ABCA7* due to the lack of LD between this SNP and the GWAS tag SNP rs3764650 (D` of 0.829 and a r^2 of 0.013). In the same genotyping project performed in the Mayo Clinic, Jacksonville (USA) discussed in Chapter 3, this variant was also genotyped, alongside the GWAS tag SNP rs3764650. A brief association test utilising this data, acquired in Chapter 3 (a Fisher's exact test as performed in Section 3.2.8) was carried out and the results are presented in Table 6.7. Here it can be seen that, although rs3764650 is strongly associated with LOAD as expected (OR = 1.32, $p = 3.28e^{-7}$), rs2020000 was not (OR = 0.9674, $p = 0.50$).

Table 6.7
Statistical results from the Fisher's Exact tests run on rs2020000 and rs3764650 from the Mayo genotyping data acquired as part of the work presented in Chapter 3. CHR = chromosome the variant is present on, SNP = SNP ID (in this case the rs number of both SNPs), BP = base pair position of the SNP, F_A = frequency of the minor allele in cases, F_U = frequency of minor allele in controls, P = exact *p* -value, OR = Odds Ratio, L95 = Lower 95% Confidence Interval, U95 = Upper 95% Confidence Interval. As can be seen, only the test for the rs3764650 variants returned a significant result.

CHR	SNP	BP	F_A	F_U	P	OR	L95	U95
19	rs2020000	1038223	0.1249	0.1286	0.5009	0.9674	0.8803	1.063
19	rs3764650	1046521	0.1106	0.08607	3.28E-07	1.32	1.187	1.468

Regardless of this, as previously discussed, the levels of ABCA7 do seem to be associated with AD risk. Could it, therefore, be this GWAS tag SNP rs3764650 that is regulatory? As already stated, the major allele of this variant has been associated with increased messenger levels of *ABCA7* (Vasquez et al., 2013). However, it only scores 2a in RegulomeDB and there is no evidence in the HaploReg database of it having regulatory potential in neurological tissue, although there may be some evidence here of its putative regulatory action in other tissues, such as blood. The minor allele of rs3764650 could, therefore, be a tag for haplotypes with actual regulatory variants or with variants coding for isoforms which are subjected to nonsense mediated decay in order to decrease *ABCA7* expression. As a consequence of this, the risk presented by the minor allele of rs3764650 may be due to the protective effect of its major allele in multiple haplotypes with relatively weak regulatory variants which increase *ABCA7* messenger levels, such as the major allele of rs2020000. These weak variants do not, therefore, show an association with AD as they present with the major allele of rs3764650, masking them.

As mentioned previously, a recent DNA sequencing project, specifically looking at *ABCA7*, sequenced 772 LOAD patients and 757 controls. The most significant association with disease was an intronic variant, rs78117248. This variant is located within a transcription factor binding region, DNase hypersensitivity site and had a RegulomeDB score of 4 indicating a small regulatory effect. The variant is also in

high LD with all of the *ABCA7* GWAS tag SNPs (see Figure 6.1), highlighting what a potentially vital role regulatory variants play in disease association (Cuyvers et al., 2015). This is not the first tentative regulatory variant highlighted in *ABCA7*. In 2012, the same dataset used to elucidate rs2020000 was used to computationally identify the minor alleles of rs7247087 and rs2072102 as being associated with elevated levels of *ABCA7* mRNA. It could be any, or all, of these putative regulatory variants in LD with rs3764650 which, therefore, alter ABCA7 levels pathologically. However, none of these variants have been confirmed to be regulatory through assays such as the ones performed here (Pers. Comm. Dr Medway).

Although this variant does appear to have regulatory effects, the mechanism behind this is unknown. Figure 6.13 demonstrates the output from the MatInspector program (Genomatix, Munich, Germany) which annotates TFBS within DNA sequences (Cartharius et al., 2005). As this demonstrates, it appears that the minor allele of rs2020000 alters the binding of several transcription factors, eliminating a TFBS and introducing a further three. This may be the cause of the regulatory effect seen in these functional assays. Future work would involve performing Electrophoretic Mobility Shift Assays (EMSAs) utilising nuclear proteins extracted from the BE(2)-C cells in order to see if there is a difference in binding of these proteins between oligonucleotides containing both the major and minor alleles. These EMSAs were started as part of this project but were not able to be optimised in the time available. Two of the TFBS introduced by the minor allele of rs2020000 are part of the X-box binding protein (XBP) family (see Figure 6.13). Interestingly in 2011, one member of this protein family (XBP1) was associated with neuroprotective actions in AD models by reducing the neurotoxicity of Aβ, causing the authors of this study to identify it as a potential therapeutic target for AD (Casas-Tinto et al., 2011).

Major Allele - G

Page 1 of 1 | 50 | View 1 - 28 of 28

Check transcription factor <-> matrix family assignment				Position			Strand	Matrix sim.	Additional lines of evidence		Sequence
Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	from	to	anchor			#	Evidence	
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$MYBL1.01	V-myb avian myeloblastosis viral oncogene homolog-like 1 (AMYB)	20	40	30	(+)	0.850	0		accggtcccgACGGtttctg
V$ZFHX	Two-handed zinc finger homeodomain transcription factors	V$AREB6.04	AREB6 (Atp1a1 regulatory element binding factor 6)	29	41	35	(+)	0.994	0		cgacgGTTTctgc
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$CMYB.02	c-Myb, important in hematopoesis, cellular equivalent to avian myoblastosis virus oncogene v-myb	26	46	36	(-)	0.960	0		ggcaagcagAAACcgtcgggg
V$MAZF	Myc associated zinc fingers	V$MAZ.01	Myc associated zinc finger protein (MAZ)	40	52	46	(-)	0.901	0		aggcGAGGcaagc
V$ZF57	KRAB domain zinc finger protein 57	V$ZFP57.01	Krueppel-associated box-containing zinc-finger protein 57 (KRAB-ZFP 57)	40	52	46	(+)	0.848	0		gctTGCCtcgcct
V$PLAG	Pleomorphic adenoma gene	V$PLAG1.02	Pleomorphic adenoma gene 1	40	62	51	(-)	1.000	0		aaGGGGgaagaggcgaggcaagc

Minor Allele - C

Page 1 of 1 | 50 | View 1 - 28 of 28

Check transcription factor <-> matrix family assignment				Position			Strand	Matrix sim.	Additional lines of evidence		Sequence
Matrix Family	Detailed Family Information	Matrix	Detailed Matrix Information	from	to	anchor			#	Evidence	
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$MYBL1.01	V-myb avian myeloblastosis viral oncogene homolog-like 1 (AMYB)	20	40	30	(+)	0.850	0		accggtcccgACGGtttctg
V$ZFHX	Two-handed zinc finger homeodomain transcription factors	V$AREB6.04	AREB6 (Atp1a1 regulatory element binding factor 6)	29	41	35	(+)	0.994	0		cgacgGTTTctgg
V$MYBL	Cellular and viral myb-like transcriptional regulators	V$CMYB.02	c-Myb, important in hematopoesis, cellular equivalent to avian myoblastosis virus oncogene v-myb	26	46	36	(-)	0.960	0		ggcaaccagAAACcgtcgggg
V$MEF3	MEF3 binding sites	V$SIX.01	Binding sites for Six1, Six4 and Six5	34	46	40	(+)	0.919	0		gttTCTGgttgcc
V$ZF57	KRAB domain zinc finger protein 57	V$ZFP57.01	Krueppel-associated box-containing zinc-finger protein 57 (KRAB-ZFP 57)	40	52	46	(+)	0.845	0		ggtTGCCtcgcct
V$XBBF	X-box binding factors	V$RFX1.01	X-box binding protein RFX1	38	56	47	(-)	0.941	0		gaagaggcgagGCAAccag
V$XBBF	X-box binding factors	V$RFX4.03	Regulatory factor X 4	38	56	47	(+)	0.774	0		ctgGTTGcctcgcctcttc
V$PLAG	Pleomorphic adenoma gene	V$PLAG1.02	Pleomorphic adenoma gene 1	40	62	51	(-)	1.000	0		aaGGGGgaagaggcgaggcaacc

Figure 6.13
Transcription factor binding sites predicted for the genomic regions both containing the major allele (G) and the minor allele (C) of rs2020000. The minor allele appears to remove a site for the Myc associated zinc finger and introduces a binding site for Myocyte enhancer factor 3 (MEF3) as well as two X-box binding (XBP) factor sites. Images taken from MatInspector (Cartharius *et al* 2005).

The importance of regulatory variants within the pathogenesis of complex disorders, such as late onset Alzheimer's disease, has already been discussed. The fact that gene expression regulation could also provide a therapeutic target, or even as a diagnostic marker, for these complex diseases should also be highlighted. Not only can the genes themselves be targeted, but also the TFs which regulate them, providing many potential therapeutic targets and diagnostic aids (Gamazon et al., 2015; Golde, 2016; Nho, 2016).

6.5 Conclusions

In conclusion, both the functional and computational analysis of rs2020000 has suggested that it is significantly associated with altered messenger levels of *ABCA7*. The functional studies suggest that the minor allele of this variant is associated with decreased levels of proteins and the rationale behind this association will need to be examined through Electrophoretic Mobility Shift Assays. It will be important to establish the role of this particular variant in relation to AD risk. However, the methodology utilised here provides a direction for further functional variant discovery within GWAS loci hits for any complex disorder.

7 General Discussion

The ABCA7 protein is a member of the ATP-Binding Cassette family, all of which are transmembrane transporter proteins. ABCA7 itself has been associated with transporting phospholipids, and to a lesser extent cholesterol, as do most of the ABCA family. ABCA7 specifically transports substrates across cell membranes to extracellular apolipoprotein A-I and apolipoprotein E (Abe-Dohmae et al., 2004; Chan et al., 2008; Wang, 2003). However, ABCA7 has also been shown to share a high degree of homology with CED-7: a vital protein in phagocytosis in nematodes. It is now recognized that ABCA7 also regulates phagocytosis in several mammalian cell lines (Iwamoto, 2006; Jehle et al., 2006).

In 2009 rs3764650-G, located in intron 13-14 of *ABCA7*, was reported to be associated with an increased risk of late onset Alzheimer's disease (LOAD) (Hollingworth et al., 2011; Adam C. Naj et al., 2011). However, the mechanisms behind the association of this SNP with LOAD, as well as the role of *ABCA7* in LOAD risk, still remain largely unknown. A large proportion of GWAS hits, including rs3764650, are in noncoding regions of the genome. It is possible, therefore, that these GWAS hits may instead play a role in the regulation of gene or isoform expression, as opposed to the direct protein structure encoded by these genes.

ABCA7 messenger levels are, in fact, elevated during the AD disease process and cognitive decline (Karch et al., 2012) and rs3764650 is itself negatively correlated with *ABCA7* mRNA levels in human brain tissue (Vasquez et al., 2013). This SNP has

also been associated with neuritic plaque burden in AD brain autopsy samples (Shulman et al., 2013) and amyloid-beta (Aβ) plaque load when assessed by Pittsburgh compound B-positron emission tomography (DiB-PET) (Hughes et al., 2014). This implies that ABCA7 plays a role in AD pathogenesis, also corroborated by genetic studies. In a recent re-sequencing project involving Icelandic, European and American populations, loss-of-function *ABCA7* variants increased AD risk with a combined odds ratio (OR) of 2.03 and *p*-value of 6.8×10^{-15} (Steinberg et al., 2015). This has also been replicated in a Belgium cohort (OR = 4.03 and $p = 0.0002$) (Cuyvers et al., 2015).

The idea that these functional variants could be one of the reasons for *ABCA7*'s association with LOAD, as well as the fact that predicting functionality of coding variants is much better understood then noncoding variants, led to the first three of the aims of this thesis. These were to create a catalogue of *ABCA7* exonic variants utilising previous sequencing projects as well as online catalogues; to analyse these variants in order to identify those predicted to affect ABCA7 function before performing genotyping assays on these variants to determine if they were associated with disease. This full catalogue of variants was created and is available in Appendix A. Based on the data collected, as well as the annotations from the prediction programmes used, a total of five variants were identified as being potentially damaging to ABCA7's function as well as having a minor allele frequency (MAF) high enough in order to genotype them in the ARUK DNA bank whilst achieving power of at least 75%. All five were therefore genotyped utilising the KASP™ genotyping system with rs3752239 (OR = 0.86, $p = 0.039$), rs59851484 (OR = 0.28, $p = 0.022$) and 19:1056958 T>C (OR = 3.53, $p = 0.38$) all presenting promising, suggestive results. However, no results were significant following correction for covariates or multiple testing and replication will therefore be required in a larger dataset, especially for the variant at genomic position 19:1056958 due to its rarity.

However, what is of interest is the location of rs778244634. This variant is located within exon 34 which is the same functional domain (transmembrane eight) as a variant in strong LD with the GWAS tag SNP; rs3752246 (Adzhubei et al., 2010; Hollingworth et al., 2011). A variant discussed below is also within this exon, rs881768, however is not within this transmembrane domain but within the large extracellular domain prior to this domain.

The final aim of this thesis was to design and carry out *in vitro* assays to analyse potential functional *ABCA7* variants, presented in Chapters 4, 5 and 6. In Chapter 4, the variant rs881768, identified through a Next Generation Sequencing (NGS) project, was predicted by *in silico* tools to affect splicing of exon 32 of *ABCA7*. The minor allele (G) was predicted to remove an SRSF6 site, introduce an SRSF1 site (by the ESEfinder programme), introduce a novel donor site (by the BDGP site), activate a cryptic donor site and alter both an ESE and an ESS site (by Human Splicing Finder). This variant was therefore analysed utilising the minigene assay where it was apparent the minor allele (G) was the only genotype which showed inclusion of exon 32, corroborating the *in silico* predictions of the minor allele strengthening the acceptor site. However, this was not replicated in expression databases such as GTEx, RNA extracted from brain or lymphoblastoid cells, even when the surveillance pathway nonsense mediated decay (NMD) was inhibited in Chapter 5; implying this truncated isoform is not produced *in vivo*. LOCATION RE rs3752246 (Hollingworth)

This data shows that exon 32 is, in fact, incorporated in all transcripts suggesting that, perhaps both the *in silico* and the *in vitro* tools used here may not be entirely accurate. Indeed, at the time the NGS project was performed, the best predictor of splicing variants was the Variant Effect Predictor (VEP) which only examines the first three base pairs in the exons and the first eight base pairs in the introns, therefore potentially missing many splicing mutations. More high-throughput programmes are

available now but it is known that numerous splicing variants have been missed or incorrectly annotated due to the inaccuracy of programmes such as these (Singh and Cooper, 2012).

In Chapter 6, the variant rs2020000 was examined for its potential to regulate the ABCA7 protein after *in silico* analysis presented it as a strong functional candidate. Both assessment of gene expression in brain tissue, through eGWAS (Allen et al., 2012), and *in vitro* assessment, performed here; indicate the minor allele (C) does alter expression of *ABCA7* mRNA although this was in opposite directions. In brain tissue, as well as in expression databases such as GTEx, the minor allele appears to increase *ABCA7* mRNA expression but within the dual luciferase cellular assays, it appears to downregulate reporter gene activity. This contradiction between the cellular assays performed here and the microarray and RNA-seq results may be due to several caveats. For example, this variant may act in a regulatory haplotype not included in the cellular assays, as well as the transformed cell line used (BE(2)-C) may not be entirely representative of the gene regulatory architecture *in vivo*.

However, there is also little evidence of association between rs2020000 and LOAD, shown upon association testing performed upon the same genotyping data acquired from the Mayo Clinic as used in Chapter 3. Despite this apparent lack of disease association, this variant does still have regulatory potential, although the mechanisms behind its action remain unknown. This will be examined through electromobility shift assays (EMSAs) in the future to observe if the difference in expression is due to a difference in transcription factor binding. Other *ABCA7* variants will also be examined in the same way - specifically rs2072102 and rs7247087, presented in Allen et al., 2012 as part of the same project which identified rs2020000.

The pipeline developed as part of this thesis will be invaluable in the future in order to identify further regulatory variants, as it is becoming increasingly obvious that gene expression plays a key role in complex disease pathology (Gamazon et al., 2015; Maurano et al., 2012; Zou et al., 2012). However, expression SNPs still remain remarkably difficult to annotate due to the fact that the linear relationship between the SNP and the nearest gene cannot be inferred and gene expression can be regulated by multiple elements (Corradin et al., 2014). More and more programmes are becoming available in order to provide these annotations, for example ENCODE and RegulomeDB. However they are not always tissue specific due to the eQTL architectures altering significantly between different tissue and cell types.

The exponential increase in sequencing projects designed to identify the functional variations associated with LOAD has led to a vast amount of sequencing data being available. However, we are still very much reliant on functional databases and prediction programmes in order to annotate this data correctly. These types of platforms have become increasingly common and have gained better accuracy over the past few years. However, as presented here, they are still not entirely representative and there is a need for better predictions. Improvements in these *in silico* databases will allow increasingly accurate functional annotation of sequencing data and, therefore, help direct future laboratory experimental design.

7.1 AD Genetics Update

Despite seven years since the original LOAD GWAS were published (Harold et al., 2009; Lambert et al., 2009), we are still not much closer to identifying the precise genetic aetiology of this devastating disease.

The hypothesis originally adopted when designing GWAS was "common disease - common variant" where common diseases, such as LOAD, are caused by a combination of common variants, present in all populations (Gibson, 2012). However, given that there is still a lack of knowledge of the full heritability of LOAD - anywhere from 44% to 60% has been said to still be missing in recent studies (Escott-Price et al., 2016; Ridge et al., 2016) - alternative hypotheses have been theorised. These include large numbers of small-effect common variants, small numbers of large-effect rare variants as well as a combination of genotypic, environmental and epigenetic factors providing the genetic risk for these diseases (Gibson, 2012). Given the vast number of sequencing projects that have been carried out on these LOAD loci (NGS, whole genome sequencing, exome sequencing etc.), it is going to be more and more problematic to identify the remainder of this "missing heritability."

In the past two years the concept of polygenic risk scores (PRS), in relation to AD genetics, have gained ground. These work on the hypothesis that there are many common variants; each having a small effect on an individual's risk of a complex disorder, and this "genetic background" can be used to give an individual a risk of developing these disorders. Most recently, 87,600 SNPs were used to give a PRS for LOAD with an accuracy of 82% which may be increased further by incorporating age, gender, environmental and clinical data into the algorithm (Escott-Price et al., 2016). However, these authors suggest that any further loci or pathways identified as being associated with LOAD will not add much more to this accuracy as the majority of

them will be so rare. Although additional rare variants have been detected to be associated with AD, for example in the *TREM2* (Guerreiro et al., 2013), *AKAP9* (Logue et al., 2013), *PLD3* (Cruchaga et al., 2014), and *UNC5C* (Wetzel-Smith et al., 2014) loci, these have been few and far between.

This "missing heritability" has also been attributed to epigenetic factors, including hard-to-detect gene interactions, as well as exogenous environmental factors. A small number of these epigenetic factors can be examined through the notoriously difficult to annotate noncoding variants. Despite this difficulty, their role in complex disorders is vital, with 90% of GWAS hits being in noncoding regions (Schaub et al., 2012), where they may in fact, still be functional. For example, up to 80% of GWAS hits have been shown to be in DNase hypersensitivity sites, possibly playing a role in transcription (Gamazon et al., 2015). Functional annotation programmes for these noncoding variants are more common (for example GTEx, RegulomeDB, HaploReg, BDGP and ESEfinder as used in this project). However, tissue and cell specificity of these interactions, as well as gene expression regulation, still need to be considered and are not in databases such as these. Therefore, integrating additional elements into these programmes will only allow increased utilisation of them in order to improve the accuracy in annotating these noncoding variants.

7.2 Future of AD Genetics

As previously discussed, identifying additional rare variants associated with LOAD may not add a huge amount more to our knowledge of this disease's genetic architecture. Further work may, therefore, focus on understanding the molecular mechanisms behind the disease process. Examples of this could include examining chromatin structure within the LOAD loci, in order to explore gene regulation, as well as the use of induced pluripotent stem cells (iPSC). These iPSC, generated from patient samples, provide an excellent alternative to *in vitro* systems currently used, for example the minigene and dual luciferase assays used in this project. The differentiation of human stem cells to neurones, as well as other neuronal cell types such as astrocytes, microglia and oligodendrocytes, will allow the identification and validation of new disease pathways, including possible therapeutic targets, within the background of AD patients' genomes. This will provide further understanding as to how the complex genetic architectures act, in order to appreciate the susceptibility and response to environmental factors these provide. This will possibly supply pharmacologically modifiable targets in these AD patients (Goldstein et al., 2015).

The idea of "risk haplotypes" is becoming more likely, as opposed to single variants defining disease risk, especially with the recent development of PRSs. The indication of an *ABCA7* risk haplotype was discussed in Chapter 6 (see Section 6.4), as well as in Cuyvers et al., 2015. Here the authors identified a four times enrichment of rare, loss-of-function mutations in LOAD patients within *ABCA7*. These mutations are predicted to result in loss of *ABCA7* transcript due to NMD, reducing ABCA7 expression, suggesting that haploinsufficiency may be the mode of action of any pathogenic *ABCA7* mutations (Cuyvers et al., 2015).

However, what is unknown is the molecular mechanism of these loss-of-function variants and why they associate with LOAD risk. Is it because of ABCA7's postulated function within phagocytosis (Jehle et al., 2006)? Does a decrease in ABCA7 lower the ability of microglia to clear apoptotic debris or amyloid plaques? Or altered lipid metabolism (Abe-Dohmae et al., 2004), especially vital in the lipid rich brain? These functionalities - impaired or not - are the kind of processes that could be examined in iPSCs carrying these loss-of-function variations.

7.3 Conclusions

This thesis has focused on identifying functional variants within the late onset Alzheimer's disease candidate gene *ABCA7*. Five potential variants (rs3752233, rs3752239, rs59851484, rs114782266 and 19:1056958 T>C) were identified in Chapter 2 and genotyped in Chapter 3. Three of these (rs3752239, rs59851484 and 19:1056958) showed tentative results. However, further patient samples are required in order to completely clarify their association with LOAD risk.

In Chapters 4 and 5, putative splicing variants (rs881768) and regulatory variants (rs2020000) were examined. Both of these variants did appear to have some functional activity. However, improved functional databases are required, for both coding and noncoding variants, in order to advance identification of such variants out of deep re-sequencing project results. Enhanced *in vitro* assays, for example the use of iPSCs, are also required, in order to assist with the accurate analysis of these functional variants in more biologically relevant systems.

Genetic studies into LOAD, such as GWAS, have provided a massive insight into the molecular mechanisms behind the disease - imparting directions for drug development, disease treatment, diagnosis and screening. With further work, these avenues can be exploited to an even greater extent, in order to put a halt on the advancement of this catastrophic disease.

8 References

Abe-Dohmae, S., Ikeda, Y., Matsuo, M., Hayashi, M., Okuhira, K., Ueda, K., Yokoyama, S., 2004. Human ABCA7 supports apolipoprotein-mediated release of cellular cholesterol and phospholipid to generate high density lipoprotein. J. Biol. Chem. 279, 604–611. doi:10.1074/jbc.M309888200

Abe-Dohmae, S., Ueda, K., Yokoyama, S., 2006. ABCA7, a molecule with unknown function. FEBS Lett. 580. doi:10.1016/j.febslet.2005.12.029

Abel, T., Zukin, R.S., 2008. Epigenetic targets of HDAC inhibition in neurodegenerative and psychiatric disorders. Curr. Opin. Pharmacol. 8, 57–64. doi:10.1016/j.coph.2007.12.002

Adzhubei, I.A., Schmidt, S., Peshkin, L., Ramensky, V.E., Gerasimova, A., Bork, P., Kondrashov, A.S., Sunyaev, S.R., 2010. A method and server for predicting damaging missense mutations. Nat. Methods 7, 248–249. doi:10.1038/nmeth0410-248

Akiyama, M., 2005. Mutations in lipid transporter ABCA12 in harlequin ichthyosis and functional recovery by corrective gene transfer. J. Clin. Invest. 115, 1777–1784. doi:10.1172/JCI24834

Albert, F.W., Kruglyak, L., 2015. The role of regulatory variation in complex traits and disease. Nat. Rev. Genet. 16, 197–212. doi:10.1038/nrg3891

Albrecht, C., Viturro, E., 2007. The ABCA subfamily--gene and protein structures, functions and associated hereditary diseases. Pflüg. Arch. Eur. J. Physiol. 453, 581–589. doi:10.1007/s00424-006-0047-8

Allen, M., Zou, F., Chai, H.S., Younkin, C.S., Crook, J., Pankratz, V.S., Carrasquillo, M.M., Rowley, C.N., Nair, A.A., Middha, S., Maharjan, S., Nguyen, T., Ma, L., Malphrus, K.G., Palusak, R., Lincoln, S., Bisceglio, G., Georgescu, C., Schultz, D., Rakhshan, F., Kolbert, C.P., Jen, J., Haines, J.L., Mayeux, R., Pericak-Vance, M.A., Farrer, L.A., Schellenberg, G.D., Petersen, R.C., Graff-Radford, N.R., Dickson, D.W., Younkin, S.G., Ertekin-Taner, N., Alzheimer's Disease Genetics Consortium (ADGC), Apostolova, L.G., Arnold, S.E., Baldwin, C.T., Barber, R., Barmada, M.M., Beach, T., Beecham, G.W., Beekly, D., Bennett, D.A., Bigio, E.H., Bird, T.D., Blacker, D., Boeve, B.F., Bowen, J.D., Boxer, A., Burke, J.R., Buros, J., Buxbaum, J.D., Cairns, N.J., Cantwell, L.B., Cao, C., Carlson, C.S., Carney, R.M., Carroll, S.L., Chui, H.C., Clark, D.G., Corneveaux, J., Cotman, C.W., Crane, P.K., Cruchaga, C., Cummings, J.L., De Jager, P.L., DeCarli, C., DeKosky, S.T., Demirci, F.Y., Diaz-Arrastia, R., Dick, M., Dombroski, B.A., Duara, R., Ellis, W.D., Evans, D., Faber, K.M., Fallon, K.B., Farlow, M.R., Ferris, S., Foroud, T.M., Frosch, M., Galasko, D.R., Gallins, P.J., Ganguli, M., Gearing, M., Geschwind, D.H., Ghetti, B., Gilbert, J.R., Gilman, S., Giordani, B., Glass, J.D., Goate, A.M., Green, R.C., Growdon, J.H., Hakonarson, H., Hamilton, R.L., Hardy, J., Harrell, L.E., Head, E., Honig, L.S., Huentelman, M.J., Hulette, C.M., Hyman, B.T., Jarvik, G.P., Jicha, G.A., Jin, L.-W., Jun, G., Kamboh, M.I., Karlawish, J., Karydas, A., Kauwe, J.S.K., Kaye, J.A., Kennedy, N., Kim, R., Koo, E.H., Kowall, N.W., Kramer, P., Kukull, W.A., Lah, J.J., Larson, E.B., Levey, A.I., Lieberman, A.P., Lopez, O.L., Lunetta, K.L., Mack, W.J., Marson, D.C., Martin, E.R., Martiniuk, F., Mash, D.C., Masliah, E., McCormick, W.C., McCurry, S.M., McDavid, A.N., McKee, A.C., Mesulam, M., Miller, B.L., Miller, C.A., Miller, J.W., Montine, T.J., Morris, J.C., Myers, A.J., Naj, A.C., Nowotny, P., Parisi, J.E., Perl, D.P.,

Peskind, E., Poon, W.W., Potter, H., Quinn, J.F., Raj, A., Rajbhandary, R.A., Raskind, M., Reiman, E.M., Reisberg, B., Reitz, C., Ringman, J.M., Roberson, E.D., Rogaeva, E., Rosenberg, R.N., Sano, M., Saykin, A.J., Schneider, J.A., Schneider, L.S., Seeley, W., Shelanski, M.L., Slifer, M.A., Smith, C.D., Sonnen, J.A., Spina, S., St George-Hyslop, P., Stern, R.A., Tanzi, R.E., Trojanowski, J.Q., Troncoso, J.C., Tsuang, D.W., Van Deerlin, V.M., Vardarajan, B.N., Vinters, H.V., Vonsattel, J.P., Wang, L.-S., Weintraub, S., Welsh-Bohmer, K.A., Williamson, J., Woltjer, R.L., 2012. Novel late-onset Alzheimer disease loci variants associate with brain gene expression. Neurology 79, 221–228. doi:10.1212/WNL.0b013e3182605801

Allikmets, R., Wasserman, W.W., Hutchinson, A., Smallwood, P., Nathans, J., Rogan, P.K., Schneider, T.D., Dean, M., 1998. Organization of the ABCR gene: analysis of promoter and splice junction sequences. Gene 215, 111–122.

Altmann, A., Weber, P., Bader, D., Preuss, M., Binder, E.B., Müller-Myhsok, B., 2012. A beginners guide to SNP calling from high-throughput DNA-sequencing data. Hum. Genet. 131, 1541–1554. doi:10.1007/s00439-012-1213-z

Alves, L., Correia, A.S.A., Miguel, R., Alegria, P., Bugalho, P., 2012. Alzheimer's disease: a clinical practice-oriented review. Front. Neurol. 3, 63. doi:10.3389/fneur.2012.00063

Alzheimer, A., Stelzmann, R.A., Schnitzlein, H.N., Murtagh, F.R., 1995. An English translation of Alzheimer's 1907 paper, "Uber eine eigenartige Erkankung der Hirnrinde." Clin. Anat. N. Y. N 8, 429–431. doi:10.1002/ca.980080612

Alzheimer's Association, 2015. 2015 Alzheimer's disease facts and figures. Alzheimers Dement. 11, 332–384. doi:10.1016/j.jalz.2015.02.003

Andreutti-Zaugg, C., Scott, R.J., Iggo, R., 1997. Inhibition of nonsense-mediated messenger RNA decay in clinical samples facilitates detection of human MSH2 mutations with an in vivo fusion protein assay and conventional techniques. Cancer Res. 57, 3288–3293.

Antonell, A., Lladó, A., Altirriba, J., Botta-Orfila, T., Balasa, M., Fernández, M., Ferrer, I., Sánchez-Valle, R., Molinuevo, J.L., 2013. A preliminary study of the whole-genome expression profile of sporadic and monogenic early-onset Alzheimer's disease. Neurobiol. Aging 34, 1772–1778. doi:10.1016/j.neurobiolaging.2012.12.026

Arriagada, P.V., Growdon, J.H., Hedley-Whyte, E.T., Hyman, B.T., 1992. Neurofibrillary tangles but not senile plaques parallel duration and severity of Alzheimer's disease. Neurology 42, 631–639.

Assmann, G., von Eckardstein, A., Brewer, H.B., 2001. Familial Analphalipoproteinemia: Tangier Disease, in: The Online Metabolic and Molecular Bases of Inherited Disease (Valle, D, Beaudet, AL, Vogelstein, B, Kinzler, KW, Antonarakis, AE, Ballabio, A, Gibson, M, Mitchell, G. Eds). McGraw-Hill, New York, pp. 2931–2960.

Bai, B., Hales, C.M., Chen, P.-C., Gozal, Y., Dammer, E.B., Fritz, J.J., Wang, X., Xia, Q., Duong, D.M., Street, C., Cantero, G., Cheng, D., Jones, D.R., Wu, Z., Li, Y., Diner, I., Heilman, C.J., Rees, H.D., Wu, H., Lin, L., Szulwach, K.E., Gearing, M., Mufson, E.J., Bennett, D.A., Montine, T.J., Seyfried, N.T., Wingo, T.S., Sun, Y.E., Jin, P., Hanfelt, J., Willcock, D.M., Levey, A., Lah, J.J., Peng, J., 2013. U1 small nuclear ribonucleoprotein complex and RNA splicing alterations in Alzheimer's disease. Proc. Natl. Acad. Sci. 110, 16562–16567. doi:10.1073/pnas.1310249110

Balding, D.J., Bishop, M., Cannings, C. (Eds.), 2007. Handbook of Statistical Genetics, 3rd ed. John Wiley & Sons, Ltd, Chichester, UK.

Bales, K.R., Verina, T., Cummins, D.J., Du, Y., Dodel, R.C., Saura, J., Fishman, C.E., DeLong, C.A., Piccardo, P., Petegnief, V., Ghetti, B., Paul, S.M., 1999. Apolipoprotein E is essential for amyloid deposition in the APP(V717F)

transgenic mouse model of Alzheimer's disease. Proc. Natl. Acad. Sci. U. S. A. 96, 15233–15238.

Bansal, V., Libiger, O., Torkamani, A., Schork, N.J., 2010. Statistical analysis strategies for association studies involving rare variants. Nat. Rev. Genet. 11, 773–785. doi:10.1038/nrg2867

Baralle, D., Baralle, M., 2005. Splicing in action: assessing disease causing sequence changes. J. Med. Genet. 42, 737–748. doi:10.1136/jmg.2004.029538

Barbet, R., Peiffer, I., Hutchins, J.R.A., Hatzfeld, A., Garrido, E., Hatzfeld, J.A., 2012. Expression of the 49 human ATP binding cassette (ABC) genes in pluripotent embryonic stem cells and in early- and late-stage multipotent mesenchymal stem cells: Possible role of ABC plasma membrane transporters in maintaining human stem cell pluripotency. Cell Cycle 11, 1611–1620. doi:10.4161/cc.20023

Barnes, M.R. (Ed.), 2007. Bioinformatics for geneticists: a bioinformatics primer for the analysis of genetic data, 2nd ed. ed. Wiley, Chichester, England ; Hoboken, NJ.

Barrett, J.C., Fry, B., Maller, J., Daly, M.J., 2005. Haploview: analysis and visualization of LD and haplotype maps. Bioinformatics 21, 263–265. doi:10.1093/bioinformatics/bth457

Barucker, C., Harmeier, A., Weiske, J., Fauler, B., Albring, K.F., Prokop, S., Hildebrand, P., Lurz, R., Heppner, F.L., Huber, O., Multhaup, G., 2014. Nuclear Translocation Uncovers the Amyloid Peptide Aβ42 as a Regulator of Gene Transcription. J. Biol. Chem. 289, 20182–20191. doi:10.1074/jbc.M114.564690

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

Beach, T.G., Monsell, S.E., Phillips, L.E., Kukull, W., 2012. Accuracy of the clinical diagnosis of Alzheimer disease at National Institute on Aging Alzheimer Disease Centers, 2005-2010. J. Neuropathol. Exp. Neurol. 71, 266–273. doi:10.1097/NEN.0b013e31824b211b

Beecham, G.W., Hamilton, K., Naj, A.C., Martin, E.R., Huentelman, M., Myers, A.J., Corneveaux, J.J., Hardy, J., Vonsattel, J.-P., Younkin, S.G., Bennett, D.A., De Jager, P.L., Larson, E.B., Crane, P.K., Kamboh, M.I., Kofler, J.K., Mash, D.C., Duque, L., Gilbert, J.R., Gwirtsman, H., Buxbaum, J.D., Kramer, P., Dickson, D.W., Farrer, L.A., Frosch, M.P., Ghetti, B., Haines, J.L., Hyman, B.T., Kukull, W.A., Mayeux, R.P., Pericak-Vance, M.A., Schneider, J.A., Trojanowski, J.Q., Reiman, E.M., the Alzheimer's Disease Genetics Consortium (ADGC), Schellenberg, G.D., Montine, T.J., 2014. Genome-Wide Association Meta-analysis of Neuropathologic Features of Alzheimer's Disease and Related Dementias. PLoS Genet. 10, e1004606. doi:10.1371/journal.pgen.1004606

Below, J., 2016. Abstract: Tissue-Specific Genome-Wide Predictions of Genetically Regulated Expression in Alzheimer's Disease. Presented at the Alzheimer's International Conference 2016, Toronto, Canada.

Blennow, K., Hampel, H., Weiner, M., Zetterberg, H., 2010. Cerebrospinal fluid and plasma biomarkers in Alzheimer disease. Nat. Rev. Neurol. 6, 131–144. doi:10.1038/nrneurol.2010.4

Bloch, K.J., Buchanan, W.W., Wohl, M.J., Bunim, J.J., 1992. Sjögren's syndrome. A clinical, pathological, and serological study of sixty-two cases. 1965. Medicine (Baltimore) 71, 386-401; discussion 401-403.

Bock, C., Paulsen, M., Tierling, S., Mikeska, T., Lengauer, T., Walter, J., 2006. CpG Island Methylation in Human Lymphocytes Is Highly Correlated with DNA

Sequence, Repeats, and Predicted DNA Structure. PLoS Genet 2, e26. doi:10.1371/journal.pgen.0020026

Bodzioch, M., Orsó, E., Klucken, J., Langmann, T., Böttcher, A., Diederich, W., Drobnik, W., Barlage, S., Büchler, C., Porsch-Özcürümez, M., Kaminski, W.E., Hahmann, H.W., Oette, K., Rothe, G., Aslanidis, C., Lackner, K.J., Schmitz, G., 1999. The gene encoding ATP-binding cassette transporter 1 is mutated in Tangier disease. Nat. Genet. 22, 347–351. doi:10.1038/11914

Borst, P., Elferink, R.O., 2002. Mammalian Abc Transporters in Health and Disease. Annu. Rev. Biochem. 71, 537–592. doi:10.1146/annurev.biochem.71.102301.093055

Borst, P., Zelcer, N., van Helvoort, A., 2000. ABC transporters in lipid transport. Biochim. Biophys. Acta BBA - Mol. Cell Biol. Lipids 1486, 128–144. doi:10.1016/S1388-1981(00)00053-6

Bots, M.L., Breteler, M.M., van Kooten, F., Haverkate, F., Meijer, P., Koudstaal, P.J., Grobbee, D.E., Kluft, C., 1998. Coagulation and fibrinolysis markers and risk of dementia. The Dutch Vascular Factors in Dementia Study. Haemostasis 28, 216–222. doi:22433

Boyle, A.P., Hong, E.L., Hariharan, M., Cheng, Y., Schaub, M.A., Kasowski, M., Karczewski, K.J., Park, J., Hitz, B.C., Weng, S., Cherry, J.M., Snyder, M., 2012. Annotation of functional variation in personal genomes using RegulomeDB. Genome Res. 22, 1790–1797. doi:10.1101/gr.137323.112

Braae, A., 2016. Exploring potential functional variants in the Alzheimer's disease associated genes, CD2AP, EPHA1 and CD33 [WWW Document]. URL http://eprints.nottingham.ac.uk/33083/ (accessed 10.3.16).

Braae, A., Thompson, C., Morgan, K., 2014. University of Nottingham - app note - LGC Group (Application Note). Univeristy of Nottingham.

Brasier, A.R., Ron, D., 1992. [34] Luciferase reporter gene assay in mammalian cells, in: Enzymology, B.-M. in (Ed.), Recombinant DNA Part G. Academic Press, pp. 386–397.

Bredesen, D.E., 2014. Reversal of cognitive decline: A novel therapeutic program. Aging 6, 707–717. doi:10.18632/aging.100690

Broccardo, C., Luciani, M., Chimini, G., 1999. The ABCA subclass of mammalian transporters. Biochim. Biophys. Acta 1461, 395–404.

Broccardo, C., Osorio, J., Luciani, M.-F., Schriml, L.M., Prades, C., Shulenin, S., Arnould, I., Naudin, L., Lafargue, C., Rosier, M., Jordan, B., Mattei, M.G., Dean, M., Denèfle, P., Chimini, G., 2001. Comparative analysis of the promoter structure and genomic organization of the human and mouse ABCA7 gene encoding a novel ABCA transporter. Cytogenet. Genome Res. 92, 264–270. doi:10.1159/000056914

Brookmeyer, R., Corrada, M.M., Curriero, F.C., Kawas, C., 2002. Survival following a diagnosis of Alzheimer disease. Arch. Neurol. 59, 1764–1767.

Bu, G., 2009. Apolipoprotein E and its receptors in Alzheimer's disease: pathways, pathogenesis and therapy. Nat. Rev. Neurosci. 10, 333–344. doi:10.1038/nrn2620

Buil, A., Brown, A.A., Lappalainen, T., Viñuela, A., Davies, M.N., Zheng, H.-F., Richards, J.B., Glass, D., Small, K.S., Durbin, R., Spector, T.D., Dermitzakis, E.T., 2014. Gene-gene and gene-environment interactions detected by transcriptome sequence analysis in twins. Nat. Genet. doi:10.1038/ng.3162

Bungert, S., Molday, L.L., Molday, R.S., 2001. Membrane Topology of the ATP Binding Cassette Transporter ABCR and Its Relationship to ABC1 and Related ABCA Transporters IDENTIFICATION OF N-LINKED GLYCOSYLATION SITES. J. Biol. Chem. 276, 23539–23546. doi:10.1074/jbc.M101902200

Butler, A.W., Ng, M.Y.M., Hamshere, M.L., Forabosco, P., Wroe, R., Al-Chalabi, A., Lewis, C.M., Powell, J.F., 2009. Meta-analysis of linkage studies for

Alzheimer's disease--a web resource. Neurobiol. Aging 30, 1037–1047. doi:10.1016/j.neurobiolaging.2009.03.013

Campbell, C.D., Ogburn, E.L., Lunetta, K.L., Lyon, H.N., Freedman, M.L., Groop, L.C., Altshuler, D., Ardlie, K.G., Hirschhorn, J.N., 2005. Demonstrating stratification in a European American population. Nat. Genet. 37, 868–872. doi:10.1038/ng1607

Campion, D., Dumanchin, C., Hannequin, D., Dubois, B., Belliard, S., Puel, M., Thomas-Anterion, C., Michon, A., Martin, C., Charbonnier, F., Raux, G., Camuzat, A., Penet, C., Mesnage, V., Martinez, M., Clerget-Darpoux, F., Brice, A., Frebourg, T., 1999. Early-onset autosomal dominant Alzheimer disease: prevalence, genetic heterogeneity, and mutation spectrum. Am. J. Hum. Genet. 65, 664–670. doi:10.1086/302553

Cardon, L.R., Palmer, L.J., 2003. Population stratification and spurious allelic association. Lancet Lond. Engl. 361, 598–604. doi:10.1016/S0140-6736(03)12520-2

Carrasquillo, M.M., Crook, J.E., Pedraza, O., Thomas, C.S., Pankratz, V.S., Allen, M., Nguyen, T., Malphrus, K.G., Ma, L., Bisceglio, G.D., Roberts, R.O., Lucas, J.A., Smith, G.E., Ivnik, R.J., Machulda, M.M., Graff-Radford, N.R., Petersen, R.C., Younkin, S.G., Ertekin-Taner, N., 2014. Late-Onset Alzheimer Risk Variants in Memory Decline, Incident Mild Cognitive Impairment and Alzheimer Disease. Neurobiol. Aging. doi:10.1016/j.neurobiolaging.2014.07.042

Carroll, M.C., Katzman, P., Alicot, E.M., Koller, B.H., Geraghty, D.E., Orr, H.T., Strominger, J.L., Spies, T., 1987. Linkage map of the human major histocompatibility complex including the tumor necrosis factor genes. Proc. Natl. Acad. Sci. 84, 8535–8539.

Cartegni, L., Wang, J., Zhu, Z., Zhang, M.Q., Krainer, A.R., 2003. ESEfinder: A web resource to identify exonic splicing enhancers. Nucleic Acids Res. 31, 3568–3571.

Carter, M.S., Doskow, J., Morris, P., Li, S., Nhim, R.P., Sandstedt, S., Wilkinson, M.F., 1995. A regulatory mechanism that detects premature nonsense codons in T-cell receptor transcripts in vivo is reversed by protein synthesis inhibitors in vitro. J. Biol. Chem. 270, 28995–29003.

Cartharius, K., Frech, K., Grote, K., Klocke, B., Haltmeier, M., Klingenhoff, A., Frisch, M., Bayerlein, M., Werner, T., 2005. MatInspector and beyond: promoter analysis based on transcription factor binding sites. Bioinforma. Oxf. Engl. 21, 2933–2942. doi:10.1093/bioinformatics/bti473

Casas-Tinto, S., Zhang, Y., Sanchez-Garcia, J., Gomez-Velazquez, M., Rincon-Limas, D.E., Fernandez-Funez, P., 2011. The ER stress factor XBP1s prevents amyloid-beta neurotoxicity. Hum. Mol. Genet. 20, 2144–2160. doi:10.1093/hmg/ddr100

Castellsagué, E., González, S., Guinó, E., Stevens, K.N., Borràs, E., Raymond, V.M., Lázaro, C., Blanco, I., Gruber, S.B., Capellá, G., 2010. Allele-specific expression of APC in adenomatous polyposis families. Gastroenterology 139, 439–447, 447.e1. doi:10.1053/j.gastro.2010.04.047

Castoreno, A.B., Wang, Y., Stockinger, W., Jarzylo, L.A., Du, H., Pagnon, J.C., Shieh, E.C., Nohturfft, A., 2005. Transcriptional regulation of phagocytosis-induced membrane biogenesis by sterol regulatory element binding proteins. Proc. Natl. Acad. Sci. U. S. A. 102, 13129–13134. doi:10.1073/pnas.0506716102

Chan, S.L., Kim, W.S., Kwok, J.B., Hill, A.F., Cappai, R., Rye, K.-A., Garner, B., 2008. ATP-binding cassette transporter A7 regulates processing of amyloid precursor protein in vitro. J. Neurochem. 106, 793–804. doi:10.1111/j.1471-4159.2008.05433.x

Chen, Z.J., Vulevic, B., Ile, K.E., Soulika, A., Davis, W., Jr, Reiner, P.B., Connop, B.P., Nathwani, P., Trojanowski, J.Q., Tew, K.D., 2004. Association of ABCA2 expression with determinants of Alzheimer's disease. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. 18, 1129–1131. doi:10.1096/fj.03-1490fje

Chouliaras, L., Mastroeni, D., Delvaux, E., Grover, A., Kenis, G., Hof, P.R., Steinbusch, H.W.M., Coleman, P.D., Rutten, B.P.F., van den Hove, D.L.A., 2013. Consistent decrease in global DNA methylation and hydroxymethylation in the hippocampus of Alzheimer's disease patients. Neurobiol. Aging 34, 2091–2099. doi:10.1016/j.neurobiolaging.2013.02.021

Chung, S.J., Kim, M.-J., Kim, Y.J., Kim, J., You, S., Jang, E.H., Kim, S.Y., Lee, J.-H., 2014. CR1, ABCA7, and APOE genes affect the features of cognitive impairment in Alzheimer's disease. J. Neurol. Sci. 339, 91–96. doi:10.1016/j.jns.2014.01.029

Cirulli, E.T., Goldstein, D.B., 2007. In vitro assays fail to predict in vivo effects of regulatory polymorphisms. Hum. Mol. Genet. 16, 1931–1939. doi:10.1093/hmg/ddm140

Clancy, S., 2008. RNA Splicing: Introns, Exons and Spliceosome. Nat. Educ. 1, 31.

Clement, N., Braae, A., Turton, J., Lord, J., Guetta-Baranes, T., Medway, C., Brookes, K., Barber, I., Patel, T., Millar, L., Azzopardi, M., Lowe, J., Mann, D., Pickering-Brown, S., Kalsheker, N., ARUK Consortium, Passmore, P., Chappell, S., Morgan, K., 2016. Investigating splicing variants uncovered by next-generation sequencing the Alzheimer's disease candidate genes, CLU, PICALM, CR1, ABCA7, BIN1, the MS4A locus, CD2AP, EPHA1 and CD33. J. Alzheimers Dis. Park. 6. doi:10.4172/2161-0460.1000276

Cooper, T.A., 2005. Use of minigene systems to dissect alternative splicing elements. Methods San Diego Calif 37, 331–340. doi:10.1016/j.ymeth.2005.07.015

Cooper, T.A., Mattox, W., 1997. The regulation of splice-site selection, and its role in human disease. Am. J. Hum. Genet. 61, 259–266. doi:10.1086/514856

Corder, E.H., Saunders, A.M., Strittmatter, W.J., Schmechel, D.E., Gaskell, P.C., Small, G.W., Roses, A.D., Haines, J.L., Pericak-Vance, M.A., 1993. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. Science 261, 921–923.

Corradin, O., Saiakhova, A., Akhtar-Zaidi, B., Myeroff, L., Willis, J., Cowper-Sal lari, R., Lupien, M., Markowitz, S., Scacheri, P.C., 2014. Combinatorial effects of multiple enhancer variants in linkage disequilibrium dictate levels of gene expression to confer susceptibility to common traits. Genome Res. 24, 1–13. doi:10.1101/gr.164079.113

Cruchaga, C., Karch, C.M., Jin, S.C., Benitez, B.A., Cai, Y., Guerreiro, R., Harari, O., Norton, J., Budde, J., Bertelsen, S., Jeng, A.T., Cooper, B., Skorupa, T., Carrell, D., Levitch, D., Hsu, S., Choi, J., Ryten, M., UK Brain Expression Consortium, Hardy, J., Ryten, M., Trabzuni, D., Weale, M.E., Ramasamy, A., Smith, C., Sassi, C., Bras, J., Gibbs, J.R., Hernandez, D.G., Lupton, M.K., Powell, J., Forabosco, P., Ridge, P.G., Corcoran, C.D., Tschanz, J.T., Norton, M.C., Munger, R.G., Schmutz, C., Leary, M., Demirci, F.Y., Bamne, M.N., Wang, X., Lopez, O.L., Ganguli, M., Medway, C., Turton, J., Lord, J., Braae, A., Barber, I., Brown, K., Alzheimer's Research UK Consortium, Passmore, P., Craig, D., Johnston, J., McGuinness, B., Todd, S., Heun, R., Kölsch, H., Kehoe, P.G., Hooper, N.M., Vardy, E.R.L.C., Mann, D.M., Pickering-Brown, S., Brown, K., Kalsheker, N., Lowe, J., Morgan, K., David Smith, A., Wilcock, G., Warden, D., Holmes, C., Pastor, P., Lorenzo-Betancor, O., Brkanac, Z., Scott, E., Topol, E., Morgan, K., Rogaeva, E., Singleton, A.B., Hardy, J., Kamboh, M.I., St George-Hyslop, P., Cairns, N., Morris, J.C., Kauwe, J.S.K., Goate, A.M., 2014. Rare coding variants in the phospholipase D3 gene confer risk for Alzheimer's disease. Nature 505, 550–554. doi:10.1038/nature12825

Cruts, M., Van Broeckhoven, C., 1998. Presenilin mutations in Alzheimer's disease. Hum. Mutat. 11, 183–190. doi:10.1002/(SICI)1098-1004(1998)11:3<183::AID-HUMU1>3.0.CO;2-J

Cuyvers, E., De Roeck, A., Van den Bossche, T., Van Cauwenberghe, C., Bettens, K., Vermeulen, S., Mattheijssens, M., Peeters, K., Engelborghs, S., Vandenbulcke, M., Vandenberghe, R., De Deyn, P.P., Van Broeckhoven, C., Sleegers, K., 2015. Mutations in ABCA7 in a Belgian cohort of Alzheimer's disease patients: a targeted resequencing study. Lancet Neurol. 14, 814–822. doi:10.1016/S1474-4422(15)00133-7

Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R., Lunter, G., Marth, G., Sherry, S.T., McVean, G., Durbin, R., Group, 1000 Genomes Project Analysis, 2011. The Variant Call Format and VCFtools. Bioinformatics btr330. doi:10.1093/bioinformatics/btr330

Davies, P., Maloney, A.J., 1976. Selective loss of central cholinergic neurons in Alzheimer's disease. Lancet 2, 1403.

Day, J.J., Sweatt, J.D., 2011. Epigenetic mechanisms in cognition. Neuron 70, 813–829. doi:10.1016/j.neuron.2011.05.019

Day-Williams, A.G., McLay, K., Drury, E., Edkins, S., Coffey, A.J., Palotie, A., Zeggini, E., 2011. An evaluation of different target enrichment methods in pooled sequencing designs for complex disease association studies. PloS One 6, e26279. doi:10.1371/journal.pone.0026279

De Jager, P.L., Srivastava, G., Lunnon, K., Burgess, J., Schalkwyk, L.C., Yu, L., Eaton, M.L., Keenan, B.T., Ernst, J., McCabe, C., Tang, A., Raj, T., Replogle, J., Brodeur, W., Gabriel, S., Chai, H.S., Younkin, C., Younkin, S.G., Zou, F., Szyf, M., Epstein, C.B., Schneider, J.A., Bernstein, B.E., Meissner, A., Ertekin-Taner, N., Chibnik, L.B., Kellis, M., Mill, J., Bennett, D.A., 2014. Alzheimer's disease: early alterations in brain DNA methylation at ANK1, BIN1, RHBDF2 and other loci. Nat. Neurosci. advance online publication. doi:10.1038/nn.3786

Dean, M., 2002. The Human ATP-Binding Cassette (ABC) Transporter Superfamily [WWW Document]. URL http://www.ncbi.nlm.nih.gov/books/NBK31/ (accessed 1.7.14).

Dean, M., Annilo, T., 2005. Evolution of the ATP-binding cassette (ABC) transporter superfamily in vertebrates. Annu. Rev. Genomics Hum. Genet. 6, 123–142. doi:10.1146/annurev.genom.6.080604.162122

Dementia Statistics - Alzheimer's Research UK [WWW Document], n.d. URL http://www.alzheimersresearchuk.org/dementia-statistics/ (accessed 4.7.14).

Desmet, F.-O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M., Béroud, C., 2009. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. Nucleic Acids Res. gkp215. doi:10.1093/nar/gkp215

Devlin, B., Risch, N., 1995. A comparison of linkage disequilibrium measures for fine-scale mapping. Genomics 29, 311–322. doi:10.1006/geno.1995.9003

Di Resta, C., Manzoni, M., Zoni Berisso, M., Siciliano, G., Benedetti, S., Ferrari, M., 2014. Evaluation of damaging effects of splicing mutations: Validation of an in vitro method for diagnostic laboratories. Clin. Chim. Acta 436, 276–282. doi:10.1016/j.cca.2014.05.026

Dietschy, J.M., Turley, S.D., 2001. Cholesterol metabolism in the brain: Curr. Opin. Lipidol. 12, 105–112. doi:10.1097/00041433-200104000-00003

Doma, M.K., Parker, R., 2006. Endonucleolytic cleavage of eukaryotic mRNAs with stalls in translation elongation. Nature 440, 561–564. doi:10.1038/nature04530

Donahue, C.P., Muratore, C., Wu, J.Y., Kosik, K.S., Wolfe, M.S., 2006. Stabilization of the tau exon 10 stem loop alters pre-mRNA splicing. J. Biol. Chem. 281, 23302–23306. doi:10.1074/jbc.C600143200

Dubois, B., Hampel, H., Feldman, H.H., Scheltens, P., Aisen, P., Andrieu, S., Bakardjian, H., Benali, H., Bertram, L., Blennow, K., Broich, K., Cavedo, E., Crutch, S., Dartigues, J.-F., Duyckaerts, C., Epelbaum, S., Frisoni, G.B., Gauthier, S., Genthon, R., Gouw, A.A., Habert, M.-O., Holtzman, D.M., Kivipelto, M., Lista, S., Molinuevo, J.-L., O'Bryant, S.E., Rabinovici, G.D., Rowe, C., Salloway, S., Schneider, L.S., Sperling, R., Teichmann, M., Carrillo, M.C., Cummings, J., Jack, C.R., 2016. Preclinical Alzheimer's disease: Definition, natural history, and diagnostic criteria. Alzheimers Dement. 12, 292–323. doi:10.1016/j.jalz.2016.02.002

Duits, F.H., Prins, N.D., Lemstra, A.W., Pijnenburg, Y.A.L., Bouwman, F.H., Teunissen, C.E., Scheltens, P., van der Flier, W.M., 2014. Diagnostic impact of CSF biomarkers for Alzheimer's disease in a tertiary memory clinic. Alzheimers Dement. doi:10.1016/j.jalz.2014.05.1753

Dykxhoorn, D.M., 2016. Abstract: ABCA7 Frameshift Deletion Associated with Alzheimer's Disease in African Americans. Presented at the Alzheimer's Association International Conference 2016.

Ebbert, M.T.W., Ridge, P.G., Wilson, A.R., Sharp, A.R., Bailey, M., Norton, M.C., Tschanz, J.T., Munger, R.G., Corcoran, C.D., Kauwe, J.S.K., 2014. Population-based Analysis of Alzheimer's Disease Risk Alleles Implicates Genetic Interactions. Biol. Psychiatry, Mechanisms of Aging and Cognition 75, 732–737. doi:10.1016/j.biopsych.2013.07.008

Edwards, N.C., Hing, Z.A., Perry, A., Blaisdell, A., Kopelman, D.B., Fathke, R., Plum, W., Newell, J., Allen, C.E., S., G., Shapiro, A., Okunji, C., Kosti, I., Shomron, N., Grigoryan, V., Przytycka, T.M., Sauna, Z.E., Salari, R., Mandel-Gutfreund, Y., Komar, A.A., Kimchi-Sarfaty, C., 2012. Characterization of Coding Synonymous and Non-Synonymous Variants in ADAMTS13 Using Ex Vivo and In Silico Approaches. PLoS ONE 7, e38864. doi:10.1371/journal.pone.0038864

Eikelenboom, P., Veerhuis, R., Scheper, W., Rozemuller, A.J.M., van Gool, W.A., Hoozemans, J.J.M., 2006. The significance of neuroinflammation in understanding Alzheimer's disease. J. Neural Transm. Vienna Austria 1996 113, 1685–1695. doi:10.1007/s00702-006-0575-6

El Hiani, Y., Linsdell, P., 2014. Metal Bridges Illuminate Transmembrane Domain Movements During Gating of the Cystic Fibrosis Transmembrane Conductance Regulator Chloride Channel. J. Biol. Chem. doi:10.1074/jbc.M114.593103

Elliott, D., 2010. Molecular Biology of RNA. OUP Oxford, Oxford ; New York.

Ellis, R.E., Jacobson, D.M., Horvitz, H.R., 1991. Genes required for the engulfment of cell corpses during programmed cell death in Caenorhabditis elegans. Genetics 129, 79–94.

Elman, J.A., Oh, H., Madison, C.M., Baker, S.L., Vogel, J.W., Marks, S.M., Crowley, S., O'Neil, J.P., Jagust, W.J., 2014. Neural compensation in older people with brain amyloid-β deposition. Nat. Neurosci. 17, 1316–1318. doi:10.1038/nn.3806

Engelhart, M.J., Geerlings, M.I., Meijer, J., Kiliaan, A., Ruitenberg, A., van Swieten, J.C., Stijnen, T., Hofman, A., Witteman, J.C.M., Breteler, M.M.B., 2004. Inflammatory proteins in plasma and the risk of dementia: the rotterdam study. Arch. Neurol. 61, 668–672. doi:10.1001/archneur.61.5.668

Escott-Price, V., Shoai, M., Pither, R., Williams, J., Hardy, J., 2016. Polygenic score prediction captures nearly all common genetic risk for Alzheimer's disease. Neurobiol. Aging. doi:10.1016/j.neurobiolaging.2016.07.018

Ewers, M., Mattsson, N., Minthon, L., Molinuevo, J.L., Antonell, A., Popp, J., Jessen, F., Herukka, S.-K., Soininen, H., Maetzler, W., Leyhe, T., Bürger, K., Taniguchi, M., Urakami, K., Lista, S., Dubois, B., Blennow, K., Hampel, H., 2014. CSF biomarkers for the differential diagnosis of Alzheimer's disease. A

large-scale international multicenter study. Alzheimers Dement. J. Alzheimers Assoc. 0. doi:10.1016/j.jalz.2014.12.006

Fan, Z., Aman, Y., Ahmed, I., Chetelat, G., Landeau, B., Ray Chaudhuri, K., Brooks, D.J., Edison, P., 2014. Influence of microglial activation on neuronal function in Alzheimer's and Parkinson's disease dementia. Alzheimers Dement. doi:10.1016/j.jalz.2014.06.016

Farfara, D., Lifshitz, V., Frenkel, D., 2008. Neuroprotective and neurotoxic properties of glial cells in the pathogenesis of Alzheimer's disease. J. Cell. Mol. Med. 12, 762–780. doi:10.1111/j.1582-4934.2008.00314.x

Farrer, L.A., 2015. Expanding the genomic roadmap of Alzheimer's disease. Lancet Neurol. 14, 783–785. doi:10.1016/S1474-4422(15)00146-5

Farrer LA, Cupples L, Haines JL, et al, 1997. Effects of age, sex, and ethnicity on the association between apolipoprotein e genotype and alzheimer disease: A meta-analysis. JAMA 278, 1349–1356. doi:10.1001/jama.1997.03550160069041

Ferri, C.P., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., Ganguli, M., Hall, K., Hasegawa, K., Hendrie, H., Huang, Y., Jorm, A., Mathers, C., Menezes, P.R., Rimmer, E., Scazufca, M., Alzheimer's Disease International, 2005. Global prevalence of dementia: a Delphi consensus study. Lancet 366, 2112–2117. doi:10.1016/S0140-6736(05)67889-0

Filippini, N., MacIntosh, B.J., Hough, M.G., Goodwin, G.M., Frisoni, G.B., Smith, S.M., Matthews, P.M., Beckmann, C.F., Mackay, C.E., 2009. Distinct patterns of brain activity in young carriers of the APOE-epsilon4 allele. Proc. Natl. Acad. Sci. U. S. A. 106, 7209–7214. doi:10.1073/pnas.0811879106

Fitzgerald, M. L., Morris, A.L., Chroni, A., Mendez, A.J., Zannis, V.I., Freeman, M.W., 2004. ABCA1 and amphipathic apolipoproteins form high-affinity molecular complexes required for cholesterol efflux. J. Lipid Res. 45, 287–294. doi:10.1194/jlr.M300355-JLR200

Fitzgerald, Michael L, Okuhira, K.-I., Short, G.F., 3rd, Manning, J.J., Bell, S.A., Freeman, M.W., 2004. ATP-binding cassette transporter A1 contains a novel C-terminal VFVNFA motif that is required for its cholesterol efflux and ApoA-I binding activities. J. Biol. Chem. 279, 48477–48485. doi:10.1074/jbc.M409848200

Fu, Y., Hsiao, J.-H.T., Paxinos, G., Halliday, G.M., Kim, W.S., 2015. ABCA5 regulates amyloid-β peptide production and is associated with Alzheimer's disease neuropathology. J. Alzheimers Dis. JAD 43, 857–869. doi:10.3233/JAD-141320

Fuhrmann, M., Bittner, T., Jung, C.K.E., Burgold, S., Page, R.M., Mitteregger, G., Haass, C., LaFerla, F.M., Kretzschmar, H., Herms, J., 2010. Microglial Cx3cr1 knockout prevents neuron loss in a mouse model of Alzheimer's disease. Nat. Neurosci. 13, 411–413. doi:10.1038/nn.2511

Funk, K.E., Mirbaha, H., Jiang, H., Holtzman, D.M., Diamond, M.I., 2015. Distinct therapeutic mechanisms of Tau antibodies: promoting microglial clearance vs. blocking neuronal uptake. J. Biol. Chem. jbc.M115.657924. doi:10.1074/jbc.M115.657924

Gamazon, E.R., Wheeler, H.E., Shah, K.P., Mozaffari, S.V., Aquino-Michaels, K., Carroll, R.J., Eyler, A.E., Denny, J.C., Nicolae, D.L., Cox, N.J., Im, H.K., 2015. A gene-based association method for mapping traits using reference transcriptome data. Nat. Genet. 47, 1091–1098. doi:10.1038/ng.3367

Geerlings, M.I., den Heijer, T., Koudstaal, P.J., Hofman, A., Breteler, M.M.B., 2008. History of depression, depressive symptoms, and medial temporal lobe atrophy and the risk of Alzheimer disease. Neurology 70, 1258–1264. doi:10.1212/01.wnl.0000308937.30473.d1

Geillon, F., Gondcaille, C., Charbonnier, S., Roermund, C.W. van, Lopez, T.E., Dias, A.M.M., Barros, J.-P.P. de, Arnould, C., Wanders, R.J., Trompier, D., Savary,

S., 2014. Structure-function analysis of peroxisomal ABC transporters using chimeric dimers. J. Biol. Chem. jbc.M114.575506. doi:10.1074/jbc.M114.575506

Genin, E., Hannequin, D., Wallon, D., Sleegers, K., Hiltunen, M., Combarros, O., Bullido, M.J., Engelborghs, S., De Deyn, P., Berr, C., Pasquier, F., Dubois, B., Tognoni, G., Fiévet, N., Brouwers, N., Bettens, K., Arosio, B., Coto, E., Del Zompo, M., Mateo, I., Epelbaum, J., Frank-Garcia, A., Helisalmi, S., Porcellini, E., Pilotto, A., Forti, P., Ferri, R., Scarpini, E., Siciliano, G., Solfrizzi, V., Sorbi, S., Spalletta, G., Valdivieso, F., Vepsäläinen, S., Alvarez, V., Bosco, P., Mancuso, M., Panza, F., Nacmias, B., Bossù, P., Hanon, O., Piccardi, P., Annoni, G., Seripa, D., Galimberti, D., Licastro, F., Soininen, H., Dartigues, J.-F., Kamboh, M.I., Van Broeckhoven, C., Lambert, J.C., Amouyel, P., Campion, D., 2011. APOE and Alzheimer disease: a major gene with semi-dominant inheritance. Mol. Psychiatry 16, 903–907. doi:10.1038/mp.2011.52

Gewirtz, A.M., Calabretta, B., 1988. A c-myb antisense oligodeoxynucleotide inhibits normal human hematopoiesis in vitro. Science 242, 1303–1306.

Gibson, G., 2012. Rare and common variants: twenty arguments. Nat. Rev. Genet. 13, 135–145. doi:10.1038/nrg3118

Golde, T.E., 2016. Abstract: From Systems Level Transcriptomics to New Immune Targets for Alzheimer's Disease. Presented at the Alzheimer's Association International Conference 2016.

Goldman, J.S., Hahn, S.E., Catania, J.W., LaRusse-Eckert, S., Butson, M.B., Rumbaugh, M., Strecker, M.N., Roberts, J.S., Burke, W., Mayeux, R., Bird, T., 2011. Genetic counseling and testing for Alzheimer disease: Joint practice guidelines of the American College of Medical Genetics and the National Society of Genetic Counselors. Genet. Med. 13, 597–605. doi:10.1097/GIM.0b013e31821d69b8

Goldstein, J.L., Brown, M.S., 2009. History of Discovery: The LDL Receptor. Arterioscler. Thromb. Vasc. Biol. 29, 431–438. doi:10.1161/ATVBAHA.108.179564

Goldstein, L.S.B., Reyna, S., Woodruff, G., 2015. Probing the Secrets of Alzheimer's Disease Using Human-induced Pluripotent Stem Cell Technology. Neurotherapeutics 12, 121–125. doi:10.1007/s13311-014-0326-6

Goll, D.E., Thompson, V.F., Li, H., Wei, W., Cong, J., 2003. The calpain system. Physiol. Rev. 83, 731–801. doi:10.1152/physrev.00029.2002

Goymer, P., 2007. Synonymous mutations break their silence. Nat. Rev. Genet. 8, 92–92. doi:10.1038/nrg2056

Griffin, W.S., Sheng, J.G., Royston, M.C., Gentleman, S.M., McKenzie, J.E., Graham, D.I., Roberts, G.W., Mrak, R.E., 1998. Glial-neuronal interactions in Alzheimer's disease: the potential role of a "cytokine cycle" in disease progression. Brain Pathol. Zurich Switz. 8, 65–72.

GTEx Consortium, 2015. Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science 348, 648–660. doi:10.1126/science.1262110

Guerreiro, R., Wojtas, A., Bras, J., Carrasquillo, M., Rogaeva, E., Majounie, E., Cruchaga, C., Sassi, C., Kauwe, J.S.K., Younkin, S., Hazrati, L., Collinge, J., Pocock, J., Lashley, T., Williams, J., Lambert, J.-C., Amouyel, P., Goate, A., Rademakers, R., Morgan, K., Powell, J., St George-Hyslop, P., Singleton, A., Hardy, J., Alzheimer Genetic Analysis Group, 2013. TREM2 variants in Alzheimer's disease. N. Engl. J. Med. 368, 117–127. doi:10.1056/NEJMoa1211851

Guo, J.L., Lee, V.M.Y., 2014. Cell-to-cell transmission of pathogenic proteins in neurodegenerative diseases. Nat. Med. 20, 130–138. doi:10.1038/nm.3457

Gusareva, E.S., Carrasquillo, M.M., Bellenguez, C., Cuyvers, E., Colon, S., Graff-Radford, N.R., Petersen, R.C., Dickson, D.W., Mahachie John, J.M., Bessonov, K., Van Broeckhoven, C., Harold, D., Williams, J., Amouyel, P., Sleegers, K., Ertekin-Taner, N., Lambert, J.-C., Van Steen, K., 2014. Genome-wide association interaction analysis for Alzheimer's disease. Neurobiol. Aging 35, 2436–2443. doi:10.1016/j.neurobiolaging.2014.05.014

Gusev, A., Ko, A., Shi, H., Bhatia, G., Chung, W., Penninx, B.W.J.H., Jansen, R., de Geus, E.J.C., Boomsma, D.I., Wright, F.A., Sullivan, P.F., Nikkola, E., Alvarez, M., Civelek, M., Lusis, A.J., Lehtimäki, T., Raitoharju, E., Kähönen, M., Seppälä, I., Raitakari, O.T., Kuusisto, J., Laakso, M., Price, A.L., Pajukanta, P., Pasaniuc, B., 2016. Integrative approaches for large-scale transcriptome-wide association studies. Nat. Genet. 48, 245–252. doi:10.1038/ng.3506

Gyimesi, G., Borsodi, D., Sarankó, H., Tordai, H., Sarkadi, B., Hegedűs, T., 2012. ABCMdb: a database for the comparative analysis of protein mutations in ABC transporters, and a potential framework for a general application. Hum. Mutat. 33, 1547–1556. doi:10.1002/humu.22138

Haass, C., Selkoe, D.J., 2007. Soluble protein oligomers in neurodegeneration: lessons from the Alzheimer's amyloid beta-peptide. Nat. Rev. Mol. Cell Biol. 8, 101–112. doi:10.1038/nrm2101

Hanenberg, M., McAfoose, J., Kulic, L., Welt, T., Wirth, F., Parizek, P., Strobel, L., Cattepoel, S., Spani, C., Derungs, R., Maier, M., Pluckthun, A., Nitsch, R.M., 2014. Amyloid- Peptide-specific DARPins as a Novel Class of Potential Therapeutics for Alzheimer Disease. J. Biol. Chem. 289, 27080–27089. doi:10.1074/jbc.M114.564013

Hardy, J., Allsop, D., 1991. Amyloid deposition as the central event in the aetiology of Alzheimer's disease. Trends Pharmacol. Sci. 12, 383–388.

Hardy, J.A., Higgins, G.A., 1992. Alzheimer's disease: the amyloid cascade hypothesis. Science 256, 184–185.

Harold, D., Abraham, R., Hollingworth, P., Sims, R., Gerrish, A., Hamshere, M.L., Pahwa, J.S., Moskvina, V., Dowzell, K., Williams, A., Jones, N., Thomas, C., Stretton, A., Morgan, A.R., Lovestone, S., Powell, J., Proitsi, P., Lupton, M.K., Brayne, C., Rubinsztein, D.C., Gill, M., Lawlor, B., Lynch, A., Morgan, K., Brown, K.S., Passmore, P.A., Craig, D., McGuinness, B., Todd, S., Holmes, C., Mann, D., Smith, A.D., Love, S., Kehoe, P.G., Hardy, J., Mead, S., Fox, N., Rossor, M., Collinge, J., Maier, W., Jessen, F., Schürmann, B., Heun, R., van den Bussche, H., Heuser, I., Kornhuber, J., Wiltfang, J., Dichgans, M., Frölich, L., Hampel, H., Hüll, M., Rujescu, D., Goate, A.M., Kauwe, J.S.K., Cruchaga, C., Nowotny, P., Morris, J.C., Mayo, K., Sleegers, K., Bettens, K., Engelborghs, S., De Deyn, P.P., Van Broeckhoven, C., Livingston, G., Bass, N.J., Gurling, H., McQuillin, A., Gwilliam, R., Deloukas, P., Al-Chalabi, A., Shaw, C.E., Tsolaki, M., Singleton, A.B., Guerreiro, R., Mühleisen, T.W., Nöthen, M.M., Moebus, S., Jöckel, K.-H., Klopp, N., Wichmann, H.-E., Carrasquillo, M.M., Pankratz, V.S., Younkin, S.G., Holmans, P.A., O'Donovan, M., Owen, M.J., Williams, J., 2009. Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease. Nat. Genet. 41, 1088–1093. doi:10.1038/ng.440

Hartley, D., Blumenthal, T., Carrillo, M., DiPaolo, G., Esralew, L., Gardiner, K., Granholm, A.-C., Iqbal, K., Krams, M., Lemere, C., Lott, I., Mobley, W., Ness, S., Nixon, R., Potter, H., Reeves, R., Sabbagh, M., Silverman, W., Tycko, B., Whitten, M., Wisniewski, T., 2014. Down syndrome and Alzheimer's disease: Common pathways, common goals. Alzheimers Dement. J. Alzheimers Assoc. 0. doi:10.1016/j.jalz.2014.10.007

Hayashi, M., Abe-Dohmae, S., Okazaki, M., Ueda, K., Yokoyama, S., 2005. Heterogeneity of high density lipoprotein generated by ABCA1 and ABCA7. J. Lipid Res. 46, 1703–1711. doi:10.1194/jlr.M500092-JLR200

He, L., Vasiliou, K., Nebert, D.W., 2009. Analysis and update of the human solute carrier (SLC) gene superfamily. Hum. Genomics 3, 195–206.

Heppner, F.L., Ransohoff, R.M., Becher, B., 2015. Immune attack: the role of inflammation in Alzheimer disease. Nat. Rev. Neurosci. 16, 358–372. doi:10.1038/nrn3880

Higgins, C.F., 1992. ABC transporters: from microorganisms to man. Annu. Rev. Cell Biol. 8, 67–113. doi:10.1146/annurev.cb.08.110192.000435

Hirsch-Reinshagen, V., Burgess, B.L., Wellington, C.L., 2008. Why lipids are important for Alzheimer disease? Mol. Cell. Biochem. 326, 121–129. doi:10.1007/s11010-008-0012-2

Hofman, A., Ott, A., Breteler, M.M., Bots, M.L., Slooter, A.J., van Harskamp, F., van Duijn, C.N., Van Broeckhoven, C., Grobbee, D.E., 1997. Atherosclerosis, apolipoprotein E, and prevalence of dementia and Alzheimer's disease in the Rotterdam Study. Lancet 349, 151–154. doi:10.1016/S0140-6736(96)09328-2

Hollingworth, P., Harold, D., Sims, R., Gerrish, A., Lambert, J.-C., Carrasquillo, M.M., Abraham, R., Hamshere, M.L., Pahwa, J.S., Moskvina, V., Dowzell, K., Jones, N., Stretton, A., Thomas, C., Richards, A., Ivanov, D., Widdowson, C., Chapman, J., Lovestone, S., Powell, J., Proitsi, P., Lupton, M.K., Brayne, C., Rubinsztein, D.C., Gill, M., Lawlor, B., Lynch, A., Brown, K.S., Passmore, P.A., Craig, D., McGuinness, B., Todd, S., Holmes, C., Mann, D., Smith, A.D., Beaumont, H., Warden, D., Wilcock, G., Love, S., Kehoe, P.G., Hooper, N.M., Vardy, E.R.L.C., Hardy, J., Mead, S., Fox, N.C., Rossor, M., Collinge, J., Maier, W., Jessen, F., Rüther, E., Schürmann, B., Heun, R., Kölsch, H., van den Bussche, H., Heuser, I., Kornhuber, J., Wiltfang, J., Dichgans, M., Frölich, L., Hampel, H., Gallacher, J., Hüll, M., Rujescu, D., Giegling, I., Goate, A.M., Kauwe, J.S.K., Cruchaga, C., Nowotny, P., Morris, J.C., Mayo, K., Sleegers, K., Bettens, K., Engelborghs, S., De Deyn, P.P., Van Broeckhoven, C., Livingston, G., Bass, N.J., Gurling, H., McQuillin, A., Gwilliam, R., Deloukas, P., Al-Chalabi, A., Shaw, C.E., Tsolaki, M., Singleton, A.B., Guerreiro, R., Mühleisen, T.W., Nöthen, M.M., Moebus, S., Jöckel, K.-H., Klopp, N., Wichmann, H.-E., Pankratz, V.S., Sando, S.B., Aasly, J.O., Barcikowska, M., Wszolek, Z.K., Dickson, D.W., Graff-Radford, N.R., Petersen, R.C., van Duijn, C.M., Breteler, M.M.B., Ikram, M.A., DeStefano, A.L., Fitzpatrick, A.L., Lopez, O., Launer, L.J., Seshadri, S., Berr, C., Campion, D., Epelbaum, J., Dartigues, J.-F., Tzourio, C., Alpérovitch, A., Lathrop, M., Feulner, T.M., Friedrich, P., Riehle, C., Krawczak, M., Schreiber, S., Mayhaus, M., Nicolhaus, S., Wagenpfeil, S., Steinberg, S., Stefansson, H., Stefansson, K., Snædal, J., Björnsson, S., Jonsson, P.V., Chouraki, V., Genier-Boley, B., Hiltunen, M., Soininen, H., Combarros, O., Zelenika, D., Delepine, M., Bullido, M.J., Pasquier, F., Mateo, I., Frank-Garcia, A., Porcellini, E., Hanon, O., Coto, E., Alvarez, V., Bosco, P., Siciliano, G., Mancuso, M., Panza, F., Solfrizzi, V., Nacmias, B., Sorbi, S., Bossù, P., Piccardi, P., Arosio, B., Annoni, G., Seripa, D., Pilotto, A., Scarpini, E., Galimberti, D., Brice, A., Hannequin, D., Licastro, F., Jones, L., Holmans, P.A., Jonsson, T., Riemenschneider, M., Morgan, K., Younkin, S.G., Owen, M.J., O'Donovan, M., Amouyel, P., Williams, J., 2011. Common variants at ABCA7, MS4A6A/MS4A4E, EPHA1, CD33 and CD2AP are associated with Alzheimer's disease. Nat. Genet. 43, 429–435. doi:10.1038/ng.803

Holmen, O.L., Zhang, H., Fan, Y., Hovelson, D.H., Schmidt, E.M., Zhou, W., Guo, Y., Zhang, Ji, Langhammer, A., Løchen, M.-L., Ganesh, S.K., Vatten, L., Skorpen, F., Dalen, H., Zhang, Jifeng, Pennathur, S., Chen, J., Platou, C.,

Mathiesen, E.B., Wilsgaard, T., Njølstad, I., Boehnke, M., Chen, Y.E., Abecasis, G.R., Hveem, K., Willer, C.J., 2014. Systematic evaluation of coding variation identifies a candidate causal variant in TM6SF2 influencing total cholesterol and myocardial infarction risk. Nat. Genet. 46, 345–351. doi:10.1038/ng.2926

Holton, P., Ryten, M., Nalls, M., Trabzuni, D., Weale, M.E., Hernandez, D., Crehan, H., Gibbs, J.R., Mayeux, R., Haines, J.L., Farrer, L.A., Pericak-Vance, M.A., Schellenberg, G.D., Alzheimer's Disease Genetics Consortium, Ramirez-Restrepo, M., Engel, A., Myers, A.J., Corneveaux, J.J., Huentelman, M.J., Dillman, A., Cookson, M.R., Reiman, E.M., Singleton, A., Hardy, J., Guerreiro, R., 2013. Initial assessment of the pathogenic mechanisms of the recently identified Alzheimer risk Loci. Ann. Hum. Genet. 77, 85–105. doi:10.1111/ahg.12000

Holtzman, D.M., Herz, J., Bu, G., 2012. Apolipoprotein E and Apolipoprotein E Receptors: Normal Biology and Roles in Alzheimer Disease. Cold Spring Harb. Perspect. Med. 2. doi:10.1101/cshperspect.a006312

Honig, L.S., Tang, M.-X., Albert, S., Costa, R., Luchsinger, J., Manly, J., Stern, Y., Mayeux, R., 2003. Stroke and the risk of Alzheimer disease. Arch. Neurol. 60, 1707–1712. doi:10.1001/archneur.60.12.1707

Hoogendoorn, B., Coleman, S.L., Guy, C.A., Smith, K., Bowen, T., Buckland, P.R., O'Donovan, M.C., 2003. Functional analysis of human promoter polymorphisms. Hum. Mol. Genet. 12, 2249–2254. doi:10.1093/hmg/ddg246

Horton, J.D., Shimomura, I., 1999. Sterol regulatory element-binding proteins: activators of cholesterol and fatty acid biosynthesis. Curr. Opin. Lipidol. 10, 143–150.

Houdayer, C., Dehainault, C., Mattler, C., Michaux, D., Caux-Moncoutier, V., Pagès-Berhouet, S., d'Enghien, C.D., Laugé, A., Castera, L., Gauthier-Villars, M., Stoppa-Lyonnet, D., 2008. Evaluation of in silico splice tools for decision-making in molecular diagnosis. Hum. Mutat. 29, 975–982. doi:10.1002/humu.20765

Hrycyna, C.A., Ramachandra, M., Germann, U.A., Cheng, P.W., Pastan, I., Gottesman, M.M., 1999. Both ATP sites of human P-glycoprotein are essential but not symmetric. Biochemistry (Mosc.) 38, 13887–13899.

Hughes, T.M., Lopez, O.L., Evans, R.W., Kamboh, M.I., Williamson, J.D., Klunk, W.E., Mathis, C.A., Price, J.C., Cohen, A.D., Snitz, B.E., Dekosky, S.T., Kuller, L.H., 2014. Markers of cholesterol transport are associated with amyloid deposition in the brain. Neurobiol. Aging 35, 802–807. doi:10.1016/j.neurobiolaging.2013.09.040

Hye, A., Riddoch-Contreras, J., Baird, A.L., Ashton, N.J., Bazenet, C., Leung, R., Westman, E., Simmons, A., Dobson, R., Sattlecker, M., Lupton, M., Lunnon, K., Keohane, A., Ward, M., Pike, I., Zucht, H.D., Pepin, D., Zheng, W., Tunnicliffe, A., Richardson, J., Gauthier, S., Soininen, H., Kłoszewska, I., Mecocci, P., Tsolaki, M., Vellas, B., Lovestone, S., 2014. Plasma proteins predict conversion to dementia from prodromal disease. Alzheimers Dement. 10, 799–807.e2. doi:10.1016/j.jalz.2014.05.1749

Igartua, C., Myers, R.A., Mathias, R.A., Pino-Yanes, M., Eng, C., Graves, P.E., Levin, A.M., Del-Rio-Navarro, B.E., Jackson, D.J., Livne, O.E., Rafaels, N., Edlund, C.K., Yang, J.J., Huntsman, S., Salam, M.T., Romieu, I., Mourad, R., Gern, J.E., Lemanske, R.F., Wyss, A., Hoppin, J.A., Barnes, K.C., Burchard, E.G., Gauderman, W.J., Martinez, F.D., Raby, B.A., Weiss, S.T., Williams, L.K., London, S.J., Gilliland, F.D., Nicolae, D.L., Ober, C., 2015. Ethnic-specific associations of rare and low-frequency DNA sequence variants with asthma. Nat. Commun. 6, 5965. doi:10.1038/ncomms6965

Iida, A., Saito, S., Sekine, A., Mishima, C., Kitamura, Y., Kondo, K., Harigae, S., Osawa, S., Nakamura, Y., 2002. Catalog of 605 single-nucleotide

polymorphisms (SNPs) among 13 genes encoding human ATP-binding cassette transporters: ABCA4, ABCA7, ABCA8, ABCD1, ABCD3, ABCD4, ABCE1, ABCF1, ABCG1, ABCG2, ABCG4, ABCG5, and ABCG8. J. Hum. Genet. 47, 285–310. doi:10.1007/s100380200041

Ikeda, Y., Abe-Dohmae, S., Munehira, Y., Aoki, R., Kawamoto, S., Furuya, A., Shitara, K., Amachi, T., Kioka, N., Matsuo, M., Yokoyama, S., Ueda, K., 2003. Posttranscriptional regulation of human ABCA7 and its function for the apoA-I-dependent lipid release. Biochem. Biophys. Res. Commun. 311, 313–318. doi:10.1016/j.bbrc.2003.10.002

Ingman, M., Gyllensten, U., 2009. SNP frequency estimation using massively parallel sequencing of pooled DNA. Eur. J. Hum. Genet. 17, 383–386. doi:10.1038/ejhg.2008.182

Inoue, F., Kircher, M., Martin, B., Cooper, G.M., Witten, D.M., McManus, M.T., Ahituv, N., Shendure, J., 2016. A systematic comparison reveals substantial differences in chromosomal versus episomal encoding of enhancer activity (No. biorxiv;061606v1).

International Genomics of Alzheimer's Disease Consortium (IGAP), International Genomics of Alzheimer's Disease Consortium IGAP, 2014. Convergent genetic and expression data implicate immunity in Alzheimer's disease. Alzheimers Dement. J. Alzheimers Assoc. doi:10.1016/j.jalz.2014.05.1757

Iwamoto, N., 2006. ABCA7 expression is regulated by cellular cholesterol through the SREBP2 pathway and associated with phagocytosis. J. Lipid Res. 47, 1915–1927. doi:10.1194/jlr.M600127-JLR200

Jacobsson, L.T., Axell, T.E., Hansen, B.U., Henricsson, V.J., Larsson, A., Lieberkind, K., Lilja, B., Manthorpe, R., 1989. Dry eyes or mouth--an epidemiological study in Swedish adults, with special reference to primary Sjögren's syndrome. J. Autoimmun. 2, 521–527.

Jehle, A.W., Gardai, S.J., Li, S., Linsel-Nitschke, P., Morimoto, K., Janssen, W.J., Vandivier, R.W., Wang, N., Greenberg, S., Dale, B.M., Qin, C., Henson, P.M., Tall, A.R., 2006. ATP-binding cassette transporter A7 enhances phagocytosis of apoptotic cells and associated ERK signaling in macrophages. J. Cell Biol. 174, 547–556. doi:10.1083/jcb.200601030

Jennelle, L., Hunegnaw, R., Dubrovsky, L., Pushkarsky, T., Fitzgerald, M.L., Sviridov, D., Popratiloff, A., Brichacek, B., Bukrinsky, M., 2014. HIV-1 Protein Nef Inhibits Activity of ATP Binding Cassette Transporter A1 by Targeting Endoplasmic Reticulum Chaperone Calnexin. J. Biol. Chem. jbc.M114.583591. doi:10.1074/jbc.M114.583591

Jiang, Q., Lee, C.Y.D., Mandrekar, S., Wilkinson, B., Cramer, P., Zelcer, N., Mann, K., Lamb, B., Willson, T.M., Collins, J.L., Richardson, J.C., Smith, J.D., Comery, T.A., Riddell, D., Holtzman, D.M., Tontonoz, P., Landreth, G.E., 2008. ApoE promotes the proteolytic degradation of Abeta. Neuron 58, 681–693. doi:10.1016/j.neuron.2008.04.010

Jin, S.C., Pastor, P., Cooper, B., Cervantes, S., Benitez, B.A., Razquin, C., Goate, A., Ibero-American Alzheimer Disease Genetics Group Researchers, Cruchaga, C., 2012. Pooled-DNA sequencing identifies novel causative variants in PSEN1, GRN and MAPT in a clinical early-onset and familial Alzheimer's disease Ibero-American cohort. Alzheimers Res. Ther. 4, 34. doi:10.1186/alzrt137

Jonsson, T., Atwal, J.K., Steinberg, S., Snaedal, J., Jonsson, P.V., Bjornsson, S., Stefansson, H., Sulem, P., Gudbjartsson, D., Maloney, J., Hoyte, K., Gustafson, A., Liu, Y., Lu, Y., Bhangale, T., Graham, R.R., Huttenlocher, J., Bjornsdottir, G., Andreassen, O.A., Jönsson, E.G., Palotie, A., Behrens, T.W., Magnusson, O.T., Kong, A., Thorsteinsdottir, U., Watts, R.J., Stefansson, K., 2012. A mutation in APP protects against Alzheimer's disease and age-related cognitive decline. Nature 488, 96–99. doi:10.1038/nature11283

Jun, G., Ibrahim-Verbaas, C.A., Vronskaya, M., Lambert, J.-C., Chung, J., Naj, A.C., Kunkle, B.W., Wang, L.-S., Bis, J.C., Bellenguez, C., Harold, D., Lunetta, K.L., Destefano, A.L., Grenier-Boley, B., Sims, R., Beecham, G.W., Smith, A.V., Chouraki, V., Hamilton-Nelson, K.L., Ikram, M.A., Fievet, N., Denning, N., Martin, E.R., Schmidt, H., Kamatani, Y., Dunstan, M.L., Valladares, O., Laza, A.R., Zelenika, D., Ramirez, A., Foroud, T.M., Choi, S.-H., Boland, A., Becker, T., Kukull, W.A., van der Lee, S.J., Pasquier, F., Cruchaga, C., Beekly, D., Fitzpatrick, A.L., Hanon, O., Gill, M., Barber, R., Gudnason, V., Campion, D., Love, S., Bennett, D.A., Amin, N., Berr, C., Tsolaki, M., Buxbaum, J.D., Lopez, O.L., Deramecourt, V., Fox, N.C., Cantwell, L.B., Tárraga, L., Dufouil, C., Hardy, J., Crane, P.K., Eiriksdottir, G., Hannequin, D., Clarke, R., Evans, D., Mosley, T.H., Letenneur, L., Brayne, C., Maier, W., De Jager, P., Emilsson, V., Dartigues, J.-F., Hampel, H., Kamboh, M.I., de Bruijn, R.F.A.G., Tzourio, C., Pastor, P., Larson, E.B., Rotter, J.I., O'Donovan, M.C., Montine, T.J., Nalls, M.A., Mead, S., Reiman, E.M., Jonsson, P.V., Holmes, C., St George-Hyslop, P.H., Boada, M., Passmore, P., Wendland, J.R., Schmidt, R., Morgan, K., Winslow, A.R., Powell, J.F., Carasquillo, M., Younkin, S.G., Jakobsdóttir, J., Kauwe, J.S., Wilhelmsen, K.C., Rujescu, D., Nöthen, M.M., Hofman, A., Jones, L., Haines, J.L., Psaty, B.M., Van Broeckhoven, C., Holmans, P., Launer, L.J., Mayeux, R., Lathrop, M., Goate, A.M., Escott-Price, V., Seshadri, S., Pericak-Vance, M.A., Amouyel, P., Williams, J., van Duijn, C.M., Schellenberg, G.D., Farrer, L.A., 2016. A NOVEL ALZHEIMER DISEASE LOCUS LOCATED NEAR THE GENE ENCODING TAU PROTEIN. Mol. Psychiatry 21, 108–117. doi:10.1038/mp.2015.23

Kalmijn, S., Launer, L.J., Lindemans, J., Bots, M.L., Hofman, A., Breteler, M.M., 1999. Total homocysteine and cognitive decline in a community-based sample of elderly subjects: the Rotterdam Study. Am. J. Epidemiol. 150, 283–289.

Kaminski, W.E., Orsó, E., Diederich, W., Klucken, J., Drobnik, W., Schmitz, G., 2000a. Identification of a Novel Human Sterol-Sensitive ATP-Binding Cassette Transporter (ABCA7). Biochem. Biophys. Res. Commun. 273, 532–538. doi:10.1006/bbrc.2000.2954

Kaminski, W.E., Piehler, A., Schmitz, G., 2000b. Genomic Organization of the Human Cholesterol-Responsive ABC Transporter ABCA7: Tandem Linkage with the Minor Histocompatibility Antigen HA-1 Gene. Biochem. Biophys. Res. Commun. 278, 782–789. doi:10.1006/bbrc.2000.3880

Kaminski, W.E., Piehler, A., Wenzel, J.J., 2006. ABC A-subfamily transporters: Structure, function and disease. Biochim. Biophys. Acta BBA - Mol. Basis Dis. 1762, 510–524. doi:10.1016/j.bbadis.2006.01.011

Kang, H.J., Kawasawa, Y.I., Cheng, F., Zhu, Y., Xu, X., Li, M., Sousa, A.M.M., Pletikos, M., Meyer, K.A., Sedmak, G., Guennel, T., Shin, Y., Johnson, M.B., Krsnik, Ž., Mayer, S., Fertuzinhos, S., Umlauf, S., Lisgo, S.N., Vortmeyer, A., Weinberger, D.R., Mane, S., Hyde, T.M., Huttner, A., Reimers, M., Kleinman, J.E., Šestan, N., 2011. Spatio-temporal transcriptome of the human brain. Nature 478, 483–489. doi:10.1038/nature10523

Karch, C.M., Jeng, A.T., Nowotny, P., Cady, J., Cruchaga, C., Goate, A.M., 2012. Expression of novel Alzheimer's disease risk genes in control and Alzheimer's disease brains. PloS One 7, e50976. doi:10.1371/journal.pone.0050976

Kauwe, J.S.K., Bailey, M.H., Ridge, P.G., Perry, R., Wadsworth, M.E., Hoyt, K.L., Staley, L.A., Karch, C.M., Harari, O., Cruchaga, C., Ainscough, B.J., Bales, K., Pickering, E.H., Bertelsen, S., the Alzheimer's Disease Neuroimaging Initiative, Fagan, A.M., Holtzman, D.M., Morris, J.C., Goate, A.M., 2014. Genome-Wide Association Study of CSF Levels of 59 Alzheimer's Disease

Candidate Proteins: Significant Associations with Proteins Involved in Amyloid Processing and Inflammation. PLoS Genet. 10, e1004758. doi:10.1371/journal.pgen.1004758

Kent, W.J., Sugnet, C.W., Furey, T.S., Roskin, K.M., Pringle, T.H., Zahler, A.M., Haussler, and D., 2002. The Human Genome Browser at UCSC. Genome Res. 12, 996–1006. doi:10.1101/gr.229102

Kielar, D., Kaminski, W.E., Liebisch, G., Piehler, A., Wenzel, J.J., Möhle, C., Heimerl, S., Langmann, T., Friedrich, S.O., Böttcher, A., Barlage, S., Drobnik, W., Schmitz, G., 2003. Adenosine Triphosphate Binding Cassette (ABC) Transporters Are Expressed and Regulated During Terminal Keratinocyte Differentiation: A Potential Role for ABCA7 in Epidermal Lipid Reorganization. J. Invest. Dermatol. 121, 465–474. doi:10.1046/j.1523-1747.2003.12404.x

Kim, J., Basak, J.M., Holtzman, D.M., 2009. The Role of Apolipoprotein E in Alzheimer's Disease. Neuron 63, 287–303. doi:10.1016/j.neuron.2009.06.026

Kim, W.S., 2004. Abca7 Null Mice Retain Normal Macrophage Phosphatidylcholine and Cholesterol Efflux Activity despite Alterations in Adipose Mass and Serum Cholesterol Levels. J. Biol. Chem. 280, 3989–3995. doi:10.1074/jbc.M412602200

Kim, W.S., Guillemin, G.J., Glaros, E.N., Lim, C.K., Garner, B., 2006. Quantitation of ATP-binding cassette subfamily-A transporter gene expression in primary human brain cells. Neuroreport 17, 891–896. doi:10.1097/01.wnr.0000221833.41340.cd

Kim, W.S., Li, H., Ruberu, K., Chan, S., Elliott, D.A., Low, J.K., Cheng, D., Karl, T., Garner, B., 2013. Deletion of Abca7 Increases Cerebral Amyloid-Accumulation in the J20 Mouse Model of Alzheimer's Disease. J. Neurosci. 33. doi:10.1523/JNEUROSCI.4165-12.2013

Kim, W.S., Weickert, C.S., Garner, B., 2008. Role of ATP-binding cassette transporters in brain lipid transport and neurological disease. J. Neurochem. 104, 1145–1166. doi:10.1111/j.1471-4159.2007.05099.x

Kitchen, R.R., Rozowsky, J.S., Gerstein, M.B., Nairn, A.C., 2014. Decoding neuroproteomics: integrating the genome, translatome and functional anatomy. Nat. Neurosci. 17, 1491–1499. doi:10.1038/nn.3829

Kivipelto, M., Ngandu, T., Fratiglioni, L., Viitanen, M., Kåreholt, I., Winblad, B., Helkala, E.-L., Tuomilehto, J., Soininen, H., Nissinen, A., 2005. Obesity and vascular risk factors at midlife and the risk of dementia and Alzheimer disease. Arch. Neurol. 62, 1556–1560. doi:10.1001/archneur.62.10.1556

Klein, I., Sarkadi, B., Váradi, A., 1999. An inventory of the human ABC proteins. Biochim. Biophys. Acta 1461, 237–262.

Knight, H.M., Pickard, B.S., Maclean, A., Malloy, M.P., Soares, D.C., McRae, A.F., Condie, A., White, A., Hawkins, W., McGhee, K., van Beck, M., MacIntyre, D.J., Starr, J.M., Deary, I.J., Visscher, P.M., Porteous, D.J., Cannon, R.E., St Clair, D., Muir, W.J., Blackwood, D.H.R., 2009. A Cytogenetic Abnormality and Rare Coding Variants Identify ABCA13 as a Candidate Gene in Schizophrenia, Bipolar Disorder, and Depression. Am. J. Hum. Genet. 85, 833–846. doi:10.1016/j.ajhg.2009.11.003

Kobayashi, Y., Fukumaki, Y., Yubisui, T., Inoue, J., Sakaki, Y., 1990. Serine-proline replacement at residue 127 of NADH-cytochrome b5 reductase causes hereditary methemoglobinemia, generalized type. Blood 75, 1408–1413.

Koepsell, T.D., Kurland, B.F., Harel, O., Johnson, E.A., Zhou, X.-H., Kukull, W.A., 2008. Education, cognitive function, and severity of neuropathology in Alzheimer disease. Neurology 70, 1732–1739. doi:10.1212/01.wnl.0000284603.85621.aa

Koldamova, R., Staufenbiel, M., Lefterov, I., 2005. Lack of ABCA1 considerably decreases brain ApoE level and increases amyloid deposition in APP23 mice. J. Biol. Chem. 280, 43224–43235. doi:10.1074/jbc.M504513200

Koldamova, R.P., Lefterov, I.M., Ikonomovic, M.D., Skoko, J., Lefterov, P.I., Isanski, B.A., DeKosky, S.T., Lazo, J.S., 2003. 22R-hydroxycholesterol and 9-cis-retinoic acid induce ATP-binding cassette transporter A1 expression and cholesterol efflux in brain cells and decrease amyloid beta secretion. J. Biol. Chem. 278, 13244–13256. doi:10.1074/jbc.M300044200

Kreutzberg, G.W., 1995. Microglia, the first line of defence in brain pathologies. Arzneimittelforschung. 45, 357–360.

Krogh, A., Larsson, B., von Heijne, G., Sonnhammer, E.L., 2001. Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. J. Mol. Biol. 305, 567–580. doi:10.1006/jmbi.2000.4315

Kuller, L.H., Lopez, O.L., 2011. Dementia and Alzheimer's disease: A new direction. The 2010 Jay L. Foster Memorial Lecture. Alzheimers Dement. 7, 540–550. doi:10.1016/j.jalz.2011.05.901

Kumar, P., Henikoff, S., Ng, P.C., 2009. Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat. Protoc. 4, 1073–1081. doi:10.1038/nprot.2009.86

Kyriakou, T., Hodkinson, C., Pontefract, D.E., Iyengar, S., Howell, W.M., Wong, Y., Eriksson, P., Ye, S., 2004. Genotypic Effect of the -565C>T Polymorphism in the ABCA1 Gene Promoter on ABCA1 Expression and Severity of Atherosclerosis. Arterioscler. Thromb. Vasc. Biol. 25, 418–423. doi:10.1161/01.ATV.0000149379.72018.20

Laird, P.W., 2010. Principles and challenges of genome-wide DNA methylation analysis. Nat. Rev. Genet. 11, 191–203. doi:10.1038/nrg2732

Lamba, J.K., Adachi, M., Sun, D., Tammur, J., Schuetz, E.G., Allikmets, R., Schuetz, J.D., 2003. Nonsense mediated decay downregulates conserved alternatively spliced ABCC4 transcripts bearing nonsense codons. Hum. Mol. Genet. 12, 99–109.

Lambert, J.-C., Heath, S., Even, G., Campion, D., Sleegers, K., Hiltunen, M., Combarros, O., Zelenika, D., Bullido, M.J., Tavernier, B., Letenneur, L., Bettens, K., Berr, C., Pasquier, F., Fiévet, N., Barberger-Gateau, P., Engelborghs, S., De Deyn, P., Mateo, I., Franck, A., Helisalmi, S., Porcellini, E., Hanon, O., de Pancorbo, M.M., Lendon, C., Dufouil, C., Jaillard, C., Leveillard, T., Alvarez, V., Bosco, P., Mancuso, M., Panza, F., Nacmias, B., Bossù, P., Piccardi, P., Annoni, G., Seripa, D., Galimberti, D., Hannequin, D., Licastro, F., Soininen, H., Ritchie, K., Blanché, H., Dartigues, J.-F., Tzourio, C., Gut, I., Van Broeckhoven, C., Alpérovitch, A., Lathrop, M., Amouyel, P., 2009. Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease. Nat. Genet. 41, 1094–1099. doi:10.1038/ng.439

Lambert, J.-C., Ibrahim-Verbaas, C.A., Harold, D., Naj, A.C., Sims, R., Bellenguez, C., Jun, G., DeStefano, A.L., Bis, J.C., Beecham, G.W., Grenier-Boley, B., Russo, G., Thornton-Wells, T.A., Jones, N., Smith, A.V., Chouraki, V., Thomas, C., Ikram, M.A., Zelenika, D., Vardarajan, B.N., Kamatani, Y., Lin, C.-F., Gerrish, A., Schmidt, H., Kunkle, B., Dunstan, M.L., Ruiz, A., Bihoreau, M.-T., Choi, S.-H., Reitz, C., Pasquier, F., Hollingworth, P., Ramirez, A., Hanon, O., Fitzpatrick, A.L., Buxbaum, J.D., Campion, D., Crane, P.K., Baldwin, C., Becker, T., Gudnason, V., Cruchaga, C., Craig, D., Amin, N., Berr, C., Lopez, O.L., De Jager, P.L., Deramecourt, V., Johnston, J.A., Evans, D., Lovestone, S., Letenneur, L., Morón, F.J., Rubinsztein, D.C., Eiriksdottir, G., Sleegers, K., Goate, A.M., Fiévet, N., Huentelman, M.J., Gill, M., Brown, K., Kamboh, M.I., Keller, L., Barberger-Gateau, P., McGuinness, B., Larson, E.B., Green, R., Myers, A.J., Dufouil, C., Todd, S., Wallon, D.,

Love, S., Rogaeva, E., Gallacher, J., St George-Hyslop, P., Clarimon, J., Lleo, A., Bayer, A., Tsuang, D.W., Yu, L., Tsolaki, M., Bossù, P., Spalletta, G., Proitsi, P., Collinge, J., Sorbi, S., Sanchez-Garcia, F., Fox, N.C., Hardy, J., Naranjo, M.C.D., Bosco, P., Clarke, R., Brayne, C., Galimberti, D., Mancuso, M., Matthews, F., Moebus, S., Mecocci, P., Del Zompo, M., Maier, W., Hampel, H., Pilotto, A., Bullido, M., Panza, F., Caffarra, P., Nacmias, B., Gilbert, J.R., Mayhaus, M., Lannfelt, L., Hakonarson, H., Pichler, S., Carrasquillo, M.M., Ingelsson, M., Beekly, D., Alvarez, V., Zou, F., Valladares, O., Younkin, S.G., Coto, E., Hamilton-Nelson, K.L., Gu, W., Razquin, C., Pastor, P., Mateo, I., Owen, M.J., Faber, K.M., Jonsson, P.V., Combarros, O., O'Donovan, M.C., Cantwell, L.B., Soininen, H., Blacker, D., Mead, S., Mosley, T.H., Bennett, D.A., Harris, T.B., Fratiglioni, L., Holmes, C., de Bruijn, R.F.A.G., Passmore, P., Montine, T.J., Bettens, K., Rotter, J.I., Brice, A., Morgan, K., Foroud, T.M., Kukull, W.A., Hannequin, D., Powell, J.F., Nalls, M.A., Ritchie, K., Lunetta, K.L., Kauwe, J.S.K., Boerwinkle, E., Riemenschneider, M., Boada, M., Hiltunen, M., Martin, E.R., Schmidt, R., Rujescu, D., Wang, L.-S., Dartigues, J.-F., Mayeux, R., Tzourio, C., Hofman, A., Nöthen, M.M., Graff, C., Psaty, B.M., Jones, L., Haines, J.L., Holmans, P.A., Lathrop, M., Pericak-Vance, M.A., Launer, L.J., Farrer, L.A., van Duijn, C.M., Van Broeckhoven, C., Moskvina, V., Seshadri, S., Williams, J., Schellenberg, G.D., Amouyel, P., 2013. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. Nat. Genet. 45, 1452–1458. doi:10.1038/ng.2802

Langmann, T., Mauerer, R., Zahn, A., Moehle, C., Probst, M., Stremmel, W., Schmitz, G., 2003. Real-time reverse transcription-PCR expression profiling of the complete human ATP-binding cassette transporter superfamily in various tissues. Clin. Chem. 49, 230–238.

Lauber, K., Blumenthal, S.G., Waibel, M., Wesselborg, S., 2004. Clearance of apoptotic cells: getting rid of the corpses. Mol. Cell 14, 277–287.

Lauer, J., Shen, C.K., Maniatis, T., 1980. The chromosomal arrangement of human α-like globin genes: Sequence homology and α-globin gene deletions. Cell 20, 119–130. doi:10.1016/0092-8674(80)90240-8

Lee, M., McGeer, E., McGeer, P.L., 2014. Activated human microglia stimulate neuroblastoma cells to upregulate production of beta amyloid protein and tau: implications for Alzheimer's disease pathogenesis. Neurobiol. Aging. doi:10.1016/j.neurobiolaging.2014.07.024

Lee, S., Abecasis, G.R., Boehnke, M., Lin, X., 2014. Rare-variant association analysis: study designs and statistical tests. Am. J. Hum. Genet. 95, 5–23. doi:10.1016/j.ajhg.2014.06.009

Lefèvre, C., Audebert, S., Jobard, F., Bouadjar, B., Lakhdar, H., Boughdene-Stambouli, O., Blanchet-Bardon, C., Heilig, R., Foglio, M., Weissenbach, J., Lathrop, M., Prud'homme, J.-F., Fischer, J., 2003. Mutations in the transporter ABCA12 are associated with lamellar ichthyosis type 2. Hum. Mol. Genet. 12, 2369–2378. doi:10.1093/hmg/ddg235

Lewis, F.I., Torgerson, P.R., 2016. The current and future burden of late-onset dementia in the United Kingdom: Estimates and interventions. Alzheimers Dement. doi:10.1016/j.jalz.2016.03.013

Lewis, R.A., Shroyer, N.F., Singh, N., Allikmets, R., Hutchinson, A., Li, Y., Lupski, J.R., Leppert, M., Dean, M., 1999. Genotype/Phenotype analysis of a photoreceptor-specific ATP-binding cassette transporter gene, ABCR, in Stargardt disease. Am. J. Hum. Genet. 64, 422–434.

Li, H., 2011. Tabix: Fast retrieval of sequence features from generic TAB-delimited files. Bioinformatics btq671. doi:10.1093/bioinformatics/btq671

Li, W.-H., Wu, C.-I., Luo, C.-C., 1984. Nonrandomness of point mutation as reflected in nucleotide substitutions in pseudogenes and its evolutionary implications. J. Mol. Evol. 21, 58–71. doi:10.1007/BF02100628

Li, Y.I., Geijn, B. van de, Raj, A., Knowles, D.A., Petti, A.A., Golan, D., Gilad, Y., Pritchard, J.K., 2016. RNA splicing is a primary link between genetic variation and disease. Science 352, 600–604. doi:10.1126/science.aad9417

Liao, L., Cheng, D., Wang, J., Duong, D.M., Losik, T.G., Gearing, M., Rees, H.D., Lah, J.J., Levey, A.I., Peng, J., 2004. Proteomic characterization of postmortem amyloid plaques isolated by laser capture microdissection. J. Biol. Chem. 279, 37061–37068. doi:10.1074/jbc.M403672200

Liao, Y.-C., Lee, W.-J., Hwang, J.-P., Wang, Y.-F., Tsai, C.-F., Wang, P.-N., Wang, S.-J., Fuh, J.-L., 2014. ABCA7 gene and the risk of Alzheimer's disease in Han Chinese in Taiwan. Neurobiol. Aging. doi:10.1016/j.neurobiolaging.2014.05.009

Lindberg, O., Stomrud, E., Zetterberg, H., Blennow, K., Hansson, O., 2016. Decreased ratio between β-amyloid 42 (Aβ42) and Aβ40 in cerebral spinal fluid is a better predictor of structural brain changes than Aβ42 alone in cognitively normal elderly people. Neurobiol. Aging, Abstracts from the 14th International Athens/Springfield Symposium on Advances in Alzheimer TherapyMarch 9–12, 2016, Athens, Greece 39, Supplement 1, S17. doi:10.1016/j.neurobiolaging.2016.01.081

Linsel-Nitschke, P., Jehle, A.W., Shan, J., Cao, G., Bacic, D., Lan, D., Wang, N., Tall, A.R., 2005. Potential role of ABCA7 in cellular lipid efflux to apoA-I. J. Lipid Res. 46, 86–92. doi:10.1194/jlr.M400247-JLR200

Liu, Q., Zerbinatti, C.V., Zhang, J., Hoe, H.-S., Wang, B., Cole, S.L., Herz, J., Muglia, L., Bu, G., 2007. Amyloid Precursor Protein Regulates Brain Apolipoprotein E and Cholesterol Metabolism through Lipoprotein Receptor LRP1. Neuron 56, 66–78. doi:10.1016/j.neuron.2007.08.008

Liu-Seifert, H., Siemers, E., Holdridge, K.C., Andersen, S.W., Lipkovich, I., Carlson, C., Sethuraman, G., Hoog, S., Hayduk, R., Doody, R., Aisen, P., 2015. Delayed-start analysis: Mild Alzheimer's disease patients in solanezumab trials, 3.5 years. Alzheimers Dement. Transl. Res. Clin. Interv. doi:10.1016/j.trci.2015.06.006

Logue, M., Schu, M., Vardarajan, B., Farrell, J., Lunetta, K., Baldwin, C., Fallin, D., Farrer, L., 2013. Two rare AKAP9 missense variants are associated with Alzheimer's disease in African-Americans. Alzheimers Dement. J. Alzheimers Assoc. 9, P516–P517. doi:10.1016/j.jalz.2013.04.236

Lopez-Longo, F.J., Rodriguez-Mahou, M., Escalona, M., 1994. Hetrogeneity of the Anti-Ro(SSA) Respoonse to Rheumatic Diseases. J. Rheumatol. 21, 1450.

Lord, J., Cruchaga, C., 2014. The epigenetic landscape of Alzheimer's disease. Nat. Neurosci. 17, 1138–1140. doi:10.1038/nn.3792

Lord, J., Lu, A.J., Cruchaga, C., 2014. Identification of rare variants in Alzheimer's disease. Front. Genet. 5. doi:10.3389/fgene.2014.00369

Luchsinger, J.A., Mayeux, R., 2004. Dietary factors and Alzheimer's disease. Lancet Neurol. 3, 579–587. doi:10.1016/S1474-4422(04)00878-6

Luciani, M.F., Chimini, G., 1996. The ATP binding cassette transporter ABC1, is required for the engulfment of corpses generated by apoptotic cell death. EMBO J. 15, 226–235.

Lueg, G., Gross, C.C., Lohmann, H., Johnen, A., Kemmling, A., Deppe, M., Groger, J., Minnerup, J., Wiendl, H., Meuth, S.G., Duning, T., 2014. Clinical relevance of specific T-cell activation in the blood and cerebrospinal fluid of patients with mild Alzheimer's disease. Neurobiol. Aging. doi:10.1016/j.neurobiolaging.2014.08.008

Lunnon, K., Smith, R., Hannon, E., De Jager, P.L., Srivastava, G., Volta, M., Troakes, C., Al-Sarraj, S., Burrage, J., Macdonald, R., Condliffe, D., Harries, L.W.,

Katsel, P., Haroutunian, V., Kaminsky, Z., Joachim, C., Powell, J., Lovestone, S., Bennett, D.A., Schalkwyk, L.C., Mill, J., 2014. Methylomic profiling implicates cortical deregulation of ANK1 in Alzheimer's disease. Nat. Neurosci. 17, 1164–1170. doi:10.1038/nn.3782

Mably, A.J., Kanmert, D., Mc Donald, J.M., Liu, W., Caldarone, B.J., Lemere, C.A., O'Nuallain, B., Kosik, K.S., Walsh, D.M., 2014. Tau immunization: a cautionary tale? Neurobiol. Aging. doi:10.1016/j.neurobiolaging.2014.11.022

MacArthur, D.G., Manolio, T.A., Dimmock, D.P., Rehm, H.L., Shendure, J., Abecasis, G.R., Adams, D.R., Altman, R.B., Antonarakis, S.E., Ashley, E.A., Barrett, J.C., Biesecker, L.G., Conrad, D.F., Cooper, G.M., Cox, N.J., Daly, M.J., Gerstein, M.B., Goldstein, D.B., Hirschhorn, J.N., Leal, S.M., Pennacchio, L.A., Stamatoyannopoulos, J.A., Sunyaev, S.R., Valle, D., Voight, B.F., Winckler, W., Gunter, C., 2014. Guidelines for investigating causality of sequence variants in human disease. Nature 508, 469–476. doi:10.1038/nature13127

Macé, S., Cousin, E., Ricard, S., Génin, E., Spanakis, E., Lafargue-Soubigou, C., Génin, B., Fournel, R., Roche, S., Haussy, G., Massey, F., Soubigou, S., Bréfort, G., Benoit, P., Brice, A., Campion, D., Hollis, M., Pradier, L., Benavides, J., Deleuze, J.-F., 2005. ABCA2 is a strong genetic risk factor for early-onset Alzheimer's disease. Neurobiol. Dis. 18, 119–125. doi:10.1016/j.nbd.2004.09.011

Malik, M., Simpson, J.F., Parikh, I., Wilfred, B.R., Fardo, D.W., Nelson, P.T., Estus, S., 2013. CD33 Alzheimer's risk-altering polymorphism, CD33 expression, and exon 2 splicing. J. Neurosci. Off. J. Soc. Neurosci. 33, 13320–13325. doi:10.1523/JNEUROSCI.1224-13.2013

Mandelkow, E.-M., Mandelkow, E., 1998. Tau in Alzheimer's disease. Trends Cell Biol. 8, 425–427. doi:10.1016/S0962-8924(98)01368-3

Mapstone, M., Cheema, A.K., Fiandaca, M.S., Zhong, X., Mhyre, T.R., MacArthur, L.H., Hall, W.J., Fisher, S.G., Peterson, D.R., Haley, J.M., Nazar, M.D., Rich, S.A., Berlau, D.J., Peltz, C.B., Tan, M.T., Kawas, C.H., Federoff, H.J., 2014. Plasma phospholipids identify antecedent memory impairment in older adults. Nat. Med. 20, 415–418. doi:10.1038/nm.3466

Maquat, L.E., 2004. Nonsense-Mediated mRNA Decay: Splicing, Translation And mRNP Dynamics. Nat. Rev. Mol. Cell Biol. 5, 89–99. doi:10.1038/nrm1310

Marchesi, V.T., 2011. Alzheimer's dementia begins as a disease of small blood vessels, damaged by oxidative-induced inflammation and dysregulated amyloid metabolism: implications for early detection and therapy. FASEB J. Off. Publ. Fed. Am. Soc. Exp. Biol. 25, 5–13. doi:10.1096/fj.11-0102ufm

Martins, I.J., Berger, T., Sharman, M.J., Verdile, G., Fuller, S.J., Martins, R.N., 2009. Cholesterol metabolism and transport in the pathogenesis of Alzheimer's disease. J. Neurochem. 111, 1275–1308.

Matlin, A.J., Clark, F., Smith, C.W.J., 2005. Understanding alternative splicing: towards a cellular code. Nat. Rev. Mol. Cell Biol. 6, 386–398. doi:10.1038/nrm1645

Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., Shafer, A., Neri, F., Lee, K., Kutyavin, T., Stehling-Sun, S., Johnson, A.K., Canfield, T.K., Giste, E., Diegel, M., Bates, D., Hansen, R.S., Neph, S., Sabo, P.J., Heimfeld, S., Raubitschek, A., Ziegler, S., Cotsapas, C., Sotoodehnia, N., Glass, I., Sunyaev, S.R., Kaul, R., Stamatoyannopoulos, J.A., 2012. Systematic localization of common disease-associated variation in regulatory DNA. Science 337, 1190–1195. doi:10.1126/science.1222794

Mayeux, R., Reitz, C., Brickman, A.M., Haan, M.N., Manly, J.J., Glymour, M.M., Weiss, C.C., Yaffe, K., Middleton, L., Hendrie, H.C., Warren, L.H., Hayden, K.M., Welsh-Bohmer, K.A., Breitner, J.C.S., Morris, J.C., 2011.

Operationalizing diagnostic criteria for Alzheimer's disease and other age-related cognitive impairment-Part 1. Alzheimers Dement. J. Alzheimers Assoc. 7, 15–34. doi:10.1016/j.jalz.2010.11.005

McCartney, M., 2015. Margaret McCartney: The "breakthrough" drug that's not been shown to help in Alzheimer's disease. BMJ 351, h4064.

McKhann, G., Drachman, D., Folstein, M., Katzman, R., Price, D., Stadlan, E.M., 1984. Clinical diagnosis of Alzheimer's disease: Report of the NINCDS-ADRDA Work Group* under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. Neurology 34, 939–939. doi:10.1212/WNL.34.7.939

Medway, C., Morgan, K., 2014. Review: The genetics of Alzheimer's disease; putting flesh on the bones: The genetics of Alzheimer's disease. Neuropathol. Appl. Neurobiol. 40, 97–105. doi:10.1111/nan.12101

Medway, C.W., Abdul-Hay, S., Mims, T., Ma, L., Bisceglio, G., Zou, F., Pankratz, S., Sando, S.B., Aasly, J.O., Barcikowska, M., Siuda, J., Wszolek, Z.K., Ross, O.A., Carrasquillo, M., Dickson, D.W., Graff-Radford, N., Petersen, R.C., Ertekin-Taner, N., Morgan, K., Bu, G., Younkin, S.G., 2014. ApoE variant p.V236E is associated with markedly reduced risk of Alzheimer's disease. Mol. Neurodegener. 9, 11. doi:10.1186/1750-1326-9-11

Meurs, I., Calpe-Berdiel, L., Habets, K.L.L., Zhao, Y., Korporaal, S.J.A., Mommaas, A.M., Josselin, E., Hildebrand, R.B., Ye, D., Out, R., Kuiper, J., Van Berkel, T.J.C., Chimini, G., Van Eck, M., 2012. Effects of deletion of macrophage ABCA7 on lipid metabolism and the development of atherosclerosis in the presence and absence of ABCA1. PloS One 7, e30984. doi:10.1371/journal.pone.0030984

Michaelson, D.M., 2014. ApoE4: The most prevalent yet understudied risk factor for Alzheimer's disease. Alzheimers Dement. J. Alzheimers Assoc. doi:10.1016/j.jalz.2014.06.015

Minati, L., Edginton, T., Bruzzone, M.G., Giaccone, G., 2009. Current concepts in Alzheimer's disease: a multidisciplinary review. Am. J. Alzheimers Dis. Other Demen. 24, 95–121. doi:10.1177/1533317508328602

Molinuevo, J.L., Blennow, K., Dubois, B., Engelborghs, S., Lewczuk, P., Perret-Liaudet, A., Teunissen, C.E., Parnetti, L., 2014. The clinical use of cerebrospinal fluid biomarker testing for Alzheimer's disease diagnosis: A consensus paper from the Alzheimer's Biomarkers Standardization Initiative. Alzheimers Dement. 10, 808–817. doi:10.1016/j.jalz.2014.03.003

Möller, S., Croning, M.D., Apweiler, R., 2001. Evaluation of methods for the prediction of membrane spanning regions. Bioinforma. Oxf. Engl. 17, 646–653.

Morgan, K., 2011. The three new pathways leading to Alzheimer's disease. Neuropathol. Appl. Neurobiol. 37, 353–357. doi:10.1111/j.1365-2990.2011.01181.x

Musiek, E.S., Holtzman, D.M., 2015. Three dimensions of the amyloid hypothesis: time, space and "wingmen." Nat. Neurosci. 18, 800–806. doi:10.1038/nn.4018

Mutis, T., Gillespie, G., Schrama, E., Falkenburg, J.H.F., Moss, P., Goulmy, E., 1999. Tetrameric HLA class I–minor histocompatibility antigen peptide complexes demonstrate minor histocompatibility antigen-specific cytotoxic T lymphocytes in patients with graft-versus-host disease. Nat. Med. 5, 839–842. doi:10.1038/10563

Naj, Adam C, Jun, G., Beecham, G.W., Wang, L.-S., Vardarajan, B.N., Buros, J., Gallins, P.J., Buxbaum, J.D., Jarvik, G.P., Crane, P.K., Larson, E.B., Bird, T.D., Boeve, B.F., Graff-Radford, N.R., De Jager, P.L., Evans, D., Schneider, J.A., Carrasquillo, M.M., Ertekin-Taner, N., Younkin, S.G., Cruchaga, C., Kauwe, J.S.K., Nowotny, P., Kramer, P., Hardy, J., Huentelman, M.J., Myers,

A.J., Barmada, M.M., Demirci, F.Y., Baldwin, C.T., Green, R.C., Rogaeva, E., St George-Hyslop, P., Arnold, S.E., Barber, R., Beach, T., Bigio, E.H., Bowen, J.D., Boxer, A., Burke, J.R., Cairns, N.J., Carlson, C.S., Carney, R.M., Carroll, S.L., Chui, H.C., Clark, D.G., Corneveaux, J., Cotman, C.W., Cummings, J.L., DeCarli, C., DeKosky, S.T., Diaz-Arrastia, R., Dick, M., Dickson, D.W., Ellis, W.G., Faber, K.M., Fallon, K.B., Farlow, M.R., Ferris, S., Frosch, M.P., Galasko, D.R., Ganguli, M., Gearing, M., Geschwind, D.H., Ghetti, B., Gilbert, J.R., Gilman, S., Giordani, B., Glass, J.D., Growdon, J.H., Hamilton, R.L., Harrell, L.E., Head, E., Honig, L.S., Hulette, C.M., Hyman, B.T., Jicha, G.A., Jin, L.-W., Johnson, N., Karlawish, J., Karydas, A., Kaye, J.A., Kim, R., Koo, E.H., Kowall, N.W., Lah, J.J., Levey, A.I., Lieberman, A.P., Lopez, O.L., Mack, W.J., Marson, D.C., Martiniuk, F., Mash, D.C., Masliah, E., McCormick, W.C., McCurry, S.M., McDavid, A.N., McKee, A.C., Mesulam, M., Miller, B.L., Miller, C.A., Miller, J.W., Parisi, J.E., Perl, D.P., Peskind, E., Petersen, R.C., Poon, W.W., Quinn, J.F., Rajbhandary, R.A., Raskind, M., Reisberg, B., Ringman, J.M., Roberson, E.D., Rosenberg, R.N., Sano, M., Schneider, L.S., Seeley, W., Shelanski, M.L., Slifer, M.A., Smith, C.D., Sonnen, J.A., Spina, S., Stern, R.A., Tanzi, R.E., Trojanowski, J.Q., Troncoso, J.C., Van Deerlin, V.M., Vinters, H.V., Vonsattel, J.P., Weintraub, S., Welsh-Bohmer, K.A., Williamson, J., Woltjer, R.L., Cantwell, L.B., Dombroski, B.A., Beekly, D., Lunetta, K.L., Martin, E.R., Kamboh, M.I., Saykin, A.J., Reiman, E.M., Bennett, D.A., Morris, J.C., Montine, T.J., Goate, A.M., Blacker, D., Tsuang, D.W., Hakonarson, H., Kukull, W.A., Foroud, T.M., Haines, J.L., Mayeux, R., Pericak-Vance, M.A., Farrer, L.A., Schellenberg, G.D., 2011. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat. Genet. 43, 436–441. doi:10.1038/ng.801

Naj, Adam C., Jun, G., Beecham, G.W., Wang, L.-S., Vardarajan, B.N., Buros, J., Gallins, P.J., Buxbaum, J.D., Jarvik, G.P., Crane, P.K., Larson, E.B., Bird, T.D., Boeve, B.F., Graff-Radford, N.R., De Jager, P.L., Evans, D., Schneider, J.A., Carrasquillo, M.M., Ertekin-Taner, N., Younkin, S.G., Cruchaga, C., Kauwe, J.S.K., Nowotny, P., Kramer, P., Hardy, J., Huentelman, M.J., Myers, A.J., Barmada, M.M., Demirci, F.Y., Baldwin, C.T., Green, R.C., Rogaeva, E., St George-Hyslop, P., Arnold, S.E., Barber, R., Beach, T., Bigio, E.H., Bowen, J.D., Boxer, A., Burke, J.R., Cairns, N.J., Carlson, C.S., Carney, R.M., Carroll, S.L., Chui, H.C., Clark, D.G., Corneveaux, J., Cotman, C.W., Cummings, J.L., DeCarli, C., DeKosky, S.T., Diaz-Arrastia, R., Dick, M., Dickson, D.W., Ellis, W.G., Faber, K.M., Fallon, K.B., Farlow, M.R., Ferris, S., Frosch, M.P., Galasko, D.R., Ganguli, M., Gearing, M., Geschwind, D.H., Ghetti, B., Gilbert, J.R., Gilman, S., Giordani, B., Glass, J.D., Growdon, J.H., Hamilton, R.L., Harrell, L.E., Head, E., Honig, L.S., Hulette, C.M., Hyman, B.T., Jicha, G.A., Jin, L.-W., Johnson, N., Karlawish, J., Karydas, A., Kaye, J.A., Kim, R., Koo, E.H., Kowall, N.W., Lah, J.J., Levey, A.I., Lieberman, A.P., Lopez, O.L., Mack, W.J., Marson, D.C., Martiniuk, F., Mash, D.C., Masliah, E., McCormick, W.C., McCurry, S.M., McDavid, A.N., McKee, A.C., Mesulam, M., Miller, B.L., Miller, C.A., Miller, J.W., Parisi, J.E., Perl, D.P., Peskind, E., Petersen, R.C., Poon, W.W., Quinn, J.F., Rajbhandary, R.A., Raskind, M., Reisberg, B., Ringman, J.M., Roberson, E.D., Rosenberg, R.N., Sano, M., Schneider, L.S., Seeley, W., Shelanski, M.L., Slifer, M.A., Smith, C.D., Sonnen, J.A., Spina, S., Stern, R.A., Tanzi, R.E., Trojanowski, J.Q., Troncoso, J.C., Van Deerlin, V.M., Vinters, H.V., Vonsattel, J.P., Weintraub, S., Welsh-Bohmer, K.A., Williamson, J., Woltjer, R.L., Cantwell, L.B., Dombroski, B.A., Beekly, D., Lunetta, K.L., Martin, E.R., Kamboh, M.I., Saykin, A.J., Reiman, E.M., Bennett, D.A., Morris, J.C., Montine, T.J., Goate, A.M., Blacker, D., Tsuang, D.W., Hakonarson, H., Kukull, W.A.,

Foroud, T.M., Haines, J.L., Mayeux, R., Pericak-Vance, M.A., Farrer, L.A., Schellenberg, G.D., 2011. Common variants at MS4A4/MS4A6E, CD2AP, CD33 and EPHA1 are associated with late-onset Alzheimer's disease. Nat. Genet. 43, 436–441. doi:10.1038/ng.801

Neher, J.J., Neniskyte, U., Brown, G.C., 2012. Primary phagocytosis of neurons by inflamed microglia: potential roles in neurodegeneration. Front. Pharmacol. 3, 27. doi:10.3389/fphar.2012.00027

Nejentsev, S., Walker, N., Riches, D., Egholm, M., Todd, J.A., 2009. Rare variants of IFIH1, a gene implicated in antiviral responses, protect against type 1 diabetes. Science 324, 387–389. doi:10.1126/science.1167728

Neumann, M., Tolnay, M., Mackenzie, I.R.A., 2009. The molecular basis of frontotemporal dementia. Expert Rev. Mol. Med. 11, e23. doi:10.1017/S1462399409001136

Nguyen-Dumont, T., Jordheim, L.P., Michelon, J., Forey, N., McKay-Chopin, S., Kathleen Cuningham Foundation Consortium for Research into Familial Aspects of Breast Cancer, Sinilnikova, O., Le Calvez-Kelm, F., Southey, M.C., Tavtigian, S.V., Lesueur, F., 2011. Detecting differential allelic expression using high-resolution melting curve analysis: application to the breast cancer susceptibility gene CHEK2. BMC Med. Genomics 4, 39. doi:10.1186/1755-8794-4-39

Nho, K., 2016. Abstract: Genome-Wide Meta-Analysis of Transcriptome Profiling Identifies Novel Dysregulated Genes Implicated in Alzheimer's Disease. Presented at the Alzheimer's Association International Conference 2016, Toronto, Canada.

Noensie, E.N., Dietz, H.C., 2001. A strategy for disease gene identification through nonsense-mediated mRNA decay inhibition. Nat. Biotechnol. 19, 434–439. doi:10.1038/88099

Nordestgaard, L.T., Tybjærg-Hansen, A., Nordestgaard, B.G., Frikke-Schmidt, R., 2015. Loss-of-function mutation in ABCA1 and risk of Alzheimer's disease and cerebrovascular disease. Alzheimers Dement. doi:10.1016/j.jalz.2015.04.006

O'Brien, R.J., Wong, P.C., 2011. Amyloid precursor protein processing and Alzheimer's disease. Annu. Rev. Neurosci. 34, 185–204. doi:10.1146/annurev-neuro-061010-113613

Ogden, C.A., deCathelineau, A., Hoffmann, P.R., Bratton, D., Ghebrehiwet, B., Fadok, V.A., Henson, P.M., 2001. C1q and Mannose Binding Lectin Engagement of Cell Surface Calreticulin and Cd91 Initiates Macropinocytosis and Uptake of Apoptotic Cells. J. Exp. Med. 194, 781–796. doi:10.1084/jem.194.6.781

Oram, J.F., Heinecke, J.W., 2005. ATP-binding cassette transporter A1: a cell cholesterol exporter that protects against cardiovascular disease. Physiol. Rev. 85, 1343–1372. doi:10.1152/physrev.00005.2005

Ott, A., Breteler, M.M.B., Bruyne, M.C. de, Harskamp, F. van, Grobbee, D.E., Hofman, A., 1997. Atrial Fibrillation and Dementia in a Population-Based Study The Rotterdam Study. Stroke 28, 316–321. doi:10.1161/01.STR.28.2.316

Pabinger, S., Dander, A., Fischer, M., Snajder, R., Sperk, M., Efremova, M., Krabichler, B., Speicher, M.R., Zschocke, J., Trajanoski, Z., 2014. A survey of tools for variant analysis of next-generation genome sequencing data. Brief. Bioinform. 15, 256–278. doi:10.1093/bib/bbs086

Pagani, F., Baralle, F.E., 2004. Genomic variants in exons and introns: identifying the splicing spoilers. Nat. Rev. Genet. 5, 389–396. doi:10.1038/nrg1327

Paloneva, J., Manninen, T., Christman, G., Hovanes, K., Mandelin, J., Adolfsson, R., Bianchin, M., Bird, T., Miranda, R., Salmaggi, A., Tranebjaerg, L., Konttinen, Y., Peltonen, L., 2002. Mutations in two genes encoding different subunits of

a receptor signaling complex result in an identical disease phenotype. Am. J. Hum. Genet. 71, 656–662. doi:10.1086/342259

Pan, Q., Shai, O., Lee, L.J., Frey, B.J., Blencowe, B.J., 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. Nat. Genet. 40, 1413–1415. doi:10.1038/ng.259

Panza, F., Solfrizzi, V., Frisardi, V., Capurso, C., D'Introno, A., Colacicco, A.M., Vendemiale, G., Capurso, A., Imbimbo, B.P., 2009. Disease-modifying approach to the treatment of Alzheimer's disease: from alpha-secretase activators to gamma-secretase inhibitors and modulators. Drugs Aging 26, 537–555. doi:10.2165/11315770-000000000-00000

Paris, D., Ait-Ghezala, G., Bachmeier, C., Laco, G., Beaulieu-Abdelahad, D., Lin, Y., Jin, C., Crawford, F., Mullan, M., 2014. The Spleen Tyrosine Kinase (Syk) Regulates Alzheimer Amyloid- Production and Tau Hyperphosphorylation. J. Biol. Chem. 289, 33927–33944. doi:10.1074/jbc.M114.608091

Piehler, A.P., Özcürümez, M., Kaminski, W.E., 2012. A-Subclass ATP-Binding Cassette Proteins in Brain Lipid Homeostasis and Neurodegeneration. Front. Psychiatry 3. doi:10.3389/fpsyt.2012.00017

Pogge, E., 2010. Vitamin D and Alzheimer's disease: is there a link? Consult. Pharm. J. Am. Soc. Consult. Pharm. 25, 440–450. doi:10.4140/TCP.n.2010.440

Priller, C., Bauer, T., Mitteregger, G., Krebs, B., Kretzschmar, H.A., Herms, J., 2006. Synapse Formation and Function Is Modulated by the Amyloid Precursor Protein. J. Neurosci. 26, 7212–7221. doi:10.1523/JNEUROSCI.1450-06.2006

Puglielli, L., Konopka, G., Pack-Chung, E., Ingano, L.A.M., Berezovska, O., Hyman, B.T., Chang, T.Y., Tanzi, R.E., Kovacs, D.M., 2001. Acyl-coenzyme A: cholesterol acyltransferase modulates the generation of the amyloid β-peptide. Nat. Cell Biol. 3, 905–912. doi:10.1038/ncb1001-905

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 81, 559–575. doi:10.1086/519795

Quail, M., Smith, M.E., Coupland, P., Otto, T.D., Harris, S.R., Connor, T.R., Bertoni, A., Swerdlow, H.P., Gu, Y., 2012. A tale of three next generation sequencing platforms: comparison of Ion torrent, pacific biosciences and illumina MiSeq sequencers. BMC Genomics 13, 341. doi:10.1186/1471-2164-13-341

Quazi, F., Molday, R.S., 2013. Differential Phospholipid Substrates and Directional Transport by ATP-binding Cassette Proteins ABCA1, ABCA7, and ABCA4 and Disease-causing Mutants. J. Biol. Chem. 288, 34414–34426. doi:10.1074/jbc.M113.508812

Ramirez, L.M., Goukasian, N., Porat, S., Hwang, K.S., Eastman, J.A., Hurtz, S., Wang, B., Vang, N., Sears, R., Klein, E., Coppola, G., Apostolova, L.G., 2016. Common variants in ABCA7 and MS4A6A are associated with cortical and hippocampal atrophy. Neurobiol. Aging 39, 82–89. doi:10.1016/j.neurobiolaging.2015.10.037

Rang, M.M.D.J.M.R., R.J. Fowler H.P., 2008. Rang and Dale's Pharmacology. Churchill Livingstone, Edinburgh.

Rapoport, M., Dawson, H.N., Binder, L.I., Vitek, M.P., Ferreira, A., 2002. Tau is essential to beta -amyloid-induced neurotoxicity. Proc. Natl. Acad. Sci. U. S. A. 99, 6364–6369. doi:10.1073/pnas.092136199

Reddy, P.H., 2011. Abnormal tau, mitochondrial dysfunction, impaired axonal transport of mitochondria, and synaptic deprivation in Alzheimer's disease. Brain Res. 1415, 136–148. doi:10.1016/j.brainres.2011.07.052

Reese, M.G., Eeckman, F.H., Kulp, D., Haussler, D., 1997. Improved splice site detection in Genie. J. Comput. Biol. J. Comput. Mol. Cell Biol. 4, 311–323.

Reisberg, B., Doody, R., Stöffler, A., Schmitt, F., Ferris, S., Möbius, H.J., 2003. Memantine in Moderate-to-Severe Alzheimer's Disease. N. Engl. J. Med. 348, 1333–1341. doi:10.1056/NEJMoa013128

Reitz, C., Jun, G., Naj, A., Rajbhandary, R., Vardarajan, B.N., Wang, L.-S., Valladares, O., Lin, C.-F., Larson, E.B., Graff-Radford, N.R., Evans, D., De Jager, P.L., Crane, P.K., Buxbaum, J.D., Murrell, J.R., Raj, T., Ertekin-Taner, N., Logue, M., Baldwin, C.T., Green, R.C., Barnes, L.L., Cantwell, L.B., Fallin, M.D., Go, R.C.P., Griffith, P., Obisesan, T.O., Manly, J.J., Lunetta, K.L., Kamboh, M.I., Lopez, O.L., Bennett, D.A., Hendrie, H., Hall, K.S., Goate, A.M., Byrd, G.S., Kukull, W.A., Foroud, T.M., Haines, J.L., Farrer, L.A., Pericak-Vance, M.A., Schellenberg, G.D., Mayeux, R., Alzheimer Disease Genetics Consortium, 2013. Variants in the ATP-binding cassette transporter (ABCA7), apolipoprotein E ϵ4,and the risk of late-onset Alzheimer disease in African Americans. JAMA J. Am. Med. Assoc. 309, 1483–1492. doi:10.1001/jama.2013.2973

Repa, J.J., Turley, S.D., Lobaccaro, J.-M.A., Medina, J., Li, L., Lustig, K., Shan, B., Heyman, R.A., Dietschy, † J. M., Mangelsdorf, D.J., 2000. Regulation of Absorption and ABC1-Mediated Efflux of Cholesterol by RXR Heterodimers. Science 289, 1524–1529. doi:10.1126/science.289.5484.1524

Rhyne, J., Mantaring, M.M., Gardner, D.F., Miller, M., 2009. Multiple splice defects in ABCA1 cause low HDL-C in a family with hypoalphalipoproteinemia and premature coronary disease. BMC Med. Genet. 10, 1. doi:10.1186/1471-2350-10-1

Rice, P.A., Correll, C.C. (Eds.), 2008. Protein-nucleic acid interactions: structural biology, RSC biomolecular sciences. RSC Pub, Cambridge.

Richens, J.L., Morgan, K., O'Shea, P., 2014. Reverse engineering of Alzheimer's disease based on biomarker pathways analysis. Neurobiol. Aging. doi:10.1016/j.neurobiolaging.2014.02.024

Ridge, P.G., Hoyt, K.B., Boehme, K., Mukherjee, S., Crane, P.K., Haines, J.L., Mayeux, R., Farrer, L.A., Pericak-Vance, M.A., Schellenberg, G.D., Kauwe, J.S.K., 2016. Assessment of the genetic variance of late-onset Alzheimer's disease. Neurobiol. Aging 41, 200.e13-200.e20. doi:10.1016/j.neurobiolaging.2016.02.024

Ritchie, M.D., Holzinger, E.R., Li, R., Pendergrass, S.A., Kim, D., 2015. Methods of integrating data to uncover genotype–phenotype interactions. Nat. Rev. Genet. 16, 85–97. doi:10.1038/nrg3868

Rivas, M.A., Beaudoin, M., Gardet, A., Stevens, C., Sharma, Y., Zhang, C.K., Boucher, G., Ripke, S., Ellinghaus, D., Burtt, N., Fennell, T., Kirby, A., Latiano, A., Goyette, P., Green, T., Halfvarson, J., Haritunians, T., Korn, J.M., Kuruvilla, F., Lagacé, C., Neale, B., Lo, K.S., Schumm, P., Törkvist, L., National Institute of Diabetes and Digestive Kidney Diseases Inflammatory Bowel Disease Genetics Consortium (NIDDK IBDGC), United Kingdom Inflammatory Bowel Disease Genetics Consortium, International Inflammatory Bowel Disease Genetics Consortium, Dubinsky, M.C., Brant, S.R., Silverberg, M.S., Duerr, R.H., Altshuler, D., Gabriel, S., Lettre, G., Franke, A., D'Amato, M., McGovern, D.P.B., Cho, J.H., Rioux, J.D., Xavier, R.J., Daly, M.J., 2011. Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. Nat. Genet. 43, 1066–1073. doi:10.1038/ng.952

Roadmap Epigenomics Consortium, Kundaje, A., Meuleman, W., Ernst, J., Bilenky, M., Yen, A., Heravi-Moussavi, A., Kheradpour, P., Zhang, Z., Wang, J., Ziller, M.J., Amin, V., Whitaker, J.W., Schultz, M.D., Ward, L.D., Sarkar, A., Quon, G., Sandstrom, R.S., Eaton, M.L., Wu, Y.-C., Pfenning, A.R., Wang, X., Claussnitzer, M., Liu, Y., Coarfa, C., Harris, R.A., Shoresh, N., Epstein, C.B., Gjoneska, E., Leung, D., Xie, W., Hawkins, R.D., Lister, R., Hong, C.,

Gascard, P., Mungall, A.J., Moore, R., Chuah, E., Tam, A., Canfield, T.K., Hansen, R.S., Kaul, R., Sabo, P.J., Bansal, M.S., Carles, A., Dixon, J.R., Farh, K.-H., Feizi, S., Karlic, R., Kim, A.-R., Kulkarni, A., Li, D., Lowdon, R., Elliott, G., Mercer, T.R., Neph, S.J., Onuchic, V., Polak, P., Rajagopal, N., Ray, P., Sallari, R.C., Siebenthall, K.T., Sinnott-Armstrong, N.A., Stevens, M., Thurman, R.E., Wu, J., Zhang, B., Zhou, X., Beaudet, A.E., Boyer, L.A., De Jager, P.L., Farnham, P.J., Fisher, S.J., Haussler, D., Jones, S.J.M., Li, W., Marra, M.A., McManus, M.T., Sunyaev, S., Thomson, J.A., Tlsty, T.D., Tsai, L.-H., Wang, W., Waterland, R.A., Zhang, M.Q., Chadwick, L.H., Bernstein, B.E., Costello, J.F., Ecker, J.R., Hirst, M., Meissner, A., Milosavljevic, A., Ren, B., Stamatoyannopoulos, J.A., Wang, T., Kellis, M., 2015. Integrative analysis of 111 reference human epigenomes. Nature 518, 317–330. doi:10.1038/nature14248

Rosén, C., Hansson, O., Blennow, K., Zetterberg, H., 2013. Fluid biomarkers in Alzheimer's disease – current concepts. Mol. Neurodegener. 8, 20. doi:10.1186/1750-1326-8-20

Rosenthal, S.L., Barmada, M.M., Wang, X., Demirci, F.Y., Kamboh, M.I., 2014. Connecting the Dots: Potential of Data Integration to Identify Regulatory SNPs in Late-Onset Alzheimer's Disease GWAS Findings. PLoS ONE 9, e95152. doi:10.1371/journal.pone.0095152

Roses, A.D., Saunders, A.M., 2006. Perspective on a pathogenesis and treatment of Alzheimer's disease. Alzheimers Dement. J. Alzheimers Assoc. 2, 59–70. doi:10.1016/j.jalz.2005.12.001

Roy, S., Zhang, B., Lee, V.M.-Y., Trojanowski, J.Q., 2005. Axonal transport defects: a common theme in neurodegenerative diseases. Acta Neuropathol. (Berl.) 109, 5–13. doi:10.1007/s00401-004-0952-x

Rust, S., Rosier, M., Funke, H., Real, J., Amoura, Z., Piette, J.-C., Deleuze, J.-F., Brewer, H.B., Duverger, N., Denèfle, P., Assmann, G., 1999. Tangier disease is caused by mutations in the gene encoding ATP-binding cassette transporter 1. Nat. Genet. 22, 352–355. doi:10.1038/11921

Rust, S., Walter, M., Funke, H., von Eckardstein, A., Cullen, P., Kroes, H.Y., Hordijk, R., Geisel, J., Kastelein, J., Molhuizen, H.O.F., Schreiner, M., Mischke, A., Hahmann, H.W., Assmann, G., 1998. Assignment of Tangier disease to chromosome 9q31 by a graphical linkage exclusion strategy. Nat. Genet. 20, 96–98. doi:10.1038/1770

Sabirzhanova, I., Lopes Pacheco, M., Rapino, D., Grover, R., Handa, J.T., Guggino, W.B., Cebotaru, L., 2015. Rescuing Trafficking Mutants of the ATP-binding Cassette Protein, ABCA4, with Small Molecule Correctors as a Treatment for Stargardt Eye Disease. J. Biol. Chem. 290, 19743–19755. doi:10.1074/jbc.M115.647685

Samson, K., 2010. NerveCenter: Phase III Alzheimer trial halted: Search for therapeutic biomarkers continues. Ann. Neurol. 68, A9–A12. doi:10.1002/ana.22249

Sanna, S., Li, B., Mulas, A., Sidore, C., Kang, H.M., Jackson, A.U., Piras, M.G., Usala, G., Maninchedda, G., Sassu, A., Serra, F., Palmas, M.A., Wood, W.H., Nj?lstad, I., Laakso, M., Hveem, K., Tuomilehto, J., Lakka, T.A., Rauramaa, R., Boehnke, M., Cucca, F., Uda, M., Schlessinger, D., Nagaraja, R., Abecasis, G.R., 2011. Fine Mapping of Five Loci Associated with Low-Density Lipoprotein Cholesterol Detects Variants That Double the Explained Heritability. PLoS Genet. 7. doi:10.1371/journal.pgen.1002198

Sasaki, M., Shoji, A., Kubo, Y., Nada, S., Yamaguchi, A., 2003. Cloning of rat ABCA7 and its preferential expression in platelets. Biochem. Biophys. Res. Commun. 304, 777–782.

Sassi, Celeste, Guerreiro, R., Gibbs, R., Ding, J., Lupton, M.K., Troakes, C., Al-Sarraj, S., Niblock, M., Gallo, J.-M., Adnan, J., Killick, R., Brown, K.S.,

Medway, C., Lord, J., Turton, J., Bras, J., Morgan, K., Powell, J.F., Singleton, A., Hardy, J., 2014. Investigating the role of rare coding variability in Mendelian dementia genes (APP, PSEN1, PSEN2, GRN, MAPT, and PRNP) in late-onset Alzheimer's disease. Neurobiol. Aging 35, 2881.e1-2881.e6. doi:10.1016/j.neurobiolaging.2014.06.002

Sassi, C., Guerreiro, R., Gibbs, R., Ding, J., Lupton, M.K., Troakes, C., Lunnon, K., Al-Sarraj, S., Brown, K.S., Medway, C., Lord, J., Turton, J., Mann, D., Snowden, J., Neary, D., Harris, J., Bras, J., Morgan, K., Powell, J.F., Singleton, A., Hardy, J., 2014. Exome sequencing identifies two novel PSEN1 mutations (p.L166V and p.S230R) in British early onset Alzheimer's disease. Neurobiol. Aging. doi:10.1016/j.neurobiolaging.2014.04.026

Sato, N., Hori, O., Yamaguchi, A., Lambert, J.-C., Chartier-Harlin, M.-C., Robinson, P.A., Delacourte, A., Schmidt, A.M., Furuyama, T., Imaizumi, K., Tohyama, M., Takagi, T., 1999. A Novel Presenilin-2 Splice Variant in Human Alzheimer's Disease Brain Tissue. J. Neurochem. 72, 2498–2505. doi:10.1046/j.1471-4159.1999.0722498.x

Satoh, K., Abe-Dohmae, S., Yokoyama, S., St George-Hyslop, P., Fraser, P.E., 2015. ABCA7 Loss of Function Alters Alzheimer Amyloid Processing. J. Biol. Chem. jbc.M115.655076. doi:10.1074/jbc.M115.655076

Saunders, A.M., Strittmatter, W.J., Schmechel, D., George-Hyslop, P.H.S., Pericak-Vance, M.A., Joo, S.H., Rosi, B.L., Gusella, J.F., Crapper-MacLachlan, D.R., Alberts, M.J., Hulette, C., Crain, B., Goldgaber, D., Roses, A.D., 1993. Association of apolipoprotein E allele ϵ4 with late-onset familial and sporadic Alzheimer's disease. Neurology 43, 1467–1467. doi:10.1212/WNL.43.8.1467

Schaub, M.A., Boyle, A.P., Kundaje, A., Batzoglou, S., Snyder, M., 2012. Linking disease associations with regulatory information in the human genome. Genome Res. 22, 1748–1759. doi:10.1101/gr.136127.111

Schellenberg, G.D., Montine, T.J., 2012. The genetics and neuropathology of Alzheimer's disease. Acta Neuropathol. (Berl.) 124, 305–323. doi:10.1007/s00401-012-0996-2

Schenk, D., Barbour, R., Dunn, W., Gordon, G., Grajeda, H., Guido, T., Hu, K., Huang, J., Johnson-Wood, K., Khan, K., Kholodenko, D., Lee, M., Liao, Z., Lieberburg, I., Motter, R., Mutter, L., Soriano, F., Shopp, G., Vasquez, N., Vandevert, C., Walker, S., Wogulis, M., Yednock, T., Games, D., Seubert, P., 1999. Immunization with amyloid-beta attenuates Alzheimer-disease-like pathology in the PDAPP mouse. Nature 400, 173–177. doi:10.1038/22124

Schierding, W., Cutfield, W.S., O'Sullivan, J.M., 2014. The missing story behind Genome Wide Association Studies: single nucleotide polymorphisms in gene deserts have a story to tell. Front. Genet. 5. doi:10.3389/fgene.2014.00039

Schlötterer, C., Tobler, R., Kofler, R., Nolte, V., 2014. Sequencing pools of individuals — mining genome-wide polymorphism data without big funding. Nat. Rev. Genet. 15, 749–763. doi:10.1038/nrg3803

Schnetz-Boutaud, N.C., Hoffman, J., Coe, J.E., Murdock, D.G., Pericak-Vance, M.A., Haines, J.L., 2012. Identification and confirmation of an exonic splicing enhancer variation in exon 5 of the Alzheimer disease associated PICALM gene. Ann. Hum. Genet. 76, 448–453. doi:10.1111/j.1469-1809.2012.00727.x

Scotti, M.M., Swanson, M.S., 2015. RNA mis-splicing in disease. Nat. Rev. Genet. 17, 19–32. doi:10.1038/nrg.2015.3

Seshadri, S., Fitzpatrick, A.L., Ikram, M.A., DeStefano, A.L., Gudnason, V., Boada, M., Bis, J.C., Smith, A.V., Carassquillo, M.M., Lambert, J.C., Harold, D., Schrijvers, E.M.C., Ramirez-Lorca, R., Debette, S., Longstreth, W.T., Jr, Janssens, A.C.J.W., Pankratz, V.S., Dartigues, J.F., Hollingworth, P., Aspelund, T., Hernandez, I., Beiser, A., Kuller, L.H., Koudstaal, P.J., Dickson, D.W., Tzourio, C., Abraham, R., Antunez, C., Du, Y., Rotter, J.I., Aulchenko, Y.S., Harris, T.B., Petersen, R.C., Berr, C., Owen, M.J., Lopez-

Arrieta, J., Varadarajan, B.N., Becker, J.T., Rivadeneira, F., Nalls, M.A., Graff-Radford, N.R., Campion, D., Auerbach, S., Rice, K., Hofman, A., Jonsson, P.V., Schmidt, H., Lathrop, M., Mosley, T.H., Au, R., Psaty, B.M., Uitterlinden, A.G., Farrer, L.A., Lumley, T., Ruiz, A., Williams, J., Amouyel, P., Younkin, S.G., Wolf, P.A., Launer, L.J., Lopez, O.L., van Duijn, C.M., Breteler, M.M.B., CHARGE Consortium, GERAD1 Consortium, EADI1 Consortium, 2010. Genome-wide analysis of genetic loci associated with Alzheimer disease. JAMA J. Am. Med. Assoc. 303, 1832–1840. doi:10.1001/jama.2010.574

Shankar, G.M., Li, S., Mehta, T.H., Garcia-Munoz, A., Shepardson, N.E., Smith, I., Brett, F.M., Farrell, M.A., Rowan, M.J., Lemere, C.A., Regan, C.M., Walsh, D.M., Sabatini, B.L., Selkoe, D.J., 2008. Amyloid-β protein dimers isolated directly from Alzheimer's brains impair synaptic plasticity and memory. Nat. Med. 14, 837–842. doi:10.1038/nm1782

Shewale, S.J., 2012. The Potential Role of Epigenetics in Alzheimer's Disease Etiology. Biol. Syst. Open Access 02. doi:10.4172/2329-6577.1000114

Shibata, M., Yamada, S., Kumar, S.R., Calero, M., Bading, J., Frangione, B., Holtzman, D.M., Miller, C.A., Strickland, D.K., Ghiso, J., Zlokovic, B.V., 2000. Clearance of Alzheimer's amyloid-ss(1-40) peptide from brain by LDL receptor-related protein-1 at the blood-brain barrier. J. Clin. Invest. 106, 1489–1499. doi:10.1172/JCI10498

Shulenin, S., Nogee, L.M., Annilo, T., Wert, S.E., Whitsett, J.A., Dean, M., 2004. ABCA3 gene mutations in newborns with fatal surfactant deficiency. N. Engl. J. Med. 350, 1296–1303. doi:10.1056/NEJMoa032178

Shulman, J.M., Chen, K., Keenan, B.T., Chibnik, L.B., Fleisher, A., Thiyyagura, P., Roontiva, A., McCabe, C., Patsopoulos, N.A., Corneveaux, J.J., Yu, L., Huentelman, M.J., Evans, D.A., Schneider, J.A., Reiman, E.M., De Jager, P.L., Bennett, D.A., 2013. Genetic susceptibility for Alzheimer disease neuritic plaque pathology. JAMA Neurol. 70, 1150–1157. doi:10.1001/jamaneurol.2013.2815

Siemers, E.R., Sundell, K.L., Carlson, C., Case, M., Sethuraman, G., Liu-Seifert, H., Dowsett, S.A., Pontecorvo, M.J., Dean, R.A., Demattos, R., 2015. Phase 3 solanezumab trials: Secondary outcomes in mild Alzheimer's disease patients. Alzheimers Dement. J. Alzheimers Assoc. 0. doi:10.1016/j.jalz.2015.06.1893

Simmons, A., Westman, E., Muehlboeck, S., Mecocci, P., Vellas, B., Tsolaki, M., Kłoszewska, I., Wahlund, L.-O., Soininen, H., Lovestone, S., Evans, A., Spenger, C., for the AddNeuroMed Consortium, 2009. MRI Measures of Alzheimer's Disease and the AddNeuroMed Study. Ann. N. Y. Acad. Sci. 1180, 47–55. doi:10.1111/j.1749-6632.2009.05063.x

Singh, R.K., Cooper, T.A., 2012. Pre-mRNA splicing in disease and therapeutics. Trends Mol. Med. 18, 472–482. doi:10.1016/j.molmed.2012.06.006

Sink KM, Espeland MA, Castro CM, et al, 2015. Effect of a 24-month physical activity intervention vs health education on cognitive outcomes in sedentary older adults: The life randomized trial. JAMA 314, 781–790. doi:10.1001/jama.2015.9617

Sjögren, H., 1935. Zur Kenntnis Der Keratoconjunctivitis Sicca. Iii. Acta Ophthalmol. (Copenh.) 13, 40–45. doi:10.1111/j.1755-3768.1935.tb04187.x

Skerrett, R., Pellegrino, M.P., Casali, B.T., Taraboanta, L., Landreth, G.E., 2015. Combined Liver X Receptor/Peroxisome Proliferator-Activated Receptor γ Agonist Treatment Reduces Amyloid-β Levels and Improves Behavior in Amyloid Precursor Protein/Presenilin 1 Mice. J. Biol. Chem. jbc.M115.652008. doi:10.1074/jbc.M115.652008

Skoog, I., Nilsson, L., Persson, G., Lernfelt, B., Landahl, S., Palmertz, B., Andreasson, L.-A., Odén, A., Svanborg, A., 1996. 15-year longitudinal study

of blood pressure and dementia. The Lancet 347, 1141–1145. doi:10.1016/S0140-6736(96)90608-X

Small, S.A., Perera, G.M., DeLaPaz, R., Mayeux, R., Stern, Y., 1999. Differential regional dysfunction of the hippocampal formation among elderly with memory decline and Alzheimer's disease. Ann. Neurol. 45, 466–472.

Smith, C.W.J., Valcárcel, J., 2000. Alternative pre-mRNA splicing: the logic of combinatorial control. Trends Biochem. Sci. 25, 381–388. doi:10.1016/S0968-0004(00)01604-2

Sonenberg, N. (Ed.), 2000. Translational control of gene expression, 2. ed. ed, Cold Spring Harbor monograph series. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Sood, S., Anthony, R., Pease, C. t., 2000. Sjögren's syndrome. Clin. Otolaryngol. Allied Sci. 25, 350–357. doi:10.1046/j.1365-2273.2000.00412.x

Stamatoyannopoulos, J., 2016. Connecting the regulatory genome. Nat. Genet. 48, 479–480. doi:10.1038/ng.3553

Steffensen, A.Y., Dandanell, M., Jønson, L., Ejlertsen, B., Gerdes, A.-M., Nielsen, F.C., Hansen, T. vO, 2014. Functional characterization of BRCA1 gene variants by mini-gene splicing assay. Eur. J. Hum. Genet. EJHG 22, 1362–1368. doi:10.1038/ejhg.2014.40

Steinberg, S., Stefansson, H., Jonsson, T., Johannsdottir, H., Ingason, A., Helgason, H., Sulem, P., Magnusson, O.T., Gudjonsson, S.A., Unnsteinsdottir, U., Kong, A., Helisalmi, S., Soininen, H., Lah, J.J., DemGene, Aarsland, D., Fladby, T., Ulstein, I.D., Djurovic, S., Sando, S.B., White, L.R., Knudsen, G.-P., Westlye, L.T., Selbæk, G., Giegling, I., Hampel, H., Hiltunen, M., Levey, A.I., Andreassen, O.A., Rujescu, D., Jonsson, P.V., Bjornsson, S., Snaedal, J., Stefansson, K., 2015. Loss-of-function variants in ABCA7 confer risk of Alzheimer's disease. Nat. Genet. advance online publication. doi:10.1038/ng.3246

Stewart, W.F., Kawas, C., Corrada, M., Metter, E.J., 1997. Risk of Alzheimer's disease and duration of NSAID use. Neurology 48, 626–632.

Strittmatter, W.J., Saunders, A.M., Schmechel, D., Pericak-Vance, M., Enghild, J., Salvesen, G.S., Roses, A.D., 1993. Apolipoprotein E: high-avidity binding to beta-amyloid and increased frequency of type 4 allele in late-onset familial Alzheimer disease. Proc. Natl. Acad. Sci. U. S. A. 90, 1977–1981.

Su, H.P., Nakada-Tsukui, K., Tosello-Trampont, A.-C., Li, Y., Bu, G., Henson, P.M., Ravichandran, K.S., 2002. Interaction of CED-6/GULP, an Adapter Protein Involved in Engulfment of Apoptotic Cells with CED-1 and CD91/Low Density Lipoprotein Receptor-related Protein (LRP). J. Biol. Chem. 277, 11772–11779.

Taggart, A.J., DeSimone, A.M., Shih, J.S., Filloux, M.E., Fairbrother, W.G., 2012. Large-scale mapping of branchpoints in human pre-mRNA transcripts in vivo. Nat. Struct. Mol. Biol. 19, 719–721. doi:10.1038/nsmb.2327

Takahashi, K., Kimura, Y., Nagata, K., Yamamoto, A., Matsuo, M., Ueda, K., 2005. ABC proteins: key molecules for lipid homeostasis. Med. Mol. Morphol. 38, 2–12. doi:10.1007/s00795-004-0278-8

Tanaka, A.R., Ikeda, Y., Abe-Dohmae, S., Arakawa, R., Sadanami, K., Kidera, A., Nakagawa, S., Nagase, T., Aoki, R., Kioka, N., Amachi, T., Yokoyama, S., Ueda, K., 2001. Human ABCA1 Contains a Large Amino-Terminal Extracellular Domain Homologous to an Epitope of Sjögren's Syndrome. Biochem. Biophys. Res. Commun. 283, 1019–1025. doi:10.1006/bbrc.2001.4891

Tanaka, N., Abe-Dohmae, S., Iwamoto, N., Fitzgerald, M.L., Yokoyama, S., 2011. HMG-CoA reductase inhibitors enhance phagocytosis by upregulating ATP-binding cassette transporter A7. Atherosclerosis 217, 407–414. doi:10.1016/j.atherosclerosis.2011.06.031

Tanaka, N., Abe-Dohmae, S., Iwamoto, N., Fitzgerald, M.L., Yokoyama, S., 2010. Helical apolipoproteins of high-density lipoprotein enhance phagocytosis by stabilizing ATP-binding cassette transporter A7. J. Lipid Res. 51, 2591–2599. doi:10.1194/jlr.M006049

Tanzi, R.E., Bertram, L., 2005. Twenty years of the Alzheimer's disease amyloid hypothesis: a genetic perspective. Cell 120, 545–555. doi:10.1016/j.cell.2005.02.008

Teter, B., 2004. ApoE-dependent plasticity in Alzheimer's disease. J. Mol. Neurosci. MN 23, 167–179. doi:10.1385/JMN:23:3:167

Thinakaran, G., Koo, E.H., 2008. Amyloid precursor protein trafficking, processing, and function. J. Biol. Chem. 283, 29615–29619. doi:10.1074/jbc.R800019200

Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakrabortty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., Guigó, R., 2012. Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. Genome Res. 22, 1616–1625. doi:10.1101/gr.134445.111

Toda, Y., Aoki, R., Ikeda, Y., Azuma, Y., Kioka, N., Matsuo, M., Sakamoto, M., Mori, S., Fukumoto, M., Ueda, K., 2005. Detection of ABCA7-positive cells in salivary glands from patients with Sjögren's syndrome. Pathol. Int. 55, 639–643. doi:10.1111/j.1440-1827.2005.01882.x

Tolnay, Probst, 1999. REVIEW: tau protein pathology in Alzheimer's disease and related disorders. Neuropathol. Appl. Neurobiol. 25, 171–187. doi:10.1046/j.1365-2990.1999.00182.x

Turner, P.R., O'Connor, K., Tate, W.P., Abraham, W.C., 2003. Roles of amyloid precursor protein and its fragments in regulating neural activity, plasticity and memory. Prog. Neurobiol. 70, 1–32.

Turton, J., 2014. Deep resequencing of ABCA7, BIN1 and the MS4A locus to identify rare variants putatively associated with Late Onset Alzheimer's disease. University of Nottingham.

Uhlen, M., Oksvold, P., Fagerberg, L., Lundberg, E., Jonasson, K., Forsberg, M., Zwahlen, M., Kampf, C., Wester, K., Hober, S., Wernerus, H., Björling, L., Ponten, F., 2010. Towards a knowledge-based Human Protein Atlas. Nat. Biotechnol. 28, 1248–1250. doi:10.1038/nbt1210-1248

Vasiliou, V., Vasiliou, K., Nebert, D.W., 2009. Human ATP-binding cassette (ABC) transporter family. Hum. Genomics 3, 281–290. doi:10.1186/1479-7364-3-3-281

Vasquez, J.B., Fardo, D.W., Estus, S., 2013. ABCA7 expression is associated with Alzheimer's disease polymorphism and disease status. Neurosci. Lett. 556, 58–62. doi:10.1016/j.neulet.2013.09.058

Vasudevan, S., Peltz, S.W., Wilusz, C.J., 2002. Non-stop decay - a new mRNA surveillance pathway. BioEssays 24, 785–788. doi:10.1002/bies.10153

Venkateswaran, A., Laffitte, B.A., Joseph, S.B., Mak, P.A., Wilpitz, D.C., Edwards, P.A., Tontonoz, P., 2000. Control of cellular cholesterol efflux by the nuclear oxysterol receptor LXRα. Proc. Natl. Acad. Sci. 97, 12097–12102. doi:10.1073/pnas.200367697

Villemagne, V.L., Fodero-Tavoletti, M.T., Pike, K.E., Cappai, R., Masters, C.L., Rowe, C.C., 2008. The ART of loss: Abeta imaging in the evaluation of Alzheimer's disease and other dementias. Mol. Neurobiol. 38, 1–15. doi:10.1007/s12035-008-8019-y

Visscher, P.M., Brown, M.A., McCarthy, M.I., Yang, J., 2012. Five years of GWAS discovery. Am. J. Hum. Genet. 90, 7–24. doi:10.1016/j.ajhg.2011.11.029

Vuong, C.K., Black, D.L., Zheng, S., 2016. The neurogenetics of alternative splicing. Nat. Rev. Neurosci. 17, 265–281. doi:10.1038/nrn.2016.27

Wadkins, R.M., 2000. Targeting DNA secondary structures. Curr. Med. Chem. 7, 1–15.

Wahrle, S.E., Jiang, H., Parsadanian, M., Legleiter, J., Han, X., Fryer, J.D., Kowalewski, T., Holtzman, D.M., 2004. ABCA1 is required for normal central nervous system ApoE levels and for lipidation of astrocyte-secreted apoE. J. Biol. Chem. 279, 40987–40993. doi:10.1074/jbc.M407963200

Wang, N., 2003. ATP-binding Cassette Transporter A7 (ABCA7) Binds Apolipoprotein A-I and Mediates Cellular Phospholipid but Not Cholesterol Efflux. J. Biol. Chem. 278, 42906–42912. doi:10.1074/jbc.M307831200

Ward, L.D., Kellis, M., 2011. HaploReg: a resource for exploring chromatin states, conservation, and regulatory motif alterations within sets of genetically linked variants. Nucleic Acids Res. gkr917. doi:10.1093/nar/gkr917

Ware, M.D., DeSilva, D., Sinilnikova, O.M., Stoppa-Lyonnet, D., Tavtigian, S.V., Mazoyer, S., 2005. Does nonsense-mediated mRNA decay explain the ovarian cancer cluster region of the BRCA2 gene? Oncogene 25, 323–328. doi:10.1038/sj.onc.1209033

Wetterstrand, K.A., 2014. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP).

Wetzel-Smith, M.K., Hunkapiller, J., Bhangale, T.R., Srinivasan, K., Maloney, J.A., Atwal, J.K., Sa, S.M., Yaylaoglu, M.B., Foreman, O., Ortmann, W., Rathore, N., Hansen, D.V., Tessier-Lavigne, M., Alzheimer's Disease Genetics Consortium, Mayeux, R., Pericak-Vance, M., Haines, J., Farrer, L.A., Schellenberg, G.D., Goate, A., Behrens, T.W., Cruchaga, C., Watts, R.J., Graham, R.R., 2014. A rare mutation in UNC5C predisposes to late-onset Alzheimer's disease and increases neuronal cell death. Nat. Med. 20, 1452–1457. doi:10.1038/nm.3736

Wimo, A., Guerchet, M., Ali, G.-C., Wu, Y.-T., Prina, A.M., Winblad, B., Jönsson, L., Liu, Z., Prince, M., 2016. The worldwide costs of dementia 2015 and comparisons with 2010. Alzheimers Dement. doi:10.1016/j.jalz.2016.07.150

Wollmer, M.A., Streffer, J.R., Lütjohann, D., Tsolaki, M., Iakovidou, V., Hegi, T., Pasch, T., Jung, H.H., Bergmann, K. von, Nitsch, R.M., Hock, C., Papassotiropoulos, A., 2003. ABCA1 modulates CSF cholesterol levels and influences the age at onset of Alzheimer's disease. Neurobiol. Aging 24, 421–426.

Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C.L., Haase, J., Janes, J., Huss, J.W., Su, A.I., 2009. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol. 10, R130. doi:10.1186/gb-2009-10-11-r130

Wu, X., Hurst, L.D., 2015. Determinants of the Usage of Splice-Associated *cis* -Motifs Predict the Distribution of Human Pathogenic SNPs. Mol. Biol. Evol. msv251. doi:10.1093/molbev/msv251

Wu, Y.C., Horvitz, H.R., 1998. The C. elegans cell corpse engulfment gene ced-7 encodes a protein similar to ABC transporters. Cell 93, 951–960.

Wyss-Coray, T., 2006. Inflammation in Alzheimer disease: driving force, bystander or beneficial response? Nat. Med. 12, 1005–1015. doi:10.1038/nm1484

Xia, L.C., 2016. Abstract: Structural Variation (SV) in Heterogenous Whole-Genome Sequencing Data from 111 Families at Risk for Alzheimer Disease: Alzheimer Disease Sequencing Project SV Study. Presented at the Alzhiemer's Association International Conference 2016, Toronto, Canada.

Yu, L., Chibnik, L.B., Srivastava, G.P., Pochet, N., Yang, J., Xu, J., Kozubek, J., Obholzer, N., Leurgans, S.E., Schneider, J.A., Meissner, A., De Jager, P.L., Bennett, D.A., 2015. Association of Brain DNA Methylation in SORL1, ABCA7, HLA-DRB5, SLC24A4, and BIN1 With Pathological Diagnosis of Alzheimer Disease. JAMA Neurol. 72, 15. doi:10.1001/jamaneurol.2014.3049

Yuyama, K., Sun, H., Sakai, S., Mitsutake, S., Okada, M., Tahara, H., Furukawa, J. -i., Fujitani, N., Shinohara, Y., Igarashi, Y., 2014. Decreased Amyloid-Pathologies by Intracerebral Loading of Glycosphingolipid-enriched

Exosomes in Alzheimer Model Mice. J. Biol. Chem. 289, 24488–24498. doi:10.1074/jbc.M114.577213

Zamrini, E., McGwin, G., Roseman, J.M., 2004. Association between statin use and Alzheimer's disease. Neuroepidemiology 23, 94–98. doi:10.1159/000073981

Zeng, L., Liao, H., Liu, Y., Lee, T.-S., Zhu, M., Wang, X., Stemerman, M.B., Zhu, Y., Shyy, J.Y.-J., 2004. SREBP2 downregulates ABCA1 in vascular endothelial cells: a novel role of SREBP in regulating cholesterol metabolism. J. Biol. Chem. doi:10.1074/jbc.M407817200

Zhang, D.-W., Nunoya, K., Vasa, M., Gu, H.-M., Cole, S.P.C., Deeley, R.G., 2006a. Mutational analysis of polar amino acid residues within predicted transmembrane helices 10 and 16 of multidrug resistance protein 1 (ABCC1): effect on substrate specificity. Drug Metab. Dispos. Biol. Fate Chem. 34, 539–546. doi:10.1124/dmd.105.007740

Zhang, D.-W., Nunoya, K., Vasa, M., Gu, H.-M., Cole, S.P.C., Deeley, R.G., 2006b. Mutational Analysis of Polar Amino Acid Residues Within Predicted Transmembrane Helices 10 and 16 of Multidrug Resistance Protein 1 (abcc1): Effect on Substrate Specificity. Drug Metab. Dispos. 34, 539–546. doi:10.1124/dmd.105.007740

Zheng, D., Gerstein, M.B., 2006. A computational approach for identifying pseudogenes in the ENCODE regions. Genome Biol. 7, S13. doi:10.1186/gb-2006-7-s1-s13

Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., Yang, J., 2016. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. Nat. Genet. 48, 481–487. doi:10.1038/ng.3538

Zoghbi, M.E., Cooper, R.S., Altenberg, G.A., 2016. The lipid bilayer modulates the structure and function of an ATP-binding cassette exporter. J. Biol. Chem. jbc.M115.698498. doi:10.1074/jbc.M115.698498

Zou, F., Chai, H.S., Younkin, C.S., Allen, M., Crook, J., Pankratz, V.S., Carrasquillo, M.M., Rowley, C.N., Nair, A.A., Middha, S., Maharjan, S., Nguyen, T., Ma, L., Malphrus, K.G., Palusak, R., Lincoln, S., Bisceglio, G., Georgescu, C., Kouri, N., Kolbert, C.P., Jen, J., Haines, J.L., Mayeux, R., Pericak-Vance, M.A., Farrer, L.A., Schellenberg, G.D., Consortium, A.D.G., Petersen, R.C., Graff-Radford, N.R., Dickson, D.W., Younkin, S.G., Ertekin-Taner, N., 2012. Brain Expression Genome-Wide Association Study (eGWAS) Identifies Human Disease-Associated Variants. PLOS Genet 8, e1002707. doi:10.1371/journal.pgen.1002707

Zovkic, I.B., Guzman-Karlsson, M.C., Sweatt, J.D., 2013. Epigenetic regulation of memory formation and maintenance. Learn. Mem. Cold Spring Harb. N 20, 61–74. doi:10.1101/lm.026575.112

Zuk, O., Schaffner, S.F., Samocha, K., Do, R., Hechter, E., Kathiresan, S., Daly, M.J., Neale, B.M., Sunyaev, S.R., Lander, E.S., 2014. Searching for missing heritability: Designing rare variant association studies. Proc. Natl. Acad. Sci. 111, E455–E464. doi:10.1073/pnas.1322563111

Appendix A

Complete catalogue of *ABCA7* exonic variants acquired from the Exome Variant Server (EVS) and Next Generation Sequencing (NGS) performed in this lab. The genomic position and rs ID number of each of these variants are presented as well as the location within the ABCA7 gene, and the position of cDNA and protein changes they code for. Also shown are the minor allele frequencies (MAF) in five databases; HapMap (showing the European population MAF), EVS (the combined "All" frequency), dbSNP and 1000Genomes (again the collective "All" population). All MAFs are shown in percentage format. Also shown are the MAFs from the CRISP programme which was calculated as part of the NGS project. The affect this protein substitution has on the protein is also shown according to three different prediction programmes: the Grantham Score; Polyphen-2 and SIFT. The origin of each of the variants are indicated by dots in the EVS and NGS columns, the only exonic *ABCA7* GWAS variant is also highlighted in the GWAS column.

						MAF						Affect Prediction					
	CHR	GENOMIC POSITION	dbSNP ID	GENE LOCATION	cDNA	HapMap (European) (%)	EVS (All) (%)	dbSNP (%)	1000Genomes (All) (%)	CRISP (%)	PROTEIN	Grantham Score	Polyphen2	SIFT	NGS	GWAS	EVS
1	19	1041337	rs375637434	Exon 2	c.-24 T>C	NA	0.0077	NA	NA	NA	-						•
2	19	1041347	rs182233998	Exon 2	c.-14 T>C	NA	1.1456	31.80	1.00	0.61	-	-	-	-	•		•
3	19	1041352	rs3752229	Exon 2	c.-9 A>G	6.70	3.4676	12.00	12.00	7.2	-	-	-	-	•		•
4	19	1041355	rs376425513	Exon 2	c.-6 C>T	NA	0.0077	NA	NA	NA	-	-	-	-			•
5	19	1041412	rs370845702	Exon 2	c.52 C>A	NA	0.0077	NA	NA	NA	R18S	Moderately Radical (110)		Tolerated (0.14)			•
6	19	1041852	rs3764644	Exon 4	c.183 G>T	NA	6.7426	7.10	7.00	1.963	L61L			Tolerated (1)	•		•
7	19	1041909	rs144546979	Exon 4	c.240 C>T	NA	0.0308	NA	NA	0.488	T80=			Tolerated (0.637)	•		•
8	19	1041922	rs146597357	Exon 4	c.253 C>A	NA	0.0926	0.05	<1.00	NA	L85M	Conservative (15)	Benign (0.42)	Tolerated (0.23)			•
9	19	1041930	rs140142327	Exon 4	c.261 G>A	NA	0.0077	NA	NA	NA	P87P			Tolerated (0.57)			•
10	19	1041935	rs370464878	Exon 4	c.266 A>C	NA	0.0077	NA	NA	NA	E89A	Moderately Radical (107)	Probably Damaging (0.99)	Tolerated (0.09)			•
11	19	1041971	rs201665195	Exon 4	c.302 T>G	NA	0.0627	0.30	<0.10	NA	L101R	Moderately Radical (102)		Tolerated (0.14)			•
12	19	1042075	rs371660184	Exon 5	c.315 G>A	NA	0.0079	NA	NA	NA	L105L			Tolerated (1)			•
13	19	1042331	rs372562028	Exon 6	c.433 A>G	NA	0.0077	NA	NA	NA	L145E	Moderately Conservative (56)	Benign (0)	Tolerated (0.95)			•
14	19	1042369	rs140240982	Exon 6	c.471 G>A	NA	0.0077	NA	NA	NA	A157A			Tolerated (0.78)			•
15	19	1042379	rs375490467	Exon 6	c.481 A>G	NA	0.0077	NA	NA	NA	T161A	Moderately Conservative (58)	Benign (0)	Tolerated (0.75)			•
16	19	1042396	rs369528566	Exon 6	c.498 G>C	NA	0.0077	NA	NA	NA	T166T			Tolerated (0.55)			•
17	19	1042809	rs3764645	Exon 7	c.563 A>G	NA	38.8667	41.28	41.00	41.025	E188G	Moderately Conservative (98)	Benign (0.244)	Tolerated (0.647)	•		•
18	19	1042814	rs138777686	Exon 7	c.568 C>G	NA	0.0154	NA	NA	NA	L190V	Conservative (32)	Benign (0.001)	Tolerated (0.14)			•
19	19	1043071	rs370774722	Exon 8	c.611 G>A	NA	0.0077	NA	NA	NA	R204Q	Conservative (43)	Benign (0.000)	Tolerated (0.25)			•
20	19	1043091	rs148324257	Exon 8	c.631 C>G	NA	0.0077	NA	NA	NA	R211G	Moderately Radical (125)	Benign (0.097)	Tolerated (0.39)			•
21	19	1043103	rs72973581	Exon 8	c.643 G>A	NA	4.6301	2.20	2.00	4.238	G215S	Moderately Conservative (56)	Benign (0.029)	Tolerated (0.625)	•		•
22	19	1043139	rs375055112	Exon 8	c.679 G>A	NA	0.0154	NA	NA	NA	V227I	Conservative (29)	Benign (0.002)	Tolerated (0.39)			•
23	19	1043193	rs377552810	Exon 8	c.733 G>T	NA	0.0077	NA	NA	NA	D245Y	Radical (160)	Possibly Damaging (0.555)	Tolerate (0.18)			•
24	19	1043238	rs200441027	Exon 8	c.778 G>T	NA	0.0077	0.05	<0.10	NA	D260Y	Radical (160)	Probably Damaging (0.999)	Damaging (0.01)			•
25	19	1043338	rs368608996	Exon 9	c.796 G>A	NA	0.0077	NA	NA	NA	A266T	Moderately Conservative (58)	Benign (0.007)	Tolerated (0.34)			•
26	19	1043350	rs202003717	Exon 9	c.808 C>T	NA	0.0923	NA	NA	NA	L270L			Tolerated (1.00)			•
27	19	1043384	rs375028339	Exon 9	c.842 G>A	NA	0.0077	NA	NA	NA	R281H	Conservative (29)	Probably Damaging (0.998)	Tolerated (0.09)			•
28	19	1043399	rs370131491	Exon 9	c.857 G>A	NA	0.0077	NA	NA	NA	R286H	Conservative (29)	Probably Damaging (0.999)	Tolerated (0.53)			•
29	19	1043747	rs149023827	Exon 10	c.954 C>T	NA	0.2999	0.10	<1.00	0.663	L318L			Tolerated (1)	•		•
30	19	1043748	rs3752232	Exon 10	c.955 A>G	4.50	11.3025	10.10	10.00	1.788	T319A	Moderately Conservative (58)	Benign (0)	Tolerated (0.882)	•		•
31	19	1043750	rs369738612	Exon 10	c.957 C>G	NA	0.0077	NA	NA	NA	T319T			Tolerated (0.43)			•
32	19	1043766	rs141983399	Exon 10	c.973 C>T	NA	0.0077	NA	NA	NA	R325W	Moderately Radical (101)	Probably Damaging (0.960)	Tolerated (0.18)			•
33	19	1043793	rs146086314	Exon 10	c.1000 C>T	NA	0.0308	NA	NA	NA	R334W	Moderately Radical (101)	Probably Damaging (0.995)	Damaging (0.02)			•
34	19	1043794	rs147846250	Exon 10	c.1001 G>A	NA	NA	NA	NA	0.938	R334Q	Conservative (43)	Benign (0.004)	Tolerated (1)	•		
35	19	1044619	rs146982710	Exon 11	c.1091C>G	NA	0.0308	NA	NA	0.45	P364R	Moderately Conservative (103)	Possibly Damaging (0.823)	Tolerated (0.185)	•		•
36	19	1044704	rs149365975	Exon 11	c.1176 T>C	NA	0.0077	NA	NA	NA	D392D			Tolerated (0.54)			•
37	19	1044712	rs3764647	Exon 11	c.1184 A>G	4.50	10.7896	9.70	10.00	1.788	H395R	Conservative (29)	Benign (0)	Tolerated (1)	•		•
38	19	1045008	rs142777510	Exon 12	c.1223 C>T	NA	0.0154	NA	NA	NA	S408F	Radical (55)	Possibly Damaging (0.836)	Tolerated (0.08)			•
39	19	1045018	rs146070541	Exon 12	c.1233 G>C	NA	0.0077	NA	NA	NA	K411N	Moderately Conservative (94)	Probably Damaging (1.000)	Tolerated (0.13)			•
40	19	1045026	rs10405305	Exon 12	c.1241 C>G	NA	3.1471	2.70	3.00	NA	A414G	Moderately Conservative (60)	Benign (0.055)	Tolerated (0.35)			•
41	19	1045055	rs142291535	Exon 12	c.1270 C>T	NA	0.0077	NA	NA	NA	R424W	Moderately Radical (101)	Possibly Damaging (0.893)	Damaging (0.03)			•
42	19	1045083	rs150867671	Exon 12	c.1298 G>A	NA	0.0461	0.10	<1.00	NA	R433Q	Conservative (43)	Probably Damaging (0.997)	Tolerated (0.59)			•
43	19	1045105	rs149897784	Exon 12	c.1320 C>A	NA	0.0077	NA	NA	NA	F440L	Conservative (22)	Probably Damaging (1.000)	Damaging (0.01)			•
44	19	1045166	rs371516353	Exon 12	c.1381 C>G	NA	0.0077	NA	NA	NA	H461D	Moderately Conservative (81)	Probably Damaging (0.996)	Tolerated (0.30)			•
45	19	1045172	rs367549958	Exon 12	c.1387 C>T	NA	0.0077	NA	NA	NA	R463C	Radical (180)	Probably Damaging (0.999)	Damaging (0.036)			•
46	19	1045173	rs3752233	Exon 12	c.1388 G>A/C	0.00	2.9837	5.90	6.00	2.263	R463H	Conservative (29)	Probably Damaging (0.997)	Tolerated (1)	•		•
47	19	1045206	rs375990700	Exon 12	c.1421 C>T	NA	0.0077	NA	NA	NA	T474M	Moderately Conservative (81)	Probably Damaging (0.966)	Tolerated (0.07)			•
48	19	1045209	rs147783767	Exon 12	c.1424 G>A	NA	0.0077	NA	NA	NA	R475K	Conservative (26)	Benign (0.004)	Tolerated (0.99)			•
49	19	1045215	rs19913679	Exon 12	c.1430 A>G	NA	0.0077	NA	NA	NA	N477S	Conservative (46)	Possibly Damaging (0.864)	Tolerated (0.72)			•
50	19	1045220	rs376748461	Exon 12	c.1435 A>T	NA	0.0154	NA	NA	NA	I479F	Conservative (21)	Probably Damaging (0.998)	Tolerated (0.30)			•
51	19	1046239	rs141428162	Exon 13	c.1456 C>G	NA	0.0231	NA	NA	NA	P486A	Conservative (27)	Probably Damaging (0.998)	Tolerated (0.09)			•
52	19	1046266	rs370047554	Exon 13	c.1483 G>T	NA	0.0077	NA	NA	NA	D495Y	Radical (160)	Probably Damaging (0.972)	Tolerated (0.21)			•
53	19	1046293	rs374228184	Exon 13	c.1510 G>A	NA	0.0077	NA	NA	NA	V504M	Conservative (21)	Probably Damaging (0.999)	Tolerated (0.18)			•

						MAF						Affect Prediction					
	CHR	GENOMIC POSITION	dbSNP ID	GENE LOCATION	cDNA	HapMap (European) (%)	EVS (All) (%)	dbSNP (%)	1000Genomes (All) (%)	CRISP (%)	PROTEIN	Grantham Score	Polyphen2	SIFT	NGS	GWAS	EVS
54	19	1046311	rs140840476	Exon 13	c.1528 G>A	NA	0.0077	NA	NA	NA	V510M	Conservative (21)	Benign (0.249)	Tolerated (0.10)			•
55	19	1046317	rs367592755	Exon 13	c.1534 C>G	NA	0.0077	NA	NA	NA	R512G	Moderately Radical (125)	Benign (0.001)	Tolerated (0.18)			•
56	19	1046387	rs376996427	Exon 13	c.1604 C>T	NA	0.0077	NA	NA	NA	P535L	Moderately Conservative (98)	Probably Damaging (1.000)	Damaging (0.003)			•
57	19	1046825	rs141323477	Exon 14	c.1647 G>A	NA	0.008	0.10	<1.00	NA	S549S			Tolerated (0.40)			•
58	19	1046944	rs144979723	Exon 14	c.1766 C>T	NA	0.1166	NA	NA	NA	A589V	Moderately Conservative (64)	Possibly Damaging (0.907)	Tolerated (0.51)			•
59	19	1046945	rs368652784	Exon 14	c.1767 G>A	NA	0.0078	NA	NA	NA	A589A			Tolerated (0.55)			•
60	19	1046950	rs138819755	Exon 14	c.1772 T>C	NA	0.0155	NA	NA	NA	L591P	Moderately Conservative (98)	Probably Damaging (0.999)	Tolerated (0.19)			•
61	19	1047002	rs3752234	Exon 14	c.1824 A>G	NA	46.6346	40.10	40.00	58.075	A608A			Tolerated (1)	•		•
62	19	1047161	rs3752237	Exon 15	c.1851 A>G	38.30	37.5905	31.80	32.00	61.988	G617G			Tolerated (1)	•		•
63	19	1047166	rs370356161	Exon 15	c.1856 T>A	NA	0.0077	NA	NA	NA	I619N	Moderately Radical (149)	Probably Damaging (1.000)	Damaging (0.00)			•
64	19	1047169	rs144852598	Exon 15	c.1859 T>C	NA	0.0308	NA	NA	0.488	L620P	Moderately Conservative (98)	Probably Damaging (1)	Deleterious (0)	•		•
65	19	1047182	rs373020856	Exon 15	c.1872 C>T	NA	0.0077	NA	NA	NA	H624H			Tolerated (0.17)			•
66	19	1047218	rs200714491	Exon 15	c.1908 G>A	NA	0.0077	NA	NA	NA	A636A			Tolerated (1.00)			•
67	19	1047253	-	Exon 15	c.1944_1946del3	NA	0.1039	NA	NA	NA	A648_F649delinsA			Damaging (0.04)			•
68	19	1047265	rs367977381	Exon 15	c.1955 G>T	NA	0.0077	NA	NA	NA	R652L	Moderately Radical (102)	Possibly Damaging (0.898)	Damaging (0.01)			•
69	19	1047266	rs371509588	Exon 15	c.1956 C>T	NA	0.0231	NA	NA	NA	R652R			Tolerated (0.97)			•
70	19	1047280	rs141937680	Exon 15	c.1970 C>T	NA	0.0077	NA	NA	NA	A657V	Moderately Conservative (64)	Probably Damaging (1.000)	Tolerated (0.07)			•
71	19	1047288	rs373714431	Exon 15	c.1978 G>A	NA	0.0077	NA	NA	NA	G660S	Moderately Conservative (56)	Probably Damaging (0.977)	Tolerated (0.29)			•
72	19	1047336	rs59851484	Exon 15	c.2026 G>A	NA	3.8354	3.00	3.00	NA	A676T	Moderately Conservative (58)	Probably Damaging (1.000)	Tolerated (0.55)			•
73	19	1047348	rs375628486	Exon 15	c.2038 C>T	NA	0.0078	NA	NA	NA	R680W	Moderately Radical (101)	Probably Damaging (0.997)	Tolerated (0.16)			•
74	19	1047372	rs376686030	Exon 15	c.2062 G>A	NA	0.0081	NA	NA	NA	A688T	Moderately Conservative (58)	Benign (0.237)	Tolerated (0.41)			•
75	19	1047478	rs372209007	Exon 16	c.2094 C>T	NA	0.0079	NA	NA	NA	G698G			Tolerated (1.00)			•
76	19	1047507	-	Exon 16	c.2124_2130del7	NA	0.2825	NA	NA	NA	E709Afs*86		FRAMESHIFT				•
77	19	1047537	rs3752239	Exon 16	c.2153 A>C	NA	2.7002	5.80	6.00	2.063	N718T	Moderately Conservative (65)	Possibly Damaging (0.529)	Tolerated (0.239)	•		•
78	19	1047555	rs368240505	Exon 16	c.2171 C>T	NA	0.0077	NA	NA	NA	T724M	Moderately Conservative (81)	Possibly Damaging (0.947)	Tolerated (0.26)			•
79	19	1048943	rs368735263	Exon 17	c.2319 C>A	NA	0.0077	NA	NA	NA	Y773*		STOP CODON				•
80	19	1048950	rs149949633	Exon 17	c.2326 G>A	NA	0.0231	NA	NA	NA	G776R	Moderately Radical (125)	Probably Damaging (0.996)	Tolerated (0.15)			•
81	19	1048961	-	Exon 17	c.2337 C>T	NA	NA	NA	NA	0.775	P779P			Tolerated (0.636)	•		•
82	19	1048963	rs377353003	Exon 17	c.2339 C>A	NA	0.0077	NA	NA	NA	P780H	Moderately Conservative (77)	Probably Damaging (0.999)	Tolerated (0.15)			•
83	19	1048969	rs369433123	Exon 17	c.2345 G>T	NA	0.0077	NA	NA	NA	S782I	Moderately Radical (142)	Possibly Damaging (0.610)	Tolerated (0.15)			•
84	19	1048982	rs9282560	Exon 17	c.2358 C>T	NA	0.1618	0.60	1.00	NA	C786C			Tolerated (0.07)			•
85	19	1048983	rs142153271	Exon 17	c.2359 C>G	NA	0.0077	0.05	<1.00	NA	P787A	Conservative (27)	Benign (0.000)	Tolerated (0.75)			•
86	19	1048988	rs146739407	Exon 17	c.2364 C>T	NA	0.0154	NA	NA	NA	T788T			Tolerated (0.64)			•
87	19	1049269	rs4147914	Exon 18	c.2385 G>A	NA	15.2863	21.90	22.00	17.875	L795L			Tolerated (1)	•		•
88	19	1049285	rs372623005	Exon 18	c.2401 C>A	NA	0.0077	NA	NA	NA	P801T	Conservative (38)	Benign (0.049)	Tolerated (0.62)			•
89	19	1049305	rs4147915	Exon 18	c.2421 C>A	NA	13.1846	21.70	22.00	9.013	V807V			Tolerated (1)	•		•
90	19	1049316	rs374944865	Exon 18	c.2432 G>T	NA	0.0077	NA	NA	NA	S811I	Moderately Radical (142)	Benign (0.179)	Damaging (0.01)			•
91	19	1049344	rs369072404	Exon 18	c.2460 G>T	NA	0.0077	NA	NA	NA	P820P			Tolerated (0.32)			•
92	19	1049360	rs199517653	Exon 18	c.2476 G>A	NA	0.0771	NA	NA	NA	G826R	Moderately Radical (125)	Possibly Damaging (0.533)	Tolerated (0.24)			•
93	19	1049378	rs369777029	Exon 18	c.2494 T>C	NA	0.0077	NA	NA	NA	Y832H	Moderately Conservative (83)	Probably Damaging (1.000)	Tolerated (0.06)			•
94	19	1049392	rs138316402	Exon 18	c.2508 C>T	NA	0.0077	NA	NA	NA	I836I			Tolerated (1.00)			•
95	19	1049395	rs371397416	Exon 18	c.2511 C>T	NA	0.0077	NA	NA	NA	T837T			Tolerated (1.00)			•
96	19	1049398	rs374298252	Exon 18	c.2514 C>T	NA	0.0077	NA	NA	NA	A838A			Tolerated (1.00)			•
97	19	1049410	rs368179689	Exon 18	c.2526 C>T	NA	0.0077	NA	NA	NA	H842H			Tolerated (1.00)			•
98	19	1049414	rs371982554	Exon 18	c.2530 G>C	NA	0.0077	NA	NA	NA	G844R	Moderately Radical (125)	Probably Damaging (0.995)	Damaging (0.00)			•
99	19	1050996	rs74176364	Exon 19	c.2629 G>A	NA	0.5167	1.50	2.00	NA	A877T	Moderately Conserved (58)		Tolerated (0.38)			•
100	19	1050999	rs139415355	Exon 19	c.2632 G>A	NA	0.0077	NA	NA	NA	A878T	Moderately Conserved (58)	Benign (0.178)	Tolerated (0.56)			•
101	19	1051006	rs143718918	Exon 19	c.2639 G>A	NA	0.0617	0.05	<1.00	NA	R880Q	Conserved (43)	Probably Damaging (0.999)	Damaging (0.00)			•
102	19	1051214	rs3752240	Exon 20	c.2745 A>G	36.90	36.5462	30.60	31.00	37.175	V915V			Tolerated (0.11)	•		•
103	19	1051293	rs200140277	Exon 20	c.2824 G>C	NA	0.0154	NA	NA	NA	G942R	Moderately Radical (125)	Probably Damaging (1.000)	Damaging (0.00)			•
104	19	1051491	rs200207400	Exon 21	c.2868 C>T	NA	0.0077	NA	NA	NA	G956G			Tolerated (1.00)			•
105	19	1051493	-	Exon 21	c.2871del1	NA	0.0122	NA	NA	NA	S958Pfs*34		FRAMESHIFT				•
106	19	1051499	rs371611880	Exon 21	c.2876 A>G	NA	0.0077	NA	NA	NA	Q959R	Conserved (43)	Benign (0.000)	Tolerated (0.50)			•
107	19	1051505	rs148962540	Exon 21	c.2882 T>C	NA	0.0077	NA	NA	NA	V961A	Moderately Conserved (64)	Probably Damaging (1.000)	Damaging (0.00)			•

						MAF						Affect Prediction					
	CHR	GENOMIC POSITION	dbSNP ID	GENE LOCATION	cDNA	HapMap (European) (%)	EVS (All) (%)	dbSNP (%)	1000Genomes (All) (%)	CRISP (%)	PROTEIN	Grantham Score	Polyphen2	SIFT	NGS	GWAS	EVS
108	19	1051524	rs202155778	Exon 21	c.2901 G>A	NA	0.0077	0.05	<1.00	NA	T967T			Tolerated (1.00)			•
109	19	1051549	rs373667015	Exon 21	c.2926 C>T	NA	0.0077	NA	NA	NA	R976C	Radical (180)	Probably Damaging (1.000)	Damaging (0.00)			•
110	19	1051550	rs151047273	Exon 21	c.2927 G>A	NA	0.0077	NA	NA	NA	R976H	Conservative (29)	Probably Damaging (1.000)	Damaging (0.02)			•
111	19	1051551	rs112434834	Exon 21	c.2928 C>T	NA	0.0309	NA	NA	NA	R976R			Tolerated (1.00)			•
112	19	1051584	rs34430899	Exon 21	c.2962del1	NA	0.0161	NA	NA	NA	G988Vfs*4	FRAMESHIFT					•
113	19	1051944	rs139214131	Exon 22	c.2966 G>A	NA	0.0694	NA	NA	NA	R989H	Conservative (29)		Tolerated (0.07)			•
114	19	1051947	rs143226982	Exon 22	c.2969 C>T	NA	0.0154	NA	NA	NA	T990M	Moderately Conserved (81)	Probably Damaging (1.000)	Damaging (0.00)			•
115	19	1052005	rs3764652	Exon 22	c.3027 C>T	44.60	40.1544	38.50	38.00	40.963	A1009A			Tolerated (1)	•		•
116	19	1052011	rs143852644	Exon 22	c.3033 G>A	NA	0.0077	NA	NA	NA	V1011V			Tolerated (0.16)			•
117	19	1052017	rs377541682	Exon 22	c.3039 T>A	NA	0.0232	NA	NA	NA	G1013G			Tolerated (0.41)			•
118	19	1052022	rs148613263	Exon 22	c.3044 G>A	NA	0.0077	NA	NA	NA	R1015H	Conserved (29)	Probably Damaging (0.982)	Damaging (0.03)			•
119	19	1052058	rs146278525	Exon 22	c.3080 G>A	NA	0.0232	NA	NA	NA	R1027H	Conserved (29)	Benign (0.007)	Tolerated (0.18)			•
120	19	1052071	rs139447165	Exon 22	c.3093 C>T	NA	0.0077	NA	NA	NA	S1031S			Tolerated (0.78)			•
121	19	1052086	rs61576791	Exon 22	c.3108 G>A	NA	5.9452	6.00	6.00	NA	T1036T			Tolerated (1.00)			•
122	19	1052275	rs199498383	Exon 23	c.3210 C>T	NA	0.1117	0.10	<1.00	0.863	G1070G			Tolerated (0.519)	•		•
123	19	1052284	rs372686297	Exon 23	c.3219 C>T	NA	0.0081	NA	NA	NA	V1073V			Tolerated (0.51)			•
124	19	1053369	rs371433266	Exon 24	c.3262 G>A	NA	0.0156	NA	NA	NA	G1088R	Moderately Radical (125)	Benign (0.188)	Tolerated (0.25)			•
125	19	1053411	rs369970150	Exon 24	c.3304 G>A	NA	0.0077	NA	NA	NA	V1102M	Conserved (21)	Benign (0.129)	Tolerated (0.11)			•
126	19	1053437	rs139301818	Exon 24	c.3330 C>T	NA	0.0077	NA	NA	NA	D1110D			Tolerated (0.70)			•
127	19	1053438	rs150012110	Exon 24	c.3331 G>A	NA	0.0308	NA	NA	NA	G1111S	Moderately Conserved (56)	Probably Damaging (0.996)	Tolerated (0.49)			•
128	19	1053464	rs145147192	Exon 24	c.3357 G>A	NA	0.0077	NA	NA	NA	E1119E			Tolerated (1.00)			•
129	19	1053512	rs375548151	Exon 24	c.3405 C>T	NA	0.0077	NA	NA	NA	S1135S			Tolerated (1.00)			•
130	19	1053524	rs3752241	Exon 24	c.3417 C>G	21.20	16.3879	16.40	16.00	16.45	L1139L			Tolerated (1)	•		•
131	19	1053793	rs149144001	Exon 25	c.3430 C>T	NA	0.0077	NA	NA	NA	L1144L			Tolerated (1.00)			•
132	19	1053811	rs146894513	Exon 25	c.3448 T>C	NA	0.0077	NA	NA	NA	C1150R	Radical (180)	Benign (0.000)	Tolerated (0.54)			•
133	19	1054024	rs370830009	Exon 26	c.3492 C>T	NA	0.0077	NA	NA	NA	H1164H			Tolerated (0.22)			•
134	19	1054052	rs374248440	Exon 26	c.3520 G>A	NA	0.0154	NA	NA	NA	V1174I	Conserved (29)	Benign (0.233)	Tolerated (0.49)			•
135	19	1054060	rs3752243	Exon 26	c.3528 A>G	42.70	47.3551	49.30	49.00	34.888	L1176L			Tolerated (0.352)	•		•
136	19	1054061	rs141885771	Exon 26	c.3529 C>T	NA	0.0077	0.05	<1.00	NA	R1177W	Moderately Radical (101)	Probably Damaging (0.992)	Damaging (0.02)			•
137	19	1054074	rs142601420	Exon 26	c.3542 C>T	NA	0.0154	0.05	<1.00	NA	P1181L	Moderately Conserved (98)	Benign (0.000)	Tolerated (0.31)			•
138	19	1054100	rs148143428	Exon 26	c.3568 G>A	NA	0.0154	NA	NA	NA	G1190R	Moderately Radical (125)	Benign (0.012)	Tolerated (0.53)			•
139	19	1054196	rs140753161	Exon 27	c.3582 G>A	NA	0.0231	NA	NA	NA	G1194G			Tolerated (0.42)			•
140	19	1054232	rs150120455	Exon 27	c.3618 C>T	NA	0.0077	NA	NA	NA	D1206D			Tolerated (0.80)			•
141	19	1054255	rs201060968	Exon 27	c.3641 G>A	NA	0.0155	NA	NA	NA	W1214*	STOP CODON					•
142	19	1054315	rs374221840	Exon 27	c.3701 G>A	NA	0.0159	NA	NA	NA	R1234H	Conserved (29)	Probably Damaging (1.000)	Damaging (0.00)			•
143	19	1054572	rs138363826	Exon 28	c.3730 G>A	NA	0.0846	NA	NA	NA	V1244M	Conserved (21)	Probably Damaging (0.985)	Damaging (0.01)			•
144	19	1054624	rs368194702	Exon 28	c.3782 C>T	NA	0.0077	NA	NA	NA	P1261L	Moderately Conserved (98)	Probably Damaging (0.997)	Tolerated (0.17)			•
145	19	1054631	rs201657292	Exon 28	c.3789 C>T	NA	0.0077	0.05	<1.00	NA	F1263F			Tolerated (0.85)			•
146	19	1054642	rs146191555	Exon 28	c.3800 C>T	NA	0.0154	NA	NA	NA	P1267L	Moderately Conserved (98)	Probably Damaging (1.000)	Damaging (0.02)			•
147	19	1054652	rs202174458	Exon 28	c.3810 G>T	NA	0.0154	NA	NA	NA	R1270R			Tolerated (0.50)			•
148	19	1054670	rs374686104	Exon 28	c.3828 C>G	NA	0.0077	NA	NA	NA	Y1276*	STOP CODON					•
149	19	1054678	rs369855340	Exon 28	c.3836 A>G	NA	0.0077	NA	NA	NA	Q1279R	Conserved (43)	Possibly Damaging (0.923)	Tolerated (0.54)			•
150	19	1054811	rs374965623	Exon 29	c.3884 G>A	NA	0.0077	NA	NA	NA	R1295Q	Conserved (43)	Benign (0.226)	Tolerated (0.57)			•
151	19	1054818	rs368024653	Exon 29	c.3891 C>T	NA	0.0077	NA	NA	NA	L1297L			Tolerated (1.00)			•
152	19	1055101	rs371279503	Exon 30	c.3956 C>T	NA	0.0231	NA	NA	NA	S1319L	Moderately Radical (145)	Benign (0.006)	Tolerated (0.32)			•
153	19	1055191	rs3745842	Exon 30	c.4046 G>A	NA	40.6467	39.50	39.00	38.375	R1349Q	Conserved (43)	Benign (0.002)	Tolerated (0.546)	•		•
154	19	1055248	rs139811629	Exon 30	c.4103 C>T	NA	0.0155	NA	NA	NA	P1368L	Moderately Conservative (98)	Probably Damaging (0.995)	Tolerated (0.59)			•
155	19	1055313	rs112719199	Exon 30	c.4168 T>C	NA	0.0158	NA	NA	NA	F1390L	Conservative (22)	Probably Damaging (0.974)	Tolerated (0.09)			•
156	19	1055333	rs141845184	Exon 30	c.4188 G>A	NA	0.0082	0.10	<1.00	NA	P1396P			Tolerated (0.52)			•
157	19	1055344	rs145321888	Exon 30	c.4199 G>A	NA	0.0666	NA	NA	NA	R1400H	Conserved (29)	Benign (0.063)	Tolerated (0.11)			•
158	19	1055907	-	Exon 31	c.4208del1	NA	0.0639	NA	NA	NA	L1403Rfs*7	FRAMESHIFT					•
159	19	1056065	rs881768	Exon 32	c.4239 A>G	45.00	45.9372	43.80	44.00	24.75	R1413R			Tolerated (1)	•		•

						MAF						Affect Prediction					
	CHR	GENOMIC POSITION	dbSNP ID	GENE LOCATION	cDNA	HapMap (European) (%)	EVS (All) (%)	dbSNP (%)	1000Genomes (All) (%)	CRISP (%)	PROTEIN	Grantham Score	Polyphen2	SIFT	NGS	GWAS	EVS
160	19	1056080	rs148214655	Exon 32	c.4254 G>T	NA	0.0077	0.05	<1.00	NA	S1418S			Tolerated (1.00)			•
161	19	1056109	rs145232000	Exon 32	c.4283 C>T	NA	0.0154	0.05	<1.00	NA	S1428L	Moderately Radical (145)	Benign (0.403)	Tolerated (0.28)			•
162	19	1056110	rs371395251	Exon 32	c.4284 G>A	NA	0.0154	NA	NA	NA	S1428S			Tolerated (1.00)			•
163	19	1056126	rs137888610	Exon 32	c.4300 C>T	NA	0.0385	NA	NA	NA	R1434C	Radical (180)	Probably Damaging (0.991)	Damaging (0.04)			•
164	19	1056148	rs142617105	Exon 32	c.4322 C>T	NA	0.0154	NA	NA	NA	A1441V	Moderately Conserved (64)	Benign (0.029)	Tolerated (0.35)			•
165	19	1056169	rs138567176	Exon 32	c.4343 G>A	NA	0.0461	0.05	<1.00	NA	G1448D	Moderately Conserved (94)	Possibly Damaging (0.494)	Tolerated (0.68)			•
166	19	1056171	rs372575675	Exon 32	c.4345 G>A	NA	0.0077	NA	NA	NA	G1449R	Moderately Radical (125)	Benign (0.064)	Tolerated (0.54)			•
167	19	1056180	rs149293300	Exon 32	c.4354 G>A	NA	0.0077	NA	NA	NA	D1452N	Conserved (23)	Benign (0.424)	Tolerated (0.44)			•
168	19	1056183	rs144562001	Exon 32	c.4357 C>T	NA	0.0615	0.05	<1.00	NA	R1453C	Radical (180)	Probably Damaging (0.991)	Damaging (0.05)			•
169	19	1056184	rs140809830	Exon 32	c.4358 G>A	NA	0.0077	NA	NA	NA	R1453H	Moderately Radical (125)	Benign (0.036)	Tolerated (0.20)			•
170	19	1056227	rs111322482	Exon 32	c.4401 T>C	NA	1.8761	1.80	2.00	NA	A1467A			Tolerated (0.66)			•
171	19	1056237	rs375977422	Exon 32	c.4411 C>A	NA	0.0077	NA	NA	NA	L1471I	Conserved (5)	Benign (0.213)	Tolerated (0.62)			•
172	19	1056349	rs370156716	Exon 33	c.4437 C>T	NA	0.0077	NA	NA	NA	G1479G			Tolerated (0.84)			•
173	19	1056378	rs151112873	Exon 33	c.4466 G>A	NA	0.0231	NA	NA	NA	R1489Q	Conserved (43)	Possibly Damaging (0.896)	Tolerated (0.44)			•
174	19	1056399	rs141113429	Exon 33	c.4487 G>A	NA	0.0154	NA	NA	NA	R1496H	Conserved (29)	Benign (0.409)	Tolerated (0.07)			•
175	19	1056410	rs201484969	Exon 33	c.4498 C>T	NA	0.0077	NA	NA	NA	P1500S	Moderately Conserved (74)	Possibly Damaging (0.743)	Tolerated (0.85)			•
176	19	1056421	rs113711363	Exon 33	c.4509 G>A	NA	1.9837	1.80	2.00	NA	P1503P			Tolerated (0.32)			•
177	19	1056426	rs113269196	Exon 33	c.4514 G>A	NA	0.7919	0.60	1.00	NA	R1505H	Conserved (29)	Benign (0.016)	Tolerated (0.18)			•
178	19	1056431	rs192694824	Exon 33	c.4519 G>A	NA	0.0154	0.05	<1.00	NA	A1507T	Moderately Conserved (58)	Benign (0.064)	Tolerated (0.60)			•
179	19	1056469	rs369037944	Exon 33	c.4557 C>T	NA	0.0231	NA	NA	NA	L1519L			Tolerated (1.00)			•
180	19	1056492	rs3752246	Exon 33	c.4580 G>C	18.30	12.7884	17.60	18.00	77.2	G1527A	Moderately Conservative (60)	Benign (0)	Tolerated (1)	•	•	•
181	19	1056918	rs267605268	Exon 34	c.4599 G>A	NA	0.0692	NA	NA	NA	S1533S			Tolerated (0.68)			•
182	19	1056931	rs140816246	Exon 34	c.4612 G>A	NA	0.0154	NA	NA	NA	V1538I	Conserved (29)	Benign (0.115)	Tolerated (0.38)			•
183	19	1056941	rs145632609	Exon 34	c.4622 G>A	NA	0.0077	0.05	<1.00	0.663	C1541Y	Radical (194)	Possibly Damaging (0.774)	Deleterious (0.002)	•		•
184	19	1056958	-	Exon 34	c.4639 T>C	NA	NA	NA	NA	0.663	S1547P	Moderately Conservative (74)	Probably Damaging (1)	Deleterious (0)	•		
185	19	1056969	rs371808878	Exon 34	c.4650 G>A	NA	NA	NA	NA	0.0077	P1550P			Tolerated (1.00)			•
186	19	1057010	rs143045355	Exon 34	c.4691 G>A	NA	0.0077	NA	NA	NA	R1564Q	Conserved (43)	Possibly Damaging (0.472)	Tolerated (0.11)			•
187	19	1057047	rs375075592	Exon 34	c.4728 C>A	NA	0.0154	NA	NA	NA	P1576P			Tolerated (0.54)			•
188	19	1057056	rs148266574	Exon 34	c.4737 C>G	NA	0.0077	NA	NA	NA	Y1579*	STOP CODON					•
189	19	1057343	rs117187003	Exon 35	c.4795 G>A	NA	0.2691	0.20	<1.00	0.638	V1599M	Conservative (21)	Probably Damaging (0.987)	Deleterious (0.001)	•		•
190	19	1057928	rs143083561	Exon 36	c.4895 C>T	NA	0.0692	NA	NA	NA	P1632L	Moderately Conservative (98)	Probably Damaging (0.999)	Tolerated (0.21)			•
191	19	1057929	rs374606460	Exon 36	c.4896 G>A	NA	0.0077	NA	NA	NA	P1632P			Tolerated (1.00)			•
192	19	1057946	-	Exon 36	c.4914_4916del3	NA	0.1519	NA	NA	NA	S1638_F1639delinsS			Damaging (0.00)			•
193	19	1057960	rs141237099	Exon 36	c.4927 G>A	NA	0.0308	0.20	<1.00	NA	V1643M	Conserved (21)	Probably Damaging (1.000)	Tolerated (0.10)			•
194	19	1057975	rs371036349	Exon 36	c.4942 T>C	NA	0.0077	NA	NA	NA	Y1648H	Moderately Conserved (83)	Probably Damaging (1.000)	Tolerated (0.22)			•
195	19	1057995	rs372378216	Exon 36	c.4962 A>G	NA	0.0077	NA	NA	NA	I1654M	Conserved (10)	Possibly Damaging (0.949)	Damaging (0.05)			•
196	19	1058011	rs368473239	Exon 36	c.4978 A>G	NA	0.0077	NA	NA	NA	I1660V	Conserved (29)	Possibly Damaging (0.771)	Tolerated (0.42)			•
197	19	1058026	-	Exon 36	c.4994del1	NA	0.04	NA	NA	NA	T1666Pfs*15	FRAMESHIFT					•
198	19	1058176	rs4147918	Exon 37	c.5057 A>G	4.50	3.014	5.60	6.00	1.638	Q1686R	Conservative (43)	Benign (0.001)	Tolerated (0.234)	•		•
199	19	1058178	rs368743296	Exon 37	c.5059 G>A	NA	0.0077	NA	NA	NA	V1687I	Conserved (29)	Benign (0.333)	Tolerated (0.67)			•
200	19	1058232	rs148795173	Exon 37	c.5113 C>T	NA	0.0154	NA	NA	NA	R1705W	Moderately Radical (101)	Probably Damaging (0.998)	Damaging (0.04)			•
201	19	1058640	-	Exon 38	c.5173_5174insA	NA	0.008	NA	NA	NA	L1725Hfs*55	FRAMESHIFT					•
202	19	1058673	rs370760111	Exon 38	c.5206 G>T	NA	0.0077	NA	NA	NA	A1736S	Moderately Conserved (99)	Possibly Damaging (0.571)	Tolerated (0.40)			•
203	19	1058685	rs146448899	Exon 38	c.5218 C>T	NA	0.0077	NA	NA	NA	Q1740*	STOP CODON					•
204	19	1058692	rs374830215	Exon 38	c.5225 C>T	NA	0.0077	NA	NA	NA	P1742L	Moderately Conserved (98)	Benign (0.274)	Tolerated (0.40)			•
205	19	1058839	rs140933796	Exon 39	c.5300 C>T	NA	0.0077	NA	NA	NA	P1767L	Moderately Conserved (98)	Benign (0.00)	Tolerated (0.12)			•
206	19	1058848	rs139765732	Exon 39	c.5309 G>A	NA	0.3076	0.32	<1.00	NA	G1770E	Moderately Conserved (98)	Benign (0.034)	Tolerated (1.00)			•
207	19	1058872	rs147558402	Exon 39	c.5333 G>A	NA	0.0077	0.05	<1.00	NA	R1778H	Conserved (29)	Benign (0.018)	Tolerated (0.08)			•
208	19	1058883	rs372373771	Exon 39	c.5344 C>T	NA	0.0231	NA	NA	NA	R1782W	Moderately Radical (101)	Probably Damaging (1.000)	Damaging (0.00)			•
209	19	1058937	rs368523491	Exon 39	c.5398 A>G	NA	0.0077	NA	NA	NA	K1800E	Moderately Conserved (56)	Probably Damaging (1.000)	Damaging (0.00)			•
210	19	1059029	rs143615723	Exon 40	c.5408 G>A	NA	0.0154	NA	NA	NA	R1803H	Conserved (29)	Benign (0.016)	Tolerated (0.14)			•
211	19	1059047	rs368737570	Exon 40	c.5426 C>A	NA	0.0077	NA	NA	NA	A1809D	Moderately Radical (126)	Probably Damaging (1.000)	Damaging (0.00)			•
212	19	1059056	rs114782266	Exon 40	c.5435 G>A	NA	2.6838	2.70	3.00	2.963	R1812H	Conservative (29)	Benign (0.108)	Deleterious (0.017)	•		•
213	19	1059079	rs143614132	Exon 40	c.5458 G>A	NA	0.0384	NA	NA	NA	G1820S	Moderately Conserved (56)	Probably Damaging (0.994)	Tolerated (0.07)			•

						MAF						Affect Prediction					
	CHR	GENOMIC POSITION	dbSNP ID	GENE LOCATION	cDNA	HapMap (European) (%)	EVS (All) (%)	dbSNP (%)	1000Genomes (All) (%)	CRISP (%)	PROTEIN	Grantham Score	Polyphen2	SIFT	NGS	GWAS	EVS
214	19	1061804	rs78320196	Exon 41	c.5487 T>C	NA	3.9142	4.30	4.00	2.2	N1829N			Tolerated (1)	•		•
215	19	1061813	-	Exon 41	c.5497del1	NA	1.9652	NA	NA	NA	K1833Rfs*	FRAMESHIFT					•
216	19	1061878	rs374577064	Exon 41	c.5561 C>A	NA	0.0077	NA	NA	NA	A1854E	Moderately Radical (107)	Benign (0.011)	Tolerated (1.00)			•
217	19	1062192	rs4147921	Exon 42	c.5592 T>C	NA	2.9755	5.60	6.00	1.6	A1864A			Tolerated (0.501)	•		•
218	19	1062198	rs368446227	Exon 42	c.5598 C>T	NA	0.0077	NA	NA	NA	H1866H			Tolerated (0.84)			•
219	19	1062234	rs372142602	Exon 42	c.5634 C>G	NA	0.0077	NA	NA	NA	I1878M	Conserved (10)	Possibly Damaging (0.952)	Damaging (0.01)			•
220	19	1062254	rs368864109	Exon 42	c.5654 G>A	NA	0.0077	NA	NA	NA	R1885H	Moderately Radical (103)	Probably Damaging (1.000)	Tolerated (0.07)			•
221	19	1062276	rs146398172	Exon 42	c.5676 G>A	NA	0.0077	NA	NA	NA	A1892A			Tolerated (1.00)			•
222	19	1063546	rs375389773	Exon 43	c.5716 G>A	NA	0.0077	NA	NA	NA	A1906T	Moderately Conserved (58)	Possibly Damaging (0.633)	Tolerated (0.25)			•
223	19	1063562	rs146619749	Exon 43	c.5732 C>A	NA	0.0077	NA	NA	NA	A1911E	Moderately Radical (107)	Benign (0.017)	Tolerated (0.59)			•
224	19	1063563	rs369378657	Exon 43	c.5733 G>A	NA	0.0077	NA	NA	NA	A1911A			Tolerated (0.42)			•
225	19	1063580	rs376845552	Exon 43	c.5150 G>T	NA	0.0077	NA	NA	NA	W1917L	Moderately Conserved (61)	Benign (0.001)	Tolerated (0.15)			•
226	19	1063584	rs201050805	Exon 43	c.5754 C>T	NA	0.0077	0.05	<1.00	NA	Y1918Y			Tolerated (1.00)			•
227	19	1063612	rs148635111	Exon 43	c.5782 G>A	NA	0.0077	NA	NA	NA	G1928R	Moderately Radical (125)	Probably Damaging (1.000)	Damaging (0.00)			•
228	19	1063624	rs114787084	Exon 43	c.5794 C>T	NA	0.0617	0.09	<1.00	NA	R1932C	Radical (180)	Probably Damaging (1.000)	Damaging (0.00)			•
229	19	1063651	rs151130083	Exon 43	c.5821 G>A	NA	0.0077	0.05	<1.00	NA	V1941I	Conserved (29)	Benign (0.064)	Tolerated (1.00)			•
230	19	1063775	rs373283316	Exon 44	c.5864 G>T	NA	0.008	NA	NA	NA	G1955V	Moderately Radical (109)	Probably Damaging (1.000)	Damaging (0.00)			•
231	19	1064193	rs4147930	Exon 45	c.5985 G>A	0.00	29.5659	36.20	36.00	73.6	L1995L			Tolerated (1)	•		•
232	19	1064218	rs142315228	Exon 45	c.6010 C>T	NA	0.0078	NA	NA	NA	R2004C	Radical (180)	Benign (0.137)	Damaging (0.02)			•
233	19	1065018	rs4147934	Exon 46	c.6133 G>T	NA	25.0256	36.80	37.00	77.338	A2045S	Moderately Conservative (99)	Benign (0.057)	Tolerated (0.962)	•		•
234	19	1065044	rs4147935	Exon 46	c.6159 C>T	NA	26.1408	23.30	23.00	44.725	G2053G			Tolerated (0.133)	•		•
235	19	1065051	rs372768841	Exon 46	c.6166 C>T	NA	0.0081	NA	NA	NA	R2056C	Radical (180)	Probably Damaging (0.978)	Tolerated (0.09)			•
236	19	1065132	rs376353179	Exon 46	c.6247 G>A	NA	0.0077	NA	NA	NA	V2083M	Conserved (21)	Benign (0.420)	Tolerated (0.09)			•
237	19	1065305	rs139706726	Exon 47	c.6233 G>A	NA	0.0154	0.05	<1.00	NA	E2108K	Moderately Conserved (56)	Benign (0.007)	Tolerated (0.62)			•
238	19	1065320	-	Exon 47	c.6338_6340del3	NA	0.3116	NA	NA	NA	Q2113_K2114delinsQ	DELETION/INSERTION					•
239	19	1065409	rs138839714	Exon 47	c.6426 C>G	NA	0.0077	NA	NA	NA	A2142A			Tolerated (0.47)			•
240	19	1065563	rs2242437	Exon 47	c.6580 G>C	29.30	NA	36.40	36.00	75.863	-	-	-	-	•		•

Appendix B - URLs Used

All URLs utilised during this project. The URL itself, as well as a description of what it was used for is presented.

Appendix B

All URLs utilised during this project. The URL itself, as well as a description of what it was used for is presented. If any of these tools are not mentioned specifically in this thesis

Name	Description	URL
1000Genomes	1000Genomes database - a deep catalogue of human genetic variation	http://browser.1000genomes.org/index.html
ABCMdb	The ABC mutations database contains mentions of ABC protein mutations extracted from the literature using automated data mining method and also annotated infromation on variants from other databases and sources.	http://abcmutations.hegelab.org/
Alzheimer's Association	Alzheimer's Association website for up-to-date statistics on AD	www.alz.org
Alzheimer's Research UK	ARUK website for information on AD specific to the UK population	www.alzheimersresearchuk.org
Berkeley Drosophila Genome Project (BDGP)	Annotation tool to identify features of DNA sequence	http://www.fruitfly.org/
BioGPS	Gene annotation portal to identify gene and protein functions	http://biogps.org/#goto=genereport&id=10347
Clustalw2	Multiple alignment sequence programme for protein and DNA sequence	https://www.ebi.ac.uk/Tools/msa/clustalo/
dbSNP	Variant annotation within the genome	https://www.ncbi.nlm.nih.gov/projects/SNP/
ENCODE	Comprehensive list of functional elements within the human genome	https://www.encodeproject.org/
Ensembl	Genome browser for vertebrate genomes	http://www.ensembl.org
ESE Finder Release 3.0	Web-based resource to identify putative ESEs responsive to the human SR proteins	http://krainer01.cshl.edu/cgi-bin/tools/ESE3/esefinder.cgi?process=home
Exome Variant Server (EVS)	Data from NGS of the protein coding regions of the human genome across diverse populations as part of the NHLBI GO Exome Sequencing Project	http://evs.gs.washington.edu/EVS/
GTEx	Computed eQTL results associated with individual genotype, expression and clinical data	http://gtexportal.org/home/
HaploReg	Tool for exploring annotations of the noncoding genome at variants on haplotype blocks	http://archive.broadinstitute.org/mammals/haploreg/haploreg.php
HaploView	Programme to analyse LD and haplotype associations within the 1000Genomes data	https://www.broadinstitute.org/haploview/haploview
HapMap (now retired)	Haplotype map of the human genomes	https://www.ncbi.nlm.nih.gov/variation/news/NCBI_retiring_HapMap/
Human Splicing Finder 3.0	Algorithmic approach to identify potential splice sites and branch points within the human genome	http://www.umd.be/HSF3/
MatInspector	Applies matrix descriptions to locate transcription factor binding sites within DNA sequences	https://www.genomatix.de/online_help/help_matinspector/matinspector_help.html
National Center for Biotechnology Information (NCBI)	Open access genomic information	http://www.ncbi.nlm.nih.gov
NHS Alzheimer's Disease Information	NHS Choices information of Alzheimer's diseases, targeted at educating the public interested in the condition	http://www.nhs.uk/Conditions/Alzheimers-disease/Pages/Introduction.aspx
NICE Dementia Recommendations	National Institute for Clinical Excellence recommendations for dementia conditions including therapeutic interventions and diagnostic pathways	https://www.nice.org.uk/guidance/conditions-and-diseases/mental-health-and-behavioural-conditions/dementia
PLINK (v1.07)	Open-source whole genome association analysis toolset	http://pngu.mgh.harvard.edu/purcell/plink/
PolyPhen-2	Polymorphism Phenotyping v2 predicts possible impact of amino acid substitution on the structure and function of human proteins	http://genetics.bwh.harvard.edu/pph2/
Primer3	Source forge hosted programme to design PCR primers	http://bioinfo.ut.ee/primer3/
Promega	Promega's manufacture's website to access protocols and vector reference sequences	https://www.promega.co.uk/
QUANTO	Program to calculate required sample size and power for genetic studies	http://biostats.usc.edu/Quanto.html
RegulomeDB	Identify DNA features and regulatory elements in non-coding regions of the human genome	http://www.regulomedb.org/
SeattleSeq Variant Annotation 141	Annotation of SNPs by pooling data from a number of sources, provides Grantham Score	http://snp.gs.washington.edu/SeattleSeqAnnotation141/
SIFT	Prediction programme for whether an amino acid substitution affects protein function	http://sift.jcvi.org/
SNPCheck3	Tool for checking for the presence of SNPs in predicted PCR primer binding sites	https://secure.ngrl.org.uk/SNPCheck/snpcheck.htm;jsessionid=169B9842F7E80DE3345BAD2C8137A544
tabix	Generic tool that indexes position sorted files in TAB-delimited formats	https://sourceforge.net/projects/samtools/files/tabix/
TMHH Server v. 2.0	Prediction of transmembrane helices in proteins	http://www.cbs.dtu.dk/services/TMHMM/
TOPO2	Transmembrane protein display software	http://www.sacs.ucsf.edu/TOPO2/
UCSC Genome Browser	On-line genome browser hosted by the University of California, Santa Cruz	http://genome-euro.ucsc.edu
UCSC in-silico PCR Programme	*In-Silico* PCR searches a sequence database with a pair of PCR primers	http://rohsdb.cmb.usc.edu/GBshape/cgi-bin/hgPcr
UniProt	Comprehensive resource of protein sequence and functional information	http://www.uniprot.org/uniprot/Q8IZY2
vcftools	Program package designed for working with Variant Call Files files	http://vcftools.sourceforge.net/
Webcutter 2.0	On-line tool for restriction mapping nucleotide sequences	http://rna.lundberg.gu.se/cutter2/