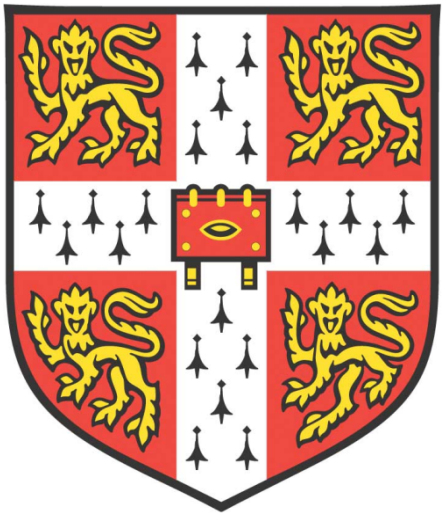
THE ROLE OF GERMLINE AND SOMATIC NUCLEAR AND MITOCHONDRIAL DNA VARIATION IN NEURODEGENERATIVE DISORDERS



Michael John Keogh
PhD candidate, Kings College

MRC Mitochondrial Biology Unit
University of Cambridge
June 2018

This dissertation is submitted to the Board of Graduate Studies in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Abstract

The role of germline and somatic nuclear and mitochondrial DNA variation in neurodegenerative disorders

Dr Michael J Keogh

Neurodegenerative disorders are a group of age-related conditions resulting in neuronal cell death and protein accumulation. It is estimated that around 5-10% of these cases are genetically mediated. Most commonly this is by pathogenic single nuclear variants (SNVs), though combinations of rare variants (termed oligogenic variation), copy-number variation (CNVs), somatic mutations in nuclear DNA, and somatically acquired mitochondrial DNA variants have all been hypothesised to increase disease risk or cause disease.

Firstly, using a combination of exome sequencing and array genotyping on 1511 *post-mortem* brain samples within the MRC Brain Bank, we detected 61 monogenic cases of disease, 349 brains carrying disease risk factors, and identified that variants in *GRN* and *PRPH* may increase the risk of developing dementia with lewy bodies (DLB) and Alzheimer's disease (AD) respectively. Secondly, we detected a previously unknown systematic bias in the interpretation of oligogenic interactions with implications for our understanding of disease mechanisms and coexistent clinical diagnostic utility. Thirdly, we detected a novel copy-number gain in *LAMA5* associated with Creutzfeldt-Jakob disease (CJD), and fourthly, we determine that at least 1% of the population carry high level somatic protein-coding mutations affecting at least 10% of cells within the brain. Subsequently, additional focussed deep-sequencing studies revealed that several regions of the brain are likely to contain clones of low-level somatic mutations that are pathogenic when present in the germline, and that age-related clonal mutations that arise in blood are present at high levels within the aging brain and are associated with Lewy Body pathology. Finally, using transgenic mice that over express human α-synuclein and which either accrue or transmit mtDNA mutations, we determine that the presence of mtDNA mutations exacerbate some phenotypic traits of Lewy body disorders, and may reduce the volume of critical neuroanatomical brain regions whilst paradoxically reducing α-synuclein accumulation.

Taken together, these data enable the first genetically stratified brain tissue resource in the UK, describe new disease genetic risk factors (both SNVs and CNVs) for neurodegenerative disorders, and also help define the somatic genetic architecture of the human brain. In addition, we describe the *in vivo* interaction between mutations in the mitochondrial genome and a progressive neurodegenerative disorder in mice.

Declaration

This dissertation is the result of my own work and includes nothing which is the outcome of work done in collaboration except where specifically indicated in the text. It has not been previously submitted, in part or whole, to any university of institution for any degree, diploma, or other qualification. In accordance with the Biology Degree Committee guidelines, this thesis is does not exceed 60,000 words.

Signed: _____

Date: _____

Michael Keogh BMedSci (Hons), BMBS (Hons), MSc, MClinRes, MRCP (UK).

PhD Student, Department of Clinical Neuroscience, Kings College, University of Cambridge.

Publications arising from this thesis

A novel de novo STXBP1 mutation is associated with mitochondrial complex I deficiency and late-onset juvenile-onset parkinsonism. Keogh MJ, Daud D, Pyle A, Duff J, Griffin H, He L, Alston CL, Steele H, Taggart S, Basu AP, Taylor RW, Horvath R, Ramesh V, Chinnery PF. Neurogenetics. 2015 Jan;16(1):65-7. PMID: 25418441

Clinical heterogeneity of primary familial brain calcification due to a novel mutation in PDGFB. Keogh MJ, Pyle A, Daud D, Griffin H, Douroudis K, Eglon G, Miller J, Horvath R, Chinnery PF. Neurology. 2015 Apr 28;84(17):1818-20. PMID: 25832657

Reply: Evaluation of exome sequencing variation in undiagnosed ataxias. Pyle A, Griffin H, **Keogh MJ**, Horvath R, Chinnery PF. Brain. 2015 Apr 4. PMID: 25842392

Frequency of rare recessive mutations in unexplained late onset cerebellar ataxia. Keogh MJ, Steele H, Douroudis K, Pyle A, Duff J, Hussain R, Smertenko T, Griffin H, Santibanez-Koref M, Horvath R, Chinnery PF. J Neurol. 2015 May 16. PMID: 25976027

Mitochondrial DNA mutations in neurodegeneration. Keogh MJ, Chinnery PF. Biochim Biophys Acta. 2015 May 23. PMID: 26014345

Voxel based analysis in neuroferritinopathy expands the phenotype and determines radiological correlates of disease severity. Keogh MJ , Aribisala BS , He J , Tulip E , Butteriss D , Morris C , Gorman G, Horvath R , Chinnery PF, Blamire AM. J Neurol. 2015 Oct;262(10):2232-40

SCP2 mutations and neurodegeneration with brain iron accumulation. Horvath R, Lewis-Smith D, Douroudis K, Duff J, **Keogh MJ,** Pyle A, Fletcher N, Chinnery PF. Neurology. 2015 Nov 24;85(21):1909-11

Prevalence of neurogenetic disorders in the North of England. Bargiela D, Yu-Wai-Man P, **Keogh M**, Horvath R, Chinnery PF.Neurology. 2015 Oct 6;85(14):1195-201

SVI: a simple single-nucleotide Human Variant Interpretation tool for Clinical Use. Missier P, Wijaya E, Kirby R, **Keogh M.** Volume 9162 of the series Lecture Notes in Computer Science pp 180-194

Exome Sequencing in Dementia with Lewy bodies. Keogh MJ, Kurzawa-Akanbi M, Griffin H, Douroudis K, Ayers KL, Hussein RI, Hudson G, Pyle A, Cordell HJ, Attems J, McKeith IG, O'Brien JT, Burn DJ, Morris CM, Thomas AJ, Chinnery PF.Transl Psychiatry. 2016 Feb 2;6:e728.

HDBR Expression: A Unique Resource for Global and Individual Gene Expression Studies during Early Human Brain Development. Lindsay SJ, Xu Y, Lisgo SN, Harkin LF, Copp AJ, Gerrelli D, Clowry GJ, Talbot A, **Keogh MJ,** Coxhead J, Santibanez-Koref M, Chinnery PF. Front Neuroanat. 2016 Oct 26;10:86. PMID: 27833533

Genetic compendium of 1511 human brains available through the UK Medical Research Council Brain Banks Network Resource. Keogh MJ, Wei W, Wilson I, Coxhead J, Ryan S, Rollinson S, Griffin H, Kurzawa-Akanbi M, Santibanez-Koref M, Talbot K, Turner MR, McKenzie CA, Troakes C, Attems J, Smith C, Al Sarraj S, Morris CM, Ansorge O, Pickering-Brown S, Ironside JW, Chinnery PF. Genome Res. 2017 Jan;27(1):165-173.

Mitochondrial DNA point mutations and relative copy number in 1363 disease and control human brains. Wei W, **Keogh MJ**, Wilson I, Coxhead J, Ryan S, Rollinson S, Griffin H, Kurzawa-Akinibi M, Santibanez-Koref M, Talbot K, Turner MR, McKenzie CA, Troakes C, Attems J, Smith C, Al Sarraj S, Morris CM, Ansorge O, Pickering-Brown S, Ironside JW, Chinnery PF. Acta Neuropathol Commun. 2017 Feb 2;5(1):13

PLP1 mutations and central demyelination: Evidence from electrophysiologic phenotyping in female manifesting carriers. Keogh MJ, Jaiser SR, Steele HE, Horvath R, Chinnery PF, Baker MR. Neurology Clinical Practice. 2017. In Press

Oligogenic genetic variation in 980 post mortem human brains. MJ Keogh, W Wei, J Aryaman, Wilson I, K Talbot, MR Turner, C-A McKenzie, C Troakes, J Attems, C Smith, S Al Sarraj, CM Morris, O Ansorge, S Pickering-Brown, N Jones, JW Ironside, PF Chinnery. Journal of Neurology Neurosurgery and Psychiatry (JNNP). 2017. In Press.

Acknowledgments

I would like to thank Professor Patrick Chinnery for the opportunity to undertake this work within his laboratory and for his help and guidance over the last few years. I would also like to thank several people who have provided technical help and assistance, primarily Dr Wei Wei at the University of Cambridge, without whom this work would not have been possible, and Dr Chris Morris and Dr Ian Wilson at the University of Newcastle for their technical expertise. I would also like to thank all the members of Professor Chinnery's lab in both Newcastle and Cambridge for putting up with me over the last few years, and all of the staff at each of the Medical Research Council (MRC) Brain Banks who provided DNA and tissue samples together with sample metadata, and for which I am forever indebted.

I would also like to thank The Wellcome Trust for awarding me the funding to undertake this Clinical Research Fellowship, and finally my parents for all their help and support over the years; I promise this is the last degree….

Table of Contents

List of Figures

List of Tables

Nomenclature

Acronym	Definition
ACMG	American College of Medical Genetics
AD	Alzheimer's Disease
ALS	Amyotrophic Lateral Sclerosis
BAF	Beta Allele Frequency
BAM	Binary Alignment Map
BSE	Bovine Spongiform Encephalopathy
BWA	Burrows Wheeler Alignment
C57/bl6	C57/black 6
CBD	Corticobasal Degeneration
CJD	Creutzfeldt-Jakob Disease
CNS	Central Nervous System
CNV	Copy Number Variation
COX	Cytochrome c Oxidase
CSF	Cerebrospinal Fluid
DLB	Dementia with Lewy Bodies
DNA	Deoxyribonucleic Acid
DRLS	Derived Log Ratio Spread
DRPLA	Dentatorubral-pallidoluysian atrophy
EOAD	Early Onset Alzhimer's Disease
ExAC	Exome Aggregation Consortium
FFI	Fatal Familial Insomnia
FSE	Fast Spin Echo
FTD	Frontotemporal Dementia
FUS	Fused in Sarcoma
GATK	Genome Analysis Tool Kit
GDP	Gross Domestic Product
GSS	Gerstmann-Straussler-Scheinker syndrome
GWAS	Genome Wide Association Study
HD	Huntington's Disease
IBD	Identity By Descent
kHz	Kilohertz
LB	Lewy Body
LBD	Lewy Body Disease
LRR	Log-R-Ratio
MAF	Minor Allele Frequency
MND	Motor Neuron Disease
MRC	Medical Research Council
MRI	Magnetic Resonance Imaging
MSA	Multiple System Atrophy
mtDNA	Mitochondrial Deoxyribonucleic Acid
NGS	Next Generation Sequencing

NORT	Novel Object Recognition Test
NPV	Negative Predictive Value
OFT	Open Field Test
OR	Odds Ratio
PCR	Polymerase Chain Reaction
PD	Parkinson's Disease
PET	Positron Emission Tomography
PLS	Primary Lateral Sclerosis
PPV	Positive Predictive Value
PrP	Prion Protein
PSP	Progressive Supranuclear Palsy
QC	Quality Control
rCRS	Revised Cambridge Reference Sequence
RNA	Ribonucleic Acid
ROS	Reactive Oxygen Species
rRNA	Ribosomal Ribonucleic Acid
RT	Room Temperature
SAM	Sequence Alignment Map
SBMA	Spinal and Bulbar Muscular Atrophy
SCA	Spinocerebellar Ataxia
SD	Standard Deviation
SEM	Standard Error of the Mean
SN	Substantia Nigra
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
SOD	Superoxide Dismutase
TDP	TAR DNA-binding Protein
TE	Echo Time
TR	Repetition Time
tRNA	Transfer Ribonucleic Acid
UPR	Unfolded Protein Response
VAF	Variant Allele Frequency
VCF	Variant Call Format
VD	Vascular Dementia
WGS	Whole Genome Sequencing

Chapter 1

Introduction

1.1 Overview of neurodegenerative disorders

Neurodegenerative disorders can be broadly considered as diseases in which selective neuronal vulnerability and degeneration occur in conjunction with deposits of abnormal proteins (Taylor, Hardy et al. 2002, Ross and Poirier 2004). Although not exclusive, several major disorders are commonly accepted to be included under the umbrella term of neurodegenerative diseases including Alzheimer's disease (AD), Parkinson's disease (PD), Frontotemporal dementia (FTD), and motor neuron disease (MND) / amyotrophic lateral sclerosis (ALS).

These diseases encompass a diverse clinical spectrum, ranging from pure cognitive syndromes (such as most cases of Alzheimer's disease (Dubois, Feldman et al. 2014)) through to isolated movement disorders (such as the Primary Lateral Sclerosis variant of ALS (Pringle, Hudson et al. 1992)), though many neurodegenerative disorders consist of a mixture of cognitive and motor features, with the clinical phenotype determined by the neurological systems and pathways affected by the disease pathology (Taipa, Pinho et al. 2012).

1.2 Epidemiology and financial burden of neurodegenerative diseases

The true prevalence of neurodegenerative disorders is difficult to establish, with most epidemiological studies focused on defining the incidence and prevalence of single diseases only, or of the broad category of 'dementia', which fails to encapsulate forms of neurodegenerative disorders in which cognitive features either have not developed or are not part of the phenotype. However, even accounting for these methodological difficulties,

estimates suggest that ~2% of the population are currently suffering from a neurodegenerative disease (Prusiner 2001).

Amongst the specific neurodegenerative disorders, Alzheimer's disease remains by far the most common, affecting approximately 10% of adults aged over 65 (Association 2015), equating to almost 5 million adults in the USA (Hebert, Scherr et al. 2003) rising to an estimated 13.2 million affected adults in the USA by 2050 (Hebert, Scherr et al. 2003). Dementia with Lewy Bodies (DLB) is the second most common neurodegenerative disorder, accounting for approximately 4% of all cases of dementia (Vann Jones 2014), with an estimated population prevalence of 0.6% in those aged over 65 years of age (Stevens, Livingston et al. 2002). Parkinson's disease is the third most common, with an overall incidence of 16-19 per 100,000 person years (Twelves, Perkins et al. 2003), rising to 160 per 100,000 person years in those aged over 65 (Hirtz, Thurman et al. 2007), and equating to 59,000 new cases of PD in the USA each year (Hirtz, Thurman et al. 2007). Motor neuron disease is less common than Alzheimer's, Parkinson's and DLB, but still has an estimated lifetime risk of 1 in 472 for women and 1 in 350 for men (Alonso, Logroscino et al. 2009). Finally, other neurodegenerative diseases are far more rare, with progressive supranuclear palsy (PSP) occurring at a rate of 3/100,000 person years (Bower, Maraganore et al. 1997), Huntington's disease (HD) at 0.7/100,000 years (Wexler, Collett et al. 2016) and Prion diseases occurring in less than 1 per million individuals per year (Holman, Belay et al. 2010).

Whilst all neurodegenerative disorders are rare in young people, they differ in incidence with age, with disorders such as Alzheimer's disease showing a progressive increase over time, where as other diseases, such as ALS show a distinct peak at the age of 75-79 years of age (Figure 1.1).

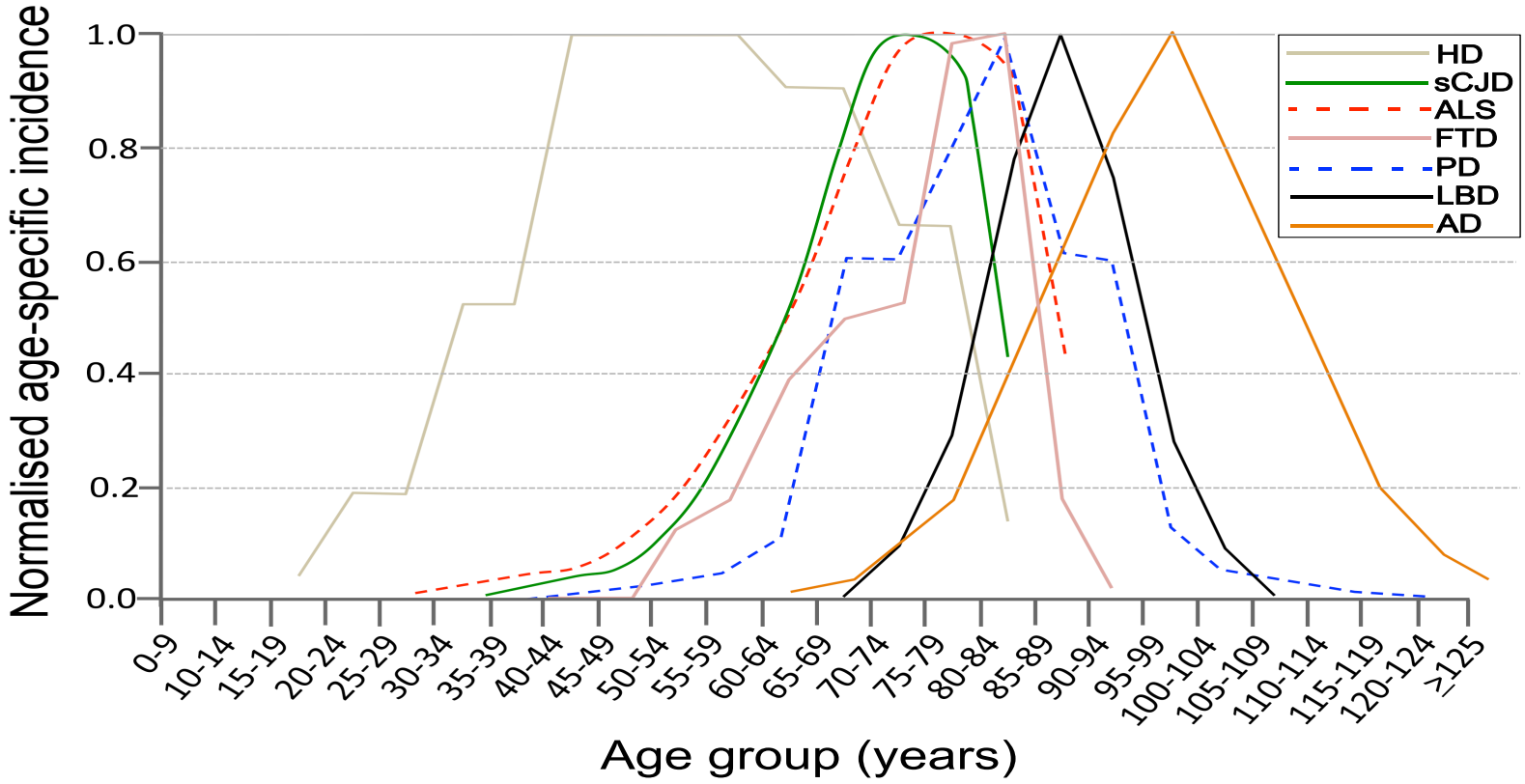


Figure 1.1. The incidence of each neurodegenerative disease by age. Each disorder can be identified by the key in the top right corner of the figure. This image is adapted from *de Pedro-Cuesta et al* (de Pedro-Cuesta, Rabano et al. 2015).

The rapid increase in life expectancy in the Western world and developing nations over the past 50 years has therefore in part led to an increased prevalence of these disorders. The increase in the number of affected individuals coupled with the increased cost of healthcare means that neurodegenerative disorders now place significant financial strains on the economies of developed and developing nations. For example, the total spending on dementia and dementia care in the UK has been estimated to be around 0.6% of the gross domestic product (GDP) (Comas-Herrera, Wittenberg et al. 2007). In the USA, current estimates are that approximately 220 billion US dollars per year are spent on caring for people with dementia (Association 2015). For individual patients, the personal financial costs can also be extensive, with personal care costs for patients with ALS averaging $32,000 dollars a year in Canada (£20,000) (Gladman, Dharamshi et al. 2014), rising to over $1.4 million dollars in the USA if home care and home ventilation is required for a long period of time (Obermann and Lyon 2015).

1.3 Major forms of neurodegenerative disorders

Whilst neurodegenerative diseases are often discussed in epidemiological studies in terms of either the major eponymous phenotypic diseases such as Alzheimer's disease or Parkinson's disease, they are not actually discrete clinical and pathological entities. Work, primarily over the last 30 years, has revealed a marked degree of phenotypic and pathological heterogeneity both within and between these ostensibly discrete conditions, resulting in the appreciation of a relative continuum of neurodegenerative disorders.

This continuum of disease is in stark contrast to initial classification systems which were based upon original isolated case reports of single cases resulting in the eponymous syndromes of Parkinson's disease, Alzheimer's disease and Pick's disease being identified in the late 19th and early 20th century (Pick 1906, Alzheimer 1907, Pearce 2003). Following

these initial descriptions, the gradual accumulation of phenotypic information and macroscopic neuropathological data enabled the appreciation of broader phenotypes and pathological findings in these disorders, beginning the era of 'neuropathological criteria' in which integrated clinical and pathological data are utilized to make a diagnosis.

The classification and diagnosis of neurodegenerative disorders took it's next major leap forward around 20 years ago with the identification of discrete specific proteins deposited within the brain of each of the major neurodegenerative disorders, for example α-synuclein deposition in Parkinson's disease (Spillantini, Schmidt et al. 1997). Neuropathological studies over the last 20 years have built on these observations by discovering additional molecular features which can further define sub-types of these disorders, many of which are genetically mediated, but clinically and pathologically previously appeared to be in keeping with sporadic cases of disease. Identifying and understanding these distinct molecular forms of disease is likely to be vital in order to understand converging molecular mechanisms that cause disease (Figure 1.2), and are likely to move neurodegenerative disorders towards a 'precision medicine' based classification system and away from either eponymous disease terminology, or grouping by the major neuropathological protein deposited (Figure 1.2).

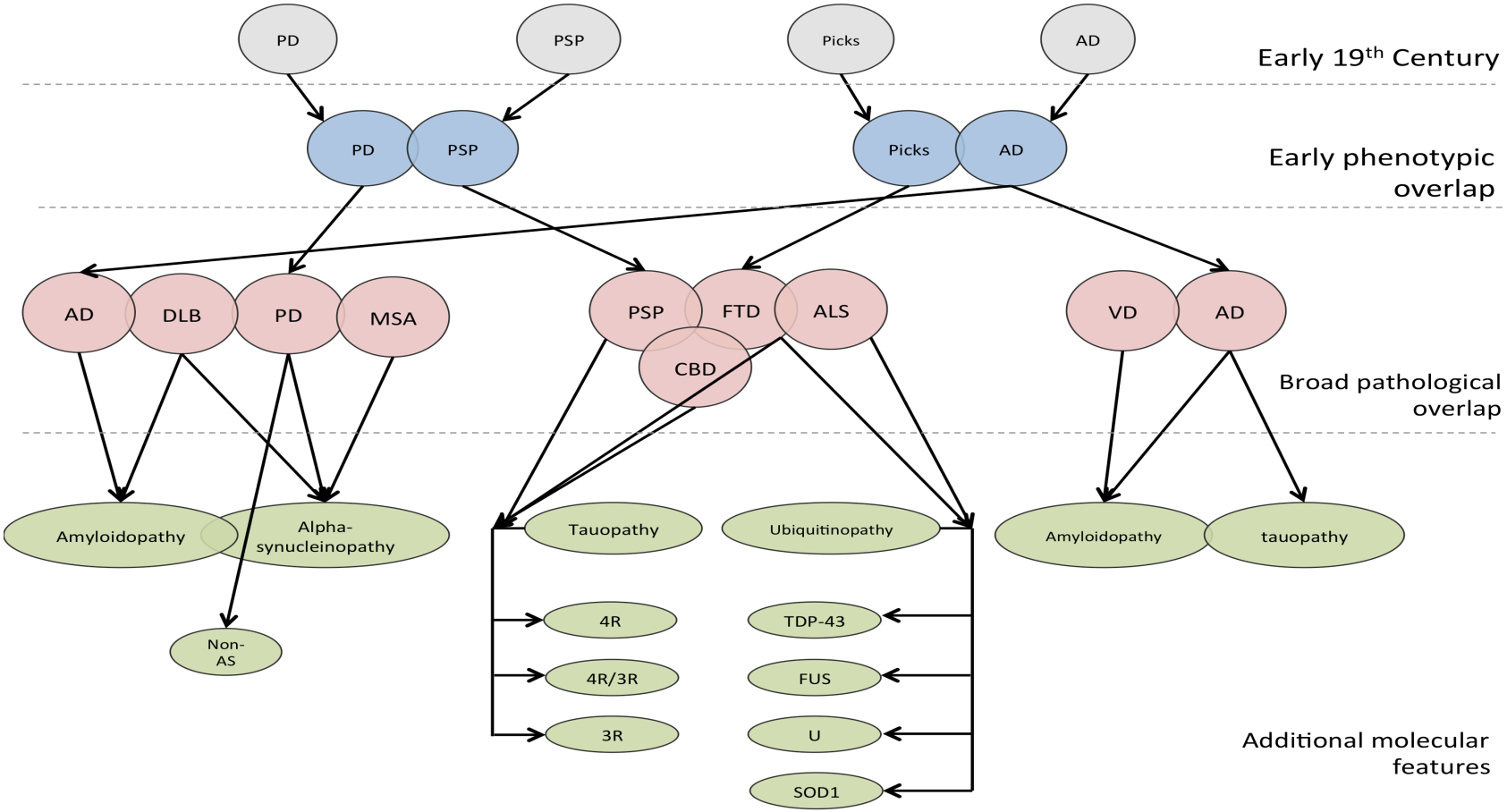


Figure 1.2. The clinical, pathological and molecular overlap between major neurodegenerative disorders. The initial clinical descriptions identified in the early 20th century are shown (grey circles). The phenotypic overlap identified following initial descriptions are shown (blue circles), and the subsequently determined pathological overlap that emerged over the next century. Finally, the recent identification of molecular pathological features that define sub-groups of these disorders are shown.

1.3.1 α-synucleinopathies – clinical features

The primary forms of α-synucleinopathy syndromes are Parkinson's disease (PD), Dementia with Lewy Bodies (DLB) and Multiple system atrophy (MSA), and each syndrome differs in both their clinical phenotype and neuropathology.

Parkinson's disease (PD): Whilst Parkinson's disease has primarily been considered as a movement disorder, it has become apparent that non-motor features of the disease are extremely common, with many beginning prior to, or early in the disease course including sleep dysfunction, hyposmia and autonomic dysfunction (Chaudhuri, Healy et al. 2006). The condition is however still largely dominated by the cardinal clinical triad of tremor, rigidity and bradykinesia (Jankovic 2008) .

Dementia with Lewy Bodies (DLB): DLB is also frequently accompanied by Parkinsonian motor features, acting as one of three core features of the disease alongside fluctuating attention and concentration and visual hallucinations (McKeith, Boeve et al. 2017). Importantly Parkinsonism can only be used for diagnosis in the presence of dementia, highlighting the very different clinical phenotype in the initial stage of disease compared to PD.

Multiple system atrophy (MSA): The predilection of α-synuclein pathology to afflict the brain stem and autonomic nervous system in MSA results in a relatively discrete phenotype within the synucleinopathies. This is characterized by autonomic failure or orthostatic hypotension, a poor response to levodopa (in contrast to PD), and, commonly, a cerebellar syndrome (which is not seen in any other α-synucleinopathy). As with all the synucleinopathies, there are no definitive diagnostic tests, but the observation of putaminal, pontine and middle cerebellar peduncle (MCP) atrophy with MRI imaging of the brain is extremely helpful in making the diagnosis (Schulz, Klockgether et al. 1994).

1.3.2 α-synucleinopathies – neuropathology

Alpha-synuclein is a 140 amino-acid protein, encoded by the *SNCA* gene, which is found abundantly in the pre-synaptic region of neurons (Maroteaux, Campanelli et al. 1988), and comprises 1% of total cytosolic protein within the brain (Stefanis 2012). Whilst it's exact function is unknown, it is highly expressed during neurodevelopment (Murphy, Rueter et al. 2000), possibly functioning as a mechanism to control neurotransmitter release in synaptic terminals (Abeliovich, Schmitz et al. 2000).

Neuropathologically, aberrant deposition of α-synuclein can occur in the form of large α-synuclein aggregates around eosinophilic cytoplasmic inclusions called Lewy Bodies or fibrilliar inclusions in the neuronal processes called Lewy neuritis (Stefanis 2012). The majority of these α-synuclein aggregates occur within the dopamine producing neurons of the substantia nigra (SN), particularly in the zona-compacta, which lies in the mid-section of the SN. Whilst there is significant overlap between the conditions, patients with DLB tend to have higher degrees of α-synuclein deposition within the cortex and striatum, together with additional amyloid plaques compared to patients with PD (Tsuboi, Uchikado et al. 2007), which may begin to explain the differences in the cognitive phenotypes of the two conditions and highlights how they exist on a spectrum of disease.

In contrast to the observation of Lewy bodies and Lewy neuritis in PD and DLB, patients with MSA exhibit markedly different pathognemonic α-synuclein based neuropathological features termed Papp-Lantos bodies (Papp, Kahn et al. 1989, Ahmed, Asi et al. 2012) which are sickle shaped aggregates of α-synuclein and ubiquitin within oligodendrocytes. In addition to differing cellular aggregation of α-synuclein, in MSA, these inclusions are dispersed widely through the cerebral white-matter primarily affecting the pons, medulla, putamen but also the

substantia nigra pars compacta, cerebellum and preganglionic autonomic structures (Papp and Lantos 1994).

It is important to note that some cases of PD occur in the absence of α-synuclein deposition, for example those arising secondary to pathogenic mutations in the *PARK2* gene (Houlden and Singleton 2012). Most Parkinsonian disorders without α-synuclein accumulation are genetically mediated, and whilst clinically similar to sporadic PD, they clearly have differing molecular aetiologies and pathological features, highlighting the importance of genetic testing in order to delineate between such disorders.

1.3.3 Tauopathies – clinical features

The tauopathies are a heterogeneous group of degenerative conditions predominantly consisting of FTD (comprising several subtypes), together with Progressive supranuclear palsy (PSP), and corticobasal degeneration (CBD). They also overlap with motor neuron disease (MND) / amyotrophic lateral sclerosis (ALS).

FTD is challenging to diagnose due to the heterogeneity of symptoms, and several subtypes have now been described relating to the anatomical location of neuronal injury; Behavioural variant FTD (BvFTD) in which social and emotional function is affected, and three types of Primary progressive aphasia (PPA) syndromes (previously known as semantic dementia); a semantic variant (sv-PPA), a non-fluent variant (nfv-PPA), and a logopenic variant (lv-PPA)(Chare, Hodges et al. 2014). Finally, FTD also has the possibility to occur in conjunction with amyotrophic lateral sclerosis (FTD-ALS)(Ferrari, Kapogiannis et al. 2011), and whilst the specific clinical features that discriminate between these sub-types are beyond the scope of this introduction, they are described extensively elsewhere (Rascovsky, Hodges et al. 2011).

In contrast to FTD, PSP predominantly begins with postural instability (~70%) (Nath, Ben-Shlomo et al. 2003), followed by Parkinsonism, vertical gaze palsy and evidence of a frontal lobe dementia (Litvan, Grimes et al. 1999).

CBD has a similar clinical phenotype to PSP, but with more asymmetrical Parkinsonian features, earlier cognitive involvement, and an absence of vertical gaze palsies. Often additional atypical symptoms such as focal myoclonus or an alien limb phenomena can be observed in contrast to other synucleinopathies (Litvan, Grimes et al. 1999).

ALS / MND is a progressive degenerative disorder primarily affecting the anterior horn cells of the central nervous system. This results in progressive motor weakness, fasciculations of striated muscles, often including the respiratory muscles (Brooks 1994, Nodera, Izumi et al. 2007). Although prognosis differs by the subtype of ALS, the mean time to death following diagnosis is 3-5 years (Kleopa, Sherman et al. 1999).

ALS/MND occurs in conjunction with frontotemporal dementia (FTD) in up to 50% of cases. Whilst ALS-MND (whether in combination with FTD or not) rarely causes tau pathology, they are included into the tauopathy section due to the clinical and genetic overlap.

1.3.4 Tauopathies – neuropathology

The tau protein is expressed by the *MAPT* gene on chromosome 17, and functions primarily to stabilize intracellular microtubules within neurons, hence having significant influence on the morphology and physiology of neurons (Avila, Lucas et al. 2004). In diseased states, tau develops or assumes a number of alterations including changes in phosphorylation (Morishima-Kawashima, Hasegawa et al. 1995), or post-translational modifications such as truncations, and aggregation into oligomers and large insoluble filaments (Arendt, Stieler et al. 2016).

In FTD, there is a predilection for the loss of neurons within the frontal and temporal lobes, and the anatomical location of neuronal vulnerability relates to the specific clinical sub-type of FTD. The pathological diagnostic subgrouping primarily consists of defining the cellular localization of tau with conditions such as Pick's disease, and Alzheimer's disease showing neuronal tau pathology, Progressive supranuclear palsy (PSP) and corticobasal degeneration (CBD) showing mixed pathology and glial tauopathies showing only glial pathology (Lee, Goedert et al. 2001). Additionally, a variety of different antibodies can be used with immunohistochemical techniques to determine the nature of tau immunoreactivities and therefore the type of tau deposition (for example neurofibrillary tangles or Pick bodies) (Kovacs 2015). In addition, the detection of the two isoforms of the tau gene; the 3 repeat (3R) or 4 repeat (4R) in the diseased state can assist in classifying these disorders (Hutton, Lendon et al. 1998, Spillantini, Murrell et al. 1998, Ingelsson, Ramasamy et al. 2007).

Around 55% of FTD and most ALS cases do not involve tau pathology, and involve either the presence of ubiquitin-positive TAR DNA-binding protein (TDP-43) pathology in 45% of cases, or ubiquitin-positive and fused in sarcoma (FUS)-positive pathology in the remaining 5-10% of patients (Mackenzie, Neumann et al. 2010) (Figure 1.2). These findings strongly suggest markedly differing molecular aetiologies of disease converging on a similar cell-specific vulnerability.

1.3.5 Alzheimer's disease – clinical features

Alzheimer's disease is the most common form of dementia, comprising over 70% of all cases of neurodegenerative disease (Prusiner 2001). The condition is primarily characterized by the insidious onset of an amnestic syndrome in which the patient struggles with learning and the recall of recently learned information (McKhann, Knopman et al. 2011). The condition like

all neurodegenerative disorders is progressive, and over time involves other cognitive domains such as visuospatial processes and language.

Over the past 5-10 years there have been a multitude of invasive and non-invasive biomarkers associated with the development and progression of AD which aim to identify pre-symptomatic markers of disease (Jack, Knopman et al. 2013). Such techniques include Positron Emission Tomography (PET) imaging, amyloid PET (Marcus, Mena et al. 2014), and Cerebrospinal Fluid (CSF) analysis to detect amyloid and tau (Hansson, Zetterberg et al. 2006). Although they are not yet incorporated into major diagnostic algorithms (McKhann, Knopman et al. 2011), this is likely to change in the near future with further refinement of their sensitivity and specificity to predict disease.

1.3.6 Alzheimer's disease – neuropathology

Dual amyloid and tau pathology is exhibited in Alzheimer's disease, with the extracellular accumulation of amyloid-β fibrils and the intraneuronal accumulation of abnormally phosphorylated tau (Jellinger 2013). Formal neuropathological assessment in AD now often includes formal semi-quantitative assessment of neurofibrillary tangles (Jellinger 2013) according to the established registry for Alzheimer's disease protocol (Mirra, Heyman et al. 1991) , the topographic staging of AD pathology (Braak and Braak 1991) and the distribution of amyloid-β deposition in comparison to tau pathology (Thal, Rub et al. 2002). Over the past 20 years the progressive development and implementation of these assessments has enabled the development of scores such as the National Institute of Aging – Reagan Institute (NIA-RI) criteria to relate dementia to AD-typical lesions as either high, intermediate, or low likelihood (Hyman and Trojanowski 1997), conferring a 90% sensitivity and specificity to diagnose AD (Jellinger 2013).

1.3.7 Prion disorders – clinical features

Prion disorders are a group of both human and non-human neurodegenerative diseases. They result from the intercellular propagation of a misfolded protein in its pathogenic form. In humans, the *PRNP* gene encodes the prion protein (PrP), which is highly expressed within the brain, although it's specific function remains unknown (Andrews and Rothnagel 2014).

These disorders are classically separated into Creutzfeldt-Jakob disease (CJD), Gerstmann-Straussler-Scheinker syndrome (GSS), familial fatal insomnia (FFI) and Kuru. The largest of these cohorts is CJD, which can be iatrogenic (iCJD), occurring when prion proteins have been inadvertently transmitted from one human to another such as through the utilization of cadaveric derived human growth hormone (Duffy, Wolf et al. 1974)), genetically mediated (fCJD) when it arises due to mutations in the Prion Protein gene (PRNP) (Mead 2006), or sporadic in nature (sCJD). In addition, new variant CJD (vCJD) can also occur in which prion strains from other animals (in this case bovine spongiform encephalopathy (BSE) (Bruce, Will et al. 1997)) transmit to humans.

Given the rarity of GSS and FFI, I will focus on CJD in terms of the clinical phenotype and neuropathology. The clinical syndromes of the different subtypes of CJD are united by their relative rapidity and inevitably fatal course. However, vCJD has a younger age of onset, peaking in the 25-29 year age-range, with a highly homogeneous clinical presentation characterized by psychiatric symptoms early in disease course, followed by cognitive impairment, ataxia and a progressive movement disorder (Will 2003). Iatrogenic CJD also shows a relatively conserved phenotype, characterized by the onset of ataxia beginning around 6-8 years after exposure to the source of prions (Brown, Preece et al. 2000, Hoshi, Yoshino et al. 2000) often with the preservation of memory until late stage disease or even until death (Will 2003).

Patients with sCJD are older than vCJD patients, with a mean age of disease onset of 64 years of age (Collins, Sanchez-Juan et al. 2006, Puoti, Bizzi et al. 2012). They most commonly present with cognitive symptoms (40%), cerebellar (22%) or behavioural problems (20%) (Rabinovici, Wang et al. 2006), and the disease is rapidly progressive with a mean survival of 6 months, and with 90% of patients dying within 1 year (Geschwind 2015).

1.3.8 Prion disorders – neuropathology

Immunohistochemistry that shows the presence of prion protein (PrP) is vital in order to make a definitive diagnosis. Subsequently, in the case of sCJD, the nature of the PrP deposits (such as whether they are diffuse, synaptic, patchy or perivacuolar) correlate well with the molecular genotype of the PRNP gene at codon 129 (Parchi, Castellani et al. 1996) (see Chapter 1.11.4).

The major molecular classification of CJD relies on both the determination of the genotype at codon 129 of the PRNP gene together with the molecular weight of protease-resistant prion protein (PrP^{Sc}) extracted at post-mortem, after partial digestion with protease K and run on a Western blot (Geschwind 2015). Type 1 proteins have a more distal cleavage site resulting in a 21 kilodalton (kDa) protein, and type 2 have a more proximal cleavage site resulting in a 19kDa protein. Taken together, this results in 6 combinations of PRNP genotype and PrP^{Sc} weight (MM1, MV1, VV1, MM2, MV2, VV2) to be attributed to any case (Puoti, Bizzi et al. 2012, Geschwind 2015).

1.3.9 Polyglutamine (PolyQ) diseases – clinical features

Whilst the techniques used in subsequent chapters were not specifically able to detect PolyQ disorders, these conditions are only discussed briefly here to provide clarity to the background literature, and are reviewed extensively elsewhere (Ross 2002, Fan, Ho et al. 2014).

The polyglutamine disorders are a group of neurodegenerative diseases caused by expanded repeats of three alleles; cytosine, adenine and guanine (CAG) encoding a long polyglutamine (polyQ) tract in their respective proteins. A total of nine polyQ disorders have been described to date: five spinocerebellar ataxias (SCA) types 1,2,6,7,17; Machado-Joseph disease (MJD/SCA3); Huntington's disease (HD); dentatorubral-pallidoluysian atrophy (DRPLA); and spinal and bulbar muscular atrophy, X-linked 1 (SMAX1/SBMA) (Fan, Ho et al. 2014).

During protein synthesis the expanded CAG repeats are translated into a series of uninterrupted glutamine residues forming polyQ tracts. This results in the induction of the ubiquitin proteasome system (Bennett, Shaler et al. 2007), and the formation of intracellular aggregates, which some groups suggest are directly injurious to vulnerable neuronal subpopulations (Gatchel and Zoghbi 2005). However, there is compelling evidence that these inclusions may also be protective (Klement, Skinner et al. 1998, Saudou, Finkbeiner et al. 1998), and therefore whether the aggregates themselves contribute to causing cellular dysfunction remains contentious.

The clinical phenotypes of the nine established polyQ disorders vary significantly. This is complicated further by the intra-genotype clinical heterogeneity that ensues from the variability in repeat length between individuals, with longer repeat lengths causing earlier onset phenotypes in almost all polyQ disorders (Fan, Ho et al. 2014).

Despite this degree of intra and inter genotype-phenotype variability, there is a general predilection for polyQ disorders to affect the brain stem and cerebellum, predominantly causing a bulbar weakness or spinocerebellar syndrome as a relatively conserved element of each disease phenotype, though individual disorders also show relatively unique phenotypic features such as myoclonus in DRPLA, retinal degeneration in SCA7 or gynaecomastia in SMBA (Schols, Bauer et al. 2004).

1.3.10 Polyglutamine (PolyQ) diseases – neuropathology

The majority of PolyQ disorders cause nuclear inclusions comprised of aggregates consisting of a multitude of proteins. These proteins accumulate within the nucleus and are widespread throughout the CNS beyond the regions and cell types that show degeneration and cell death, highlighting that these disorders are both multisystem proteinopathies and further suggesting that the aggregates themselves may not be primarily injurious (Yamada, Tsuji et al. 2000). Recent work has also shown the presence of TDP-43 immunostaining within several polyQ disorders, suggesting mechanistic overlap with other TDP-43 related disorders such as ALS (Schwab, Arai et al. 2008).

Specific neuropathological features of almost all of the nine subtypes of PolyQ diseases have been reviewed extensively elsewhere (Yamada, Tsuji et al. 2000).

Table 1.1. The major clinical and pathological features of the most common neurodegenerative disorders. Each disorder is shown together with the common spatial profile of neuropathology, the proteins involved in the formation of aggregates within each disorder, the intracellular location of the aggregates, and the underlying genetic composition of the disease as either sporadic disease or inherited.

Disease	Clinical Features	Affected brain region	Proteins involved	Location of aggregates	Transmission
Alzheimer's disease	Progressive dementia	Hippocampus, cerebral cortex	Amyloid-β and hyperphosphorylated tau	Amyloid: extracellular, tau: cytoplasmic	Sporadic (95%), inherited (5%)
Parkinson's disease	Primarily a movement disorder	Brain stem, substantia nigra	α-synuclein	Cytoplasmic	Sporadic > 95%, inherited <5%
Amyotrophic lateral sclerosis	Movement disorder	Anterior horn cells, brainstem, motor cortex	Superoxide dismutase, TDP-43, Bonina bodies	Cytoplasmic	Sporadic ~ 75%, inherited ~ 25%
Frontotemporal dementia	Progressive dementia	Frontal and temporal cortex, hippocampus	Hyperphosphorylated tau, 'Pick bodies'	Cytoplasmic	Sporadic ~ 90%, inherited ~10%
Dementia with Lewy Bodies	Progressive dementia and subsequent movement disorder	Neocortex, locus coeruleus, substantia nigra	α-synuclein, 'Lewy bodies', amyloid, hyperphosphorylated tau	Amyloid: extracellular, tau and α-synuclein: cytoplasmic	Sporadic ~ 75%, 25% have a genetic component.
Progressive Supranuclear Palsy	Progressive movement disorder and dementia	Basal ganglia, brainstem, neocortex	Hyperphosphorylated tau within astrocytes	Cytoplasmic	Sporadic
Huntington's disease	Progressive movement disorder and dementia	Striatum, basal ganglia, neocortex, occasionally cerebellum	Huntingtin with polyglutamine expansion	Intranuclear and cytoplasmic	Genetic – 100%
Prion disorders	Progressive dementia, ataxia, psychiatric disturbance, insomnia	Varies significantly, generally an extensive pathological load	Prion protein	Extracellular	Sporadic – 90%, inherited ~ 8%, infectious ~2%

1.4 The structure of nuclear DNA

Deoxyribonucleic acid (DNA) is composed of a long sequence of DNA nucleotides (referred to as bases), comprising Adenine (A), Thymine (T), Guanine (G) and Cytosine (C). Together, over 3.3 billion of these bases constitute the human reference genome (Lander, Linton et al. 2001, Venter, Adams et al. 2001).

Bases within the human genome can be broadly classified as either being within 'coding regions' - which code for proteins (and comprise 1% of all DNA bases), or non-coding regions - which do not directly encode for proteins (Lander, Linton et al. 2001). It must be noted that 'non-coding' does not mean 'non-functioning' with several functions of non-coding regions of DNA having been established, including their transcription into non-coding RNA molecules, or a role in transcriptional or translational regulation of protein structures. Taken together, it is estimated that approximately 8-10% of the human genome possesses a functional role (Rands, Meader et al. 2014).

Coding regions are sub-composed of exons, which refer to DNA bases that directly encode a part of the final mature RNA (Figure 1.3). These bases are translated into RNA, with non-coding sections subsequently removed through splicing, and actively translated into proteins (Graveley 2001). Therefore mutations within the human exome have the ability to significantly alter protein structure and function, and confer the ability to cause or contribute to human diseases (Lek, Karczewski et al. 2016).

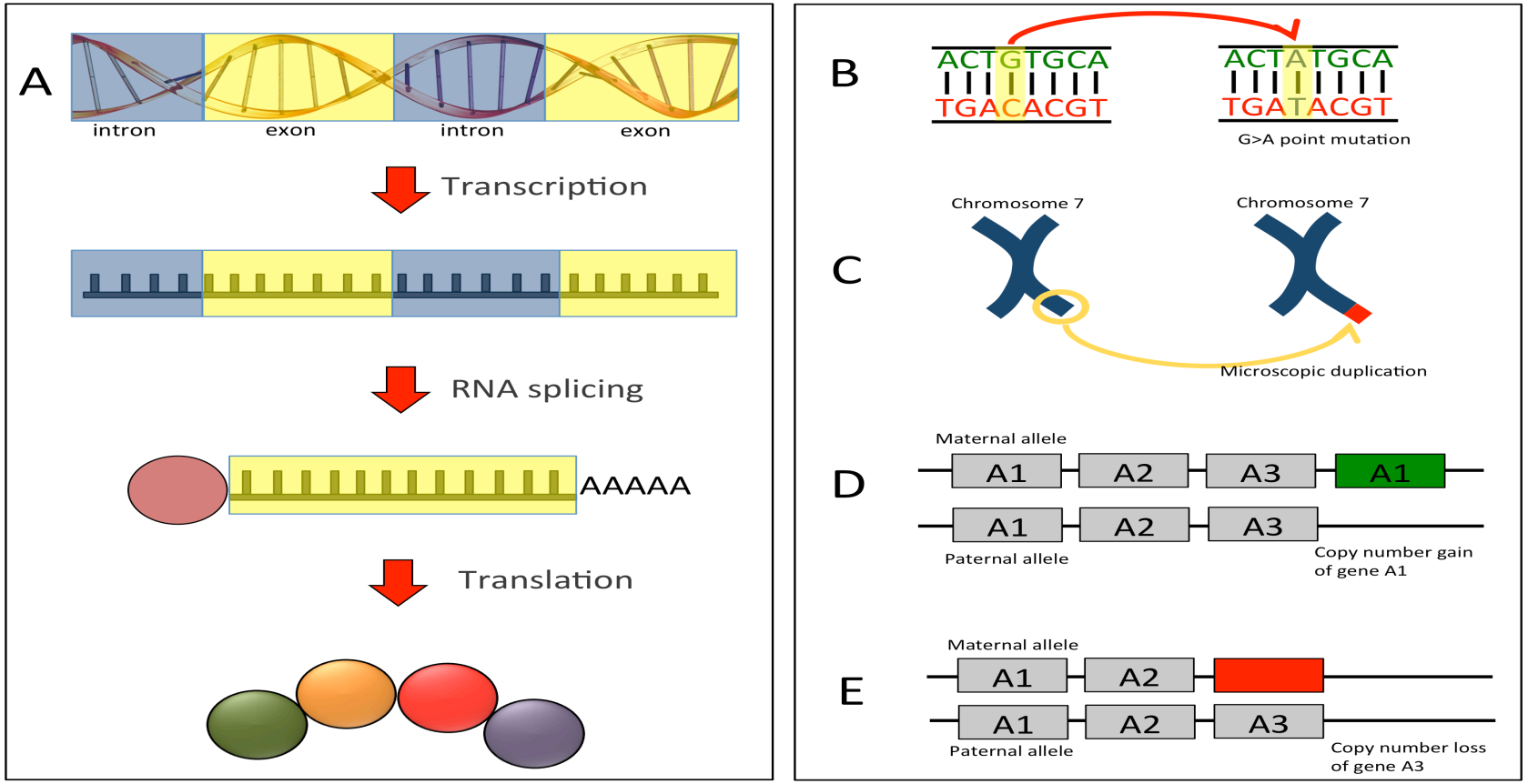


Figure 1.3. Mechanisms of genetic variation causing neurodegenerative disease. A: The mechanism of protein translation is shown. Genes, separated into intronic regions (blue) and exonic regions (yellow) are transcribed into RNA, spliced, and then translated into an amino-acid sequence (each amino-acid represented by a different coloured circle). B: A point mutation in DNA is shown. A guanine (G) (yellow) is shown on one strand to mutate to an Adenine (A), with corresponding changes on the alternate strand shown. C: Shows a microduplication of a region of a region of chromosome 7 (red). D: An example sequence of three genes (A1,A2,A3) are shown to be present on both alleles, with a duplication (copy number gain) of the A1 gene on the maternal allele. E: The same three genes are shown, with a deletion (copy number loss) of the A3 gene on the maternal allele leaving only one copy of the gene.

1.5 Mutations in nuclear DNA

Point mutations: Coding regions of the human genome can be divided up into codons – sets of three bases that specify an amino acid or signal the end of the protein sequence (Ramakrishnan 2002). If any base of DNA within the exon is mutated, then it may therefore alter the amino acid which it encodes. For example, the codon GCA encodes for the amino acid Alanine. If the first allele of that codon (G) mutates to a C (which can be written as G>C, with the first allele referring to the reference allele and the second the mutated or 'alternate' allele), then the codon will be coded CCA which encodes for the amino acid Proline. When such a mutation changes an amino-acid it is referred to as a non-synonymous mutation. When no change to the protein structure occurs as a result of a mutation it can be considered a 'synonymous' mutation. In addition, a mutation can change the codon composition from one that encodes for a particular amino-acid (for example TAT which encodes for Tyrosine) into a premature stop codon (eg TAA which induces a stop codon). This results in the premature termination of gene translation and a shortened protein product.

Additional types of mutation can occur within the coding region of DNA in addition to the substitution of DNA bases. For example, DNA bases can be inserted or deleted potentially resulting in a shift of the reading frame of DNA transcription, and therefore protein translation. For example, the removal of 2 bases within a gene will result in a completely different translation from the original coding sequence. A non-frameshift mutation is divisible by 3 bases (ie the insertion of deletion of one codon and therefore one amino-acid in a sequence). Whilst this will either insert or delete one amino-acid, the remainder of the translated protein will be the same as the original.

The ability of non-synonymous, frame-shift, or premature stop-codon inducing mutations to significantly alter a protein structure means that they can cause significant functional impact.

Such mutations have been estimated to be responsible for ~85% of all Mendelian disorders (Botstein and Risch 2003).

In contrast, whilst mutations in non-coding DNA do not change the structural conformation of a protein, variation in the non-coding region of DNA has the ability to alter cellular function by altering transcriptional or translational activity. For example mutations within enhancer regions of the genome (Corradin and Scacheri 2014), or at alleles that affect gene splicing (Zhang, Joehanes et al. 2015) show the ability of non-coding mutations to significantly alter gene expression. Occasionally these mutations can cause Mendelian forms of disease (Zhang and Lupski 2015), however, non-coding mutations usually mildly increase disease risk for complex disorders such as neurodegenerative diseases (Wellcome Trust Case Control 2007, Lill and Bertram 2011).

Structural variation: In addition to single base mutations within the human genome, structural variation can also occur, which refers to genomic alterations that involve segments of DNA that are larger than 1000 bases (1 kb) in length (Feuk, Carson et al. 2006). These structural variants can be further sub-classified by size into microscopic structural variation (which is greater than 3 million bases and includes whole chromosome duplications or losses), and submicroscopic structural variation, which is less than 3 million bases. At present at least 12% of the human genome has been observed in association with copy number variation (Hastings, Lupski et al. 2009), though greater resolution sequencing platforms are likely to drastically increase these estimates.

Types of structural variation include copy number variants (CNVs) in which a segment of DNA larger than 1kb is present at a variable copy number compared to the reference genome, for example either a deletion or duplication of such a length of DNA (Figure 1.3). Common copy number variants that occur in greater than 1% of the population are often referred to as

copy-number polymorphisms and are being increasingly appreciated (Sebat, Lakshmi et al. 2004). Additional forms of structural variation such as an inversion or translocation of the genome can occur in which the orientation of a segment of DNA is reversed, or a region of DNA has moved from it's established position within the genome into another location respectively (Feuk, Carson et al. 2006).

The mechanisms underlying the formation of CNVs have been investigated for over a decade and they are believed to occur via two primary mechanisms; homologous recombination and non-homologous recombination events (Hastings, Lupski et al. 2009). While the exact mechanisms do not require extensive elaboration, the basic premise of homologous recombination is that during meiotic recombination the pairing of homologous chromosomes from both parents may aberrantly misalign and result in the transfer of additional copies or deletions of the genome to the offspring (Paques and Haber 1999). Non-homologous recombination is likely to result from ineffective DNA repair mechanisms that result in either the deletion of a section of DNA or inadvertent duplication (Lieber 2008).

1.6 The genetic landscape of neurodegenerative disorders

Neurodegenerative disorders have both a monogenic and complex genetic aetiology. Monogenic disorders arise as the result of single defective gene within the genome, for example a single non-synonymous mutation that significantly alters the structure and function of a critical gene. This can usually be observed by the inheritance pattern of that disorder within a family, for example, a clear autosomal dominant family history of disease. It must be noted that in cases of either recessive disorders (in which a prior family history is often not visible), or in cases of late-onset disorders (in which parents may be deceased prior to the onset of disease), then a family history is not always observed. Monogenic disorders are estimated to comprise between 5 and 15% of all cases of neurodegenerative disease, but their

exact proportions vary significantly between the major neurodegenerative disorders (Lill and Bertram 2011).

The majority of cases of neurodegenerative disease are therefore believed to be complex traits, resulting from genetic variation in multiple genes that interact with behavioural and environmental factors to influence the risk of disease (Lill and Bertram 2011). In this setting, the majority of cases show no readily predictable pattern of inheritance, and the associated genetic variation is usually non-coding in nature and confers a small increase in the risk of that individual developing the disease (with the size of this risk often quantified by the 'effect size') (Lohmueller, Pearce et al. 2003, Wellcome Trust Case Control 2007). These variants can also be relatively common in the population given that they, in isolation, are unlikely to cause disease, and are unlikely to be selected against by the laws of natural selection (unlike many highly penetrant alleles that cause monogenic cases of disease).

The rather dichotomous separation of neurodegenerative diseases into a strict Mendelian forms of disease (caused by highly penetrant single alleles) or a complex genetic aetiology (contributed to by common non-coding alleles that confer a small risk of disease) is likely to be an over simplification. For example, in the case of late-onset disorders, incomplete penetrance (a setting in which only a proportion of individuals carry the mutation cause disease) is being increasingly recognized. In addition, the recent identification of protein coding variants that moderately increase risk or cause an 'intermediate risk' for several disorders including neurodegenerative diseases (Guerreiro, Wojtas et al. 2013) mean that rather than a black and white divide between Mendelian and complex forms of disease, that a spectrum of genetic risk may be more likely to exist (Figure 1.4).

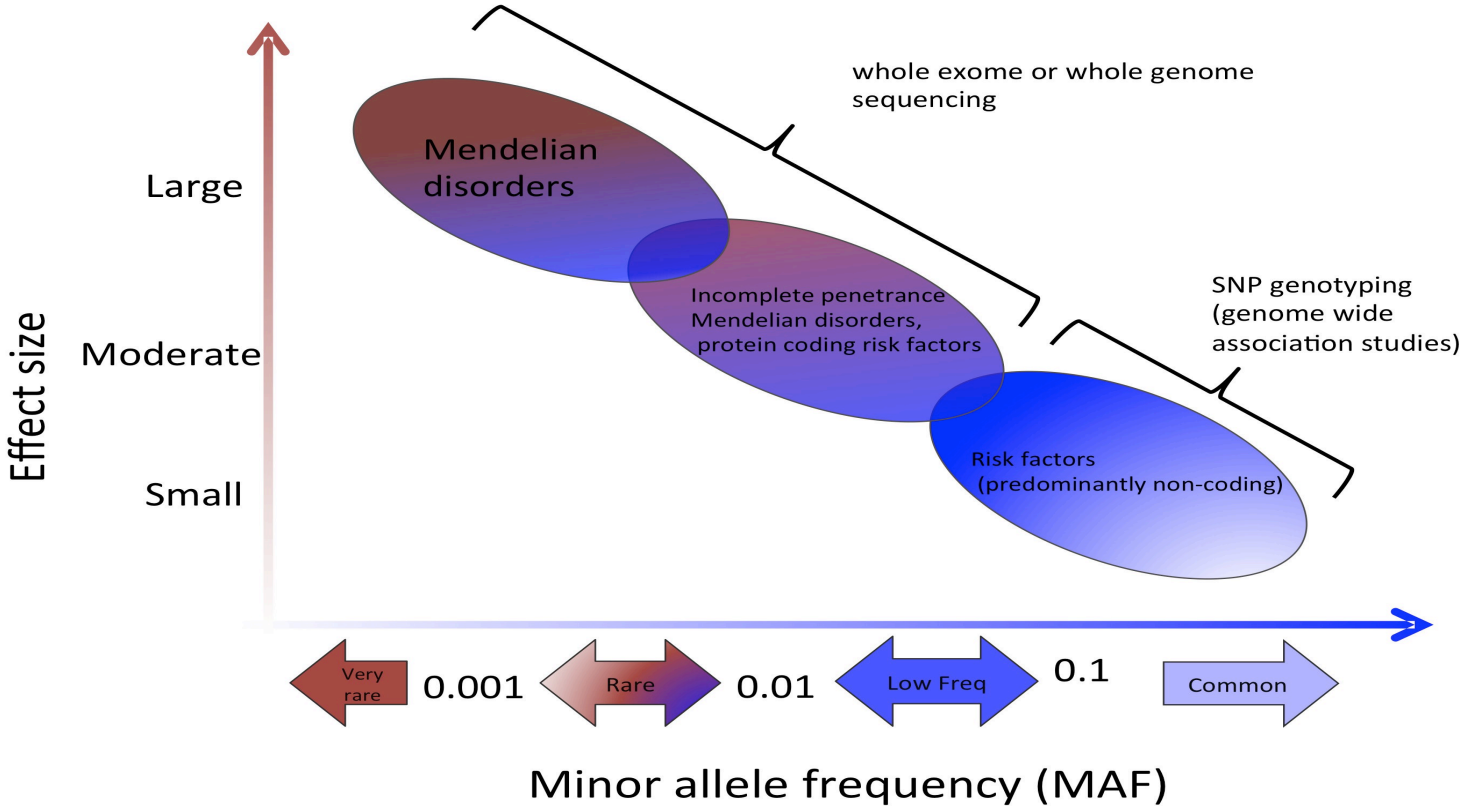


Figure 1.4. The mechanism of genetic variation causing neurodegenerative disease. The effect size (the magnitude of effect that an allele has on a phenotype) is shown on the Y axis, and the minor allele frequency (MAF) (the frequency of the allele within the population) is shown on the X axis. The relevant contribution to disease is shown within the X-Y grid (Mendelian, Incomplete penetrance or protein coding risk factors, and risk factors) are shown, together with the appropriate sequencing methodology best placed to capture the relevant types of genetic variation.

1.7 Approaches to discovering the molecular basis of Mendelian disorders

In order to detect rare, often private mutations that cause disease within single families, the approach until around 2010 would have utilized linkage-analysis. This involved using a SNP genotyping platform that would determine the genotype at alleles spaced tens or even hundreds of thousands of bases apart. Determining regions of shared alleles between affected individuals within a family would enable the detection of a chromosomal region in which a likely pathogenic mutation would lie (given that genes that reside physically close on a chromosome remain linked during meiosis) (Pulst 1999). Subsequently, genes within the shared region were prioritized based on their putative function in relation to the disease mechanisms, and subsequently Sanger sequenced manually. Sanger sequencing is a technique which can read a DNA sequence of up to~ 1000 bases, however, this can only be within the pre-defined region or gene of interest, and inevitably this approach is limited, and made sequencing the entirety of multiple genes extremely cumbersome and expensive to perform in this fashion (Pulst 1999).

Similarly in terms of clinical diagnostics, then until recently, clinical testing in a family affected by a dominant form of disease (for example) would have taken place sequentially, with the most common genetic mutation for that particular disorder tested first (often by Sanger sequencing), before sequential testing of progressively less common genotypes associated with disease took place. This process was laborious, expensive, and often resulted in significant diagnostic delays for patients.

Since 2010, second generation sequencing, or ‘next-generation’ sequencing has become increasingly employed to detect rare variants causing Mendelian forms of disease given it’s ability to sequence vast quantities of DNA in a short period of time, up to, and including, the whole genome. This has delivered a step-change in genetic analysis. For example, it cost $2.7

billion and took 13 years to read the 3.3 billion bases in the first human genome performed by massive laboratories on both sides of the Atlantic (Lander, Linton et al. 2001, McPherson, Marra et al. 2001, Sachidanandam, Weissman et al. 2001). In contrast, the human genome can now be sequenced in 2 weeks for < £1000. The precise chemistries have been reviewed elsewhere (Bamshad, Ng et al. 2011), but the principles are briefly summarised below.

1.8 Next generation sequencing

Next-generation sequencing, also known as high-throughput sequencing, is a term used to describe the processes in which large quantities of DNA can be fragmented, sequenced and aligned against the human reference genome to determine genetic variation.

Following DNA extraction from the tissue of interest, the entire quantity of DNA is fragmented (sheared into pre-defined lengths) usually using physical methods (i.e. acoustic shearing / sonication), before the ends of the sheared DNA fragments are blunted and the 5' end is phosphorylated using a mixture of enzymes. Next, a non-template nucleotide (A-tail) is added to the 3' end to facilitate ligation of sequencing adaptors, and this prevents concatemer formation whilst also introducing a sequence for the annealing primer used by reverse transcriptase for the first strand cDNA synthesis (Head, Komori et al. 2014). This multi-step process can also be performed as a single 'tagmentation' reaction using a transposase enzyme to simultaneously fragment the DNA and attach the adaptors (Figure 1.5).

Subsequently, after clean-up to remove the transposases, and amplification to increase the quantity of adaptor tagged DNA sequences, primers required for sequencing and indexing (sample identification) are added as one of three PCR enrichment steps. After this, short sequences of DNA (termed 'probes') hybridize to the genomic regions of interest, which can either be the whole human exome, or a specific set of genes in the case of diagnostic panels for example. These probes (and accompanying regions of DNA of interest) are then selected

or ‘pulled down’ washing away excess uncaptured DNA. This process is called ‘in-solution capture’, and although an alternative paradigm called array-based capture is also available, in-solution techniques are most commonly utilized at the present time.

At this point in the process, isolated DNA fragments from the human exome or panel of interest are present in solution with attached sequencing primer binding sites and appropriate sample indexes added at each end, and are ready for sequencing. The first step of massively parallel sequencing is the generation of clusters on a flow cell. Essentially this involves the prepared DNA being captured by a lawn of oligonucleotides that are able to capture a specific DNA sequence that was incorporated into the PCR annealed sequences. Captured DNA is used to generate a second strand, which can then also in turn be copied itself in close proximity by bridge amplification, generating clusters of similar reads within the flow-cell. At this point, numerous copies of each DNA fragment have been generated and are now able to undergo sequencing by synthesis. This sequencing technology is similar to first generation ‘Sanger’ sequencing but employs the incorporation of dNTPs containing a reversible terminator which then blocks further polymerization. Each termination event contains a fluorescent label, with each type of base (A,T,C or G) having a unique colour. This means that the incorporation of each base to each chain can be detected by a highly specialised camera to determine the base sequence. A second critical difference between NGS and Sanger sequencing is that millions of different strands of DNA are copied/sequenced at the same time, hence the term ‘massively parallel sequencing’.

The reads produced from the above process will include the index sequences incorporated in the library preparation steps in order to identify which sample/individual each sample has arisen from. It is important to note that the samples are sequenced from the 5’ to 3’ end in a bidirectional manner producing read 1 and read 2, which can subsequently be amalgamated to

improve the calling of genetic variants at the bioinformatics stage (see below). In essence though, all reads from the same individual are now able to be aligned to the reference genome and the proportion of reads that have non-reference alleles determined, and therefore the genotype derived (see below).

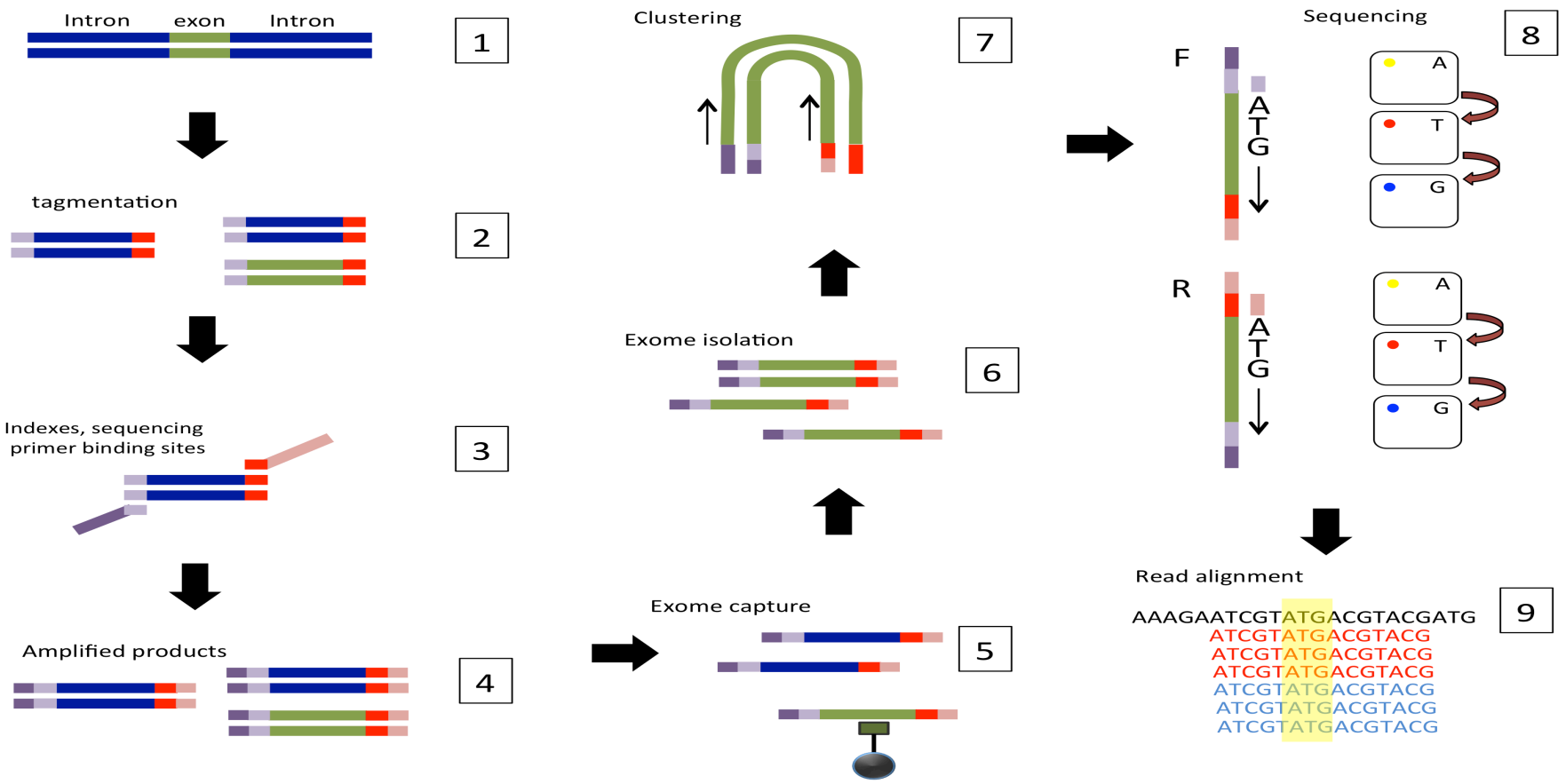


Figure 1.5. The preparation and processing of DNA in next generation sequencing (NGS). Step1-3: A schematic representation of a fragment of DNA (introns: blue, exons: green) undergoing tagmentation and the annealing of indexes and primer binding sites (pink and purple regions). Step 4-6: indexed fragments are amplified, and then selected regions of the exome are captured, pulled down, and purified. Step 7: shows bridge amplification of the captured regions of DNA to generate copies of the original strands, before step 8 shows the process of sequencing by synthesis in which three bases of DNA (ATG) are sequenced on the forward (top) and reverse (bottom) strand as fluorescent nucleotides are incorporated into the strand. Step 9 shows how forward reads (red) and reverse reads (blue) are compiled in order to determine the genotype at each base position (again the ATG sequence is highlighted yellow).

1.9 Bioinformatic approaches in NGS

Pre-processing and quality control of the raw data: Raw data that emerges from the sequencer arrives in a 'fastq' format. This essentially is a text-based representation of sequences that incorporates the Phred-scaled base quality score of each read. This first needs to undergo quality control to check the overall number of bases sequenced, base quality scores, and the average read length for example. This involves trimming reads to the desired length, removal of regions of poor quality (generally at the end of reads), and removing adaptor sequences that were added in the sample preparation steps (Bao, Huang et al. 2014).

Alignment to reference genome: The second step in the analysis of NGS data is alignment to the reference genome. This involves a series of steps in order to determine the position of the read within the reference human genome (in this thesis – human genome reference build 19 (hg19)). Given that reads are on average 100-150 bases in length, compared to the 3.3 billion bases in the genome, then a complex process of algorithms is employed to map these reads to the reference genome as accurately as possible. This results in the generation of a mapped set of reads to the reference genome in a SAM (sequence alignment map) format.

Post alignment processing: Whilst some degree of quality control was employed in the initial stages to trim the end of reads (which are prone to the development of errors), subsequent steps must be taken to ensure that the highest quality reads remain aligned to the human genome to ensure that the most accurate list of variant calls possible can be determined. This primarily involves the removal of 'duplicate reads' (which are reads that have the same start and end point and are presumed to have arisen from sequencing the same fragment of DNA twice), the re-alignment of reads of DNA in which a potential insertion or deletion of bases has occurred (which is more computationally intensive), and finally base score quality calibration (the removal of bases in which poor quality DNA calls are present and are likely to

lead to false positive results). Whilst some metric of this is provided within the initial QC steps (using programmes such as fastQC (Andrews 2010)), they are often inaccurate and suffer from a degree of systematic bias (Minoche, Dohm et al. 2011), and thus a recalibration process after alignment should significantly improve the quality of base calling.

Variant analysis: This consists of determining the genotype at each base (i.e. whether there is a heterozygous or homozygous mutation). This process can either be performed by considering each sample as a single sample (using callers such as Varscan (Koboldt, Chen et al. 2009) or Freebayes (Garrison 2012)), or performed across multiple samples at the same time (for example using programmes such as the Genome Analysis Tool Kit (GATK) (McKenna, Hanna et al. 2010)). This latter approach helps to correct for both false positive and negatives by providing a comparative sample to call against from a homogeneous dataset, however it is also computationally far more taxing.

Subsequently, variants are annotated, which primarily involves determining whether it is a synonymous or non-synonymous variant, or by whether the mutation is likely to significantly alter the protein structure to cause detrimental biological effects (primarily determined by the use of *in-silico* modeling software). Comparison of mutations with well-established databases of several thousand humans also enables the determination of how common the minor allele frequency (MAF) may be within the population (Lek, Karczewski et al. 2016).

Clinical interpretation of variant pathogenicity: For any single individual, whole-exome sequencing (for example) will identify roughly 20 – 25,000 single nucleotide polymorphisms and several thousand base insertions or deletions, highlighting the intrinsic variability of the human exome. The real challenge is subsequently to determine which, if any of these mutations are responsible for the disease (Figure. 1.6).

Until recently the approach to defining pathogenicity in NGS data was variable, but several general rules applied; firstly if the disorder was rare within the population, and proposed to be caused by a highly penetrant allele, then the mutation causing disease must also be proportionally rare. In addition, the mutation must either be non-synonymous, affect splice-sites, stop codons or reading frames. Finally, the mutation should be predicted by *in silico* modelling tools to significantly alter protein function, and that there must be a relevant biological mechanism by which the mutation would affect a biological pathway involved in that disease. This relatively loose approach however led to a multitude of false positive reports of pathogenicity within the medical literature (Bell, Dinwiddie et al. 2011, Xue, Chen et al. 2012). Therefore in 2014 and 2015 a team of leading geneticists (MacArthur, Manolio et al. 2014) and the American College of Medical Genetics (ACMG) (Richards, Aziz et al. 2015) published new stringent guidelines for the interpretation of variant pathogenicity. These data offer a framework for the degree of weighting to be applied to each of the elements of variant pathogenicity (for example MAF, *in silico* prediction), and also placed a large degree of emphasis on the requirement for *in vitro* or *in vivo* functional data together with a careful assessment of the location of the mutation with relevance to known functional domains of proteins. These data are likely to help reduce the false positive burden of pathogenicity and homogenise the criteria to report pathogenicity (MacArthur, Manolio et al. 2014, Richards, Aziz et al. 2015) (Figure 1.6).

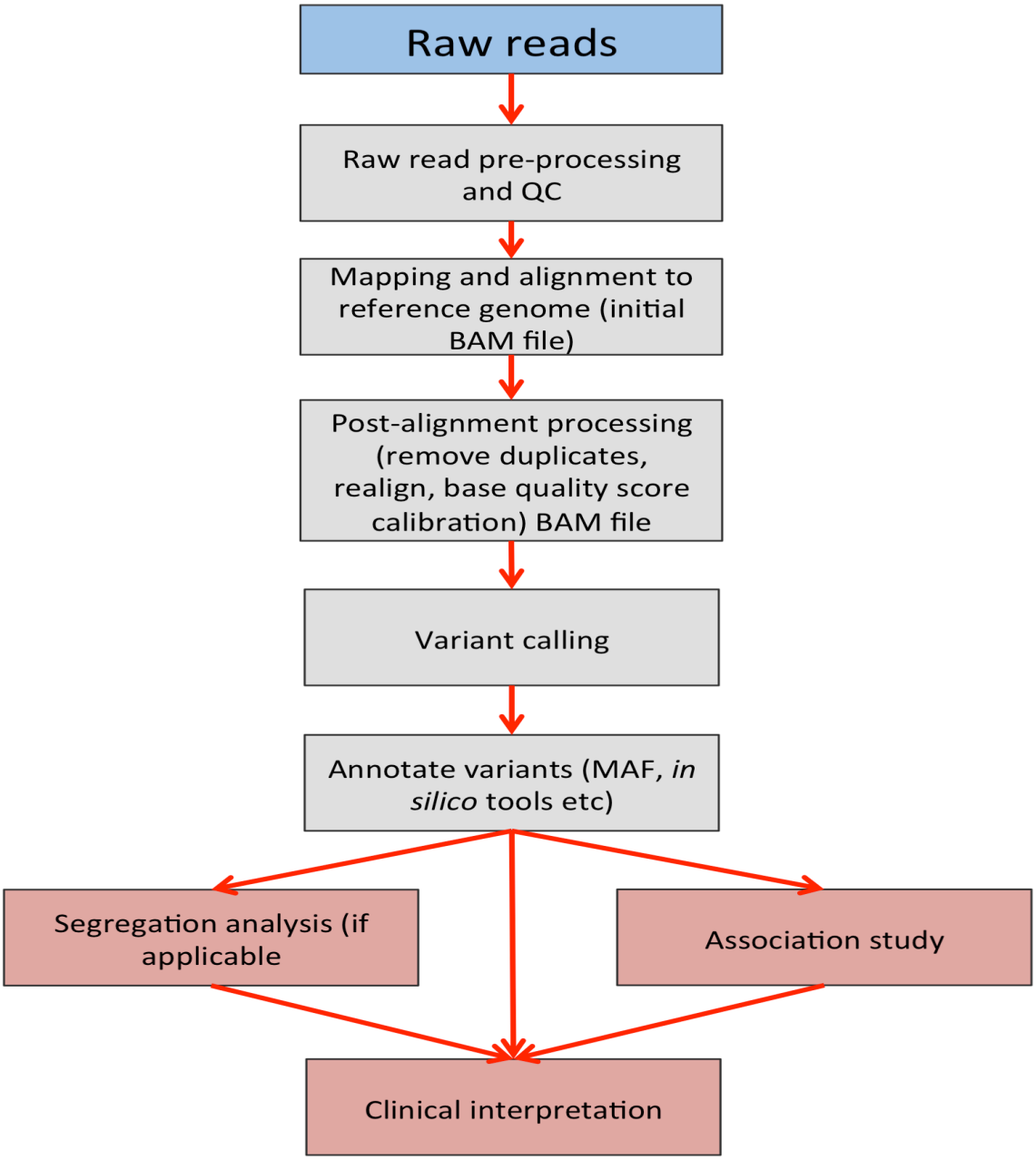


Figure 1.6. The bionformatic approaches to generating and interpreting NGS data.
Firstly raw reads from the sequencer are taken (blue) in a fastq format, and pre-processed to check read quality and then trim poor quality regions of the reads. They are then mapped and aligned against the human reference genome (such as hg19) in order to generate the first BAM file (binary version of the Sequence Alignment Map (SAM file)). Subsequent steps involve improving alignment and data quality before the variants can be called based upon primarily the allelic ratios in the BAM file. After that, the variant called file (in a variant called format (VCF) file) can be annotated by numerous parameters including the minor allele frequency (MAF). Finally, the boxes in red show that in the context of multiple samples from the same family segregation analysis can be performed, or in the context of unrelated individuals an association study. After that, clinical interpretation of remaining variants is vital to determine likely candidates or established pathogenic alleles.

1.10 Approaches to detect novel variants associated with disease

Two primary approaches can be employed in order to detect either highly penetrant alleles that cause monogenic forms of disease, or the identification of an allele or alleles associated with disease from multiple unrelated individuals.

To detect highly penetrant alleles, the most common approach is to identify rare variants that segregate with disease within an affected family, and are either absent or extremely rare within the general population. This approach can be especially promising when including distantly related family members (thus increasing the number of meiosis between affected individuals) which will significantly reduce this list of variants to a few hundred or even lower. Thereafter, the clinical interpretation of the remaining variants is of vital importance (as described above).

To detect variants that confer risk to develop a neurodegenerative disease involves studying unrelated individuals with the same disease phenotype rather than a family. This has been the primary paradigm of genome-wide association studies (GWAS) which utilize SNP genotyping arrays of tens, or even hundreds of thousands of SNPs to detect specific regions of the human genome that are associated with a disease phenotype, by determining the alleleic ratios between cases and controls for any given disorder, and then correcting for the number of tests performed. GWAS studies often require several thousand samples to detect these variants due to the number of statistical tests performed and the relatively low risk that each allele confers for disease. To date these approaches have only helped explain a relatively small proportion of the heritability of many complex disorders including neurodegenerative diseases (Keller, Saad et al. 2012, Keller, Ferrucci et al. 2014).

The detection of rare variants identified through NGS that may confer risk for neurodegenerative disorders, if employing the same statistical tests as GWAS, would be

extremely problematic unless either sample sizes were in the tens of thousands, or the effect size of any variant was remarkably large (Li and Leal 2008). Therefore, in order to detect rare variants that may cause a modest risk of disease, new statistical methods employing 'burden tests' have been designed to surmount these problems (Madsen and Browning 2009, Wu, Lee et al. 2011). These tests collapse and summarise variants within a particular region of the genome (usually a gene) into a single value which can then be tested for association with disease. This makes it possible to identify rare genetic variants from only a few hundred patients and offer a promising mechanism to find new genetic risk factors.

1.11 Established genes associated with neurodegenerative disorders

Over the past 20 years, numerous genes have been identified as causing or contributing to the development of the major neurodegenerative disorders, with differing levels of evidence supporting both true pathogenicity as either monogenic alleles or disease risk factors.

1.11.1 Alzheimer's disease

Early onset Alzheimer's disease (EOAD) is defined by disease onset < 60 years of age, and accounts for around 10% of all cases of AD. EOAD is believed to be almost entirely genetically mediated, with a heritability of 92-100% (Wingo, Lah et al. 2012). In contrast, late-onset Alzheimer's disease is likely to be a highly polygenic disease, with a heritability of ~70% (95% CI: 64.6-75.0%) (Wingo, Lah et al. 2012).

***Monogenic alleles**:* Despite the prevalence of AD, only three genes have been robustly identified as causing monogenic forms of the disease, with all three integral to β-amyloid production and processing. The first is the amyloid precursor protein (*APP*) on Chromosome 21, in which both copy number duplications and point mutations have been described as causing disease (Rovelet-Lecrux, Hannequin et al. 2006). Over 30 different point mutations in APP can cause familial forms of Alzheimer's disease (Cruts and Van Broeckhoven 1998), and

alone they account for about 10-15% of all early onset cases of AD (Bird 2008). The second gene causing monogenic forms of AD is Presenilin 1 (*PSEN1*), which is present on chromosome 14 and is the most common cause of genetically mediated AD. The presenilin family of genes perform a critical role in the γ-secretase cleavage of APP (De Strooper, Saftig et al. 1998, Wolfe, Xia et al. 1999), and almost 200 pathogenic mutations have been described in *PSEN1*, which together are responsible for ~50% of EOAD (Theuns, Del-Favero et al. 2000). Another member of the Presenilin family, *PSEN2*, also causes familial AD, often with a slightly later onset, variable penetrance, and significant clinical heterogeneity (Sherrington, Froelich et al. 1996) (Figure 1.7).

***Moderate risk factors**:* Two genes have been identified in which mutations (both protein coding) may act as moderate risk factors for Alzheimer's disease; Apolipoprotein E (*APOE*) and Triggering Receptor Expressed on Myeloid Cells 2 (*TREM2*). The presence of the *APOE* ε4 allele confers an increased odds ratio for AD of approximately 3, and the presence of the allele in the homozygous state of over 30 (Kukull, Schellenberg et al. 1996). Specific variants in *TREM2* also double or triple the risk of an individual developing Alzheimer's disease (Guerreiro, Wojtas et al. 2013). The fascinating aspect of these variants is that they appear to have drastically differing biological functions, with *APOE* involved in the catabolism of triglyceride rich lipoproteins (Hatters, Peters-Libeu et al. 2006), and *TREM2* a gene involved in inflammatory cell signalling (Ulrich and Holtzman 2016). Such significant functional variability is a consistent theme of the results of genetic association studies over the last 20 years in neurodegenerative diseases, and it remains unknown how genetic variants in these genes all appear to converge on singular common biochemical pathways.

GWAS loci: Sixteen genes have been identified as increasing or deceasing the risk of developing AD when incorporating data from over 74000 patients (Lambert, Ibrahim-Verbaas

et al. 2013), with identified variants conferring an odds ratio of disease of between 0.93 and 1.22 (Lambert, Ibrahim-Verbaas et al. 2013). Identified genes function to alter cholesterol metabolism, immune mediated processes, endocytosis, cytoskeleton formation, and epigenetic regulation of genes (Robinson, Lee et al. 2017), further supporting the diverse array of cellular processes in which genetic variants confer risk for AD (Figure 1.7).

1.11.2 Parkinson's disease

Monogenic alleles: Around 10% of patients with Parkinson's disease report a positive family history of disease (Thomas and Beal 2007, Klein and Westenberger 2012), and it is estimated that nine genes can now explain around 50% of familial cases (Gasser 2015). Unlike Alzheimer's disease in which all identified monogenic cases are heterozygous mutations resulting in dominant forms of the disease, around 50% of genetically mediated forms of PD are caused by recessive mutations (Gasser 2015). In addition, unlike many of the other neurodegenerative disorders, several monogenic causes of PD, particularly the recessive forms, whilst causing a clinical syndrome in keeping with idiopathic Parkinson's disease, do not show the characteristic neuropathological hallmark of the disease - Lewy bodies (LBs) (Cookson, Hardy et al. 2008). For example, LBs have not been observed in most studies of *Parkin/PARK2* mutation carriers (Yokochi 1997, van de Warrenburg, Lammens et al. 2001), *VPS* mutation carriers (Tsika, Glauser et al. 2014), and some *PLA2G6* carriers (Paisan-Ruiz, Li et al. 2012).

Dominant monogenic alleles: The first discovery of a monogenic form of PD was made in 1997 when a missense mutation in *SNCA* (the gene encoding α-synuclein) was found in a large family with autosomal dominant PD (Polymeropoulos, Lavedan et al. 1997). Shortly afterwards, α-synuclein was determined to be the predominant aggregated protein in the brain of patients (Spillantini, Schmidt et al. 1997), resulting in the beginning of the PD genetic era.

Since then, several point mutations and both duplications and triplications of *SNCA* have been observed as causing familial forms of PD, with triplication carriers suffering disease onset around a decade before duplication carriers consistent with a gene dosage effect (Singleton and Gwinn-Hardy 2004). Clinically, young onset PD (YOPD) is almost invariable in SNCA mutation carriers, with significant levo-dopa responsiveness (Golbe, Di Iorio et al. 1990), early cognitive symptoms of dementia and hallucinations, and widespread LB deposition in the brainstem and cortex (Farrer, Kachergus et al. 2004).

In 2004, the leucine-rich repeat kinase gene (*LRRK2*) was identified as another dominant cause of PD (Paisan-Ruiz, Jain et al. 2004, Zimprich, Biskup et al. 2004). The exact function of the LRRK2 protein remains unclear, with some functional domains consistent with performing a kinase activity and others a scaffold protein (Nuytemans, Theuns et al. 2010). Mutations in *LRRK2*, unlike other genetic forms of PD, have frequently been observed in late onset cases, with approximately 1-2% of late onset PD cases arising due to mutations in *LRRK2* in Northern Europe (Paisan-Ruiz 2009), and up to 40% in some regions of North Africa, with the G2019S mutation by far the most common genotype (Lesage, Durr et al. 2006). Determining pathogenicity of *LRRK2* mutations is fraught with difficulty given that the gene comprises over 2000 amino-acids and 51 exons, and therefore the incidence of benign but rare polymorphisms in this gene is extremely high. There is also strong evidence for reduced penetrance arising with the G2019S mutation, with the mutation having 28% penetrance by the age of 59 years of age, but 74% by 79 years of age (Healy, Falchi et al. 2008), leading some authors to regard this mutation rather uniquely as both a monogenic allele and disease risk factor (Gasser 2015) (Figure 1.7).

***Recessive monogenic forms alleles*:** Recessive mutations in *Parkin/PARK2* comprise about 50% of established monogenic cases of PD, primarily causing disease through a loss-of-

function effect (Nuytemans, Theuns et al. 2010). It must be noted that both point mutations, and copy number variation / small deletions within both coding and promoter regions of this gene are able to cause disease (Hattori, Kitada et al. 1998). This gene functions as an E3 ubiquitin ligase, tagging dysfunctional proteins for degradation (Shimura, Hattori et al. 2000) whilst also facilitating the maintenance of mitochondria (Deng, Dodson et al. 2008).

Homozygous point mutations and deletions within P-TEN induced putative kinase 1 *(PINK1)* also cause young onset PD (Valente, Abou-Sleiman et al. 2004), accounting for approximately 25% of all proven cases of monogenic PD (Nuytemans, Theuns et al. 2010). Mutations in *DJ-1* also account for 5% of monogenic PD (Nuytemans, Theuns et al. 2010), and again, disease can be caused by both point mutations and deletions, with the mechanism of cell death likely to be secondary to increased oxidative stress (Canet-Aviles, Wilson et al. 2004).

Other recessive forms of PD are caused by mutations in *PLA2G6, ATP13A2* and *FBX07*. It should be noted that these disorders are better characterized as complex syndromes with Parkinsonism rather than PD, as in all these genotypes of disease, Parkinsonism is just one of several clinical features, and LBs are not always seen pathologically (Paisan-Ruiz, Li et al. 2012). However, these genes should be considered for clinical testing in cases of young onset PD given the phenotypic overlap (Gasser 2015) (Figure 1.7).

Moderate risk factors: In addition to the *LRRK2* G2019S mutation, the only other major category of protein coding variants that confer a moderate risk of PD is mutations in glucocerebrosidase (GBA), in which carriers of risk factors have a cumulative risk of 11% to develop PD by the age of 85 (Rana, Balwani et al. 2013) (Figure 1.7).

GWAS loci: To date at least 28 genetic risk factor loci for PD have been identified, with each allele relatively common in the population, but only conferring a small increased risk for disease (1.2-1.5 fold) (Nalls, Pankratz et al. 2014). As with Alzheimer's disease, they are present in genes associated with a huge array of biological functions. In addition, as with other neurodegenerative disorders, even if all known risk factors are combined they only explain a small proportion of the heritability of PD (Keller, Saad et al. 2012) (Figure 1.7).

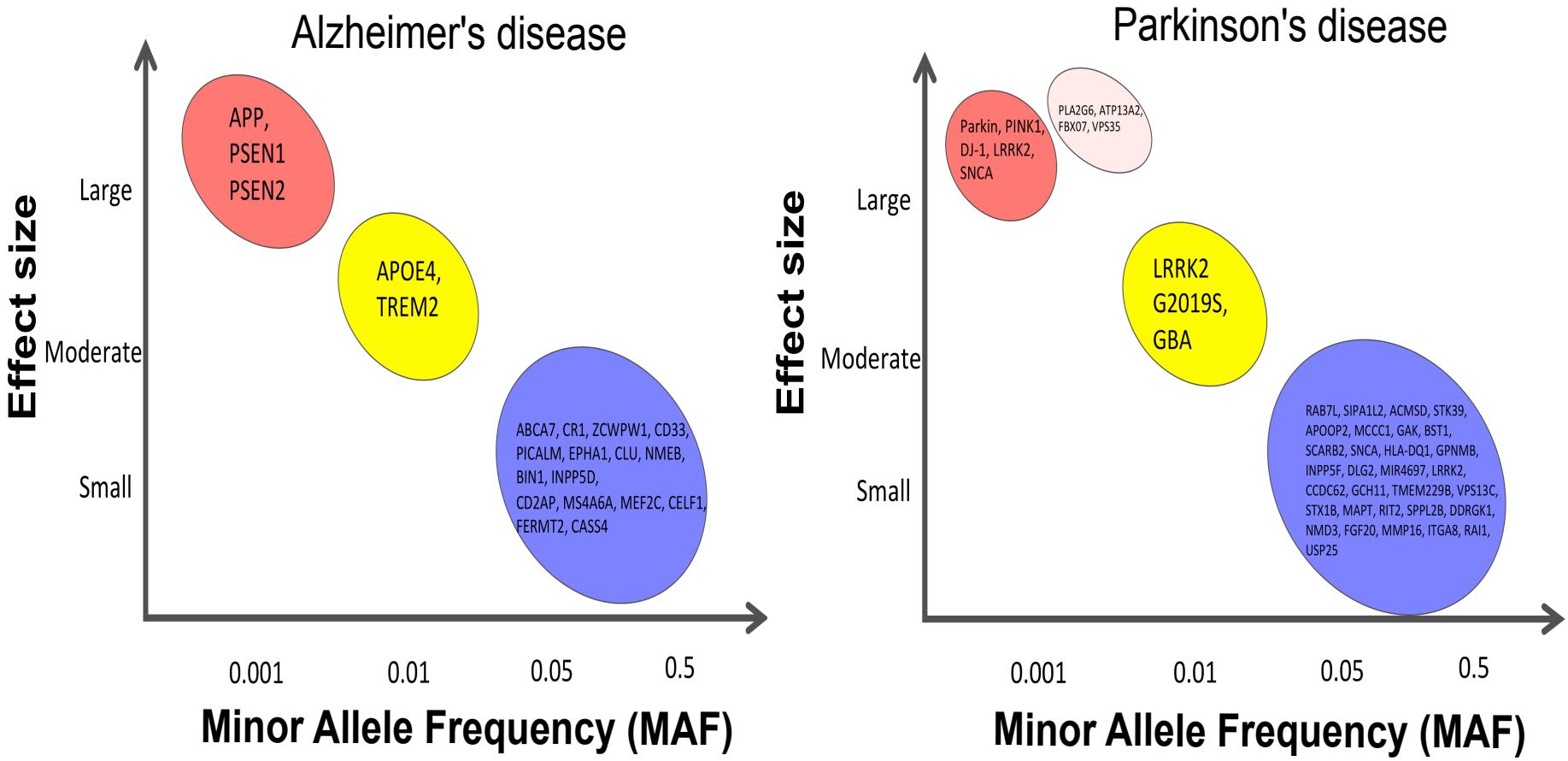


Figure 1.7. Genes associated with Alzheimer's disease and Parkinson's disease. Genes in which genetic variants are associated with both disorders are shown with Alzheimer's disease on the left and Parkinson's disease on the right. The relevant effect size and the average minor allele frequency within the population are shown. Genes in the red and pink circles are likely to cause monogenic and familial forms of disease. Those in the yellow and blue boxes act as disease risk factors.

1.11.3 Frontotemporal dementia – Amyotrophic lateral sclerosis (FTD-ALS)

***Monogenic alleles**:* It is now both clinically and genetically apparent that Frontotemporal dementia (FTD) and Amyotrophic lateral sclerosis (ALS) can no longer be considered to be distinct clinical or pathological entities. Clinically, up to 50% of ALS cases show some features of FTD, with 15% fulfilling diagnostic criteria (Ringholz, Appel et al. 2005), and conversely up to 27% of patients with FTD have features of ALS (Burrell, Kiernan et al. 2011). It is now apparent that FTD-ALS is most likely to have the highest genetic contribution of all neurodegenerative disorders, with an estimated 60% overall heritability for ALS observed in twin studies (Al-Chalabi, Fang et al. 2010, McLaughlin, Vajda et al. 2015), and with clinical studies showing that at least 10%-20% of all patients carry monogenic alleles, even in ostensibly sporadic cases (Al-Chalabi, van den Berg et al. 2017, Blauwendraat, Wilke et al. 2017, Turner, Al-Chalabi et al. 2017).

There have now been over 10 genes identified causing monogenic forms of FTD-ALS. These genetically mediated forms of disease also frequently show significant clinical pleiotropy, meaning that mutations in the same genes may cause significantly different phenotypes between individuals. For example, the *C9orf72* hexanucleotide repeat expansion (Renton, Majounie et al. 2011, Simon-Sanchez, Dopper et al. 2012) can cause both FTD, ALS or FTD-ALS.

It is now almost 25 years since mutations in superoxide dismutase (*SOD1*) were identified as the first familial forms of ALS (Rosen, Siddique et al. 1993). SOD1 is pivotal in controlling apoptotic signalling and reactive oxygen species (ROS) generation (Bunton-Stasyshyn, Saccon et al. 2015). Over the past 20 years point mutations in either the heterozygous or homozygous state have been identified as causing disease (Rosen, Siddique et al. 1993, Orrell, Marklund et al. 1997). The gene has now been established to cause about 12% of familial

cases of ALS, and 1% of sporadic cases (Chio, Traynor et al. 2008). The phenotype is most commonly an isolated ALS phenotype (Li and Wu 2016).

A hexanucleotide repeat expansion in Open reading frame 72 on chromosome 9 (*C9orf72*) was, in 2011, identified as the cause of up to 40% of familial FTD and 25% of familial ALS (Renton, Majounie et al. 2011, Majounie, Renton et al. 2012). Remarkably, *C9orf72* mutations also account for around 7% of all cases of sporadic ALS in Europeans (Majounie, Renton et al. 2012). Unlike most other highly penetrant mutations, a hexanucleotide expansion will not be identified by either SNP genotyping or current methods of NGS and at present requires a specific PCR followed by Southern blotting to detect and confirm the expansion. In addition, despite it's prevalence, the mechanism of disease remains largely unknown, with RNA toxicity, and proteotoxicity suggested (Gitler and Tsuiji 2016).

A further direct link between the genetic aetiology of disease and the neuropathological features of FTD-ALS was made with the discovery that heterozygous mutations in TAR DNA-binding protein (*TARDBP*), which encodes TDP-43 (a commonly observed protein in the nervous system of patients with FTD-ALS), was identified as a cause of ALS in 2008 (Neumann, Sampathu et al. 2006, Sreedharan, Blair et al. 2008). Mutations in *TARDBP*, like *SOD1*, predominantly cause a relatively pure ALS phenotype and rarely show a significant frontotemporal component. To date it is estimated that around 4% of familial ALS cases occur due to mutations in *TARDBP* (Chio, Calvo et al. 2012).

Mutations in the fused in sarcoma (FUS) protein accounts for a similar proportion of familial ALS as TARBP mutations (~4%) (Renton, Chio et al. 2014). The FUS protein shows significant functional homology with TDP-43, and FUS-immunoreactive cytoplasmic inclusions can be observed in cases carrying mutations. Despite the functional homology, the phenotype of *FUS* mutations is far broader than *TARDBP* with variants associated with

phenotypes including a rigid-akinetic parkinsonism that presents late in the course of FTD (Deng, Gao et al. 2014), a corticobasal syndrome (Baizabal-Carvallo and Jankovic 2016), and potentially even essential tremor (Merner, Girard et al. 2012).

There are also a multitude of additional genes that have been reported as causing monogenic forms of FTD-ALS, each with differing levels of supporting evidence. There are at least 4 genes reported as causing multisystem proteinopathies (MSP), which are pleotropic degenerative disorders affecting muscle, bone and the nervous system. These genes include the Valosin-containing protein (VCP) (in which mutations are responsible for 1-2% of familial ALS (Johnson, Mandrioli et al. 2010)), *hnRNPA1*, *hnRNPA2B1* (Kim, Kim et al. 2013) and *SQSTM1* (Fecto, Yan et al. 2011). Other genes such as homozygous mutations in optineurin (OPTN) have been observed in a Japanese family as causing ALS (Maruyama, Morino et al. 2010), and Profilin (PFN1) mutations have been observed segregating with ALS in several pedigrees (Wu, Fallini et al. 2012) though appear to be a very rare cause of ALS (Lattante, Le Ber et al. 2013). Finally, mutations in *DCTN1, CHMP2B, SETX, ANG, SPG11* and *VAPB* have all been observed in small pedigrees of FTD-ALS, or in related phenotypes and should be considered as appropriate for clinical testing (Renton, Chio et al. 2014).

GWAS loci*:* There have been a relative paucity of GWAS loci identified in FTD-ALS in comparison to the number of highly penetrant alleles, with only 6 GWAS loci having been successfully replicated; *UNC13A, SARM1, C21orf2, MOBP, SCFD1* and *TBK1* (Al-Chalabi, van den Berg et al. 2017) (Figure 1.8).

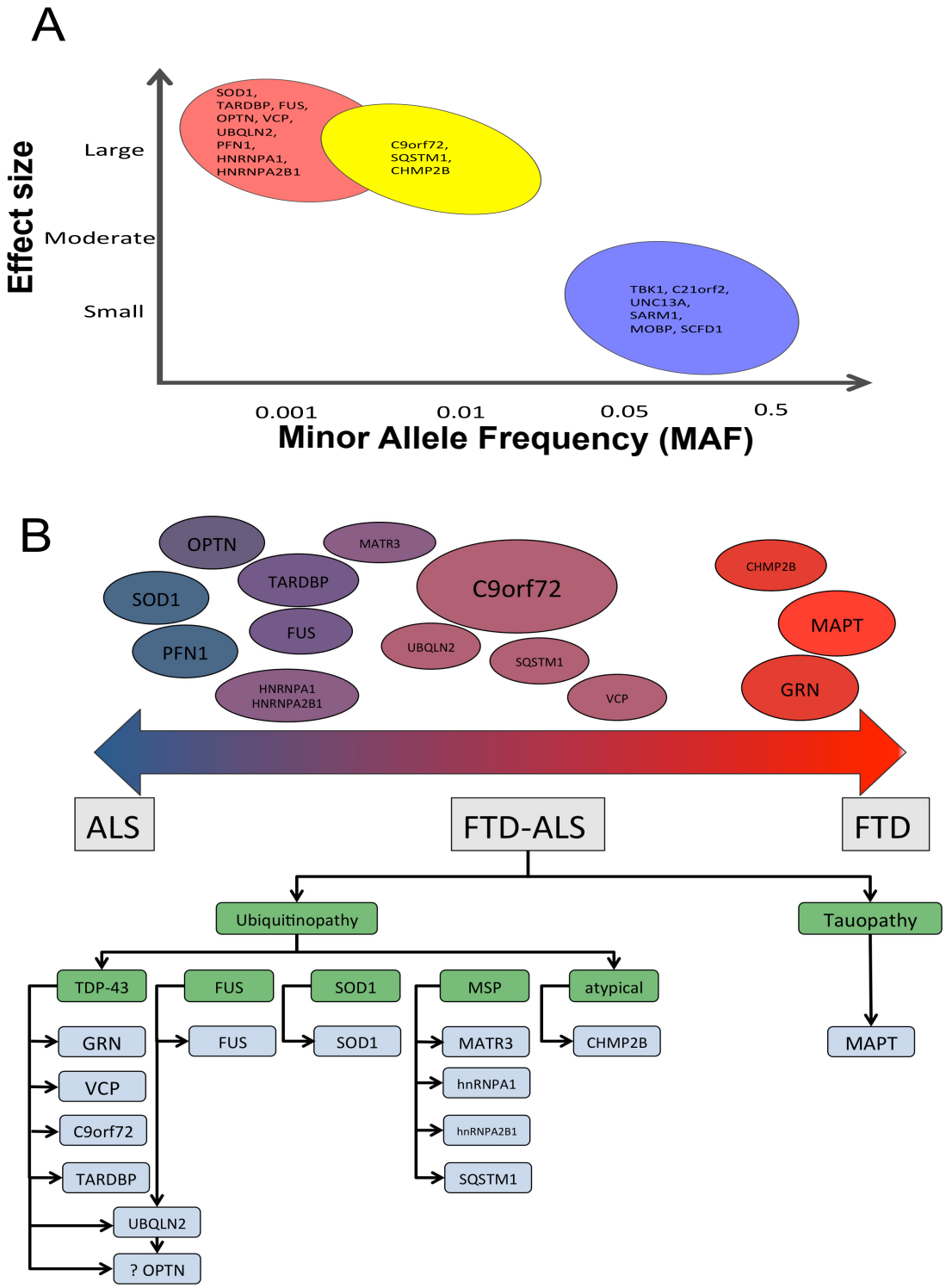


Figure 1.8. Genes associated with Frontotemporal dementia – amyotrophic lateral sclerosis (FTD-ALS) . A: Genes in which genetic variants are associated with FTD-ALS and their effect size and the average minor allele frequency within the population are shown (top). Genes in the red and pink circles are likely to cause monogenic and familial forms of disease. Those in the yellow and blue boxes act as disease risk factors. B: The top section shows where on the phenotypic spectrum of FTD-ALS monogenic forms of disease lie. The bottom section shows the molecular findings associated with each genotype of disease.

1.11.4 CJD

To date, only mutations in the Prion Protein gene (PRNP) have been conclusively shown to both cause monogenic forms of CJD and increase the risk of disease (Mead 2006). Within PRNP there have been a multitude of mutations that segregate in families and appear to act as highly penetrant alleles causing monogenic forms of disease (Mead 2006). The most common of these is the E200K mutation which has been described in over 100 cases causing CJD with a median age of onset of 58 years and a mean survival of 7 months (Bertoni, Brown et al. 1992, Collinge, Palmer et al. 1993, Inoue, Kitamoto et al. 1994, Chapman, Arlazoroff et al. 1996, Mead 2006).

Whilst the coding structure of the short, two-exon PRNP gene is highly conserved, a relatively common polymorphism at codon 129 between methionine and valine both increases disease risk and modifies the phenotype (Palmer, Dryden et al. 1991). Remarkably, the codon 129 polymorphism affects different forms of CJD in different ways. For example, all variant CJD cases (vCJD) have been homozygous for the methionine amino-acid at codon 129 (129M) (Zeidler, Stewart et al. 1997). Homozygous 129M and 129V individuals are also more susceptible to sporadic CJD (Palmer, Dryden et al. 1991).

Surprisingly, despite extensive efforts, there have been no other genes identified that either cause monogenic cases of disease nor act as significant risk loci for the disorder.

1.12 Somatic mosaicism

Genetically mediated disorders (other than cancer) have, until recently, been thought to exclusively result from genetic variants either inherited, or arising *de novo* within the germline. Such variants are therefore present in every cell within the human body.

However, genetic variants in the nuclear genome also arise during mitotic cell divisions in embryogenesis and human development due to the inherent error rate of DNA polymerases (Hoang, Kinde et al. 2016, Milholland, Dong et al. 2017). Such mutations are therefore only present in a portion of cells and are termed somatic mutations. The earlier in embryogenesis that a mutation occurs, then the higher the proportion of cells within the human body that would be expected to carry that mutation (assuming symmetrical cell division), and conversely, the later the mutation arose during development then the fewer the number of cells that would be expected to carry the mutation.

For example, the trilaminar blastocyst forms during week 3 of human embryological development, separating the developing embryo into three different germ layers (the ectoderm, mesoderm and endoderm) (Schoenwolf, Bleyl et al. 2015). Mutations arising prior to this stage are therefore likely to become present in all tissues within the body, albeit only in a proportion of cells (Figure 1.9). Conversely, those arising after this stage, would only be present within their subsequent cell lineages (e.g. tissues arising from that germ layer). Those arising even later during organogenesis, would only be present within that particular organ or cells derived from it.

Mutations that arise early in development, despite being present within all tissues are not easy to detect from blood; the most commonly sampled human tissue. This is because current sequencing methodologies and bioinformatic techniques are designed to detect germline variation (i.e. mutations present in 50% or 100% of reads – representing heterozygous or homozygous mutations respectively). In addition, mutations arising later in development, and which may only be present within a single tissue (such as brain) would be impossible to detect in blood (given that they occurred after the developmental separation of these tissue types) and would require direct sampling of brain tissue itself to detect the mutations, which is

particularly problematic for non *post-mortem* cases. Therefore, given the difficulty in detecting somatic mutations, it is conceivably possible that they may be an underappreciated cause of neurodegenerative disorders.

Outside the CNS, somatic mutations have already been identified as causing other forms of non-neurological disorders, albeit predominantly early onset neurodevelopmental abnormalities. For example, there are numerous reports of somatic mutations causing focal anatomical overgrowth syndromes such as Proteus syndrome which occur due to somatic mutations in *AKT1* (Lindhurst, Sapp et al. 2011), Maffucci syndrome caused by mutations in *IDH1* (Pansuriya, van Eijk et al. 2011) and Surge Weber syndrome occurring due to somatic mutations in *GNAQ* (Shirley, Tang et al. 2013). Somatic mutations do not have to be point mutations, and can also include somatic duplications of whole chromosomes (termed aneuploidy) (Pangalos, Avramopoulos et al. 1994), chromosomal regions (CNVs) (O'Huallachain, Karczewski et al. 2012), and somatic movement of retrotransposable elements (Callinan and Batzer 2006, Baillie, Barnett et al. 2011).

Within the CNS, it is now also becoming clear that somatic mutations arising through a variety of mechanisms may predispose to several, though predominantly young-onset neurological disorders. Firstly, somatic point mutations have recently been described as causing morphological abnormalities within the brain, with somatic mutations in *LIS1* and *DCX* identified as causing two differing forms of lissencephaly (an absence of the normal folding pattern of the brain (Gleeson, Minnerath et al. 2000, Sicca, Kelemen et al. 2003)). In both cases the mutations were able to be detected in blood suggesting that it occurred early in development before the formation of the trilaminar plate. More recently, a case of Dravet's syndrome was also detected due to a mutation in *SCN1A* in very early development, and which was present in the majority of cells (Vadlamudi, Dibbens et al. 2010).

Developmental somatic mutations arising later in development, and which are present solely within the brain, have been detected in association with several focal brain abnormalities. For example, somatic mutations in *MTOR* have been detected in numerous cases of focal cortical dysplasias (Poduri, Evrony et al. 2013, Jamuar, Lam et al. 2014, Lim, Kim et al. 2015).

Neurodegenerative disorders have only extremely rarely been described in association with early somatic mutational events. For example, somatic aneuploidy of Chromosome 21 (which carries the APP gene) has been observed in a case of Alzheimer's disease (Geller and Potter 1999), and a somatic point mutation in *PRNP* was observed causing CJD (Alzualde, Moreno et al. 2010). Finally, a single case of EOAD has been described secondary to a somatic mutation in *PSEN1* (Beck, Poulter et al. 2004). These isolated case reports suggest that somatic mutations may be a more prevalent cause of neurodegenerative disorders, but focused studies of human brain tissue cases on a large scale has yet to be undertaken due to technical limitations and access to appropriate tissue.

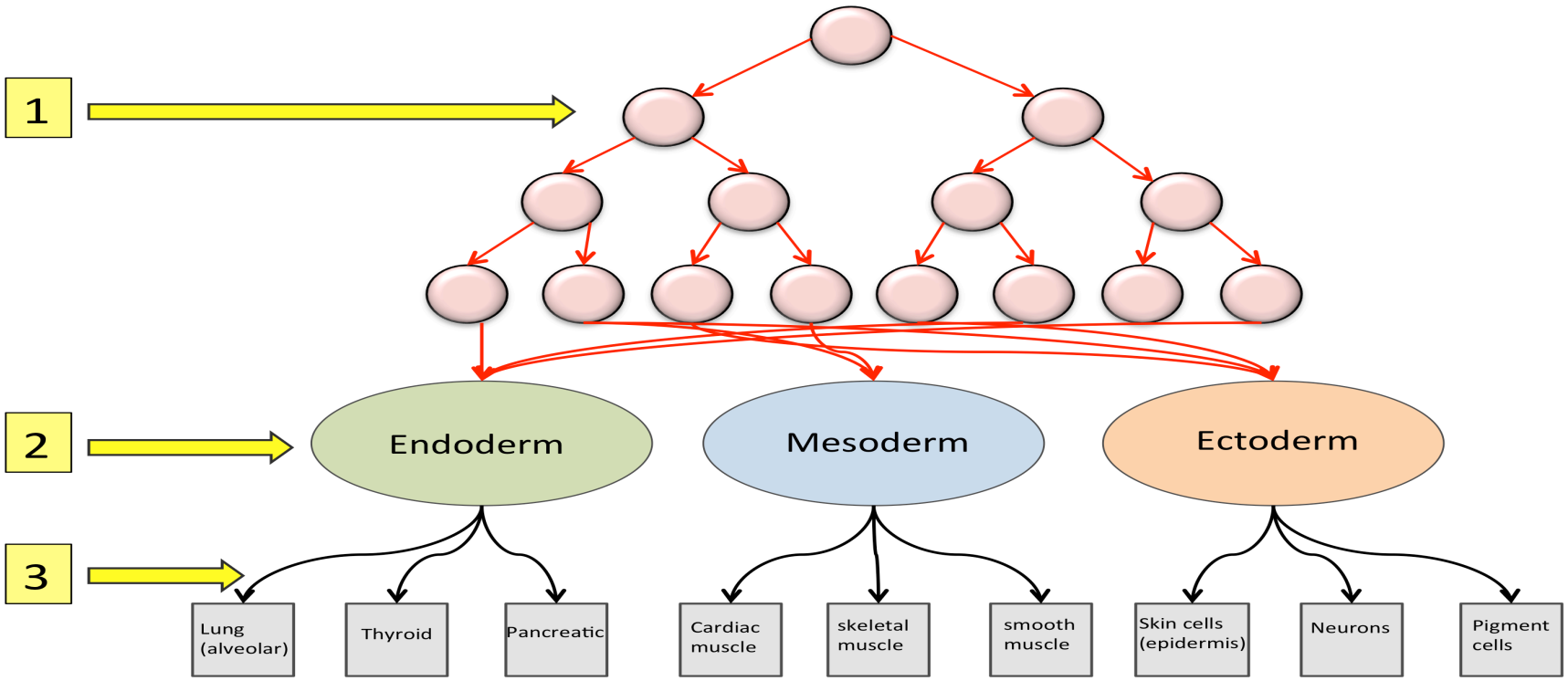


Figure 1.9. The mechanism and distribution of somatic mutations. Cellular proliferation from the initial zygote are show (top line). Subsequent daughter cells divide in a symmetrical pattern until cells coalesce into the three germ cell layers. Subsequently these germ cell layers go on to form the tissues shown in grey boxes. If a mutation were to occur at point 1 (yellow arrow), then approximately 50% of all cells would carry the mutation. If the mutation occurred at point 2 then all tissues derived from that germ cell layer could carry the mutation. If it occurred at point three then only the single tissue would carry the mutation.

1.13 Mitochondrial DNA and neurodegenerative disorders

Whilst somatic mutations in the nuclear genome have not been comprehensively studied in the context of neurodegenerative disorders, somatic mutations in the mitochondrial genome have been for over a decade.

The central pathway of oxidative phosphorylation in the human body is conducted through the mitochondrial respiratory chain, and many of the proteins within this chain are encoded by genes within mitochondrial DNA (mtDNA). Mitochondrial DNA, unlike nuclear DNA, is exclusively maternally inherited. From the initial pool of mitochondrial genomes in the fertilized oocyte stem all subsequent copies of the mitochondrial genome. Crucially, and with particular relevance to aging processes, the mitochondrial genome also undergoes life long replication, even in post mitotic cells (such as neurons and muscle).

Human mtDNA is a circular, double-stranded DNA molecule that is 16569 base pairs in length (Andrews, Kubacka et al. 1999). The two strands of mitochondrial DNA are termed the heavy (H-strand) and light (L-strand); the former being guanine rich and the latter cytosine rich. Unlike the nuclear genome which is only present in two copies in a post-mitotic cell, each cell contains between 100-10,000 mitochondria, with each mitochondrion containing between 2 and 10 copies of mitochondrial DNA (Wiesner, Ruegg et al. 1992). The overall cellular content of mtDNA generally correlates with the underlying energy demand of the cell.

Of the 37 genes encoded within the mitochondrial genome, 28 are situated on the H-strand and 9 are present on the L-strand. The most common function of these genes is to encode transfer RNA (22 genes) and polypeptide components of the mitochondrial respiratory chain (RC) (13 genes). Two genes encode a 16s rRNA (large ribosomal subunit), and one a 12s rRNA (small ribosomal subunit) (Andrews, Kubacka et al. 1999). Unlike nuclear DNA of

which 1-2% of the genome is coding (Venter, Adams et al. 2001), most mtDNA genes are contiguous, generally separated by one or two non-coding base pairs, resulting in approximately 93% of mtDNA bases encoding proteins. The only significant non-coding region is within the displacement loop (D-loop) which contains the site for mitochondrial DNA replication initiation (origin of heavy strand synthesis, OH) (Andrews, Kubacka et al. 1999). Just as a single human reference genome for nuclear DNA is used as a template to record and annotate genetic variation (Venter, Adams et al. 2001), so is a single mtDNA genome (the revised Cambridge Reference Sequence (rCRS)) (Andrews, Kubacka et al. 1999). Genetic variation within the mitochondrial genome is therefore generally described with reference to this sequence.

Mitochondrial DNA also differs from nuclear DNA in its organization. Nuclear DNA is associated with histone complexes which are responsible for the packaging of nuclear DNA into nucleoids (Bogenhagen 2012). In comparison, mtDNA is located on the inner mitochondrial membrane and associated with several proteins forming it's own nucleoid such as mtDNA polymerase gamma (encoded by *POLG*), mtDNA transcription initiating factors (encoded by *TFAM*) together with mtDNA binding proteins and helicases (encoded by *mtSSB* and *twinkle*) (Bogenhagen 2012). Mitochondrial DNA also differs from nuclear DNA in its base composition, encoding only two stop codons 'AGA' and 'AGG' in comparison to 4 encoded by nDNA (Temperley, Richter et al. 2010).

It is now well recognized that in tissues including the brain, impairment of mitochondrial respiratory chain function occurs with age (Trounce, Byrne et al. 1989, Bender, Krishnan et al. 2006) suggesting that a deficiency of oxidative energy production may contribute to the process of aging. In addition, it is now increasingly recognized that mitochondria also play an important role in several other key intracellular pathways such as calcium signalling, lipid

biosynthesis and programmed cell death (apoptosis) (van der Giezen and Tovar 2005), all of which are increasingly recognized as central processes in the development of several neurodegenerative diseases (Bossy-Wetzel, Barsoum et al. 2003, Celsi, Pizzo et al. 2009). Taken together, impairment of mitochondrial function is now placed at the heart of many established and emerging theories of aging and neurodegeneration (Beal 2005).

Whilst not all mitochondrial dysfunction stems from mutations in mitochondrial DNA (mtDNA), a significant proportion of age related mitochondrial dysfunction is hypothesized to be due to the accumulation of mitochondrial DNA mutations as a consequence of their on-going replication in life. Such mutations can be unique within individual cells and expand over time resulting in an impairment of cellular function.

In addition, recent studies have begun to change our understanding that all of the observed mtDNA mutations in aged tissues have exclusively developed in life. As we inherit several copies of the mitochondrial genome rather than a single copy, it is now apparent that a small number of the inherited copies may contain mutations in mtDNA (termed heteroplasmy), and these mutations may subsequently expand during life rather and did not form *de novo* during life (Ross, Stewart et al. 2013).

The idea that mitochondrial dysfunction in aging may, at least in part, result from an interplay of both heritable and acquired molecular damage to mtDNA, generates important questions regarding the role of mtDNA mutations in aging and disease, and the degree to which these factors may be heritable and modifiable.

1.13.1 Mitochondrial DNA and neurodegenerative disorders

Point mutations: MtDNA, like nDNA, is damaged by intracellular events such as nucleases, reactive oxygen species, and spontaneous hydrolytic processes. The single stranded nature of

mtDNA replication, and the lack of co-existent histone complex proteins were hypothesized to make it exquisitely more vulnerable to such insults (Caldecott 2008). However, oxidative damage as measured by 7,8-dihydro-8-oxo-deoxyguanosine (8-oxodG) predominantly induce G:C to T:A transversion mutations (Pinz, Shibutani et al. 1995) which are inconsistent with the pattern of transitional mutations most commonly seen in mammalian aged brain (Baines, Stewart et al. 2014). In contrast, mice harbouring mutations in DNA polymerase γ (*POLG)* predominantly generate transitional mutations within the germline (Trifunovic, Wredenberg et al. 2004) (Stewart, Freyer et al. 2008) and somatic point mutational spectra similar to those observed in aged human tissue (Baines, Stewart et al. 2014). These data suggest that impaired replication may be the primary event in somatic point mutation formation rather than oxidative damage. Until recently it was hypothesized that a contributing factor to mtDNA point mutation generation was a relative inefficiency of mtDNA maintenance. Over the last 2 decades it has however become clear that a more complex and efficient DNA maintenance system exists than previously suggested (Simsek, Furda et al. 2011), and therefore mtDNA replication errors remain the most likely primary mechanism inducing somatic mtDNA point mutations.

The recent observations of low level inherited heteroplasmy now suggest that mutations previously attributed as being somatically generated, and therefore occurring *de novo* within the brain and other tissues, may actually be clonal expansion of the inherited low level heteroplasmic variants not visible with previous sequencing technologies (Khrapko 2011, Payne, Wilson et al. 2013, Ross, Stewart et al. 2013). Determining whether similar mechanisms occur in brain is of paramount importance given the finding in *POLG* mutant mice that low-level inherited heteroplasmic variants can clonally expand over subsequent generations aggravating aging and inducing neurodevelopmental abnormalities (Ross, Stewart et al. 2013).

1.13.2 Heteroplasmic mitochondrial DNA mutations in Lewy body diseases

The observation of a mosaic pattern of respiratory chain deficiency in human PD post-mortem brain tissue (Itoh, Weis et al. 1996) provided the first evidence that mitochondrial DNA mutations may be implicated in respiratory chain impairment in PD. In 2006 Bender *et al* confirmed that PD patients had a greater proportion of Cytochrome c oxidase (COX) deficient neurons in the substantia nigra than controls (~3% vs 1%, p=0.003) (Bender, Krishnan et al. 2006), and although mtDNA deletion levels in both PD and controls was high in SN tissue homogenates (52.3% $\pm$ 9.3% vs 43.3% $\pm$ 9.3%) this did not reach statistical significance (p=0.06). Using single cell real time PCR, they showed that mtDNA deletion levels were higher in the COX-deficient neurons than those with normal COX activity (66.9% $\pm$ 19% vs 47.7% $\pm$ 24%, p<0.00001) underlying their respiratory chain defect. Subsequently using PCR cloning techniques within both COX deficient and non COX deficient cells, they detected discrete break points in mtDNA implying intracellular clonal expansion as the underlying mechanisms of deletion formation (Bender, Krishnan et al. 2006). Intriguingly, in the hippocampus, a region with relatively sparse Lewy body deposition, patients also had significantly higher deletion levels than controls (17.8% $\pm$ 12.9% vs 14.3% $\pm$ 6.7%, p = 0.0002). Taken together, these data imply that mtDNA deletions are higher in PD patients, possibly throughout the brain, somatically generated, and clonally expand in some cases to levels able to cause COX deficiency. Whilst this may imply that these deletions contribute to the pathophysiology of PD, this actually remains unproven. The presence of mtDNA deletions in the substantia nigra, rather than being deleterious, may actually trigger adaptive mechanisms such as an increase mtDNA copy number, improved respiration and greater striatal dopamine levels (Perier, Bender et al. 2013). Additionally, there is relatively little evidence to support the notion that mitochondrial respiratory chain (RC) deficiency promotes Lewy Body formation. At least two studies have shown that RC complex activity (as

determined by protein complex expression) is normal in Lewy body positive cells (Reeve, Park et al. 2012) (Muller, Bender et al. 2013).

Over the past 5 years, at least 2 studies have shown that mtDNA mutations, independent of mitochondrial respiratory chain function, correlate with Lewy Body formation, suggesting that they may be involved in the onset of Lewy Body pathology. In 2012, Lin *et al* used cloned-PCR of single cells to show that mean somatic mtDNA point mutations in neurons were 250 mutations/10^6 bp higher in early PD and incidental Lewy Body disease (a presumed precursor to PD) compared to both controls and late stage disease (Lin, Cantuti-Castelvetri et al. 2012). They found no difference between point mutations in established PD and controls. This suggests that mtDNA point mutations may predispose to the early propagation of Lewy bodies, potentially predisposing these neurons to early cell death enabling the subsequent survival of neurons without mtDNA mutations. More recently, LB positive neurons were shown to have greater levels of mtDNA deletions than LB negative neurons in post mortem cases ($40.5 \pm 16.8\%$ vs $31.8 \pm 14.3\%$, $p < 0.05$) (Muller, Bender et al. 2013). However, the majority of both LB positive and negative cells had similar deletion levels, arguing against causality, but implying that they may confer an alteration in susceptibility to LB development, though the mechanism remains unclear.

1.14 Mouse models of Lewy body diseases

Animal models offer a unique opportunity to study the fundamental mechanisms of neurodegenerative disorders to determine factors that may influence disease progression in a complex organism. Since the advent of the first transgenic mouse in 1974 (Jaenisch and Mintz 1974) manipulation of the murine genome to create mouse models of human disorders has become a primary paradigm to this effect.

Several aspects make mouse models a compelling strategy to study the interaction between mitochondrial DNA mutations and Lewy Body diseases. Several models are already available that offer validated and reproducible phenotypes and pathology that are robust and cost-effective compared to other organisms (Vandamme 2014). An ideal model will show an adequate replication of the clinical and, ideally, pathological phenotype, with age-progressive phenotypic features and neuropathology which are highly reproducible with minimal variability (Janus and Welzl 2010). However, caveats must always be borne in mind: For example, some forms of genetic manipulation are not suitable or reliable for behavioural phenotypes (Silva 1997), especially if they are not on a consistent genetic background (Wahlsten, Cooper et al. 2005).

1.14.1 Experimental design of mouse models of Lewy body diseases

When designing experiments involving transgenic models of human neurodegenerative disorders, it is therefore important to design and conduct appropriate experiments that are adequately powered and enable, as best as possible, not only a thorough understanding of the mouse model, but that also enable insight into human disease (Janus and Welzl 2010). It is therefore vital to design cognitive tasks that not only test cognitive functions in mice, but also require the utilization of neuroanatomical systems that are conserved between species; for example the hippocampus which is involved in memory formation across almost all mammals (Morris 1990, Squire 1992, Janus and Welzl 2010). As above, efforts must be made to ensure a concordant genetic background, and experimental design and crosses adjusted appropriately.

To evaluate a potentially novel phenotype, a battery of tests should be performed to characterise the physical and motor development and possible degeneration of mice. This may include specific motor and cognitive tests, or specific procedures which have an established

neurobiological or anatomical basis in an appropriately powered experiment (Scott, Kranz et al. 2008, Janus and Welzl 2010).

Not only does the nature of the experiment need to be carefully co-ordinated and controlled, but also the natural behaviour and housing conditions of the mice. For example, mice should ideally be tested at the same phase of their circadian cycle, and receive the same housing conditions (temperature, diet, light-dark cycle, exercise) together with a similar amount of handling (Janus and Welzl 2010). In addition, care should be taken to mix genotypes of mice within cages in order to reduce the possibility of litter or 'cage effects' arising due to slightly differing housing conditions and the risk of the Hawthorn effect and systematic subconscious bias when the experimenter is aware of the animal genotype (Wahlsten, Metten et al. 2003, Janus and Welzl 2010).

1.14.2 Transgenic models of Lewy body diseases

There are several transgenic models that aim to recapitulate Parkinson's disease by different mechanisms, but they can loosely be grouped as those that induce Lewy body deposition and those that do not induce Lewy body deposition (for example transgenic *Parkin* knockout mice (Perez and Palmiter 2005)).

Considering only those that induce α-synuclein deposition, these genetic models essentially vary by both the genetic composition of the SNCA gene (e.g. the type or presence of mutation), the type of upstream promoter, and the background strain of the mice. Variation in any of these components appears to modify the subsequent neuropathology to differing degrees. Whilst not exhaustive, several common transgenic mouse models are shown in table 1.2 (adapted from Chesselet *et al* (Chesselet and Richter 2011)).

Only a couple of transgenic mice that generate mitochondrial DNA mutations have been described, with all caused by mutations in mitochondrial DNA polymerase γ (*POLG*) resulting in impaired POLG activity and the development of somatic point mutations in the homozygous state (Trifunovic, Wredenberg et al. 2004, Kujoth, Hiona et al. 2005). In the heterozygous state mice with the same mutation are also able to generate low level mtDNA heteroplasmy and are able to transmit these mutations to their offspring, though with a significantly lower overall somatic and transmitted levels of heteroplasmy (Ameur, Stewart et al. 2011).

Taken together, these mice offer the possibility to begin to understand the interaction between the generation and inheritance of mtDNA heteroplasmy and Lewy body deposition *in vivo*, to begin to decipher the role of mtDNA mutations in Lewy Body diseases such as PD.

Table 1.2. Transgenic mouse models causing α-synuclein deposition. The relevant mouse model, promoter, and cellular expression of mutant or wild-type (WT) synuclein is shown. In addition, the presence of dopaminergic (DA) cell loss in the substantia nigra (SN) and additional non-motor features are shown. Finally whether the model is known to be levodopa (LD) responsive is provided.

Mouse model	Promoter	Cells in which transgene expressed	Corresponding human genetics	DA neuron loss in SN	Striatal DA loss	aSyn aggregation in SN	Motor phenotype	Non-motor phenotype					LD responsive	Reference
								Olfactory	Cognitive	GI	Autonomic	Anxiety		
WT SNCA	PDGFB	Neurons	Duplication / triplication	NP	Mild	Mild	Mild	-	Mild	-	-	-	-	(Masliah, Rockenstein et al. 2000)
WT SNCA	mThy-1	Neurons	Duplication / triplication	NP	Mod	Strong	Strong	Strong	Mod	Mod	Mod	Mod	Mod	(Rockenstein, Mallory et al. 2002)
A53T SNCA	mThy-1	Neurons	Dominant	NP	NP	NP	Mod	-	-	-	-	-	-	(van der Putten, Wiederhold et al. 2000)
A53T	mPrion	Neurons	Dominant	NP	NP	NP	Mod	-	Mild	-	-	Mod	-	(Giasson, Duda et al. 2002)
A30P	mThy-1	Neurons	Dominant	NP	NP	NP	Mod	-	Mild	-	-	Mod	-	(Freichel, Neumann et al. 2007)

Chapter 2

A genetic compendium of 1511 human brains available through the UK Medical Research Council Brain Banks

2.1 Aims

1. To define individual cases and controls possessing highly penetrant alleles causing monogenic forms of neurodegenerative disease.
2. To identify cases and controls carrying protein coding risk-factors for any major neurodegenerative disease.
3. To determine novel highly penetrant alleles or protein coding risk factors.
4. To use the MRC Brain Bank to help refute previous claims of variant pathogenicity.
5. To provide a data resource for researchers to genetically stratify their case and control tissue selection from the MRC Brain Bank Centres.

The work in this chapter was published in Genome Research in 2017: **Genetic compendium of 1511 human brains available through the UK Medical Research Council Brain Banks Network Resource. Keogh MJ**, Wei W, Wilson I, Coxhead J, Ryan S, Rollinson S, Griffin H, Kurzawa-Akanbi M, Santibanez-Koref M, Talbot K, Turner MR, McKenzie CA, Troakes C, Attems J, Smith C, Al Sarraj S, Morris CM, Ansorge O, Pickering-Brown S, Ironside JW, Chinnery PF. Genome Res. 2017 Jan;27(1):165-173.)

2.2 Introduction

The past 20 years have witnessed major advances in our understanding of the genetic landscape of common neurodegenerative disorders including Alzheimer's disease (AD), Parkinson's disease (PD), motor neuron disease (MND)/amyotrophic lateral sclerosis (ALS), frontotemporal dementia (FTD), and prion disorders such as Creutzfeldt-Jakob Disease (CJD). Highly penetrant alleles causing familial forms of these diseases have now been identified in approximately 50 genes (Reviewed in (Mead 2006, Chen, Sayana et al. 2013, Guerreiro and Hardy 2014, Lin and Farrer 2014)), and both genome wide association and candidate gene studies in sporadic cases have detected numerous genetic risk alleles (Reviewed in (Ramanan and Saykin 2013)).

The complex genetic architecture of neurodegenerative diseases presents several challenges. First, the implications of a novel genetic variant are particularly difficult to define (MacArthur, Manolio et al. 2014). The clinical phenotype often does not correspond to the anticipated neuropathology, and specific disease alleles are associated with a wide phenotypic spectrum (Caslake, Moore et al. 2008, Grau-Rivera, Gelpi et al. 2015). This makes it extremely difficult to establish whether a novel genetic variant is responsible for the disorder - an issue that is gaining importance, given the widespread use of clinical exome and genome sequencing.

Second, our understanding of the functional consequences of specific mutations is far from clear. This is partly because the effects of genetic variants can vary between tissue and cell type (Dimas, Deutsch et al. 2009), and also because several neurodegenerative diseases are modulated by age-related co-morbidity. Both factors are extremely challenging to model *in vitro* and in animals.

The analysis of *post mortem* human brain tissue circumvents some of these difficulties, but understanding the genetic background is of critical importance. This knowledge enables the stratification of mechanistic studies, and the dissection of disease pathways downstream from specific genetic lesions. It also avoids the inadvertent enrichment of a study group for specific genetic disorders, thus preventing results being inappropriately generalized across a broad disease category.

To address these issues, we performed exome sequencing and array-based gene dosage analysis of 1511 frozen brain tissue from a brain tissue resource: the UK Medical Research Council Brain Bank. Our analysis provides a genetic compendium for this accessible tissue resource, including 61 brains with highly penetrant mutations in 17 known disease genes, and 349 with established risk alleles for neurodegenerative disease. We also show that duplications of *PRPH* are enriched in AD, and show an association between *GRN* and DLB for the first time. All VCF files and associated metadata for all exome sequenced cases within the cohort is available for download on-line, enabling researchers to access fixed and frozen *post mortem* brain tissue for informed clinicopathological studies.

2.3 Methods

2.3.1 Clinical and pathological data

Ethical approval for the genetic analysis of *post mortem* brain tissue was obtained from the ethical review board of each participating centre. 1511 brains were selected from four centres within the MRC Brain Bank Network to include a representative range of neurodegenerative disorders, and controls who died from unrelated causes (Table 2.1 and Appendix 1). Demographic data (age of disease onset and death, disease duration and family history of disease), together with the *ante mortem* clinical diagnosis and *post mortem* neuropathological diagnosis were recorded for all cases.

The *ante mortem* diagnosis was defined by the treating clinician's working diagnosis in life. The *post mortem* diagnosis was consensus neuropathological diagnosis at death. A clinical vignette of the case, together with a synopsis of the neuropathological report and all quantitative neuropathological criteria (defined below), were also requested and provided when available. All data was reviewed, and each disease cohort included brains with an *ante mortem* clinical diagnosis consistent with a neurodegenerative disorder and clinical criteria fulfilling the specific neurodegenerative disease at the time of diagnosis. In cases in which a broad phenotype (e.g. 'dementia') was recorded, and neuropathology was consistent with a specific diagnosis, they were included within the cohort as defined by their neuropathology. Control cases were defined as those in which a Braak neurofibrillary tangle stage was 2 or lower and there were no other features suggestive of a neurodegenerative disease in either *ante mortem* data, or *post mortem* assessment. Cases ascribed to 'Vascular disease' were those in which either there was a history in life suggestive of cognitive impairment with at least one large territorial vascular anomaly such as stroke had occurred, and where Braak stage was greater than 2 or unknown. Those ascribed to 'vascular/control' showed no large territorial infarcts, but have evidence of small vessel disease together with no history suggestive of *ante mortem* cognitive impairment and a Braak stage of less than 2. Given the known neuropathological and genetic overlap, cases with either frontotemporal dementia (FTD), motor neuron disease (MND)/amyotrophic lateral sclerosis (ALS), were included in a combined FTD-ALS sub-group.

All rare neurological diseases, those without *ante mortem* data, or those in which the neuropathology did not fulfil criteria for a specific diagnosis were included into the 'other' category and defined by a discrete descriptive term, individual to each case, summarizing the overall clinical and neuropathological phenotype (Appendix 1-2).

Quantitative neuropathological scores and stages (e.g. Braak neurofibrillary tangle stage (Braak, Alafuzoff et al. 2006)) were also supplied for each case as appropriate. For non-Creutzfelt-Jakob Disease (CJD) cases, additional quantitative neuropathological criteria were obtained when available. These included: Thal phase for Amyloid beta (Thal, Rub et al. 2002), Braak neurofibrillary tangle stage (Braak, Alafuzoff et al. 2006), CERAD score (Mirra, Heyman et al. 1991), ABC score (Montine, Phelps et al. 2012), Braak stage for alpha-synuclein (Braak, Sandmann-Keil et al. 2001), McKeith Lewy body stage (McKeith, Dickson et al. 2005), Cerebral Amyloid Angiopathy (CAA) staging (Attems, Jellinger et al. 2011), the presence of TDP-43 pathology and stage (Mackenzie, Neumann et al. 2011), and presence of 3R or 4R tauopathy, fused in sarcoma (FUS) or ubiquitin staining.

For cases conforming to diagnostic criteria for Creutzfeld Jakob Disease (CJD) (Budka, Aguzzi et al. 1995), the clinical subgroup (sporadic, familial, variant, or iatrogenic) was also recorded (Appendix 2), together with the presence of cerebrospinal fluid (CSF) 14-3-3 protein, and the s100b protein together with MRI imaging features were also requested and provided where available.

2.3.2 Molecular genetic and bioinformatics analysis

2.3.2.1 DNA extraction

Tissue samples (20-30mg) were extracted from Cerebellum (n=1323), Cerebral cortex (lobe undefined) (n=54), Frontal cortex (n=30), Temporal cortex (n=8), Occipital cortex (n=4), Basal ganglia (n=4), Caudate (n=3), *Substantia Nigra* (n=2), Muscle (n=1), and undefined brain region (n=81). Automated DNA extraction was performed using a DNA extraction robot (Qiasymphony SP robot; Qiagen, Hilden, Germany). Tissue was lysed in 180 µl of ATL buffer (Qiagen, Hilden, Germany) and 20 µl of Proteinase K (Qiagen, Hilden, Germany). Lysates were incubated overnight at 56 °C and at 900 rpm before being loaded onto the

Qiasymphony robot. Subsequent extraction was performed using the Qiasymphony DNA mini kit reagents (Qiagen, Hilden, Germany), as per manufacturers protocol. DNA yield was measured using the Nanodrop-8000 Spectrophotometer (NanoDrop Technologies). DNA extraction was performed by NBTR staff at each centre and not entirely by myself.

2.3.2.2 Exome sequencing and analysis

Genomic DNA was fragmented, exome enriched and sequenced (Nextera Rapid Exome Capture 62Mb and HiSeq 2000, 100 bp paired-end reads) by staff at AROS (Copenhagen, Denmark). Bioinformatic analysis was performed using an in-house pipeline including alignment (human reference genome hg19, UCSC) using Burrows-Wheeler Aligner (BWA) (Langmead, Trapnell et al. 2009). Samples with a mean coverage of less than 30 were excluded.

Variant calling in remaining samples was performed using FreeBayes (Garrison 2012). Subsequent analysis was restricted to on-target homozygous, heterozygous, and compound heterozygous variants with a minimum read depth of 10 in any case or control, and base quality score of 20 within the cohort. Further analysis was performed on frameshift, in-frame indel, or start/stop codon change, missense variants, and splice site loss variants with a minor allele frequency <0.5% or <3% (for familial forms of disease or risk factor alleles respectively) in the 1000 Genome Project Database (Genomes Project, Abecasis et al. 2012), European American cases from the NHLBI ESP exomes database (Exome_Variant_Server 2016), and ExAC server (Lek, Karczewski et al. 2016), using Qiagen Ingenuity Variant Analysis software. Variants in genes known to cause familial forms of neurodegenerative disease (Appendix 3) with the appropriate inheritance pattern were assessed in all cases according to both the 2015 American College of Medical Genetics (ACMG) (Richards, Aziz et al. 2015), and the MacArthur Criteria (MacArthur, Manolio et al. 2014), irrespective of

phenotype or neuropathological diagnosis (Figure 2.2). The ACMG criteria remain the world's leading criteria for the interpretation of sequencing variants in a clinical context, and the MacArthur criteria are a second stringent criteria that propose that researchers summarise and present a spectrum of evidence in order to truly attribute pathogenicity of identified alleles in sequencing studies. The same criteria were also applied to a specific assessment of a small number of cases with rare neurological diseases that was undertaken independently (adult onset gangliosidosis, cerebello-olivary atrophy, chorea-acanthocytosis, Huntington-like disease, Kuf's disease, infantile onset mitochondrial disease, neurodegeneration with brain iron accumulation and primary familial basal ganglia calcification) (Appendix 4).

To exclude inadvertent duplicate samples, variant calls with a base quality score of 30 and read depth of 10 were converted into PLINK v2.050 (Purcell, Neale et al. 2007) binary genotyping format using in-house scripts. Pairwise relationships were subsequently determined using KING (Manichaikul, Mychaleckyj et al. 2010) allowing for the existence of population structure. Duplicate samples were determined by kinship estimates restricted to first degree or second-degree relatives only using the –related option.

Associations at the gene level: Initially, known pathogenic and likely pathogenic cases of disease were removed. Known risk factor associations at the gene level were then tested through case burden testing using Sequence Kernel Association Test (SKAT-O) (Wu, Lee et al. 2011) against a control group of control cases aged 55 or over with no features of neurodegenerative disease (n=244) within Ingenuity Variant Analysis. Subsequent allele level associations were assessed through Chi-squared testing with a Mantel-Haenszel correction when a count of zero occurred in either group. Subsequent analysis for novel associations between disease cohorts and controls were performed against an expanded cohort of controls

across the whole age range (n=380 including n=244 aged controls, plus n=136 young controls and vascular controls).

2.3.2.3 Array genotyping and analysis

Array genotyping was performed on all samples using the Illumina HumanOmniExpress-12 BeadChip array (Illumina Inc., San Diego, CA) and subsequently utilized to perform Identity by Descent (IBD) to determine related individuals within the dataset. The calling of SNP genotypes and CNVs was performed by Dr Ian Wilson (University of Newcastle) as described below.

We followed a modified method of that described by Cooper et al (Cooper, Shtir et al. 2015) utilizing Log-R-ratio (LRR) and B-allele frequency (BAF) scores from raw allele probe intensities. In this process, samples and SNPs with call rates < 95% were first removed. No subsequent quality control (QC) was applied based on the Minor Allele Frequency (MAF) as SNVs with rare alleles still provide LRR information even when monomorphic. Subsequently several levels of QC were undertaken, with three QC statistics calculated: mean LRR; Derivative log-ratio spread; and GC Wave Factor; each using custom in-house scripts in R (*www.R-project.org*). These were calculated simultaneously and samples failing any one were excluded from subsequent analysis.

Mean Exclusion: LRR-means for all samples passing initial QC were calculated, and samples outside the upper or lower bounds (1.5x the interquartile range) of the overall mean were excluded.

Derivative log-ratio spread: derivative log-ratio spread (DRLS) was calculated as the standard deviation (SD) of the differences between successive array SNV markers (according

to genome position), divided by the square root of 2. A sample exclusion threshold of 3.5 SD above the cohort mean was applied.

GC wave exclusion: In keeping with the methods described by Cooper et al (Cooper, Shtir et al. 2015), guanine and cytosine (GC) wave was evaluated to remove samples with substantial wave intensity fluctuations. GC score was calculated by the 'total wave factor' (as described previously (Diskin, Li et al. 2008)) and is defined by the median of the absolute median absolute deviation (MAD) of LRR, with sign determined by correlation with GC percentage. Samples with a total wave factor > 3.5 SD above the mean were excluded.

Plate exclusion: As plate effects can induce some of the strongest bias in CNV estimates, plates in which 40% of samples failed QC, or where the number of failed samples was > 3.5 SDs above the study mean were removed.

PCA correction: 10% of post QC autosomal SNVs were selected and a PCA was performed in R using the bigPCA R library. The 6 largest linear components were used to correct the data. In addition, X-chromosomal calls were corrected based on autosomal SNVs.

QuantiSNP and further filtering: Using the corrected LRR from the original BAF samples, quantiSNP was performed as previously described (Colella, Yau et al. 2007). Subsequently, SNVs spanning telomeric or centromeric regions were excluded together with genes in major histocompatibility complexes (MHC) and immunoglobulin genes (Cooper, Shtir et al. 2015). In addition, samples for which the total number of CNVs exceeded 3 SDs above the mean number of CNVs per sample were excluded from the analysis.

Restriction to rare CNVs: To exclude our analysis to rare CNVs, we excluded any CNV which had an 80% or greater overlap with CNVs known to occur in greater than 1% of individuals from reference databases (MacDonald, Ziman et al. 2014).

Additional allele-specific genotyping: Specific genotyping to determine the *APOE* genotype was also performed independently because of poor exome and SNV array coverage. APOE genotyping was performed by competitive allele-specific PCR, using KASP^{TM} genotyping assays (LGC, UK) for both rs7412 and rs429358. Subsequent genotype data was converted into APOE e4 allele status (Weisgraber, Innerarity et al. 1982). Full methods for the KASP^{TM} genotyping platform are available from LGC (http://www.lgcgenomics.com/genotyping/kasp-genotyping-reagents/).

Screening for the *C9orf72* hexanucleotide repeat expansion was also performed as described (Renton, Majounie et al. 2011). Positive results were confirmed by Southern blot (Tomasetti, Vogelstein et al. 2013) allowing an estimate of the repeat size. All wet-lab *C9orf72* work was performed by Professor Stuart Pickering-Brown's laboratory at Manchester University.

2.3.2.4 Data access

The sequencing data from this study have been submitted to the European Genome Archive (*https://ega-archive.org/studies/EGAS00001001599)* alongside appropriate metadata following ethical approval by the Research Ethics Committee (REC) proportionate review process through the Health Research Authority (HRA) Approvals system (REC reference 17/EE/0033).

2.4 Results

2.4.1 Demographics

The mean age at death across all cases was 68.3 (sd=17.9) and ranged from 3 to 103 (Table 2.1 and Appendix 1). Only 2.8% of cases had a family history of disease, as defined by a recorded first or second degree relative with the same or a related neurodegenerative disease (Table 2.1). The CJD cohort comprised sporadic (n=126), variant (n=62), and other subtypes

as shown in Appendix 2. All demographic data is available online (*https://ega-archive.org/studies/EGAS00001001599*).

Table 2.1. Clinical data for the 1511 brains in this study. All ages are given in years. Key; AD – Alzheimer's disease, CBD – Corticobasal degeneration, CJD – Creutzfeldt-Jakob Disease, DLB – Dementia with Lewy Bodies, FTD-ALS – Frontotemporal dementia – Amyotrophic Lateral Sclerosis, HD – Huntington's Disease, MSA – Multiple System Atrophy, PD – Parkinson's Disease, PSP – Progressive Supranuclear Palsy. All cases in which they did not conform to a specific group were included in 'other'. NA – Not available, FH – Family history.

	Number of cases	Age onset (years)		Age of death (years)		Female		Male		FH of disease	
		Age	SD	Age	SD	n	%	n	%	n	%
AD	289	65.5	10.4	77.6	11.6	151	52.2%	138	47.3%	11	3.8%
CBD	14	60.1	11.4	69.6	12.4	4	21.4%	10	61.5%	0	0.0%
CJD	239	52.6	19.9	53.5	19.5	111	46.7%	128	53.3%	4	1.7%
Control	368			63.2	18.8	131	35.5%	237	64.2%	0	0.0%
Control (High Braak)	38			87.6	8	24	63.2%	14	36.8%	0	0.0%
DLB	58	66.7	8.4	76.7	7	22	37.9%	36	62.1%	2	3.4%
FTD-ALS	252	59.5	11.7	64.6	11.6	103	40.9%	149	59.1%	14	5.6%
HD	7	59.8	13.2	66.7	10.3	3	42.9%	4	57.1%	1	14.3%
MSA	10	58.4	6.5	68.2	9.1	1	10.0%	9	90.0%	0	0.0%
Other disorders	80	57.4	25.3	72.4	18.3	35	43.8%	45	56.3%	9	11.3%
PD	39	59.9	10.9	72.3	9.2	11	28.2%	28	71.8%	0	0.0%
PSP	17	68.7	12.1	77.1	10.3	9	52.9%	8	47.1%	0	0.0%
Cerebrovascular disease	65	79.9	8.6	85.7	6.1	32	49.2%	33	50.8%	1	1.5%
Cerebrovascular disease / AD	17	82.5	6.3	84.5	6.6	10	58.8%	7	41.2%	0	0.0%
Vascular disease / Control	18	81.2	13.7	89.8	7.2	8	44.4%	10	55.6%	0	0.0%
Total	**1511**	**59.6**	**17.7**	**68.3**	**17.9**	**655**	**43.3%**	**856**	**56.7%**	**42**	**2.8%**

2.4.2 Monogenic alleles

The mean depth of sequencing of the whole exome was 51.9 (sd=12.9), and 49.1 (sd=13.8) in genes associated with neurodegenerative disease (Appendix 3 and 5). Following quality control (QC) 1241 brains had data available using all 4 genetic approaches (exome sequencing, array SNV genotyping, *C9orf72* PCR and *APOE* genotyping, Figure. 2.1, Appendix 6).

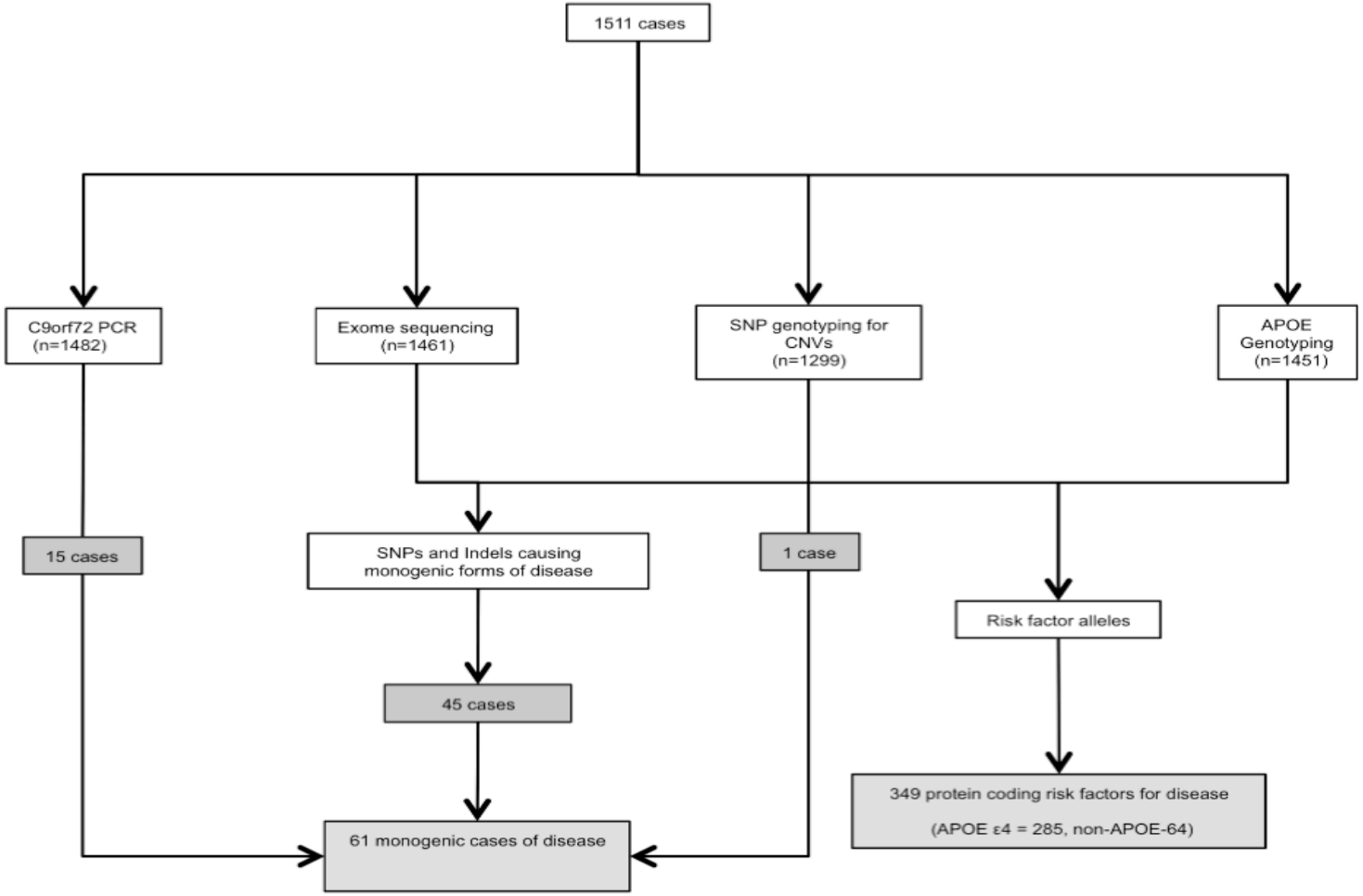


Figure 2.1. The molecular genetic approach and the data passing quality control (QC) in 1511 post mortem brains. The algorithm for sequencing approaches, variant interpretation and key results are shown.

Single heterozygous SNVs, small insertions or deletions in genes known to cause common neurodegenerative diseases*:* We initially identified rare heterozygous variants present in genes known to cause autosomal dominant forms of common neurodegenerative diseases (Appendix 7). In total, 313 variants were seen in 817 cases (Figure. 2.2), with 261 (83.4%) absent from the 1000G (Genomes Project, Abecasis et al. 2012), 212 (67.7%) absent from ESP6500 databases (Exome_Variant_Server 2016) and 149 (47.6%) absent from the ExAC database (Lek et al. 2016). 62.0% of variants (n=194) were initially determined to be of uncertain significance given the lack of any prior available data, and 39 variants in 57 of the 1461 cases (3.9%) undergoing exome sequencing were considered to be either pathogenic or likely pathogenic according to the American College of Medical Genetics criteria for variant pathogenicity (Richards, Aziz et al. 2015) (Figure. 2.2).

Following comparison and correlation of variants with the case diagnoses in our dataset, 28 pathogenic and likely pathogenic heterozygous variants were confirmed in 38 cases (3.46% of 1099 cases of disease) (Figure.2.3, Appendix 7-12). Importantly, cross-correlating our genomic data with clinical and pathological data enabled the reclassification of 149 of the 194 variants that were initially classified as being of uncertain significance. The initial classification of the 194 variants as being of uncertain significance was because, prior to our study, they had either never been reported in human reference databases, nor had had they been described in either clinical cases of disease, nor studied in vitro or in vivo. However, by detecting 149 of these in control subjects or in cases with a discordant phenotype, we were able to conclude that they were likely to be benign (Appendix 13-18). Forty-five variants in 52 cases remained classified as being of uncertain significance after initial assessment (3.6% of 1461 cases undergoing exome sequencing), with these variants present in phenotypically appropriate cases for the genotype or in young control subjects (Appendix 19-24). Over 50% of these variants are entirely novel (n=23), having never been previously detected in human

subjects not in any other cases in our dataset, nor the three on-line databases accessed in this study (n=69,660) (Appendix 19-24).

Homozygous or compound heterozygous SNVs, insertions or deletions in genes known to cause common neurodegenerative diseases: Homozygous or compound heterozygous variants known to cause disease were seen in only four cases, and were considered to be pathogenic (Figure. 2.1, Appendix 25-27).

Single heterozygous SNVs, small insertions or deletions in genes known to cause rare neurodegenerative diseases*:* In 10 cases with rare neurological disorders, pathogenic variants were detected in three brains: Case 24, adult onset ganglosidosis (*HEXA*; p.R499C); Case 29, Kuf's disease (*DNAJC5*; p.L115R); and Case 44; chorea- acanthocytosis (*VPS13A*; p.L1841*/p.W2347fs*36) (Figure. 2.1 and Appendix 4 (Gene lists/panels), Appendix 7-12 (dominant disorders), Appendix 25-27 (Recessive disorders)).

2.4.3 C9of72 screening

The *C9orf72* hexanucleotide repeat expansion screening was performed by Professor Stuart Pickering-Brown's laboratory, detecting positive repeat expansions in 15/1482 brains (Appendix 28) (1% of the total disease cohort, and 6.14% of all FTD or ALS cases (n=244) which had *C9orf72* analysis). Of note, the 'FTD-ALS' cohort combined all isolated FTD cases, isolated MND cases and FTD-ALS cases into one cohort due to the genetic overlap between these disorders (Appendix 28). These findings were confirmed by Southern blot within Professor Pickering-Brown's laboratory.

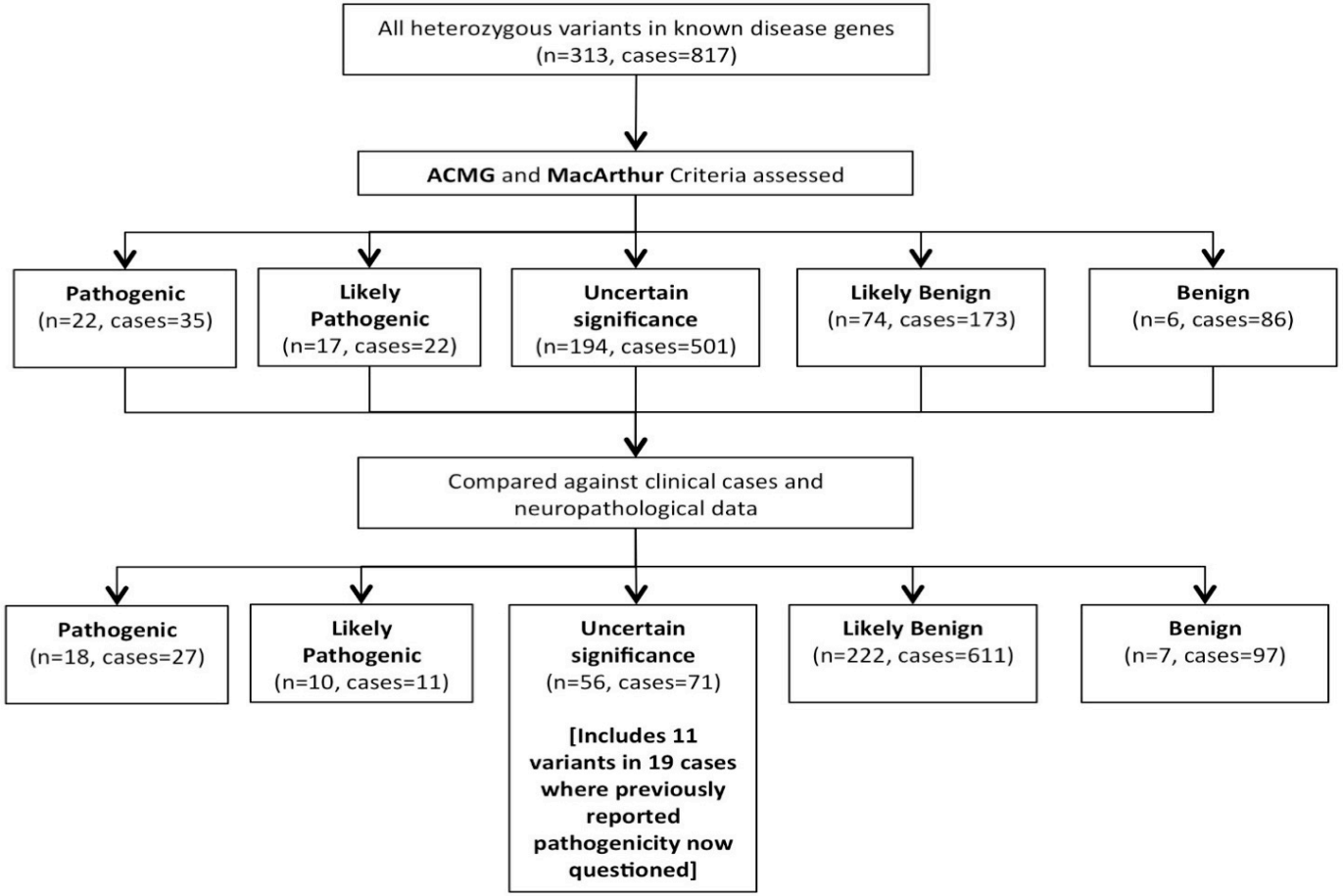


Figure 2.2. The clinical evaluation of 1461 post mortem brains. Assessment of heterozygous variants in genes known to cause familial forms of neurodegenerative disease. All variants were initially assessed against the American College of Medical Genetics (ACMG) criteria and all evidence relating to pathogenicity was recorded according to the guidelines of MacArthur et al (MacArthur, Manolio et al. 2014). Comparing variants to clinical and neuropathological data in their respective cases enabled a significant refinement of likely pathogenicity, and in particular, an increase in the number of variants likely to be benign, allowing the re-classification of previous variants considered pathogenic.

2.4.4 Identity by descent (IBD) analysis

IBD using KING revealed 9 pairs of individuals with cryptic 1st-4th degree ancestry (Appendix 29 & 30). This led to a molecular diagnosis of spinocerebellar ataxia (SCA) because one of two first-degree relatives within the cohort was diagnosed with SCA7 in parallel study, leading to the diagnosis in the remaining relative.

2.4.5 Copy number variation (CNV) analysis

Rare CNVs spanning the coding region of any of the 49 common neurodegenerative disease genes included in our study (Appendix 3) were seen in 1.46% of cases (n=19) (Appendix 31). Only one CNV occurred in a gene known to convey pathogenicity through copy number variation (Case 61: Alzheimer's disease with an *APP* triplication, Appendix 31) and was considered to be pathogenic. Copy number gains were also seen in *APOE* (n=1), *DAO* (n=1), *SCARB2* (n=1), and *SPG11* (n=1) and copy number loss in *LRRK2* (n=1) and *PARK2* (n=4). Nine cases showed a copy number gain in *PRPH* comprising three different CNVs, with a mean length of 13.8kb (SD=0.9kb) (Figure. 2.4. & Appendix 31). Six of the 9 cases had AD, resulting in *PRPH* copy number gains having a significant association with AD vs all controls with CNV data (n=383, p=0.015, Fisher's exact test), and when compared to all forms of neurodegenerative disease (n=1060, p=0.002). The three other brains with *PRPH* copy number gains showed: pathologically atypical tauopathy with AD like features; ALS, and control case with an age at death two standard deviations below the mean of the confirmed AD cases with PRPH copy number gains.

2.4.6 Genetic risk factors

Alleles previously described as conferring a risk of AD, DLB, FTD-ALS or PD were identified (Appendix 32), with 38 cases possessing established risk factors in *TREM2, GBA, LRRK2* (Appendix 32). We also compared the total burden of rare coding variants in genes

known to act as risk factors for disease for each case cohort compared to aged (> 55 years) neuropathologically normal controls. We confirmed associations between AD and variants in *TREM2* (cases; n=20 (6.9%), controls n=8 (3.3%); p=0.017), DLB with *GBA* (cases; n=8 (13.8%), controls: n=10 (4.1%); p=0.0075) and *TREM2* (cases; n= 7 (12.1%), controls: n=8 (3.3%); p=0.0084), and FTD-ALS with *SQSTM1* (cases; n= 29 (13.2%), controls: n=16 (6.6%); p=0.0076; Appendix 32). We also identified a putative novel risk factor for disease: the *TREM2* p.R62H variant with DLB (p=0.0024, OR;3.2 [95% CI 1.7-27]) which was seen in 7/58 cases (12.1%) and 5/244 controls (2.05%). We also found supporting evidence that the *SQSTM1* p.P392L variant is a risk factor for FTD-ALS (p=0.05). APOE e3/4 alleles and e4/4 alleles were strongly associated with AD (p<0.0001), and DLB (p<0.001) (Appendix 33). In total, 349 cases or controls had at least one protein-altering variant in a gene previously associated with an appropriate neurodegenerative disease (Figure. 2.2).

***Novel risk-factor associations*:** To determine whether variants in any established risk factor gene conferred risk for alternative forms of neurodegenerative disease we performed burden testing of all risk factor genes across all large cohorts (AD, FTD-ALS, CJD, DLB) compared to an extended version of all controls (n=380). Cases possessing highly penetrant pathogenic variants (Figure.2.1, Appendix 34), were removed from each group before testing. Variants in *GRN* were associated with DLB (p=0.005) with 8 variants seen in 12.1% of patients and only 3.2% of controls, though no single allele was significantly associated with disease (Appendix 34 & 35).

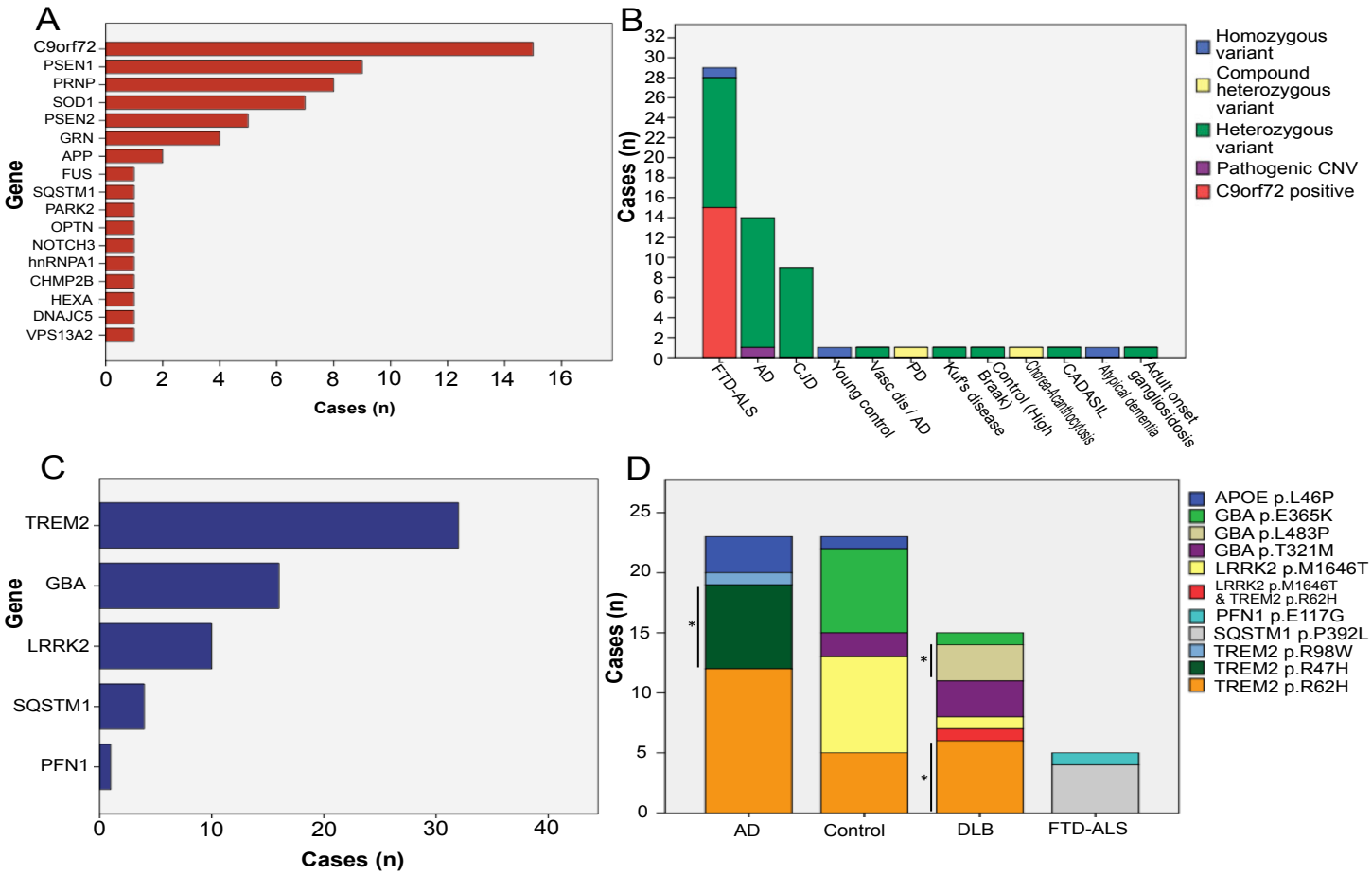


Figure 2.3. The frequency of known pathogenic mutations and risk alleles in 1511 *post mortem* brains. (a) Number of brains with mutations in known familial neurodegenerative disease genes; and (b) The nature of pathogenic and likely pathogenic variants in each disease category. (c) Number of brains with a known genetic risk factors for neurodegenerative disease; and (d) Known genetic risk factors in each sub-group (*APOE e4* alleles excluded but including rare coding variants). Variants which were significantly associated with each disease cohort are highlighted by an asterisk ($p<0.05$, Fisher's exact test). All variant associations with disease can be seen in Supplementary Tables 7 and 8**.**

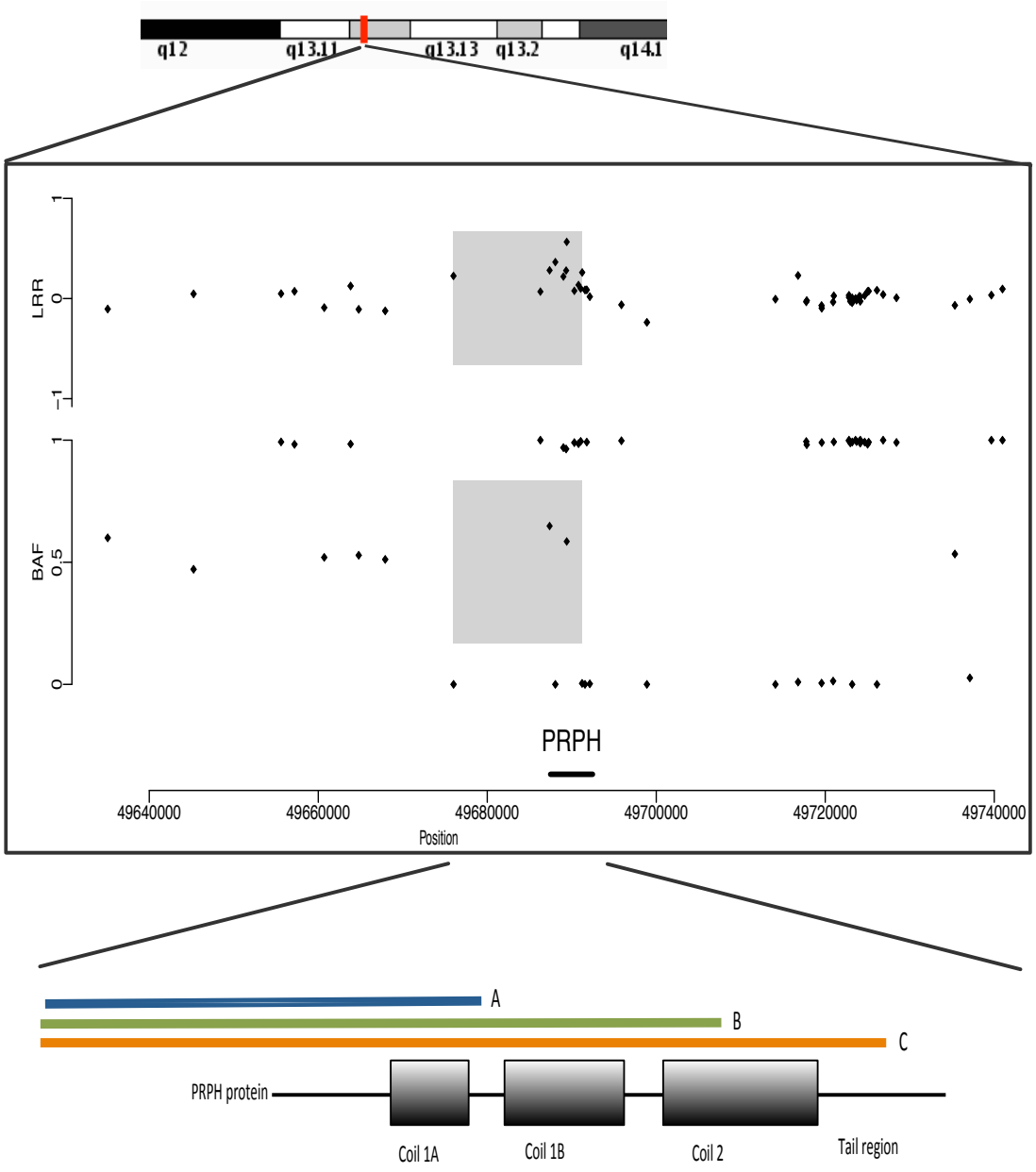


Figure 2.4. Copy number gains in *PRPH*. Upper: q13 region of chromosome 12 is shown. Middle: Individual points show the Log R Ratio (LRR) and beta allele frequency of SNV genotyping SNPs between base position 49640000 and 49740000 on Chromosome 12 (GRCh37 build). Positive LRR values in the grey shaded region are consistent with a copy number gain in that region (> 3 SDs above the mean of the cohort, see Supplementary Methods). The *PRPH* gene is within that genomic region is identified on the X-axis. Lower: 5' position of the final SNP in the copy number region in the 3 different CNVs identified and their relevant position in the PRPH protein (orange, green, blue), highlighting that three different CNV end positions were identified in patients within the study within the PRPH gene.

2.4.7 Genotype-phenotype correlations

There was no association between the presence of variants in any risk-factor gene (*TREM2, GBA, LRRK2, SQSTM1, GRN*) and any neuropathological assessment score or any clinical parameter (age of onset, death or disease duration). When sub-stratifying by disease, *TREM2* variants in AD cases were associated with a shorter disease duration (Kaplan-Meier Mantel-Cox p=0.027, Appendix 36), but there were no other associations between any risk factor variants and any clinical or pathological feature.

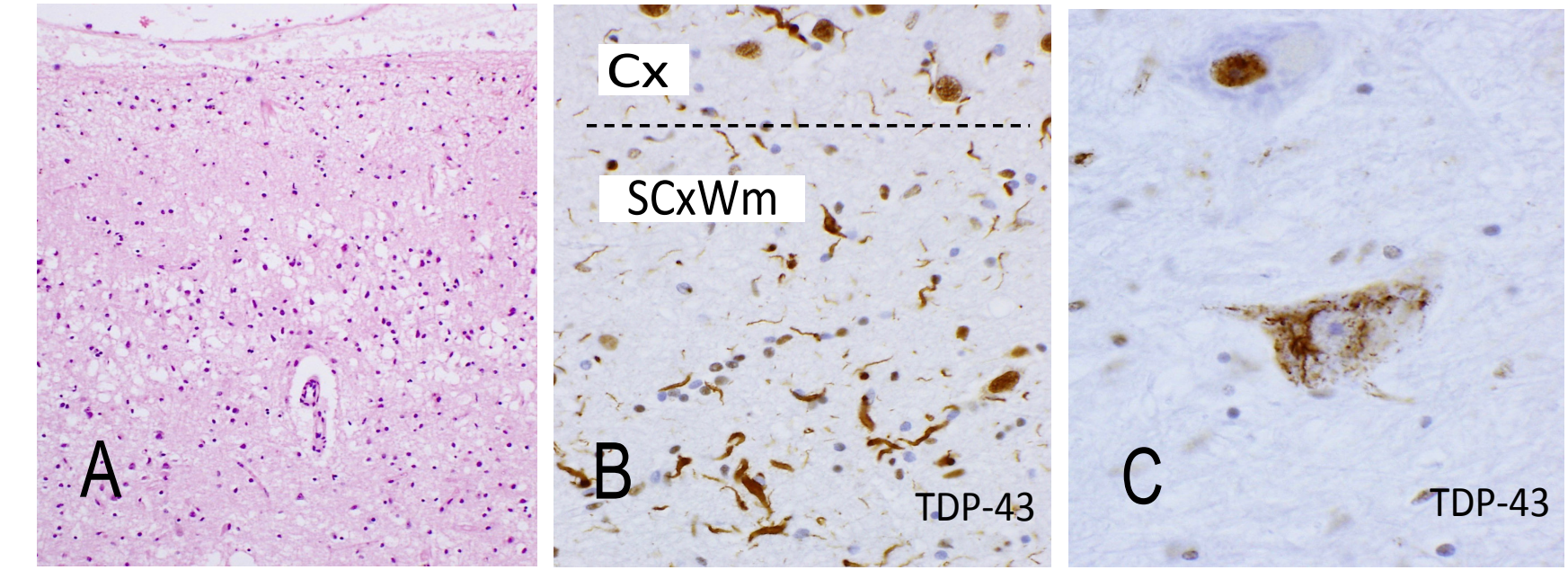


Figure 2.5. Neuropathology in the brain donors found to harbour an *OPTN* (A, B) and *HNRNPA1* mutation (C). A and B - Severe degeneration with spongiosis of the primary motor cortex (haematoxylin and eosin, and atypical, predominantly subcortical white matter TDP-43 proteinopathy, A) in the patient with the *OPTN* mutation. In contrast, in C, the *hnRNAPA1* mutation was associated with a lower motor neuron phenotype with classical skein-like cytoplasmic mislocalisation of TDP-43 (C, hypoglossal nucleus). (A x200, B and C x400). (Cx = cortex, SCxWm = subcortical white matter) (These images were provided by Dr Olaf Ansorge at Oxford University).

2.5 Discussion

Using a multi-modal sequencing approach we have identified highly-penetrant alleles causing disease in 61 individuals (heterozygous alleles n=40, homozygous or compound heterozygous alleles n=5, *C9orf72* hexanucleotide repeat expansion n=15, and pathogenic CNV n=1) at *post mortem* with clinical and pathological evidence of neurodegenerative disease and a further 349 individuals with established protein-coding heterozygous genetic risk factors for neurodegeneration (Figure. 2.1). This integrated genomic and pathological dataset provides a unique resource to genetically stratify human brain tissue research, and will ensure that future work using this tissue resource is not confounded by hitherto unknown genetic factors. The genetic data is available to all researchers, enabling the rapid identification and provision of brain tissue containing genetic variants which emerge as likely risk factors for neurodegeneration in future years, but are currently not know to be important.

Using current guidelines (Richards, Aziz et al. 2015), 62% (n= 194) of rare variants (MAF < 0.5%) were initially classified as being of uncertain significance. Just under half of these variants were not present in international reference databases (n=95, 48.9%), and with no previous clinical or functional *in vitro* modelling to support or refute pathogenicity (MacArthur, Manolio et al. 2014) (Appendix 7-28). However, analysis of our clinical and pathological data showed that 149 of these 194 variants (76.8%) were found in control brain donors, or in patients with incongruous clinical phenotypes, enabling the reclassification of 149 variants as likely to be benign. These findings illustrate the importance of studying *post mortem* tissue in late-onset disorders where the clinical phenotype alone may be unreliable.

In addition to the reclassification of uncertain variants as benign, our data also shows the converse; that specific alleles previously dismissed as benign, may be pathogenic. For example the *PSEN2* p.D439A variant was first described in patients with AD, but

subsequently found in asymptomatic controls. This led some to conclude that *PSEN2* p.D439A was not pathogenic (Sassi, Guerreiro et al. 2014). In this study we saw *PSEN2* p.D439A in both an AD and a control brain. However, the control brain had moderate Alzheimer-type pathology (Braak stage 3/4), supporting possible pathogenicity with age related penetrance. Our findings therefore show that the presence of a variant in a clinically defined 'healthy control' is not reliable, and it is unwise to reject a putative pathogenic variant as benign without a comprehensive assessment of the neuropathology in an ostensibly healthy control subject. For neurodegenerative diseases, this means a *post mortem* examination.

Related to this, another unique and valuable feature of the resource is the inclusion of young clinically unaffected controls (n=135 under age 55). Genotyped control brains provide an opportunity to explore the preclinical effects of neurodegenerative risk alleles, before significant cell loss and secondary pathological change. In this study, this was only seen once, in a young control harboring a known pathogenic mutation (compound heterozygous *PARK2* mutations). Larger cohorts of control brain tissue are likely to yield more examples, and thus provide valuable insight into the pre-clinical pathology of late onset neurological disorders, potentially revealing early targets for therapeutic intervention.

We also found exceptionally rare genotypes of disease, including a homozygous p.R217* *OPTN* mutation and a novel p.G216R mis-sense variant in *HNRNPA1,* both causing ostensibly sporadic cases of motor neuron disease (Figure. 2.5). Other than the original case series (Maruyama, Morino et al. 2010), this is one of only a handful of known pathogenic homozygous *OPTN* cases, and *HNRNPA1* variants have failed to be detected in several large clinical cohorts (Seelen, Visser et al. 2014). Identification of brain tissue from these rare forms of disease provides an invaluable opportunity to further define the molecular pathogenesis of these disorders and establish genotype-phenotype relationships. For example,

although both mutations were associated with a TDP-43 proteinopathy (Figure. 2.5), the neuropathological pattern was atypical in the homozygous p.R217* *OPTN* patient, demonstrating dominant subcortical and glial proteinopathy.

Analyzing the coding region variants across the entire cohort enabled a compilation of a comprehensive database of genotype-phenotype correlations, and facilitated the detection of rare phenotypes and pre-symptomatic cases of disease. This is perhaps best highlighted by the homozygous *SOD1* p.D91A variant, which classically presents with distal lower motor-neuron features and progresses slowly (Andersen, Forsgren et al. 1996). Here we detected the same homozygous variants in a patient with atypical dementia and some clinical features overlapping with multiple system atrophy (MSA). In addition, we found a 35-year-old clinical control male with a previously identified pathogenic compound heterozygous mutation in *PARK2*. Subsequent repeat neuropathological assessment showed no evidence of neuronal loss within the substantia nigra. Given his age at post mortem, it is not clear whether this individual was pre-symptomatic, or whether the alleles have a reduced clinical penetrance in this case. Larger cohorts of neuropathologically stratified cases are likely to assist in determining pathogenicity in such cases.

We also made number of novel observations. Firstly, p.R62H *TREM2* variant appears to be a risk factor for DLB (p=0.0024, OR;3.2 [95% CI 1.7-27]). This variant was previously reported in association with AD (Jin, Benitez et al. 2014), and although these findings need replication, these observations further emphasize the overlapping genetic aetiology of DLB and AD, and endorse the view that *TREM2* increases the risk of developing several neurodegenerative diseases (Rayaprolu, Mullen et al. 2013). Second, we show the first association of *GRN* variants with DLB. Again, this will require replication, but suggests a novel genetic overlap between DLB and FTD. Thus, a combination of genetic risk factors

contributes to the prevalence of DLB, which affects ~5% of those aged over 80 years, but rarely affected more than one individual in a family (Meeus, Theuns et al. 2012). Finally, we provide the first evidence of an association between a 13.8kb copy number gain within *PRPH* and AD, found in 6 patients with typical AD pathology (Figure. 2.4). Although point mutations have previously been found in association with ALS/MND (Gros-Louis, Lariviere et al. 2004), the recent finding that peripherin may regulate amyloid metabolism (Muresan, Villegas et al. 2014) provides a potential mechanism linking amyloid deposition and the first possible overlapping genetic mechanism between ALS/MND and AD.

There are several explanations for the absence of a relevant family history in most (73.8%) of the 61 likely genetically determined cases we identified (n=45), including early death in previous generations, incomplete penetrance, *de novo* mutation, and false paternity. Although we cannot resolve this issue for individual cases, our findings do highlight the importance of considering highly penetrant single gene defects in patients with an ostensibly sporadic neurodegenerative disease.

Although a number of known genetic causes of neurodegeneration were not included in our analysis, including large insertions, deletions and large-scale inversions, these are only likely to be relevant for a small minority of cases. In addition, being an analysis of a legacy collection, newer neuropathological assessment criteria were not available in brains collected some time ago. However, the fine detailed genetic characterization of specific alleles and contemporary histopathological assessments will be added in the near future. This genetic compendium of 1511 brains includes 40 brains with highly penetrant SNVs in 13 known disease genes, 5 cases with homozygous or heterozygous variants in 4 genes, together with pathological expansions of *C9orf72* in 15 cases, and a copy number gain of *APP* in one case, resulting in a total of 61 cases of genetically determined disease. The compendium also

includes 349 cases and aged controls with established risk protein coding alleles for neurodegenerative disease (Figure. 2.1). The whole dataset will facilitate variant interpretation, and provides a framework for the genetic stratification of future human *post mortem* analysis studies.

Chapter 3

Oligogenic genetic variation in 980 neurodegenerative diseased brains

3.1 Aims

1. To determine the frequency of apparent oligogenic variation within the MRC Brain Bank.
2. To determine the nature of oligogenic variation with reference to clinical diagnostic panels for major neurodegenerative disorders.
3. To determine whether oligogenic variation represents a burden of rare but non-highly penetrant alleles, or whether it represents the combination of a highly penetrant allele with rare but benign variation.
4. To determine whether oligogenic variation alters any parameter of disease phenotype.

This work is awaiting publication as: **Oligogenic genetic variation in 980 post mortem human brains.** MJ Keogh, W Wei, J Aryaman, Wilson I, K Talbot, MR Turner, C-A McKenzie, C Troakes, J Attems, C Smith, S Al Sarraj, CM Morris, O Ansorge, S Pickering-Brown, N Jones, JW Ironside, PF Chinnery. Journal of Neurology Neurosurgery and Psychiatry (JNNP). 2017. In Press.

3.2 Introduction

Genetic variation in over 50 genes contributes to the risk of developing neurodegenerative diseases (Tsuji 2010, Guerreiro, Bras et al. 2015, Singleton and Hardy 2016). Some of the known risk alleles are common in the general population, raising the possibility that multiple interacting genetic variants might enhance the risk of developing disease or modify the disease phenotype. In keeping with this, some familial cases of frontotemporal dementia – amyotrophic lateral sclerosis (FTD-ALS) appear to have a greater 'burden' of variants when compared to controls (van Blitterswijk, van Es et al. 2012), which may explain an earlier age of onset (Cady, Allred et al. 2015). However, it is currently not clear whether this also occurs in non-familial cases of FTD-ALS or other major neurodegenerative disorders, where previously reported associations could either be due to a single highly penetrant monogenic alleles co-associated with benign non-functioning variants, or whether there is a genuine synergistic interaction between two or more functional genetic variants.

3.3 Methods

We studied the following: Alzheimer's disease, AD, n=277; FTD-ALS n=244; Parkinson's disease or Dementia with Lewy Bodies, PD-DLB, n=97; and neuropathologically normal controls, n=362 (Appendix 37), with 97.2% of all individuals studied having no family history of a neurodegenerative disorder (Appendix 37). Demographic data including the age of disease onset and death, disease duration and family history of disease, together with the *ante mortem* clinical diagnosis and *post mortem* neuropathological diagnosis were available (See Chapter 2.3.1) (Table 3.1).

Exome sequencing was restricted to on-target homozygous, heterozygous, and compound heterozygous variants with a minimum read depth of 10, and base quality score of 20 across the 980 subjects, where the variant allele frequency (VAF) was <5% in the Exome

Aggregation Consortium (ExAC) (Lek, Karczewski et al. 2016). Ingenuity Variant Analysis™ was used to study 49 genes known to be associated with neurodegenerative disorders (Appendix 38). The 49 genes were subsequently grouped into six gene panels: AD panel (n=8), PD-DLB panel (n=16), Full FTD-ALS panel (n=28), Medium FTD-ALS panel based on that previously described (Singleton and Hardy 2016) (n=12), and a small FTD-ALS panel as previously described (van Blitterswijk, van Es et al. 2012) (n=5), together with the entire panel (n=49 genes). All panels were filtered for variants present at VAF ≤1% and ≤5%. *C9orf72* genotypes (See Chapter 2.4.3, Appendix 28) were incorporated as stated.

Pathogenic (P) or likely pathogenic (LP) variants were defined using American College of Medical Genetics (ACMG) criteria (Richards, Aziz et al. 2015) as described (See Chapter 2.3.1) together with known genetic risk factors. Other variants identified as Benign (B), Likely Benign (LB) or of Uncertain Significance (US) based on ACMG criteria, and the remaining variants (VAF 0.5-5% in monogenic genes, or non-risk factor variants in risk-factor genes) were annotated as unclassified (UC). Oligogenic individuals were defined as those who had two or more non-synonymous, frameshift, or stop-loss or gain inducing point mutations in the relevant panel (as stated), or those who tested positive for the *C9orf72* hexanucleotide repeat expansion plus had at least one of the point mutation within the panel.

3.4 Results

Across the entire cohort of 980 subjects we observed a total of 57 genetic variants in the AD gene panel, 141 variants in the primary FTD-ALS gene panel, and 140 in the PD-DLB gene panel. Six AD cases (2.17%) had >1 variant in the AD panel, and 19 cases (7.79%) of primary FTD-ALS had >1 variant in the primary FTD-ALS panel. These proportions were no different to control subjects (Control subjects for the AD panel: 5/362, 1.38%, p=0.545; and full FTD-ALS panel: 26/362, 7.18%, p=0.14) (Appendix 39). In contrast, twenty-three cases of PD-

DLB (23.71%) had >1 variant in PD-DLB genes which was greater than controls (Controls: 37/362, 10.22%, p=0.004) (Appendix 39).

Based on ACMG criteria for pathogenicity (Richards, Aziz et al. 2015) (see Chapter 2.3.2.2) only 3 individuals in the entire study (0.38% of n=980) harboured >1 pathogenic, likely pathogenic or known risk factor for a neurodegenerative disease. One patient with DLB (age of onset at 65, and death at 70) who had a *LRRK2* p.M1646T mutation associated with PD, and a *TREM2* p.R62H mutation associated with AD (Chun and Fay 2009). A 70 year old patient with Alzheimer's disease had a *PSEN2* p.L204I mutation and the *TREM2* p.R62H risk factor. A third patient who had early onset PD (onset age 40) due to a compound heterozygous mutation in *PARK2* (p.G430D/pR275W) also had the p.R98W *TREM2* possible risk factor for AD (Guerreiro, Wojtas et al. 2013), but displayed no evidence of any amyloid deposition at *post-mortem* (Appendix 40 & 41).

We observed a significant enrichment of highly penetrant alleles or risk factors within 'oligogenic' cases in all disease cohorts (Appendix 42). In FTD-ALS, 11 of the 19 oligogenic cases contained one highly penetrant allele or risk factor within the primary panel, giving the presence of oligogenic variation a positive predictive value (PPV) to identify an individual as someone carrying a pathogenic mutation or known risk factor at 57.9% (95% CI: 33.5-79.8%) (Appendix 43). We subsequently varied the panel size to reflect published approaches (van Blitterswijk, van Es et al. 2012, Singleton and Hardy 2016), raised the MAF to 5% within each panel, and removed *C9orf72* data from the analysis. In all of these permutations there was a significant over representation of highly penetrant allele or risk factor carriers within the oligogenic cohort (Figure 3.1, Appendix 43). The same enrichment for highly penetrant alleles within 'oligogenic' cases was seen in the AD panel at 1% (PPV 100%, 95% CI:54.1-100.0%) and PD-DLB panel (PPV: 43.5% (95%CI: 23.2-65.5%) (Appendix 43).

We then investigated whether the enrichment of monogenic alleles or risk factors within oligogenic cases was due to a greater overall background mutation rate in these individuals as previously suggested in some genotypes of PD (Lubbe, Escott-Price et al. 2016), but found no evidence of such an association (Appendix 44).

Finally, we removed all cases possessing a highly penetrant allele or risk factor, and compared remaining oligogenic cases of PD-DLB and FTD-ALS with controls (n=362). Based on this analysis there was no difference in either the proportion of 'oligogenic' cases, nor the mean pathogenicity defined by both SIFT or Polyphen2 score (Appendix 45-51), between any study group. We also observed no difference in the age of onset, age of death, or disease duration between remaining oligogenic cases compared to those with <2 variants (Appendix 52-53), including the *C9orf72* expansion in the presence of additional variants (Appendix 54).

Table 3.1. Clinical and demographic data for the major cohorts within the study. 'Oligogenic' was defined by the presence of >1 variant within the relevant disease panel at <1% MAF in the ExAC database. Monogenic or cases harbouring genetic risk factors were defined as outlined in the Supplementary Methods as previously(van Blitterswijk, van Es et al. 2012).

.

Phenotype	Number of cases	Male (number)	Female (number)	Mean age onset (years) (SD)	Mean age death (years) (SD)	Number with FH	Cases with highly penetrant allele or RF	Oligogenic cases (N (%))	Oligogenic cases possessing a penetrant allele or RF (N(%))	Fisher's test (p-value)
Control	362	232 (64.1)	130 (35.9)	N/A	63.3 (18.8)	N/A	N/A			
FTD-ALS	244	143 (58.6)	101 (41.4)	59.4 (11.8)	64.6 (11.7)	14	33	19 (7.78%)	11 (57.9%)	0.0001
AD	277	131 (47.3)	146 (52.7)	65.4 (10.2)	77.7 (11.7)	11	36	6 (2.17%)	6 (100%)	0.0001
DLB	58	36 (62.1)	22 (37.9)	66.7 (8.4)	76.7 (7.0)	2	16	25 (25.78%)	10(62.5%)	0.0007
PD	39	28 (71.8)	11 (28.2)	59.9 (10.9)	72.3 (9.2)					

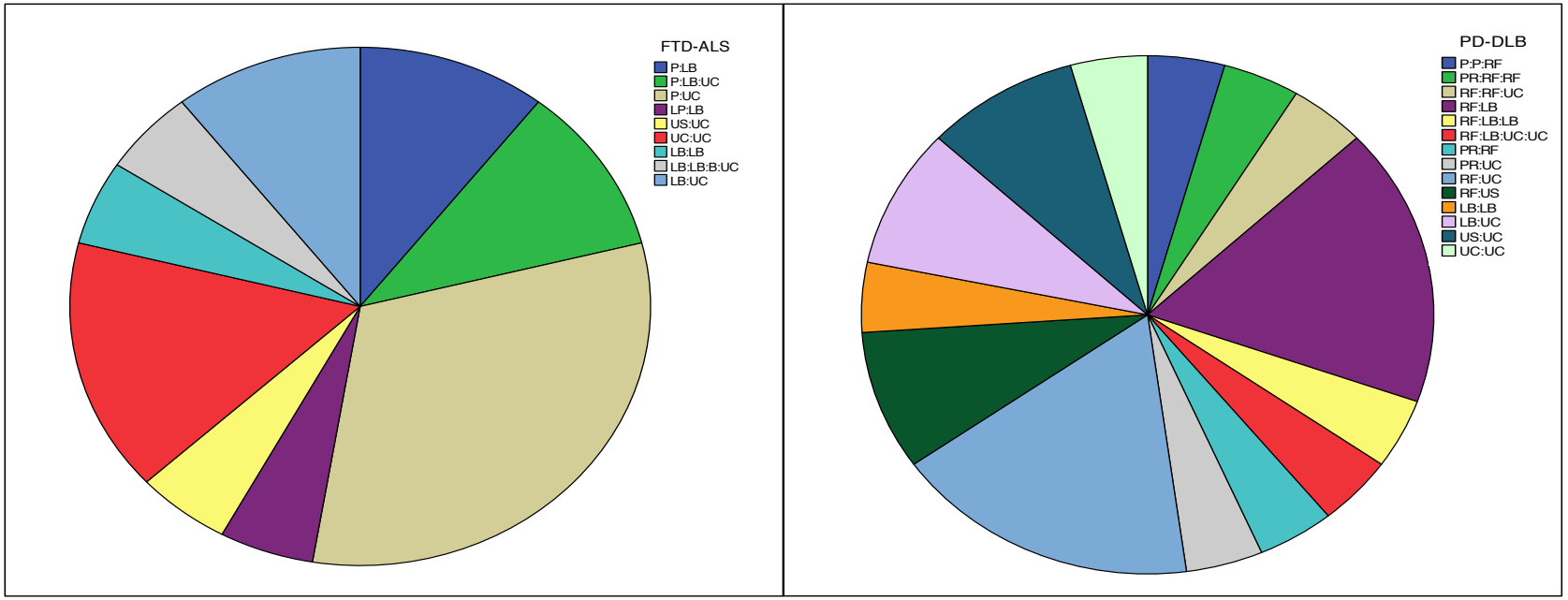


Figure 3.1. ACMG variant criteria for all cases within the FTD-ALS and PD-DLB cohorts that had >1 variant within their respective disease panels at 1% MAF. The relative combination of alleles can be seen in the top right of each cohort's chart. Key references to ACMG classification of each variant; P=Pathogenic, LP = Likely Pathogenic, RF = Risk Factor, PR = Pathogenic in the recessive state (but considered likely benign in the heterozygous state), LB = Likely Benign, US = Uncertain Significance, UC = Un categorized.

3.5 Discussion

With ever more comprehensive panels of genetic testing in neurodegenerative disorders, the possibility of detecting more than one rare variant in an individual will become increasingly likely, posing significant diagnostic challenges and difficulties for genetic counselling. Our data shows the observed frequency of 'oligogenic' variation is linked to the size of the gene panel and MAF threshold, ranging from 1.4% (AD panel) to 13.3% (PD-DLB panel) in both affected and unaffected individuals (Appendix 39). This highlights that, whilst each allele is in itself rare, it is not uncommon for any individual to have more than one rare variant across a small disease panel. This should be borne in mind when investigating the possibility of an oligogenic mechanism, particularly given the increasing number of genes identified as causing or contributing to neurodegenerative disorders.

Why are our conclusions different to previous studies that were of a similar size? In order to be defined as 'oligogenic', an individual must have >1 variant in a known relevant risk gene. This introduces a systematic bias, whereby affected individuals are more likely to harbour one of these alleles than healthy aged individuals. We reviewed the cases presented by Cady et al (Cady, Allred et al. 2015) in which they identified 18 cases which had oligogenic variation in their study in 2015. By the authors own criteria two of these cases were likely to be homozygous recessive mutations; *SOD1* p.D91A and *SETX* p.I2547T, and two potential compound heterozygous cases; *SETX* p.C1554G and p.R168Q, and *SETX* p.I2547T and p.T14I were also present. They also had 3 individuals with the pathogenic *C9Orf72* hexanucleotide repeat expansion (Renton, Majounie et al. 2011), one individual with the pathogenic heterozygous *FUS* p.P525L mutation (Conte, Lattante et al. 2012, Leblond, Webber et al. 2016), and one with the pathogenic heterozygous *SOD1* p.G38R mutation (Rosen, Siddique et al. 1993). Taken together, we therefore suggest that at least 10 of their putative 18 oligogenic cases have genetically determined forms of disease caused by a single

allele rather than though a synergistic effect. Without knowing the nature of the rest of their cohort, given that the vast majority of cases were clinically sporadic (89.3%) it is highly likely that the 55.6% of monogenic cases within the oligogenic cohort was significantly in excess of that observed in the non-oligogenic cohort.

The presence of these known risk alleles, in conjunction with a background rate of polymorphic variation, inevitably results in individuals with a known highly penetrant allele or risk factor being more likely to fall into the 'oligogenic' group. In keeping with this, our analysis shows that the vast majority of individuals defined as having 'oligogenic' variation do indeed have a known risk allele or highly-penetrant variant, explaining the initial association we observed between oligogenic variants and PD-DLB. Importantly, after excluding the known major variant in individual cases there was no association between the benign oligogenic variation and neurodegenerative disease or the age of onset.

This same systematic bias will lead to the apparent enrichment of 'oligogenic' variants in familial cases. By being familial, these individuals are more likely to harbour a known risk genetic factor, which when combined with the background variant carrier rate, makes them more likely to be classified as oligogenic than healthy controls. Thus, *a priori,* being a familial case will make it more likely for an individual to have oligogenic variants. This does not necessarily mean that the additional variants are having an effect on the risk of being a familial case. Given the frequency of any individual harbouring two or more variants, and the likely diminishing impact of each variant on the phenotype and disease risk, substantially larger datasets (eg. n>10,000) will be required to definitively resolve this complex issue with robust variant pathogenicity interpretation.

Chapter 4

The frequency and signature of somatic mutations in 1461 human brains

4.1 Aims

1. To develop a technique to identify protein-coding somatic mutations within the brain from single-sample whole exome sequencing data.
2. To determine the nature and prevalence of somatic protein coding mutations within the brain in individuals with a neurodegenerative disease and controls.
3. To begin to understand the likely aetiology and origin of somatic mutations within the brain.
4. To aim to understand the potential for identified somatic mutations to cause disease within the MRC Brain Bank.

4.2 Introduction

Mutations affecting over 50 nuclear genes contribute to the pathogenesis of late onset neurological disorders (Tsuji 2010). Present in every cell in the body, these genetic variants are either inherited or arise through a *de novo* mutation in the gamete. In contrast, some age-related disorders such as cancer arise through the accumulation of somatic mutations within a cell lineage during life, creating genetic heterogeneity within a tissue or organ (somatic mosaicism). Almost half of these mutations arise decades before tumour initiation (Reya, Morrison et al. 2001, Tomasetti, Vogelstein et al. 2013, Genovese, Kahler et al. 2014), raising the possibility that somatic mutations acquired by a similar process during development are also present within non-malignant human tissues. Within the nervous system, somatic mutations have been identified in rare, early onset, focal neurological disorders such hemimegalencephaly and lissencephaly (Gleeson, Minnerath et al. 2000, Sicca, Kelemen et al. 2003, Lee, Huynh et al. 2012, Poduri, Evrony et al. 2012), demonstrating that protein-coding variants with mosaic allelic fractions as low as 8% in the brain can cause macroscopically overt structural neurological diseases (Lee, Huynh et al. 2012), though even lower allelic fractions of around 1% may cause milder, and microscopically observable pathology such as focal cortical dysplasia (Lim, Kim et al. 2015). To date however, the frequency of somatic mutations in the human brain, and particularly in those late-onset neurological disorders has not been studied systematically.

4.3 Methods

4.3.1 Tissue samples

DNA extracted from 1461 human brains (cerebellum: n=1281 (87.7%), cerebral cortex: n=94 (6.5%), basal ganglia: n=8 (0.5%), not classified: n=78 (5.3%)) from 1099 patients with neurodegenerative diseases including Alzheimer's disease (AD), Frontotemporal dementia or

Amyotrophic lateral sclerosis (FTD-ALS), Creutzfeldt Jackob disease (CJD), Parkinson's disease and Dementia with Lewy bodies (PD-DLB) and 362 age-matched controls within the Medical Research Council (MRC) UK Brain Bank Network were utilized. As previously, Controls were defined as having no *ante mortem* history of neurological disease, no neuropathological features of any neurodegenerative disease and a Braak neurofibrillary tangle stage of ≤ 2 (Chapter 2.3.1) (Figure. 4.1a, b; Appendix 55).

4.3.2 Whole-exome sequencing (WES) and somatic variant calling

Exome sequencing data was utilized as previously described (Section 2.4.3.2). Sequencing data was aligned against the UCSC hg19 human reference genome using Burrows-Wheeler Aligner (BWA) (Li, Handsaker et al. 2009). This time, GATK's Haplotype Caller from Genome Analysis Toolkit (GATK version 3.4) was used to determine allelic counts and genotypes across the genome (McKenna, Hanna et al. 2010) using in-house scripts. We subsequently excluded the following regions:

1) Low complexity regions of the genome and sites containing markers failing Hardy Weinberg equilibrium tests in the 1000 Genomes Project phase 1 (Li 2014) (https://github.com/lh3/varcmp/blob/master/scripts/LCR-hs37d5.bed.gz and https://github.com/lh3/varcmp/blob/master/scripts/1000g.hwe-bad.bed).
2) Sites with greater than 1000 fold coverage within Phase 1 of the 1000 Genomes Project (Genovese, Handsaker et al. 2013).
3) Known segmental human genome duplications (Bailey, Yavor et al. 2001, Bailey, Gu et al. 2002).
4) Regions harboring common large inserts in 1000 Genomes Project Phase 1 (Utilized by Genovese et al(Genovese, Kahler et al. 2014) and obtained directly from the author).
5) Regions excluded from the strict mask of the 1000 Genomes Project Phase 1 (Genomes Project, Abecasis et al. 2012) (ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/phase1/analysis_results/supporting/accessible_genome_mask s/20120824_strict_mask.bed)

6) Specific additional regions of Copy Number Gain or Loss in each individual called from SNP genotyping as described above.
7) Sites with read depth < 30x in any sample (Figure. 4.2a, b; Appendix 56) were finally removed. This resulted in a total of 5,906,849 base pairs (bp) per individual available for subsequent analysis.

To detect putative somatic mutations, we used a modified work-flow that was initially described by Genovese et al (Genovese, Kahler et al. 2014), but this time using a pan-exome approach. Firstly, we restricted variants to single nucleotide variants and excluded all variants with VAF >50% or <10% (Figure. 4.2a). VAF were subsequently identified which significantly differed from the mean VAF for heterozygous variants (47% in our dataset, $p < 1 \times 10^{-5,}$ Binomial test, Figure. 4.2c). We also excluded those variants present more than once in the cohort, and those with a minor allelic frequency (MAF) >0.5% within the ExAC database of Human Exome Variation (Appendix 56) (Lek, Karczewski et al. 2016). The alignment of data against the human reference genome and the removal of sites with a read depth of < 30 was performed by Dr Wei Wei (University of Cambridge).

In order to confirm that detected putative somatic alleles also significantly differed from the base error rate in addition to the mean allelic frequency for a heterozygous variant, we utilized deepSNV (Gerstung, Papaemmanuil et al. 2014) to compare the nucleotide counts for each putative somatic variant against 328 random samples within the same dataset. Relative read counts were retrieved from the BAM file of each case, and the individual of interest was compared against the variant allele counts for the other 328 individuals using a betabinomial distribution. Variants with a p-value < 0.001 were included as putative somatic variants. This ensured putative somatic alleles passing both thresholds differed from both the observed VAF of heterozygous variants, and from the local base error rate (Figure. 4.2c). The performance of the deepSNV analysis was also performed by Dr Wei Wei. All putative somatic variants were confirmed by inspection in Integrative Genomic Viewer (Robinson, Thorvaldsdottir et al.

2011, Thorvaldsdottir, Robinson et al. 2013) and were annotated using ANNOVAR (Wang, Li et al. 2010) (Appendix 56).

4.3.3 Variant validation

Variants remaining after the above filtering strategy were then validated by targeted amplicon sequencing to confirm a somatic mutation in cases, together with their absence from controls (VAF <1%). Specific primers spanning 66 of the 70 putative somatic alleles were designed using NCBIPrimerBLAST (*https://www.ncbi.nlm.nih.gov/tools/primer-blast/*) (Appendix 57).

PCR reactions for the identified case containing the putative somatic variant, and a control case with DNA extracted from the same brain region were performed using MyTaq HS polymerase (Bioline, USA) for each amplicon. A standard master-mix was utilized (5ul of MyTaq reaction buffer, 0.2ul of MyTaq^{TM} HS DNA Polymerase, and 1 ul of forward and reverse primers (diluted to 10uM), together with 16.8ul of water). Reactions were incubated at 95°C for one minute before denaturing for 15 seconds (95 C), annealing at 58-62°C for 15 seconds, and 10 seconds of extension (at 72°C) for 33 cycles before a final extension step of 10 minutes. Amplified products were assessed by gel electrophoresis against DNA +ve and –ve controls, before quantification using a Qubit 2.0 fluorimeter (Life Technologies, Paisley, UK). Each amplicon was individually purified using Agencourt AMPure XP beads (Beckman-Coulter, USA), pooled in equimolar concentrations and re-quantified. Pooled amplicons were tagged, amplified, cleaned, fragmented to 300bp (Covaris, USA), before pooling with sample multiplexes using the Illumina Nextera XT DNA sample preparation kit (Illumina, CA, USA). Multiplex pools were sequenced using MiSeq Reagent Kit v3.0 (Illumina, CA, USA) in paired-end, 150 bp reads. Note that the amplification, pooling and tagging of amplicons was performed by an external sequencing provider.

FASTQ files were analysed using in-house bioinformatic pipelines. Reads were aligned to the UCSC hg19 human genome reference using Burrows-Wheeler Aligner (BWA) (Li, Handsaker et al. 2009). Variant calling was performed using GATK's Haplotype Caller (McKenna, Hanna et al. 2010) (minimum depth = 500x, minimum supporting reads = 40, base quality ≥ 30 and mapping quality ≥ 20), and variant to reference allelic frequencies manually extracted from BAM files (this extraction was performed by Dr Wei Wei). Subsequently, all validated variants were manually inspected and confirmed in Integrative Genomic Viewer (IGV) (Robinson, Thorvaldsdottir et al. 2011, Thorvaldsdottir, Robinson et al. 2013) (Appendix 56).

Five variants from five cases fulfilling the above criteria were randomly selected for validation by pyrosequencing. Primers were designed using PyroMark® AQ Software (Qiagen Inc) to generate a PCR product encompassing each variant of interest between 250 and 350 base pairs (bp) in size. The Polymerase Chain Reaction primer anti-parallel to the sequencing primer was designed to be biotinylated at the 5' end, purified by High Protein Liquid Chromatography (HPLC). All pairs of forward and reverse primers were checked for sequence specificity to the region encompassing the variant of interest using Primer-BLAST (Ye, Coulouris et al. 2012). Initial 50 ul PCR reactions were conducted using each primer pair and PCR master mix containing 50ng of DNA from extracted cerebellar brain tissue from each relevant individual and a control subject. The appropriate primer pairs for each variant were used in the amplification process using MyTaq™ HS as described above, with reactions incubated at 58-62°C. Subsequently, 10ul of each biotinylated PCR products were immobilized using 2ul of Streptavidin-coated Sepharose high-performance beads (GE Healthcare, Buckinghamshire, UK) in a solution of 40ul binding buffer (Qiagen Inc) and 28ul of nuclease free water. Samples were then agitated at 1400 rpm and at room temperature (20 C) using an orbital shaker. The immobilized PCR products were captured using the

PyroMark® Q24 vacuum work station and then washed with 50 mL of 70% ethanol for 5 s, denatured with 40 mL of PyroMark® denaturation solution for 5 s, and washed with 50 mL of PyroMark® wash buffer for 10 s. Biotinylated single stranded PCR products were subsequently released into the relevant well on a PyroMark® Q24 sequencing plate containing the relevant sequencing primer diluted to 0.3 uM in PyroMark® Annealing Buffer. Annealing of the sequencing primers to DNA strands was performed at 80 °C in a pre-warmed heat block for 2 minutes. Samples were then cooled for 7 minutes and processed in the PyroMark® Q24. Run specific concentrations of dNTPs, enzyme and substrate were loaded, as calculated by the PyroMark Q24 software for each run. Data was analysed using the PyroMark Q24 software for AQ quantitation, with relevant alleleic frequencies determined from the sequencing pyrogram. Each sample and control was run in duplicate and the mean of the VAF determined for each allele in each sample and control (Appendix 58). After performing validation on both platforms,

4.3.4 Occurrence of somatic mutations at methylated bases

We downloaded whole genome bisulphite sequencing (WGBS) data from the Inner Cell Mass (ICM) of an early developmental human embryo (Guo, Zhu et al. 2014). In total, 476,286,624 of 3,095,693,981 total bases were methylated (15.4%). We subsequently sought to determine whether there was enrichment of somatic mutagenesis at methylated sites by performing a binomial test using 15.4% as the background probability against the proportion of validated variants that occurred at methylated bases.

4.3.5 Mutational spectra and signatures

Mutational spectra were derived directly from the reference and alternative allele at each somatic variant allele. To understand the potential mechanisms of somatic mutagenesis we compared the somatic mutation spectrum and triplet allele (reference allele either side of the

somatic allele) against 30 previously defined mutational signatures in cancer (Alexandrov, Nik-Zainal et al. 2013) and against the mutational signatures to *de novo* genetic variants derived from trio studies in the population (Turner, Yi et al. 2017).

4.3.6 Variants in the brain proteome

All gene expression data was downloaded from the Human Protein Atlas (Uhlen, Fagerberg et al. 2015), and each gene containing a somatic variant was annotated according to the expression classification within the brain. Genes were classed as either; (1) Elevated in brain, (2) Expressed in all, (3) Mixed expression pattern, (4) Not detected in brain, (5) Not detected in any tissue as determined by the Human Protein Atlas. Binomial testing was performed in R to determine whether genes containing somatic variants were significantly different from the expression profile of all genes across the human genome within these 5 categories (Appendix 59).

4.3.7 Variants in conserved genes

To determine the relative constraint for mis-sense variation within the germline for each gene containing a somatic variant, we annotated each gene with the mis-sense Z score as determined by The Exome Aggregation Consortium (ExAC) (Lek, Karczewski et al. 2016). Binomial testing was performed to compare the proportion of genes within each quartile of the spectrum of mis-sense constraint as determined by ExAC.

4.4 Results

4.4.1 Characteristics of mutations

Exome sequencing was performed on 1461 human brain samples from 1099 patients with neurodegenerative diseases and 362 age-matched controls (Figure. 4.1a, b; Appendix 55). Mean sequencing depth of WES from 1461 samples was 51.9-fold (SD=12.9), with no

significant difference between any disease or controls (Appendix 60). Following the aforementioned filtration steps (Figure. 4.2a, b; Appendix 56), we detected 56 somatic variants in 46 brains (3.2% of 1461) (Appendix 61). Specific short primer sequences were adequately generated and sequenced for 40 of the 56 variants using two orthogonal methods **(**Appendix 57 & 58**),** and confirmed the presence of a somatic mutation in 22 (55.0%) of tested alleles; a confirmation rate in keeping with other studies of somatic mutations (Ju, Martincorena et al. 2017) (Table 4.1, Figure. 4.3a, Appendix 61). The majority of validated variants were transitions (86.4%, n=19) with 23.4% (n=3) transversions. C>T mutations were by far the most common (59.1%) (Ostrow, Barshir et al. 2014), and 27.2% (n=6/22) of the validated mutations occurred at bases methylated in the inner cell mass (Table 4.1) (Guo, Zhu et al. 2014). In addition, 11 of the 13 C>T mutations (84.6%) were present at CpG sites within the genome. None of the identified somatic variants were seen in the heterozygote state in the 1461 brains, and all were extremely rare in the background population (Lek, Karczewski et al. 2016) (Table 4.1) (Appendix 62). There was also no difference in the frequency of somatic variants between the different disease and control groups (Figure. 4.3b) (Bernoulli trial testing performed by Juvid Aryaman, Imperial College London), indicating that, whilst mutational rates may not be increased in patients with neurodegenerative diseases compared to healthy aged individuals, somatic mutations at high variant allele frequencies are relatively common in the human brain.

Table 4.1. Validated somatic variants in 1461 human brains. Variant data shows the chromosome, base position and reference and alternate allele (hg19 build), together with the amino-acid change, frequency in the ExAC population dataset (Lek, Karczewski et al. 2016), SIFT annotation score and classification (Sim, Kumar et al. 2012), expression cohort in the Human Proteome Atlas (Uhlen, Fagerberg et al. 2015), the quartile of genetic conservation within the human genome (Lek, Karczewski et al. 2016), presence of methylation at that base in the Inner Cell Mass (ICM) of an early developmental human embryo(Guo, Zhu et al. 2014), and the Variant Allele Frequency (VAF) in the WES data. Clinical data for each individual comprising sample ID, gender, brain region and disease group are shown. Abbreviations: AD, Alzheimer's disease; CJD, Creutzfeldt Jakob Disease; FTD-ALS, Frontotemporal dementia – Amyotrophic lateral sclerosis; PD-DLB, Parkinson's disease – Dementia with Lewy Bodies; PSP, Progressive Supranuclear Palsy; Syn, Synonymous; Non-syn, Non-synonymous; D, Deleterious; T, Tolerated; ICM, Inner Cell Mass; N/A, Not-applicable.

Variant data														Clinical data			
Chromosome	Base position	Ref allele	Alt allele	Mutation	Gene	AA change	ExAC	SIFT score	SIFT	Human Proteome Expression	Conservation quartile	Methylated in ICM	Exome VAF	Sample ID	Gender	Brain region	Disease group
chr1	24822434	C	T	N	OR2L3	p.R121	4.94E-	0.02	D	Not detected	2	N	15.2%	1	M	Cerebellum	Control
chr7	15081567	C	T	N	AGAP3	p.S81L	N/A	0.02	D	Expressed in all	4	N	22.3%	2	M	Temporal	Control
chr11	10490510	T	G	N	CASP1	p.K37Q	4.10E-	0.59	T	Expressed in all	2	N	16.9%	3	M	Cerebellum	Control
chr17	39502849	T	G	N	KRT33A	p.R316S	1.10E-	0.67	T	Not expressed	2	N	22.1%	4	M	Cerebellum	CJD
chr3	45837911	T	C	S	SLC6A20	p.M1V	8.77E-	0.43	T	Not expressed	3	N	20.3%	5	M	Cerebellum	AD
chr3	12262974	T	C	N	SEMA5B	p.H59R	N/A	0.02	D	Elevated brain	4	N	19.3%	5	M	Cerebellum	AD
chr19	36275201	G	A	N	ARHGAP	p.A517	N/A	0.13	T	Mixed expression	3	N	26.0%	6	M	Cerebellum	Other (PSP)
chr12	6138596	C	T	N	VWF	p.R960P	8.24E-	0.26	T	Mixed expression	3	N	19.5%	7	F	Cerebellum	Control
chr11	56344581	G	T	N	OR5M10	p.T206	1.20E-	1	T	Not detected	2	N	13.0%	8	F	Cerebellum	Other (Epilepsy)
chr16	4833750	A	G	N	SETP12	p.I131T	1.68E-	0.01	D	Not expressed	1	N	22.8%	9	M	Frontal cortex	Control
chr7	1535876	C	T	N	INTS1	p.D671	8.26E-	0	D	Expressed in all	3	N	14.7%	9	M	Frontal cortex	Control
chr8	14492155	T	C	N	NRBP2	p.I171V	3.32E-	0.15	T	Expressed in all	2	N	30.4%	9	M	Frontal cortex	Control
chr2	85991195	C	T	N	ATOH8	p.R284	8.30E-	0	D	Mixed expression	4	Y	28.3%	10	F	Temporal	Control
chr1	24125194	G	A	N	GALE	p.R50W	1.66E-	0.02	D	Expressed in all	2	Y	21.7%	11	F	Cerebellum	Control
chr1	17570577	T	C	N	PADI1	p.C126	N/A	0.01	D	Not expressed	1	N	25.3%	12	M	Cerebellum	Other (Dementia)
chr17	76499013	G	A	S	DNAH17	N/A	4.05E-	N/A	N/A	Not expressed	N/A	Y	28.2%	12	M	Cerebellum	Other (Dementia)
chr11	1718844	T	C	S	KRTAP5-	N/A	1.65E-	N/A	N/A	Not detected	N/A	N	14.9%	13	F	Cerebellum	FTD-ALS
chr19	9361855	G	A	N	OR7E24	p.A46T	2.50E-	0	D	Not expressed	1	Y	30.2%	13	F	Cerebellum	FTD-ALS

Variant data														Clinical data			
Chromosome	**Base position**	**Ref allele**	**Alt allele**	**Mutation**	**Gene**	**AA change**	**ExAC**	**SIFT score**	**SIFT**	**Human Proteome Expression**	**Conservation quartile**	**Methylated in ICM**	**Exome VAF**	**Sample ID**	**Gender**	**Brain region**	**Disease group**
chr20	60888258	G	A	S	LAMA5	N/A	N/A	N/A	N/A	Expressed in all	N/A	N	23.0%	13	F	Cerebellum	FTD-ALS
chr16	88712548	G	A	S	CYBA	N/A	N/A	N/A	N/A	Expressed in all	2	Y	23.5%	14	F	Cerebellum	Control
chr22	50752254	G	A	N	DENND6	p.R398	1.66E-	0	D	Mixed expression	1	Y	20.2%	15	M	Cerebellum	FTD-ALS
chr6	5004177	G	A	S	RPP40	N/A	4.12E-	N/A	N/A	Expressed in all	N/A	N	22.9%	15	M	Cerebellum	FTD-ALS

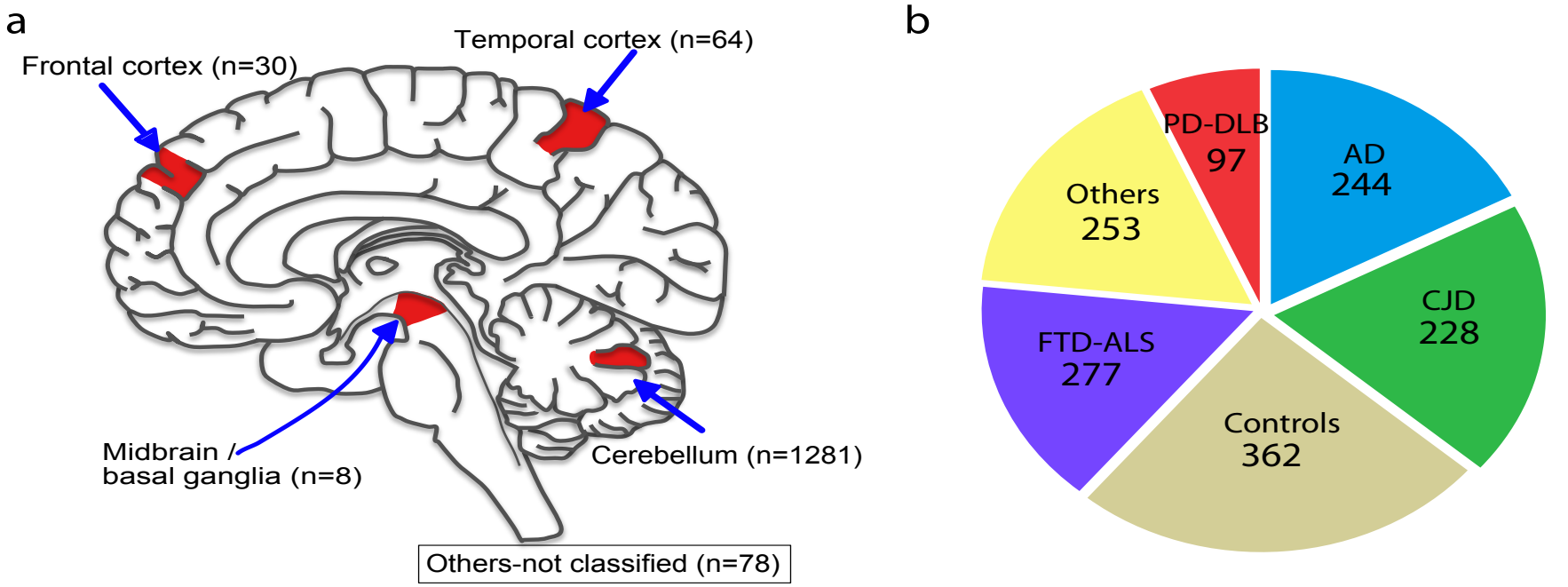


Figure 4.1. Brain regions and sample groups in 1461 *post mortem* human brains. The relative combination of alleles can be seen in the top right of each cohort's chart. Key references to ACMG classification of each variant; P=Pathogenic, LP = Likely Pathogenic, RF = Risk Factor, PR = Pathogenic in the recessive state (but considered likely benign in the heterozygous state), LB = Likely Benign, US = Uncertain Significance, UC = Un categorized.

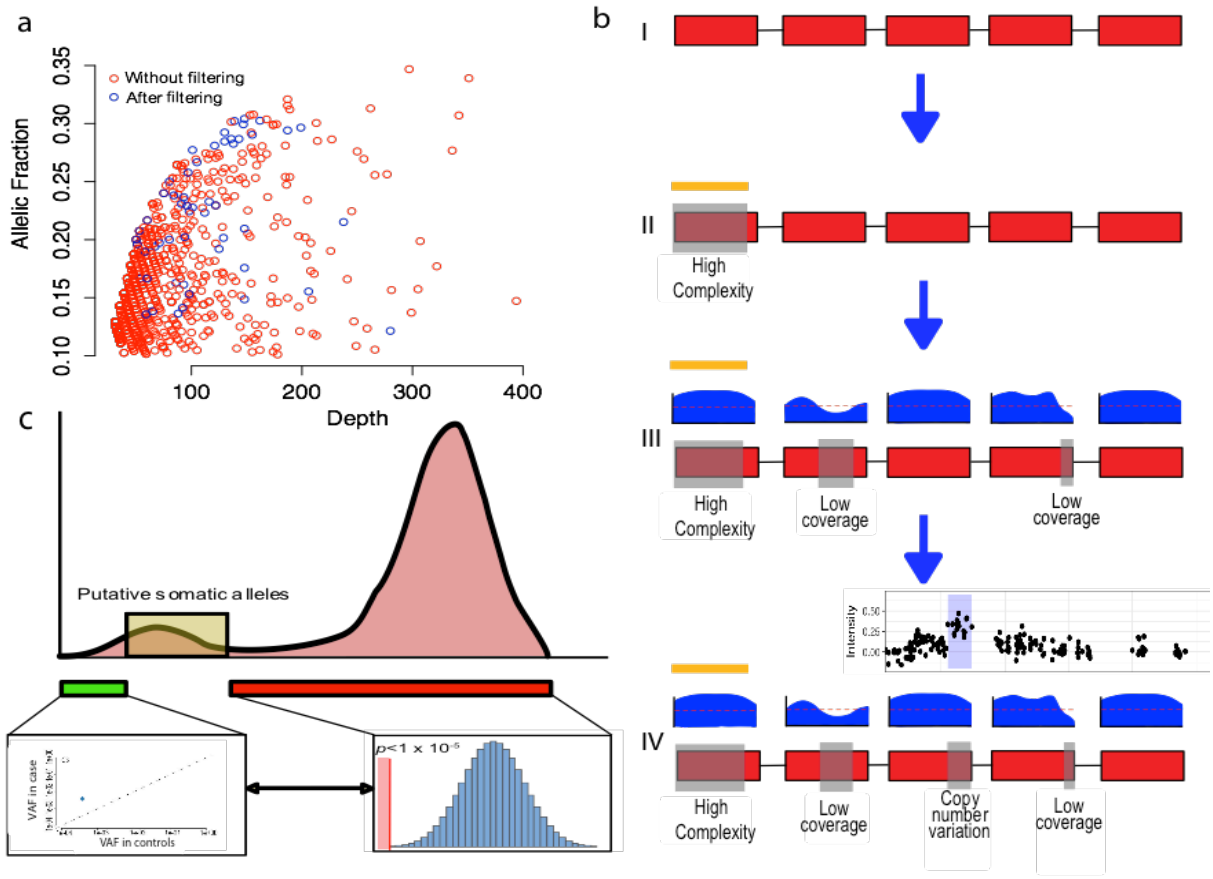


Figure 4.2. The detection of somatic variants in 1461 *post mortem* human brains. The Unfiltered variant allele frequencies (VAF) with between 10% and 35% against relative exome sequencing depth. Those that were present before and after filtering are shown (red and blue respectively). b Variant detection pipeline. Section I - Exons are shown in red, with intergenic and intronic regions as a black line. II - Regions of high genomic complexity and common structural variants (determined from population databases and previous studies) were removed (yellow line / grey box). III- relative sequencing depth of each exon is shown in blue above the relevant exon. Bases in which the sequencing depth was below 30 (as depicted by the red dashed line) in an individual were removed. These regions are then shown by grey boxes on the schematic exome and were also removed. IV – Finally, regions in which copy number variants (gains or losses) were called from array genotyping (see Chapter 2.3.2.3) were also removed from the overall panel. An example plot of the array genotyping in which a copy number gain has been detected is shown. Again the corresponding region was removed from the exome depicted by a grey box on the exome panel. After these steps, remaining regions were subsequently subjected to analysis by deepSNV (Gerstung, Papaemmanuil et al. 2014) and a binomial test against the mean VAF for heterozygous variants (47%). c Schematic representation of the putative somatic alleles in the dataset. A distribution of VAF in the whole dataset is show (pink histogram). Putative somatic alleles were those in which the VAF was greater than base error rate (as determined from DeepSNV (green box and linked inset)), and those that also differed from the binomial threshold ($<1x10^{-5}$) compared to an assumed VAF of 47% for heterozygosity.

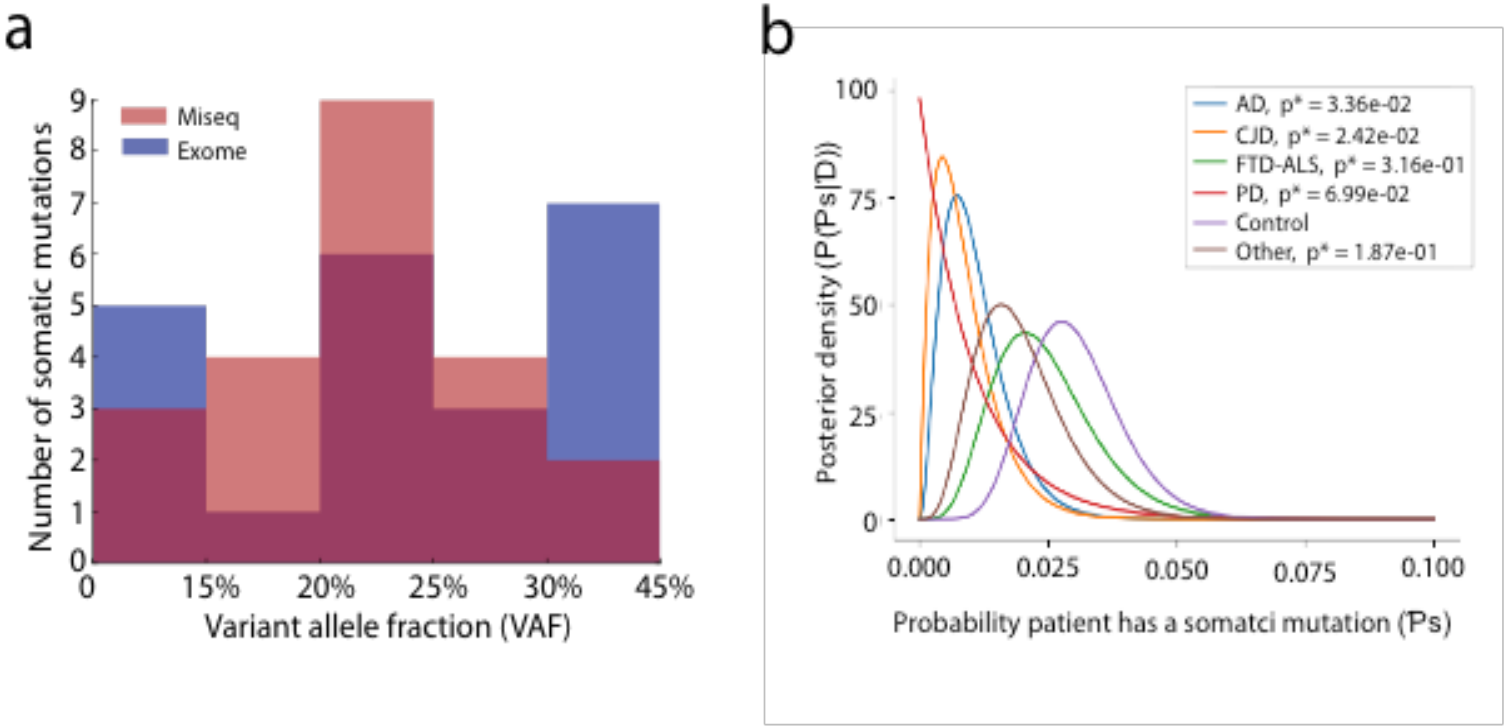


Figure 4.3. The distribution of validated somatic mutations. (a) a Distribution of allele frequencies for the validated variants in the study are shown, with the relative VAF for each allele as detected on both the MiSeq (pink), and Exome sequencing (purple) shown. b Probability of a mutation occurring in each cohort assuming a uniform prior probability and that each person is a Bernoulli trial with probability *Ps* of developing a mutation (Performed by Juvid Aryaman, Imperial College London).

4.4.2 Mutational spectrum and signatures

We further examined the correlation between the observed signature of base mutagenesis with the signature observed in cancer (Alexandrov, Nik-Zainal et al. 2013), observing the strongest correlation with mutations thought to be due to mis-match repair errors occurring during DNA replication and recombination (r^2=0.61, P value = $5.02x10^{-11}$, Pearson's product moment Figure. 4.4a, b). The data were also compared to the mutational profile of *de novo* germline variants in the population derived from the de-novo db mutation database (Turner, Yi et al. 2017), also revealing a strong association with the mutational profile of de *novo* variation (r^2=0.62, P value = $2.74x10^{-11}$, Pearson's product moment) (Figure. 4.4a, b).

4.4.3 Pattern of tissue expression and selection pressure

We subsequently determined the tissue expression pattern of each gene in which a somatic mutation was observed, and saw that ten (58.8%) of the non-synonymous or start-loss variants were present in genes expressed within the brain. These data are consistent with the notion that the somatic mutations were not selected against based on tissue expression, and were equally distributed across the expression profile of the human genome. This raises the possibility that somatic mutations contribute to disease pathogenesis in several human tissues, including the brain (Figure. 4.4c; Appendix 59). Although speculative, the relatively high variant allele fraction (VAF, the ratio of variant allele : total allele) of the observed somatic mutations could actually reflect positive selection of some mutations, particularly if they arose in later stages of development.

We also found no evidence that the selection pressures seen within the germline also act on the somatic mutations we observed in the brain, with non-synonymous somatic variants evenly distributed across conserved and non-conserved regions of the human genome (p=NS, Binomial test, Figure. 4.4d).

Finally, we determined that 58.8% of the non-synonymous or start lost variants (10/17) were predicted to be deleterious by SIFT (Sim, Kumar et al. 2012) suggesting that they are highly likely to have detrimental effects on gene expression (Table 4.1). When taken together, these findings suggest that somatic mutations in the brain may not been subject to the same constraints as genetic variation in the germ-line (Milholland, Dong et al. 2017), rendering all regions of the brain exome vulnerable to somatic mutagenesis, and therefore potentially conferring the possibility of causing a wide range of neurodegenerative diseases.

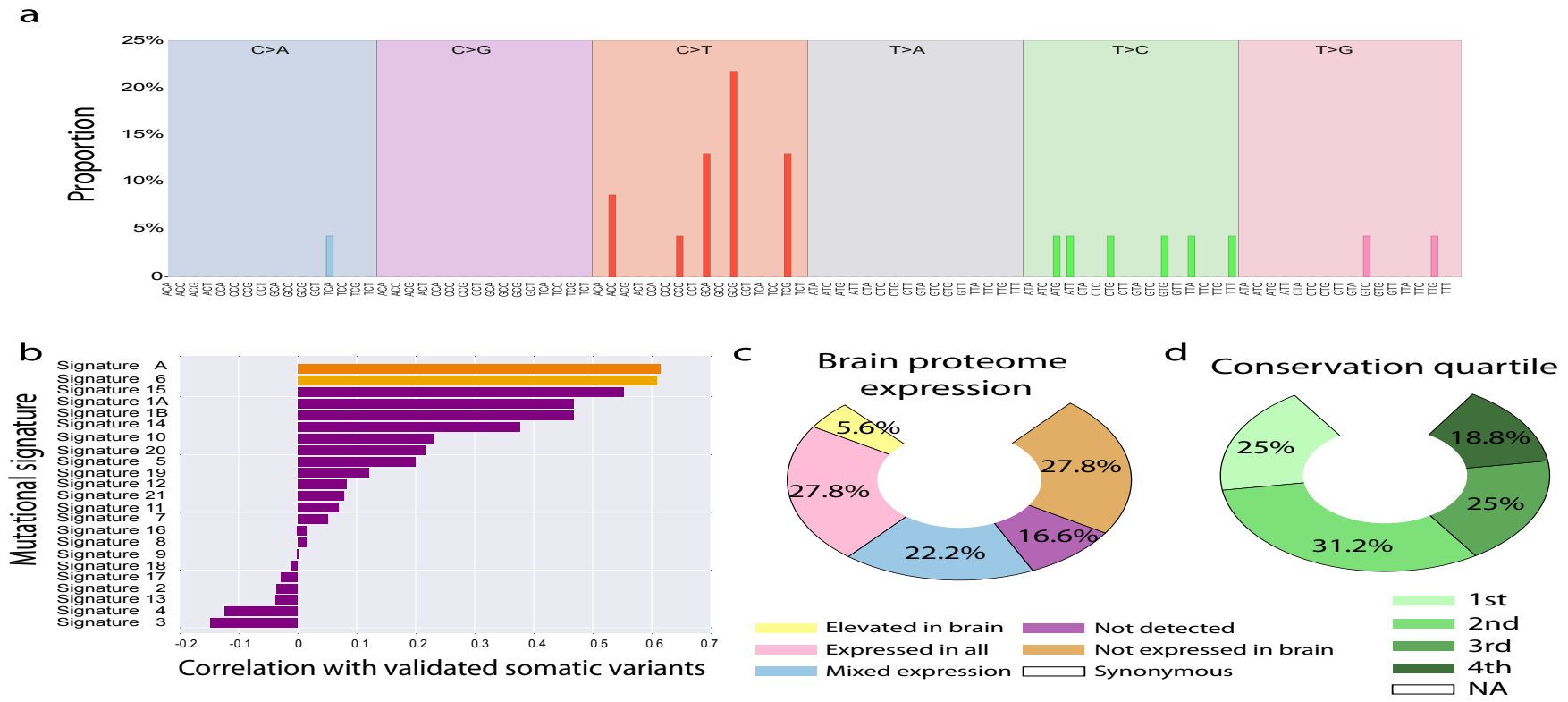


Figure 4.4. The mutational profile of all validated somatic variants. The mutational signature of all validated somatic variants. The mutated allele plus the flanking 3' and 5' base are shown. (b) correlation between the mutational signature of validated somatic variants and the mutational profiles observed in de novo germline variants detected in the population (Turner, Yi et al. 2017) (top bar) and 21 forms of cancer (Alexandrov, Nik-Zainal et al. 2013). The probable disease associations, or type of cancer in which the signature was detected by Alexandrov et al are shown next to the signature number. The Pearson's correlation coefficient is shown for each signature. c Proportion of validated variants within genes grouped by brain proteome expression (Uhlen, Fagerberg et al. 2015). (d) the proportion of validated variants based on each quartile of the gene conservation scores within the germline (4th quartile being the most conserved in the germline).

4.4.4 Estimate of mutation rate in human brains

To determine the somatic mutation rate observed within the human brain we first assumed that the mutations occurring within the first 2 cell divisions of the human zygote would give rise to VAF of 10-30%, and would likely be present in all human tissues, having arisen before tissue differentiation (Yadav, DeGregori et al. 2016) (Figure. 4.5). In this study, after QC and the removal of structural variation, we analysed 5,906,849 nucleotide bases in each individual brain (see Methods). Across the whole cohort (n=1461 cases), this resulted in the analysis of 8,629,906,389 nucleotide bases which contained 22 validated somatic mutations. This equates to a mutation rate of 2.55×10^{-9}. Assuming that the detectable mutations occur at either the first or second cell divisions (corresponding in an allelic fraction of 0.25 and 0.125 respectively, and arising from a total of 6 cells Figure. 4.3a, 4.5), this results in a minimum somatic mutational rate across the human exome of 4.25×10^{-10} per base pair per individual in the first two cell divisions of the human zygote. This is slightly lower than previously calculated human somatic mutation rates of 2.67×10^{-9} (Milholland, Dong et al. 2017), endorsing the sensitivity of our approach. Finally, assuming 3 billion bases in the full human genome, our data suggest that ~ 1.3 somatic mutations across the whole genome will occur during the first 2 cell divisions (3×10^{9} multiplied by 4.25×10^{-10}). This is slightly lower than recent estimates using whole genome sequencing (WGS) where ~3 mutations were estimated to occur per cell per division in very early development (Ju, Martincorena et al. 2017). This difference could reflect methodological differences such as the particularly conservative nature of our validation algorithm, or be due to a lower mutation across the human exome when compared to non-coding regions.

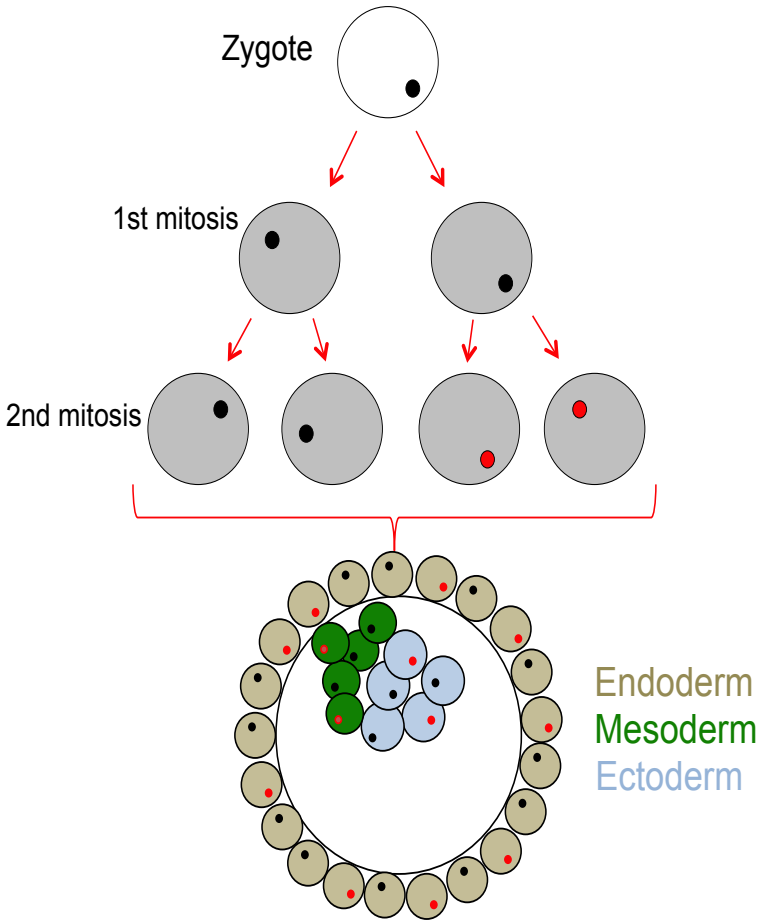


Figure 4.5. A figure showing a somatic mutation arising in early cell division. An example of somatic mutation (red) is shown, with the subsequent distribution of this mutation within the embryo.

4.5 Discussion

These data are the first to quantify the degree of high level (VAF > 10%) somatic mosaicism within the human brain, and show that at least 1% of people possess a somatic protein coding mutation within the central nervous system. Given the close correlation between our observed somatic mutation rate and previous estimates, when extrapolated across the whole genome (of 3 billion bases), our data suggests that each human brain may possess at lease ~1.3 high frequency (>10% VAF) somatic mutations which have arisen during the first two embryonic cell divisions. When considered alongside the slightly higher mutation rates within the male germline of $1.28x10^{-8}$, which confers an average of 76.9 *de novo* germline mutations in each individual (Rahbari, Wuster et al. 2016), then the degree of non-anticipated inherited or acquired genetic variation within an individual can be extensive (~80 alleles). This has important implications in considering the potential genetic aetiology of human neurological diseases.

Whilst the number of validated somatic protein coding mutations in our study was small at 22, we saw no evidence of the same selective constraints seen within the germline, which would otherwise limit the number of potentially detrimental germline alleles acquired during development (Lek, Karczewski et al. 2016). Given the predominance of C>T somatic mutations, the observation that 27.2% (n=6/22) of the validated mutations occurred at bases methylated in the inner cell mass (Table 4.1) (Guo, Zhu et al. 2014) implicates the deamination of methylated cytosines as one potential mechanism, particularly given the enrichment for C>T mutations at CpG sites. It was also surprising that there was a relatively strong association with the mutational signatures seen with *de novo* mutagenesis within the germline (Turner, Yi et al. 2017), suggesting that similar mechanisms of mutagenesis may be involved in the formation of these mutations (Ju, Martincorena et al. 2017), albeit that they do not appear to be selected against in the brain.

A second possibility is that the detected mutations were truly focal within the human brain, having arisen during corticogenesis, and subsequent to tissue differentiation during embryogenesis. For example, *Poduri et al.,*(Poduri, Evrony et al. 2012) detected a focal somatic mutation with a VAF of 17% within the brain causing hemimegancephaly which was not present in the patient's blood. Without additional tissue samples from other organs we cannot exclude this possibility in the cases we studied here. However, the lack of bias for detectable mosiacism in any of the brain regions samples (Cerebellum; 17/22 (p=0.18, Fisher's exact test *vs* other brain regions) (Figure. 4.1a; Table 4.1), together with the lack of focal morphological abnormalities such as those observed by Poduri *et al.,* point towards an early developmental origin rather than a late focal origin for the mutations we report here. However, we do appreciate that we cannot confirm this directly. These problems are likely to be overcome by large scale, higher depth sequencing which will detect lower levels of mosaicism. This will refine the mutation rates and clarify the origin of mutations within individuals with neurodegenerative disorders. However, based on the data we report here, mosaicism should also be considered as a potential source of unexpected genetic findings following diagnostic exome and genome sequencing in neurological disorders.

Taken together, we have demonstrated that at least 1% of human brain samples contain high level somatic mutations present in at least 10% cells. Many of these mutations were extremely rare in the germline of the population, were highly expressed within the brain, and conferred the ability to markedly alter protein function.

Based on the observed mutational signatures, we determine that they are likely to be driven by DNA mis-match repair, and assuming an early developmental origin, are consistent with a somatic mutation rate in the human exome of at least 4.25×10^{-10} per base pair per individual. Taken together these data determine the frequency, nature and likely origin of high frequency

somatic mutations in the human brain and show how they confer the potential to contribute to a range of neurological disorders.

Chapter 5

Copy number variation in 1342 neurodegenerative diseased brains

5.1 Aims

1. To determine the nature and prevalence of germline exonic copy-number variants in control individuals and those with neurodegenerative diseases.
2. To identify novel copy-number variants associated with major neurodegenerative disorders.
3. To validate any novel genetic associations.
4. To determine the potential for novel copy-number variants to influence clinical phenotype or neuropathology.

5.2 Introduction

Copy number variation (CNV) constitutes a substantial portion of total genetic variability in the human genome (Feuk, Carson et al. 2006), and can contribute to the development of several human diseases (Zhang, Gu et al. 2009). However, with the exception of triplications in the Amyloid Precursor Protein (*APP*) (Rovelet-Lecrux, Hannequin et al. 2006) and α-synuclein (*SNCA*) (Singleton, Farrer et al. 2003), which cause Alzheimer's disease and Parkinson's disease respectively, no CNVs have been observed which either cause, or confer a significant risk factor for any neurodegenerative disorder (Sebat, Lakshmi et al. 2004). The limited detection of structural genomic variation that contributes to these disorders has in part been mediated by the utilization of relatively low-density genotyping arrays, or arrays that are predominantly enriched within non-coding regions in case-control studies of neurodegenerative disorders. Recent advances in genotyping platforms now enable us to observe structural variation of the genome to a much greater resolution, and the detection of a moderate correlation between the intolerance of genes to both copy number variation and single-nucleotide variation in control individuals within the population have strongly implicated small CNVs as putative mechanisms of human disease, and particularly for neurological disorders (Ruderfer, Hamamsy et al. 2016).

Herein, we performed a combination of high-density exonic SNP genotyping and exome sequencing in cases of human post mortem brain tissue from individuals with neurodegenerative disorders and controls, establishing three rare novel copy-number gains associated with all forms of CJD, particularly sporadic CJD (sCJD), in which CNVs in *LAMA5* were present in ~ 16% of cases and were strongly associated with sCJD (OR: 442, 95% CI: 222-850). These data suggest that copy number gains in *LAMA5* are strongly associated with all forms of CJD, and open new avenues for the pathogenesis of prion diseases and potentially other neurodegenerative disorders.

5.3 Methods

5.3.1 Case selection

Demographic data (age of disease onset and death, disease duration and family history of disease), together with *ante mortem* clinical diagnosis and *post mortem* neuropathological diagnosis were recorded for all cases (See Chapter 2.3.1)). For cases conforming to diagnostic criteria for Creutzfeld Jakob Disease (CJD) (Budka et al. 1995), the clinical subgroup (sporadic, familial, variant, or iatrogenic) was also recorded, together with *PRNP* genotype within codon 129 and the Prion Protein (PrP) isoform (Appendix 63-67). PrP isoforms were determined by the team at The University of Edinburgh and were not directly determined by myself within this study.

5.3.2 Array genotyping

Array genotyping was performed on all samples using the Illumina HumanOmniExpress-12 BeadChip array (Illumina Inc., San Diego, CA). CNVs were called as previously described (Cooper, Shtir et al. 2015) (See Chapter 2.3.2.3). In addition, exome sequencing data (see Chapter 2.3.2.2) was used to determine related individuals within the dataset and kinship co-efficient. Subsequently, CNVs were filtered to those present in less than 1% of dbVar CNV (Lappalainen, Lopez et al. 2013).

5.3.3 Copy number variation calling from whole-exome data

Exome sequencing data from 100 individuals within the study was also utilized to validate CNV calls (Exome sequencing was performed as previously described in Chapter 2.3.2.2). Briefly, utilizing ExomeDepth (Plagnol, Curtis et al. 2012) we compared the mapped number of reads in each exon in an individual against the expected number of reads calculated from 20 samples which had the closest technical match to the tested sample using an established R

script (Plagnol, Curtis et al. 2012). CNVs with a Bayes Factor (log10 of the likelihood ratio) of > 5 were retained and compared to those also called from the array genotyping platform.

5.3.4 RNA extraction and RNA-sequencing

Appropriate ethical approvals were granted for RNA extraction from the cerebellum of 3 individual who had CNVs in *LAMA5*, *EXD3* and *TRPM2* detected with both array genotyping and exome sequencing. All RNA extraction procedures were performed a Category 3* high-risk laboratory at the University of Edinburgh by their staff, though the protocol was adapted and designed by myself in conjunction with the team in Edinburgh. Briefly, 30-50mg of cerebellum was extracted from each individual and an appropriate volume of QIAzol Lysis Reagent added. Samples were then manually homogenized using an Eppendorf micro-pestle. Samples were then incubated at room temperature (RT) (15-25°C) for 5 minutes before 0.2ml per of chloroform per 1ml of lysis solution was added. Samples were further incubated for 2-3 minutes at room temperature before centrifugation at 12000g for 15 minutes at 4°C.

Subsequently, 70% ethanol was added to the solution, and 700ul of the sample was then transferred into an RNeasy Mini Spin column and centrifuged at 8000g for 15 seconds at RT. Subsequently, 700ul of RW1 was added to the spin column and the sample was again centrifuged at 8000g for 15 seconds before 500ul of RPE buffer was added to the column and a further centrifugation was performed for 15 seconds at 8000g. RPE (500ul) was again added to the column and centrifuged at 8000g for 2 minutes, before a final spin at 14000g for 1 minute.

Elution was performed by adding 80ul of RNase free water to the membrane and centrifuged at 8000g for one minute. All flow through solutions were disposed of in 2N sodium hydroxide (NaOH).

The 80ul of RNA was subsequently mixed with an equal volume of phenol:chloroform, and gently mixed at room temperature for 10 minutes. This was subsequently microfuged for 2 minutes at 13,000rpm for 2 minutes. Subsequently the top layer of the of the solution was manually pipetted into a fresh tube, and 80ul chloroform was added to this solution, before a further 1 minute of centrifugation at 13,000rpm. Again, the top layer was manually extracted into a fresh tube before $1/10^{th}$ volume of 3M sodium acetate (pH 5.2 + 2.5 volumes 100% ethanol (at -20°C)) were added, and then mixed before freezing at -20°C for 16 hours.

Finally, the sample was allowed to defrost to 4°C, and then centrifuged at 13,000rpm for 10 minutes. Supernatant was removed, and the pellet washed with 1ml 70% ethanol, vortexed, and then microfuged again at 13,000rpm for 10 minutes. The supernatant was again discarded, and the pellet dried at room temperature for 10 minutes. The pellet was then re-suspended with 50ul of RNase free water and stored at -70 °C.

RNA was subsequently sequenced following sample preparation using the Kapa Hyper Prep kit and sequenced on an Illumina HiSeq 2000. FASTQ files were aligned and normalized read counts provided by Dr Wei Wei (University of Cambridge).

5.3.5 Statistical analysis

Was performed in R. *PRNP* codon 129 genotypes were grouped into MM (homozygous methionine), MV (methionine-valine) and VV (valine-valine) for statistical analyses. All statistical tests were performed as stated. P-values are reported at both the corrected and uncorrected levels as stated. The penetrance of detected CNVs was determined by calculating the proportion of individuals in the UK who turn 50 years old each year and who carry the CNV of interest divided by the number of individuals aged over 50 who are expected to develop the disease based on epidemiological data.

5.4 Results

Extensive copy number variation was observed within the 993 brains with neurodegenerative disorders, and 349 neuropathologically normal controls remaining after QC (total cohort = 1342) (Figure 5.1, Appendix 63 & 64). On average individuals possessed 21 (SD=5.4) CNVs (>1kb), with a mean =length of 51015 (+/- SD=142584 bases), and a total CNV burden of 1069384 bp (SD= 876026 SD (n=1342 cases). Deletions were more common than gains (mean number of deletions = 12.2, mean number of gains = 8.74, $p=9.5x10^{-113}$, t-test), but deletions were shorter (mean deletion length = 32332, mean gain length =77162, $p=1.60x10^{-128}$, t-test), and rarely spanned more than 4 genes (0.1% of deletions). In contrast, copy gains were more frequently observed in > 4 genes (0.53% of gains, $p=6.57x10^{-11}$), and were observed spanning 10 genes in 0.094% of patients (Figure 5.1) echoing the pattern of exonic CNVs observed recently (Koboldt, Zhang et al. 2012). CNVs (both deletions and gains) were also more common within non-coding regions compared to coding regions (78.6% vs 21.4%) with partial gene deletions and duplications more common than full gene, or multiple gene CNVs (Figure 5.1).

There was no difference in total CNV burden (after correction for multiple testing) (p = 0.018, ANOVA (uncorrected)), mean CNV length (p=0.11 ANOVA), or the proportion of copy number deletions or gains between any disease group and controls (p=0.31, ANOVA) (Figure 5.2). CNVs were also annotated by their relative intolerance score (Koboldt, Zhang et al. 2012), revealing no difference in the mean predicted intolerance of deletions (p=0.56), duplications (p=0.42) or all CNVs (p=0.93) between the major cohorts.

Our calling algorithm was validated by CNV calls from whole exome sequencing (WES) in 100 individuals within the dataset. The array-called CNVs were filtered by the number of

markers, length, and Bayes factor as described (see Chapter 2.3.2.3), observing that for CNVs spanning at least 3 exons, over 64% of array-CNV calls were a validated by WES data (Appendix 68 & 69).

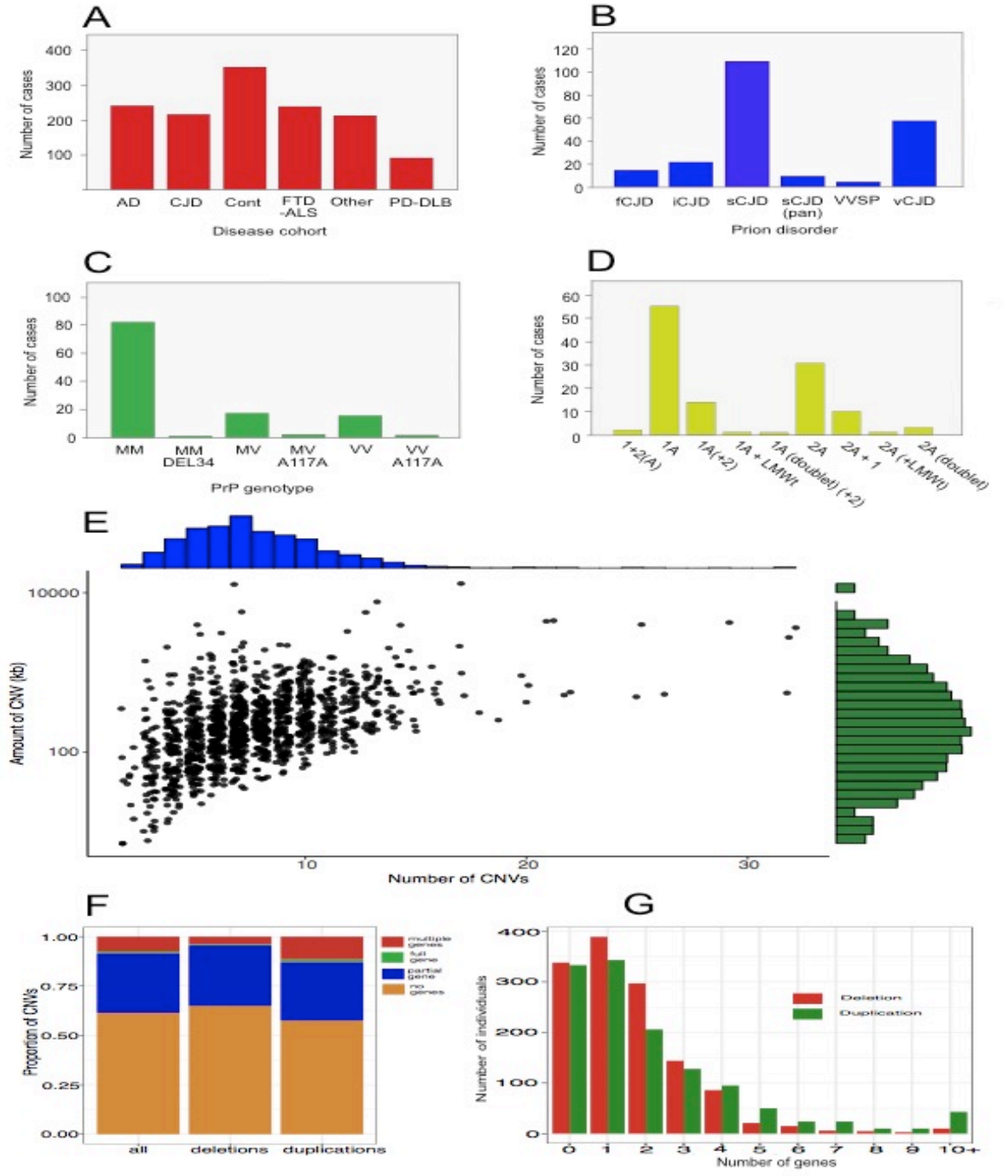


Figure 5.1. The clinopathological, demographic, and CNV profiles of 1342 cases within the study. A: The number of cases within each disease cohort in the study (post-QC). B: The number of cases of CJD (by sub-type) within the study. C: the number of sCJD cases by PRNP genotype within the study. D: The number of sCJD cases by PrP isoform wthin the study. E: A scatter plot and histogram of the number of CNVs per individual (minimum length 1000bases; X axis) against the total amount of copy number variation (Y axis). These values are also represented in a histogram above their respective axis. F: The proportion of CNVs that were partial gene, full gene and not within genes across the whole cohort. G: The number of individuals with duplications and deletions that spanned a particular number of genes.

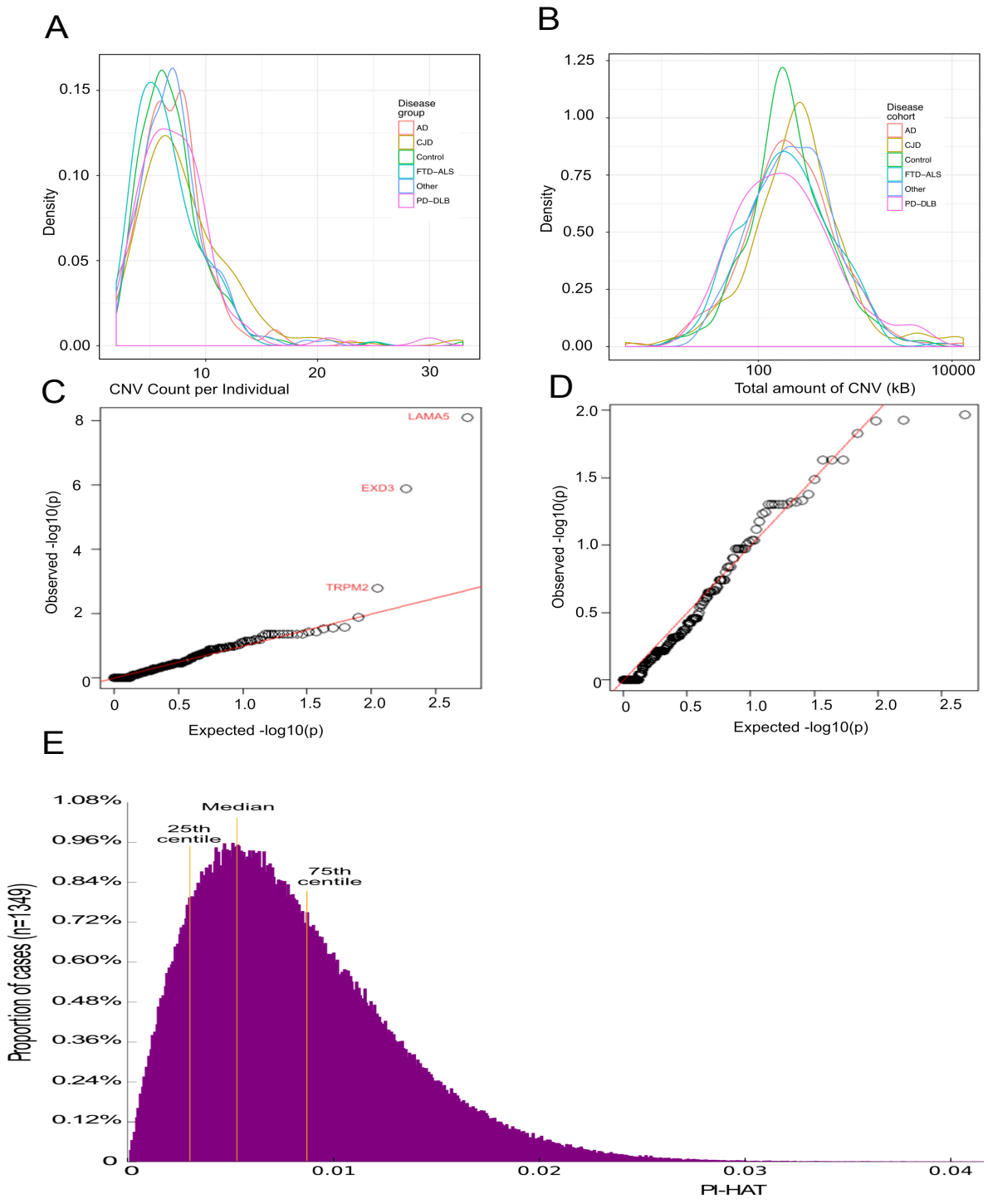


Figure 5.2. The profile of observed CNVs by cohort. A: The overall CNV count per individual by cohort (minimum CNV length 1000 bases) (p=NS between cohorts). B: The total amount of CNV per group (p=NS). C: A Q-Q plot of the observed vs expected number of CN gains by gene for CJD (n=215) vs all other samples (n=1134). D: The number of CN losses for CJD (n=215) compared to all other cases and controls (n=1134). E: Rare CN gains (<1% in the population) in sCJD (n=118) vs all other samples. F: A histogram of Pi_hat scores between 922761 tests between all 1349 individuals in the study. Those with Pi_Hat scores of zero were removed from graphical representation (n=400123). The median and Inter Quartile Range (IQR) of the Pi_hat scores is shown.

Restricting the analysis to rare CNVs (present in <1% of dbVar CNV (Uddin, Tammimies et al. 2014)) we detected a strong statistical association between copy number gains in 3 genes and all forms of CJD (n=212) compared to controls (n=349); *LAMA5* (CJD: 22/212 (10.4%), Controls: 0/349, $p=2.5x10^{-10}$ (Fisher's exact test), *EXD3* (CJD:8/212 (3.8%), Controls: 0/349) $p=3.8x10^{-4}$, and *TRPM2* (CJD:5/212 (2.4%), Controls: 0/349 $p=7.5x10^{-3}$. These CNVs were also not seen in any other non-CJD individual within the dataset (n=779) (Appendix 70). In all cases in which the *EXD3* or *TRPM2* CN gain was observed, they were always present in combination with the *LAMA5* CN gain. There was no evidence of an association between any other copy number gain or loss and any other disease cohort (Appendix 71).

Individuals with any of the three CNVs were enriched within sporadic CJD (sCJD) cases compared to other forms of CJD ($p=1.36\ x10^{-3}$, Fisher's test), with only three of the 22 individuals having a non-sporadic CJD disorder (Single cases of iatrogenic CJD (iCJD) (Case 11) familial CJD (fCJD) (Case 12), and variant CJD (vCJD) (Case 13), Table 5.1, Appendix 70).

Restricting the analysis therefore to sCJD cases (n=116), and comparing against all other neurodegenerative disease cases and controls in our cohort that carried rare CNVs (n= 1128) we determined an even stronger association between *LAMA5* CNVs and sCJD ; *LAMA5* ($p=6.4\ x10^{-21}$), *EXD3* ($p=5.8x10^{-7}$) and *TRPM2* ($p=7.21x10^{-5}$). We further determined that these 3 CNVs are exceptionally rare in the population, present in less than 0.2% of 59,898 individuals (*LAMA5*: 26/58996 (0.044%), *EXD3*: 13/58957 (0.022%), *TRPM2:* (79/59898 (0.13%)) (Ruderfer, Hamamsy et al. 2016). Taken together, these CNVs confer an OR of up to 850 for sCJD (*LAMA5* : OR; 442, 95% CI 222-850, $p=1.8x10^{-40}$ (Fisher's exact test), *EXD3*: OR; 247 95%CI 75-710 $p=1.3x10^{-12}$, *TRPM2*: OR; 27.0, 95% CI 7.1-73.7, p=2.16x-10^{-5}) (Table 5.2)

Table 5.1. All clinical and pathological data for individuals carrying a copy number gain in EXD3, LAMA5 or TRPM2. Variant Highlighted copy numbers were detected both SNP genotyping and exome sequencing. Additional pathological data is shown; PrP (Prion protein) genotype (determined by codon 129 aminoacids) are shown, together with PrP isoform. The age of onset and death of the patient is also shown together with any additional neuropathology observed at post-mortem.

Patient number	Number of CNVs	Validated CNVs	Pathological diagnosis	Co-pathology	PrP Genotype	PrP Isoform	Subtype Class	Age onset (years)	Age death (years)	LAMA5 Copy Number	EXD3 Copy Number	TRPM2 Copy Number	WGS (Cerebellum)	WGS (Spleen)	RNA (cerebellum)
1	1	1	Sporadic CJD	0	VV	2A	sCJD VV2	70	71	3					
2	1	1	Sporadic CJD	0	MM	2A	sCJD MM2	37	39	3					Y
3	1	1	Sporadic CJD	0	MM	1A	sCJD MM1	74	74	3					
4	1	1	Sporadic CJD	0	MM	1A(+2)	sCJD MM1	65	65	3					
5	2	1	Sporadic CJD	0	MM	1A(+2)	sCJD MM1	64	64	3	3				
6	3	3	Sporadic CJD	0	MM	1A	sCJD,	71	71	3/4	3	3			
7	2	1	Sporadic CJD, panencephalopathic	0	MM	1A	sCJD, panencephalopathic MM1	56	57	3	3				
8	1	1	Sporadic CJD	0	MM	1A(+2)	sCJD MM1	60	61	3			Y	Y	Y
9	3	1	Sporadic CJD	0	MM	1A	sCJD MM1	70	71	3	3	3			
10	3	1	Sporadic CJD, panencephalopathic	0	MM DELR34	1A	sCJD, panencephalopathic MM1	45	45	3/4	3	3			
11	3	3	Iatrogenic CJD	0	VV	2A	iCJD VV	29	30	3	3	3			
12	1	1	Familial CJD	0	MM E211Q	1A	fCJD E211Q mutation	53	53	3					
13	2	1	Variant CJD	0	MM	2B	vCJD MM2B	18	19	3	3				
14	1	0	Sporadic CJD	0	VV	2A	sCJD VV2	67	68	3					
15	1	0	Sporadic CJD	0	MM	1A	sCJD MM1	68	68	3					
16	3	2	Sporadic CJD	AD	MM	1A	sCJD MM1	66	67	3	3	3	Y	Y	Y
17	1	0	Sporadic CJD	0	MV	2A(+1)	sCJD MV2	61	61	3					
18	1	0	Sporadic CJD	0	MM	1A	sCJD MM1	73	73	3					
19	1	0	Sporadic CJD	0	MM	1A	sCJD MM1	67	68	3					
20	1	0	Sporadic CJD	0	MM	1A	sCJD MM1	66	66	3			Y		
21	1	0	Sporadic CJD	0	MM	1A	sCJD MM1	79	79	3					
22	1	0	Sporadic CJD	0	VV	2A	sCJD VV2	45	46	3					

Table 5.2. All clinical and pathological data for individuals carrying a copy number gain in EXD3, LAMA5 or TRPM2. Variant Highlighted copy numbers were detected both SNP genotyping and exome sequencing. Additional pathological data is shown; PrP (Prion protein) genotype (determined by codon 129 aminoacids) are shown, together with PrP isoform. The age of onset and death of the patient is also shown together with any additional neuropathology observed at post-mortem.

CNV	Num plat	CJD						sCJD vs other cases in the study (n=1128)				sCJD vs other cases in the population (n=59898)		
		Total (n=212)	sCJD (n=116)	vCJD (n=57)	iCJD (n=21)	fCJD (n=15)	VProt (n=4)	Other cases in study (total=1128) carrying rare CNVs (n)	p-value (Fisher's test) in study	Odds Ratio (95% CI) in study		Other cases in population (n)	p-value (Fisher's test) in population	Odds Ratio (95% CI) in population
LAMA5	**Total**	22	19	1	1	1	0	0	$9.3x10^{-19}$	N/A		26	$1.8x10^{-40}$	442 (222-850)
LAMA5	**Validated**	10	10	1	1	1	0	0	$3.4x10^{-11}$	N/A		26	$1.4x10^{-19}$	213 (89-467)
EXD3	**Total**	8	6	0	1	0	1	0	$5.8x10^{-7}$	N/A		13	$1.3x10^{-12}$	247 (75-710)
EXD3	**Validated**	3	2	0	1	0	0	0	$8.6x10^{-3}$	N/A		13	$3.9x10^{-4}$	79 (9-355)
TRPM2	**Total**	5	4	0	1	0	0	0	$7.2x10^{-5}$	N/A		79	$2.2x10^{-5}$	27 (7-74)
TRPM2	**Validated**	2	2	0	1	0	0	0	$8.6x10^{-5}$	N/A		79	0.01	13 (2-51)

There was however no association between the presence of any of the three CNVs and PrP genotype, or PrP isoform (Appendix 72 & 73), nor did any CNV either in isolation or combination affect either the age of onset or death in sCJD, including when stratified by *PRNP* codon 129 genotype or PrP isoform (Appendix 74 - 77), suggesting that our CNVs predispose to the development of disease, but not specific genotypes or isoforms thereof.

The detected CNVs ranged between 1013 and 49292 bases in length (*LAMA5*: mean length; 5344 bases (SD=3055), *EXD3*: mean length; 15225 bases (SD=14515), *TRPM2*; mean length 6570 bases (SD=4916)), and 13 of the 22 *LAMA5* CNVs (59%) were further validated by ExomeDepth data, with a mean of 76.2% (SD=28.1%) of the CNV region also detected within the CNV determined by exome sequencing (Figure 3,4, Appendix 70). All CNVs were also partial gene duplications (residing entirely within the coding region of the gene), and start and end positions overlapped key functional regions of the protein in all cases (Figure 5.3-5.5); *LAMA5* CNVs span domain I and II of *LAMA5*, *EXD3* CNVs span the 3'-5' exonuclease domain and *TRPM2* CNVs span at least one extracellular domain and two transmembrane helical domains in all cases (Figure 5.3 & 5.4).

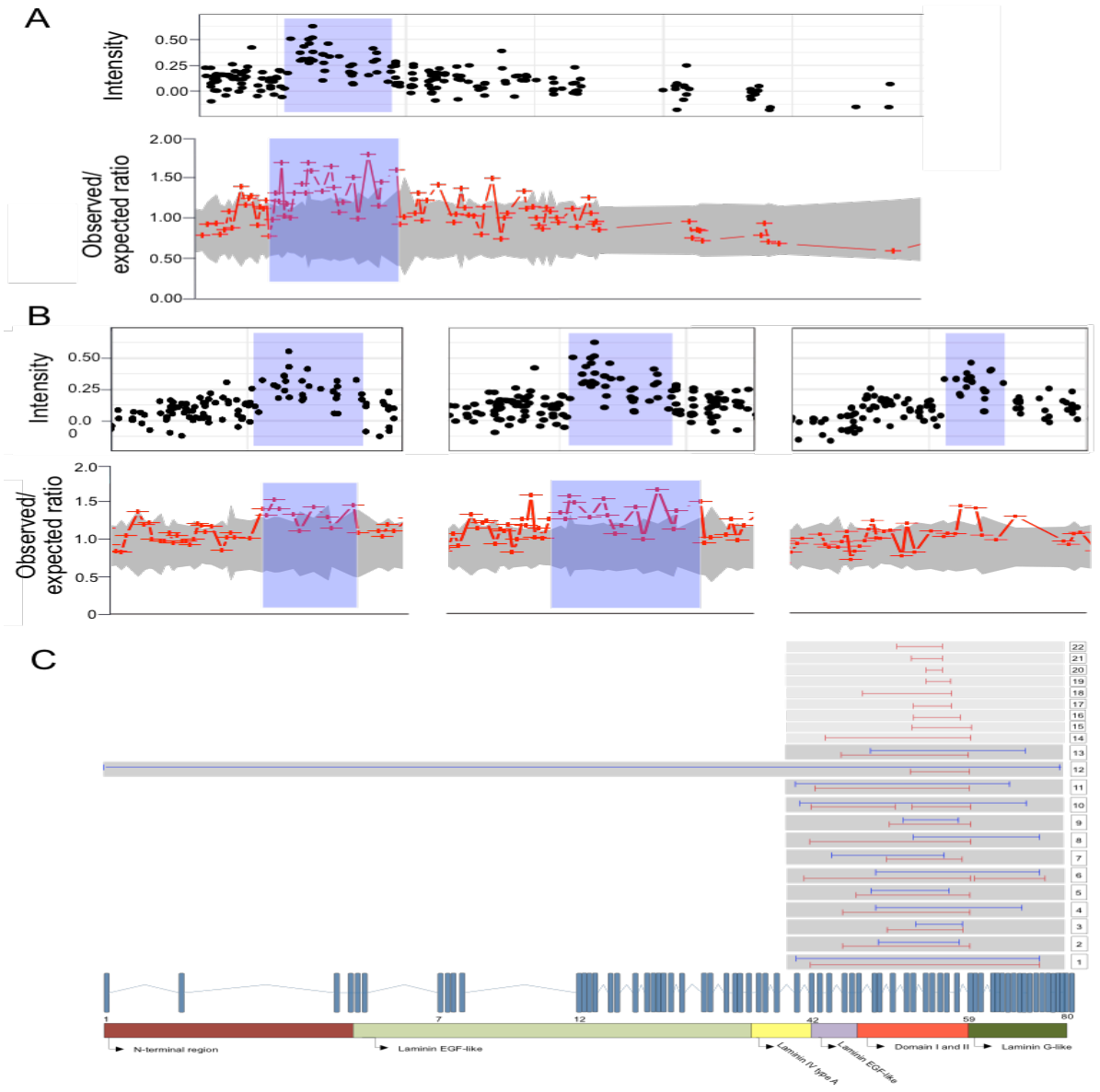


Figure 5.3. Copy number variants in LAMA5. A: SNP genotyping intensity data for Case 10 (Table 1) encompassing base position 60,883,000 on Chromosome 20 (GrCh37 build) to 60,940,000, and thus including the majority of the *LAMA5* gene. The bottom panel of figure A shows the observed:expected read depth from exome sequencing data in the same individual for the same region. Regions highlighted purple were those called as a copy number gain by each respective platform. B: Further examples of SNP genotyping intensity and exome sequencing depth plots for a 22kb region in LAMA5, and of which all detected CNVs were detected. Left; Case 6, middle; Case 10, Right; Case 17. Exome sequencing did not call a CNV within the respective region falling just below the calling threshold. C: A schematic representation of the LAMA5 gene with reference to the cDNA position. Start and end positions of CNVs called by SNP genotyping are shown in red and those validated by ExomeDepth in the same individual shown in blue. Case numbers for individuals are shown in the right hand column, and clinical information for these individuals can be seen in Case 1 and Supplementary Table 8. Exons of the LAMA5 gene are depicted in blue below the CNV start and end positions, and the functional protein domains are shown below.

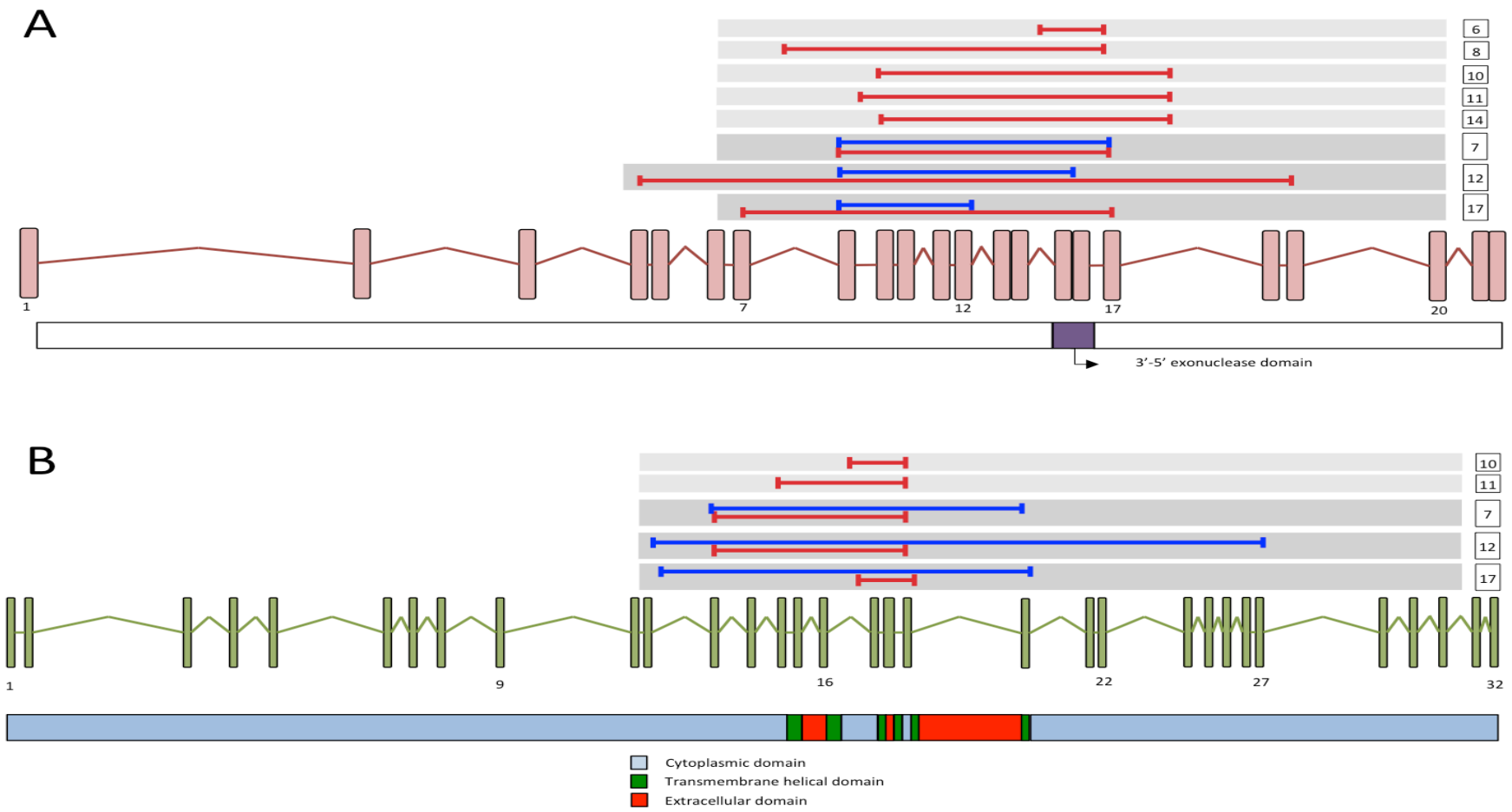


Figure 5.4. Copy number variants seen in EXD3 and TRPM2. A: Start and end positions of CNVs with reference to the cDNA position of EXD3 and called by SNP genotyping are shown in red, while those validated by ExomeDepth in the same individual are shown in blue. Case numbers for individuals are shown in the right hand column, and clinical information for these individuals can be seen in Case 1 and Supplementary Table 8. B: The same representation as above for EXD3 is shown for TRPM2

Given the rarity of each CNV in the population, and assuming independent inheritance (given their location on 3 different chromosomes), the likelihood of an individual having both a *LAMA5* and *EXD3* CNVs or *LAMA5* and *TRPM2* CNV are 1 in 14.07million and 1 in 3.58 million respectively. Therefore, the likelihood of having the observed 8 cases with both *LAMA5* and *EXD3* CNVs for example is less than 2 $x10^{-16}$ (Binomial test). This is therefore suggestive of a common microduplication that is likely to have originally contained all three genes, and which has subsequently contracted in size in some individuals over time rather than multiple independent events involving multiple CNVs. We also performed Identity by descent (IBD) therefore to look for potential common ancestry across all 1342 samples to see if any duplication event was likely to be recently acquired, performing 899811 paired tests between all individuals in the dataset (Figure 5.2, See Chapter 2.3.2.2). We detected no evidence of close ancestry (defined as a Pi_Hat of >0.0625, and equivalent of 4^{th} degree relatives) between any individuals carrying the *LAMA5* CNV or additional CNVs. We also found no evidence of more distant ancestry, with all pairs of individuals with the *LAMA5* CNV present within the bottom 35% of Pi_Hat scores within the cohort, and with mean Pi_hat scores which were actually significantly lower than those between the remaining sCJD cases who did not have the *LAMA5* CNV (*LAMA5* positive cases: mean Pi_hat score: 0.0024 (sd= 0.004), *LAMA5* negative cases: mean Pi_hat score: 0.0049 (sd= 0.0073), p= $1.22x10^{-10}$ (95% CI: $1.7x10^{-5}$ - $5.3x10^{-6}$ Wilcoxon rank sum test).

We also determined the frequency of these CNVs in 8019 individuals within the National Institute of Health Research BioResource funded Next Generation Sequencing project (The BRIDGE consortium), identifying that 25 individuals carried one of the three CNVs of interest (*EXD3*: 16/8019 (0.20%), *TRPM2*: 7/8019 (0.087%), and *LAMA5*: 3/8019 (0.037%), which are consistent with previously observed frequencies in the population (Ruderfer, Hamamsy et al. 2016) (This data was extracted by Dr Wei Wei, University of Cambridge) .

We also surprisingly identified one additional individual within the dataset who also carried both the *LAMA5* and *TRPM2* CN gain, which has an expected frequency of 1 in 3.06 million. Taken together these data are strongly supportive that the cases carrying multiple CNVs which include *LAMA5* most probably share a very distant ancestry, though we cannot exclude that isolated *LAMA5* CN gains occurred independently in a *de novo* fashion (Figure 5.2).

Utilising epidemiological data, we calculated the penetrance of the *LAMA5* CN gain in causing sCJD in the UK. Using the carrier frequency of *LAMA5* in the general population of 0.05%, assuming that all cases of sCJD in the UK occur in individuals over 50 years of age, and assuming that the UK has around 1 million people turning 50 years of age per year (Statistics 2017), then based on current incidence of sCJD (~100 case per year in the UK) (Unit 2017), we calculate the penetrance of the LAMA5 CN gain to be 0.67%.

Finally, to try and determine the mechanism by which LAMA5 CN gains may mediate disease risk, total RNA sequencing from extracted brain tissue in three individuals revealed no evidence significant changes in total gene expression arising as a result of the partial gene duplications (Figure 5.5).

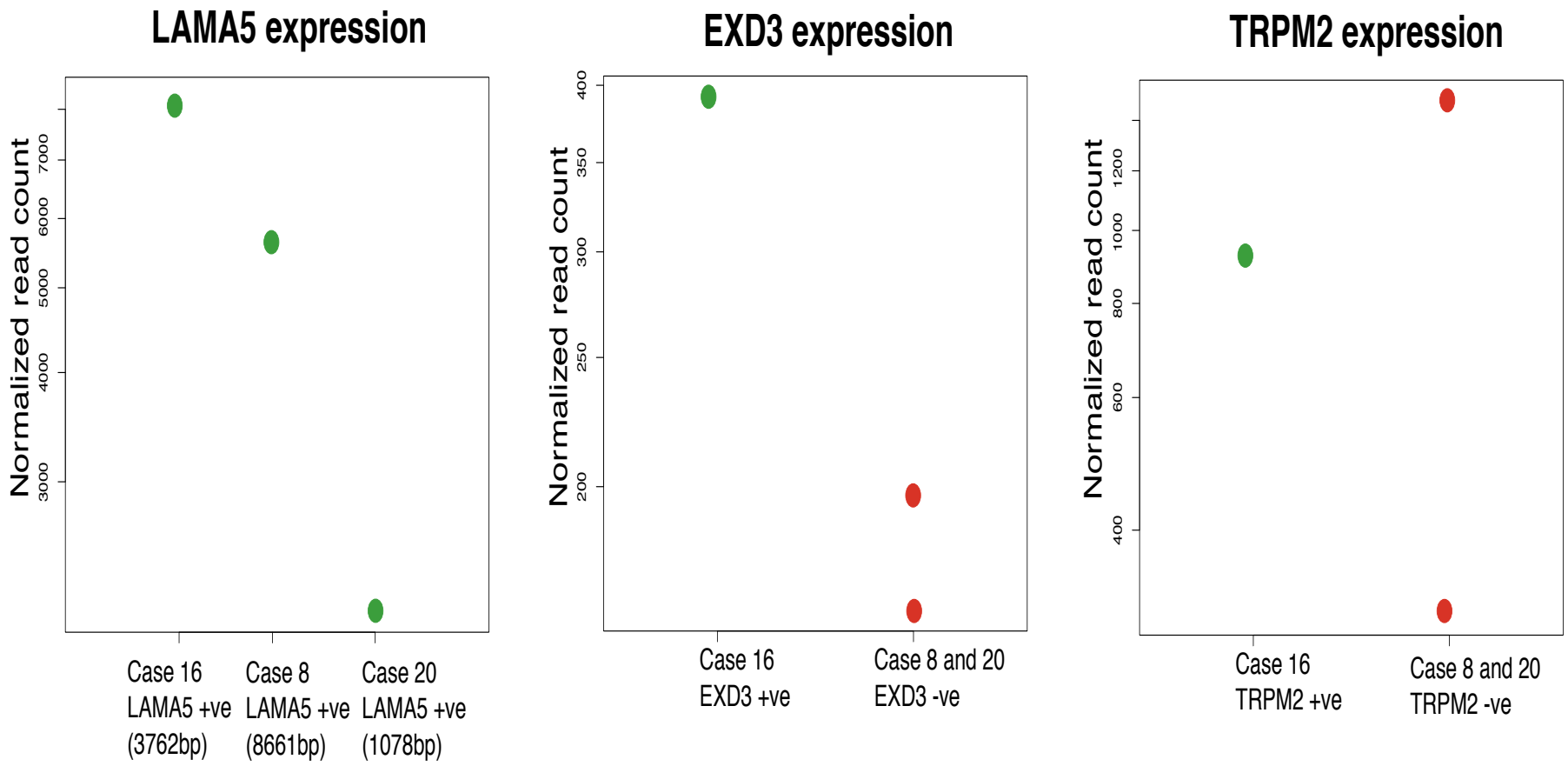


Figure 5.5. Total gene RNA expression for carriers and non-carriers of CNVs. LEFT: the total gene normalized read count for 3 carriers of the LAMA5 copy-number gain. The case number is given (corresponding to data in Table 5.2), carrier status is shown in text and in colour (green=CNV present, red=CNV not present), and the length of the CNV is also shown. MIDDLE: EXD3 expression. Case number and carrier status are shown again together with normalized read count. RIGHT: TRPM2 expression. Case number and carrier status are shown again.

5.5 Discussion

Using a combination of high-density exonic SNP genotyping alongside exome sequencing we detected and validated three small exonic CNVs (mean length; 10,008 bp (SD=10126)) with two of the identified CNVs (*EXD3* and *TRPM2*) always present in combination with the most common CNV in *LAMA5*. The CN gains in *LAMA5* always spanned domains I and II of the protein and were present in 16% of sCJD cases compared to 0.05% of the population, thus conferring an estimated disease penetrance of 0.67% (~1/1500) for sCJD. This places *LAMA5* CN gains as one of the strongest genetic risk factors for any human neurodegenerative disorder discovered to date.

Whilst the majority of rare CNVs arise due to *de novo* events (Kloosterman, Francioli et al. 2015, Yuen, Merico et al. 2016), the frequent co-association of up to three rare CNVs in our study which all reside on differing chromosomes is difficult to rationalize by multiple independent *de novo* events. One potential mechanism explaining this co-association is that the CNVs are somatically generated, with LAMA5 CNVs arising particularly frequently in sCJD patients within the brain. Further work is ongoing to exclude or confirm this possibility. The second possible mechanism explaining the observed co-association is a common distal ancestral microduplication event involving all 3 three CNVs which may underlie the co-association of CNVs seen in some individuals in our study, together with the additional case within 8019 whole genome sequences. Such a duplication event, although speculative, may have been mediated by AluSp mediated retrotransposition. AluSp transposable elements are part of the SINE family of transposable elements and confer a propensity to cause unequal recombination events resulting in the majority of gene-rich segmental duplications in the human genome (Bailey, Liu et al. 2003) (Cordaux and Batzer 2009). Given that an AluSp retrotransposable element is situated only 1000 bases from the mean end position of the *LAMA5* CNV (Hubley, Finn et al. 2016), it is possible that *LAMA5* CNVs may have been

caused by AluSp mediated recombination, although further studies will be required to prove this directly and to track it's ancestral history.

The mechanism by which a partial gene duplication of LAMA5 may confer such an increased risk for disease also remains elusive. *LAMA5* encodes the laminin α5 chain, which when in combination with β and γ laminin chains forms two heteromeric isoforms of laminin (laminin-511 or Laminin-521). These laminin complexes are essential components of the basement membrane, and are particularly highly expressed in vascular endothelium including cerebral vasculature (Wu, Ivars et al. 2009). Laminin proteins have previously been identified to interact with cellular prion protein (PrP^C) through sequence homology between the laminin receptor protein (LRP) between amino acids 161-180 and the PrP protein. This results in a significant overlap of the functional binding domains causing LRP to act as a receptor or co-receptor for the prion protein in mammalian cells (Rieger, Edenhofer et al. 1997).

The LAMA5 CNV is unlikely to cause disease through overexpression of the protein as we observed significant heterogeneity in expression levels between 3 carriers, which did not correlate with either CNV length or age of onset (Figure 5.5). In addition, the single carrier of a TRPM2 gain also showed lower expression that the case carrying the gain, although the EXD3 carrier did show a trend consistent with increased expression of the gene in that individual compared to the other two (Figure 5.5). Taken together, the lack of observed overall gene expression did not appear to correlate with the presence or absence of a CNV within that gene. This perhaps may be expected with partial gene duplications in which the regulatory region of the gene is not duplicated (Lek, Karczewski et al. 2016). In addition, the RIN values varied markedly (between 3 and 8), meaning that the interpretation of relative expression is difficult to interpret (Copois, Bibeau et al. 2007).

Further interrogation of the RNA data will enable us to determine whether there is evidence of novel short transcripts expressed by LAMA5 which would be consistent with active replication of the partial gene duplication resulting in abnormal shortened forms of the LAMA5 protein. Proteomic studies are of course extremely challenging in CJD given the likely transfer of prion proteins within any protein isolation or the significant denaturing effect that prion neutralization steps may have on remaining proteins. Therefore we are at present unable to determine whether the *LAMA5* CNV disrupts the expression of co-located genes or gene networks near to their insertion site within the genome, but is a necessary future avenue for exploration.

Taken together, we have detected the first genetic risk factor for sCJD and potentially all forms of prion disease. We show that small exonic CN gains in *LAMA5, EXD3* and *TRPM2* are likely to occur either from a distant common ancestry, or potentially have arisen *de novo* within several individuals resulting in duplication of key elements of these protein structures. Further functional work is vital in order to understand how these copy-number changes confer such an increased risk for disease. As prion disease pathogenesis results from the combination of prion protein propagation which is determined by the interaction between the prion strain (PrPSc polymer and ensemble) and the host environment (PrP sequence, expression level, modifier genes, and clearance mechanisms) (Collinge and Clarke 2007) we propose these CNVs as the first potential genetic risk factor for prion disorders, transforming our understanding of prion disease aetiology, and opening up new avenues for clinical diagnostics, screening, and treatment.

Chapter 6

High prevalence of focal and multi-focal somatic variation in the human brain

6.1 Aims

1. To develop a highly sensitive and specific method of deep-sequencing to detect and validate low-level somatic mutations within the brain
2. To determine the spatial distribution of these mutations within the brain
3. To calculate the rate of somatic mutagenesis within the brain of control individuals and those with a neurodegenerative disorder
4. To understand the aetiology and mechanisms predisposing to the formation of somatic mutations
5. To use this robust dataset to calculate the likely frequency of pathogenic somatic mutations within the brain

6.2 Introduction

Common neurodegenerative disorders, including Parkinson's disease (PD) and Alzheimer's disease (AD), are characterised by toxic protein aggregation and cell loss in defined brain regions (Goedert 2001, Goedert and Spillantini 2006). The majority of patients have no family history, but in ~5% of cases, germ line genetic variants in one of ~50 genes either cause or contribute to disease risk, with a clinically indistinguishable phenotype (Guerreiro, Bras et al. 2015, Singleton and Hardy 2016). For most neurodegenerative disease genes, a single mutated allele is required to cause disease through haploinsufficiency or a dominant negative effect. This raises the possibility that somatic mutations arising in the same genes within a specific cell lineage contribute to the pathogenesis of non-familial cases. Islands of cells containing these mutations could synthesise misfolded proteins with the potential to spread throughout the brain during human life (Brundin, Melki et al. 2010). Technological limitations have prevented this question from being addressed to date, but if correct, then developmental mutagenesis could be a major cause of sporadic neurodegenerative diseases. The hypothesis can be tested by sequencing an extensive template from a large number of single neurons, or by ultra high-depth re-sequencing from pooled cells in a brain tissue sample. Since the overall low frequency of somatic mutation events approximates the intrinsic sequencing error rate, some form of additional validation is critical, which, because of the destructive character of any one sequencing platform, cannot easily be performed on single cells.

6.3 Methods

6.3.1 Post mortem brains and histopathology.

Frozen brains were identified from the Newcastle Brain Tissue Resource (NBTR) fulfilling both *pre* and *post-mortem* criteria for either Alzheimer's disease (AD, n=20); Parkinson's disease dementia (PDD) or Dementia with Lewy bodies (DLB, n=20); and healthy controls

>65 years old with no clinical *ante mortem* history of cognitive impairment or movement disorder, no family history (>1 first degree relative) with any neurodegenerative disease, and neuropathological features consistent with normal aging (n=15, Appendix 80). 54 cases remained after quality control of the sequencing data (Appendix 80 & 81). $1cm^3$ blocks of grey matter were manually dissected from each region (frontal cortex, entorhinal Cortex, cingulate gyrus, and medulla) after excluding macroscopically identifiable white matter or vascular tissue (Appendix 80 and 81). Quantitative neuropathological data was available on fixed sections from the same regions using AT8 (for phospho-tau), 4G8 (for β-amyloid) and α-synuclein antibody staining on slides from all brain regions except the cerebellum .

6.3.2 Ultra-high depth sequencing

Two panels were defined incorporating 102 genes. Panel 1 included 56 genes known to cause or predispose to, common neurodegenerative disorders (132,617 base pairs, bp, Appendix 78, left) identified through a previous systematic review we performed (Keogh, Kurzawa-Akanbi et al. 2016). In addition, 46 control genes expressed at low levels in the brain and associated with cancer (152,519bp**,** Appendix 78, right, and subsequently referred to as 'cancer' genes).

The primary sequencing platform for the study was the Accuracy and Content Enhanced, (ACE) platform (Personalis Inc, USA) which required the extraction of 1µg of genomic DNA from brain, blood or HapMap controls (see Chapter 2.3.2.1 for methods). Subsequently, this DNA was Covaris sheared, end repaired and ligated with adaptors. Adaptor ligated DNA fragments were amplified by PCR with 6 cycles, and subjected to SureSelect enrichment with probe panels for all 102 genes. An additional 10 cycles of PCR were performed with enriched material. For the Haloplex^{HS} platform, in parallel, but independently, equal quantities of DNA to those utilised for the ACE platform were enriched by the same method. Sequencing of enriched DNA from both capture protocols was performed using HiSeq 2500 (Illumina, San

Diego, CA, USA) sequencers with single lane, paired-end 2 × 101 bp reads and Illumina's proprietary Reversible Terminator Chemistry (v3). The steps following DNA extraction were performed by the team at Personalis Inc, USA.

Variant discovery was performed using the ACE platform. The design of probes for this panel were performed by the technical team at Personalis Inc, USA. This enabled 99.6% of coding bases to be covered at > 1000 fold depth across both gene panels, some of which are notoriously difficult to sequence comprehensively (eg. *MAPT;* Figure. 6.1a, Appendix 79). The overall mean depth for the 102 genes using the ACE platform was 5,374-fold (SD = 745; Figure. 6.1b). Variant validation was performed using an amplicon based barcode-tagged platform (Haloplex^{HS}) with a mean depth of 6,830-fold (SD =1,549; Figure.6.1b) (Appendix 79). This panel was designed using the Agilent SureDesign tool to capture all exons and 25bp of intronic flanking regions of all 102 genes.

6.3.3 Sample Dilutions for the Level of Detection (LOD) testing

The limits of detection of minor alleles was determined for both platforms using HapMap CEPH cell line DNA (Coriell Institute) to simulate allele frequencies (AF) ranging 0.2%, 0.5%, 1%, 2% and 5%. DNA from cell line NA12878 was spiked into DNA from cell line NA12877 by using 2ng, 5ng, 10ng, 20ng or 50ng of NA12878, making up to 1µg total DNA with NA12877 thus achieving the stated simulated VAFs. High confidence, pedigree consistent germ-line variants, concordantly called on several platforms from high-depth after PCR-Free sample preparation were obtained for each sample from http://www.illumina.com/platinumgenomes/, and sequencing reads were processed using the same bioinformatics pipelines as described above. Duplicate HapMap dilution samples at

0.2%, 0.5%, 1%, 2% and 5% VAF were called against a 0% 'pure' sample by MuTect2 (Cibulskis, Lawrence et al. 2013), Varscan2 (Koboldt, Zhang et al. 2012),(Koboldt, Chen et al. 2009) and deepSNV (Gerstung, Papaemmanuil et al. 2014). The sensitivity and specificity of variant callers were determined using the following filtering formulae:

Sensitivity = True Positive /(True Positive + False Negative)

Specificity = 1 – False Positive Rate = 1 – False Positive/(False Positive + True Negative).

6.3.4 Bioinformatic pipeline

The bioinformatic pipeline is shown in Figure 6.1d and in detail in Appendix 83. In overview, the primary calling pipeline included variants called by either MuTect2 (Cibulskis, Lawrence et al. 2013) or Varscan2 (Koboldt, Zhang et al. 2012) at a minimum VAF of 0.5%, before DeepSNV (Gerstung, Papaemmanuil et al. 2014) confirmed the presence of the detected variants in identified samples, and also confirmed that the VAF of detected somatic variants alleles in other brain regions was no different to the base error rate of other samples (where appropriate, Figure. 6.1d, Appendix 83). The primary analysis was performed on the ACE data, and subsequently validated by the Haloplex^{HS} platform.

ACE platform - Somatic Single Nucleotide Variants (SNVs) and small indels were called using MuTect2 (Cibulskis, Lawrence et al. 2013) and Varscan2 Somatic Calling (Koboldt, Zhang et al. 2012) with the default parameters within the BED file for neurodegenerative and cancer 'control' genes. To detect Single Region Mutations (SRMs) we ran two callers on all possible sample pairs from one individual allocating each sample as the 'germline' or 'tissue' sample in turn. Multiple Region Mutations (MRMs) were called using Varscan single-sample Calling with VAF >0.1% which may not be detected by paired calling, particularly in equivocal VAF between samples. We subsequently excluded variants: (1) with <1000 total

read depth; (2) with <10 mutant reads; (3) <4 reads from the end of either forward or revise strand; (4) those called as germ-line variants; (5) variants with a Minor Allele Frequency (MAF) >1% in 1000 genome project database (Genomes Project, Abecasis et al. 2012), NHLBI ESP-6500 or ExAC database (Lek, Karczewski et al. 2016); and; (6) those within simple tandem repeats, segmental duplications, and microsatellites. All candidate mutations were subsequently annotated by ANNOVAR (Wang, Li et al. 2010). The same approach was used to analyse brain and blood combinations with a VAF > 0.5% based on spiked control sample data. Please note that the alignment and calling was performed by Dr Wei Wei (University of Cambridge), but designed by myself and Dr Wei Wei.

To ensure SRMs were truly focal, we utilised an additional caller to maximise the value of our large homogenous dataset (n=173 samples at a mean coverage of 5,374x) which ensured that identified SRMs and MRMs did differ from the base error rate seen in other samples. The DeepSNV (Gerstung, Beisel et al. 2012) caller enabled us to build a separate error model for each base, and test whether the variant allele detected in the sample is greater than that expected from a beta-binomial distribution with an associated over dispersion factor which captures the observed degree of variation within the control samples. Testing this variant caller in spiked control samples (Figure.6.1c&e) showed this caller to be the most specific caller of all those tested. To validate the focal variants, we ran deepSNV using other samples from the same individual as reference samples with a Benjamini-Hochberg corrected p-value for the number of samples tested. This ensured SRMs were: (1) Identified by MuTect2 and VarScan 2 Somatic Calling as present in only one region; (2) Significantly different to all other samples within that individual (eg in the cerebellum and frontal cortex when detected in medulla); (3) Significantly different from all other samples from all other individuals (n=170 or n=171) at the corrected threshold; and, (4) any other samples (e.g. frontal cortex and cerebellum) within that individual did not differ to all other samples from other individuals.

MRMs were defined as variants with a VAF between 0.5% - 20% called in a single sample by Varscan2, and which were present in more than one region within the same individual. These putative MRMs were confirmed with deepSNV using the samples from other individuals as reference samples, and therefore MRMs were defined as follows: (1) as any variant identified by VarScan 2 (v2.4.0) single-sample calling to be present in more than one region; and which, (2) did again significantly differ to all other samples from other individuals at the corrected threshold.

Haloplex platform - To determine the accuracy of this approach, we validated identified variants using the Haloplex^{HS} system data. In total 89.5% bases within the target region were covered above 1,000x and 94.3% bases were covered above 500x (99.4% bases with coverage above 1,000x covered by ACE). Alignment and calling was performed by Dr Wei Wei. We also manually reviewed and confirmed the read alignments for all somatic mutations detected on the ACE platform and those also covered by Haloplex^{HS} using IGV software (v.2.3.30 31) (Robinson, Thorvaldsdottir et al. 2011) confirming their presence.

Comparison of the detected VAF from each platform showed a strong correlation for detected somatic variants ($r^2 = 0.953$, $p<2.2e-16$)(Figure.6.1f). These data indicate that our detection of somatic mutations was highly specific, and given that the DNA was independently amplified and sequenced, are highly unlikely to be due to amplification artefact.

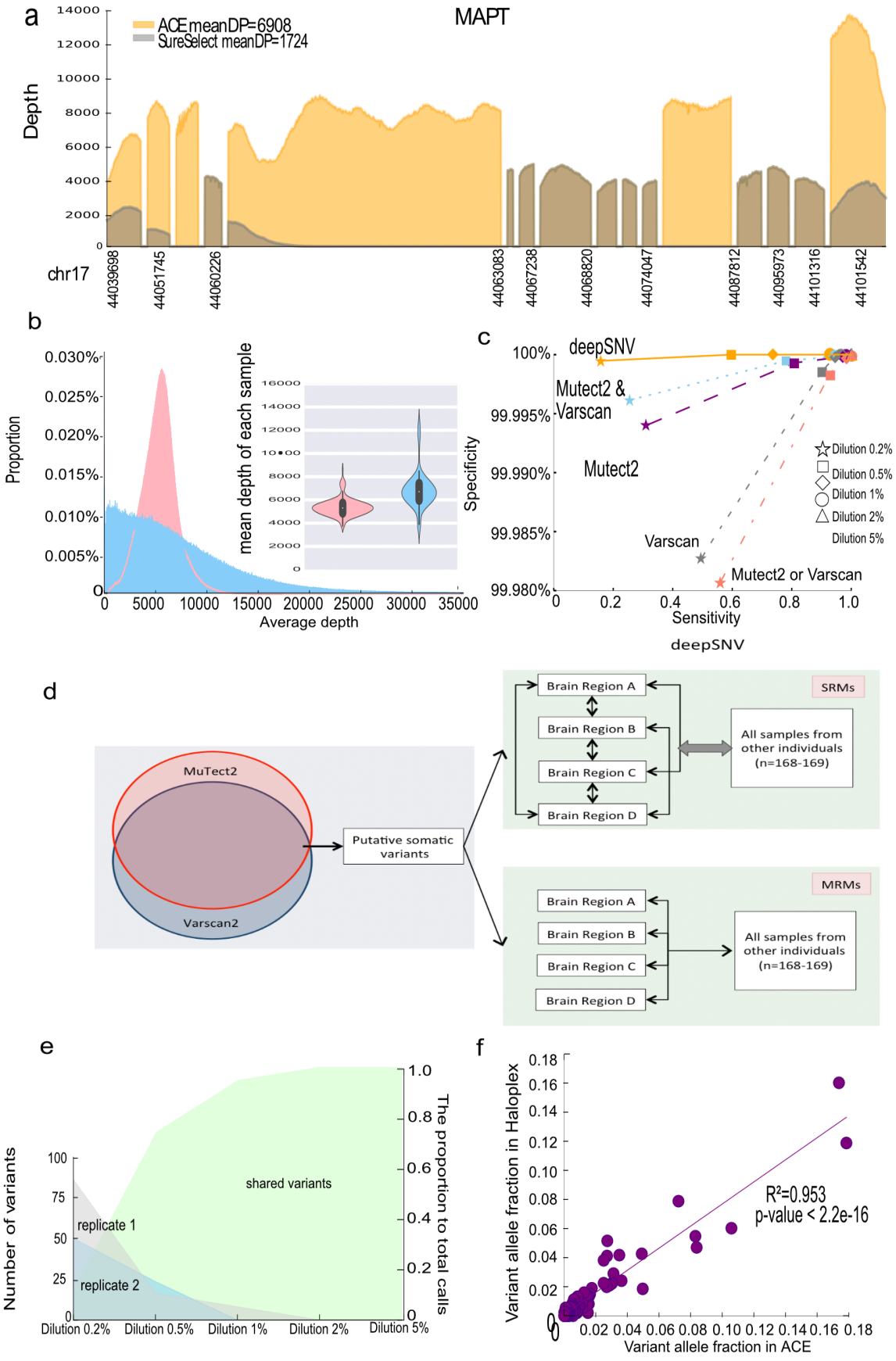

a
MAPT
ACE meanDP=6908
SureSelect meanDP=1724
Depth
chr17
b
Proportion
Average depth
mean depth of each sample
c
deepSNV
Mutect2 & Varscan
Mutect2
Varscan
Mutect2 or Varscan
Specificity
Sensitivity
Dilution 0.2%
Dilution 0.5%
Dilution 1%
Dilution 2%
Dilution 5%
d
MuTect2
Varscan2
Putative somatic variants
Brain Region A
Brain Region B
Brain Region C
Brain Region D
All samples from other individuals (n=168-169)
SRMs
MRMs
e
Number of variants
The proportion to total calls
shared variants
replicate 1
replicate 2
Dilution 0.2%
Dilution 0.5%
Dilution 1%
Dilution 2%
Dilution 5%
f
Variant allele fraction in Haloplex
Variant allele fraction in ACE
R²=0.953
p-value < 2.2e-16

Figure 6.1. Genotyping platform performance and quality. (a) Coverage plot of the *MAPT* gene on the Accuracy and Content Enhanced (ACE) platform highlighting the augmented coverage over and above that seen within the SureSelect platform alone (yellow – regions covered by custom augmented probes, brown regions - covered by ACE probes without augmentation). (b) The average depth per base for each sequencing platform: ACE (pink) or Haloplex^{HS} (blue). The mean sequencing depth across the whole panel per sample is shown in the inset violin plot for both platforms again using the same colour scheme. (c) Sensitivity and specificity of 5 different combinations of variant callers at five different VAF of HapMap control mixes. DNA from cell line NA12878 was spiked into DNA from cell line NA12877 by using 2ng, 5ng, 10ng, 20ng or 50ng of NA12878, making up to 1µg total DNA with NA12877 to create relative VAF of variants exclusive to sample NA12878 of 0.2%-5% (as indicated by varying symbols in the figure). DeepSNV(Gerstung, Beisel et al. 2012) , MuTect2 (Cibulskis, Lawrence et al. 2013), Varscan2 somatic caller (Al-Chalabi, Fang et al. 2010) and Varscan un-paired calling (Koboldt, Chen et al. 2009) were employed in different combinations (as shown) to determine the sensitivity an specificity to detect variants at each VAF (as denoted by differing symbols). The optimum caller pipeline was set at a VAF of 0.5% using a dual calling approach for variants called by either MuTect2 or Varscan2 which had a 92.98% sensitivity and 99.9984% specificity. (d) Schematic overview of the calling algorithm used in the study. Variants called from paired sample calling by either MuTect2 or Varscan2 were selected. All brain regions from those individuals were then compared to both each other (for intra-regional variation), and then to all other individuals (for inter individual variation) for those alleles to ensure that detected variants from Varscan2 or MuTect2 were truly focal or Multifocal in nature. (e) Number of observed variants called by either MuTect or Varscan at each VAF in biological replicates highlighting consistent performance of the sequencing pipeline. (f) Correlation of variant allele frequency (VAF) for the 39 detected variants in this study called by MuTect2 or Varscan2 in the SureSelect platform, and by Varscan in the Haloplex^{HS} platform.

6.4 Results

After quality control and validation, we detected 39 somatic variants in case and control genes with confidence (35 single nucleotide variants, SNVs; and 4 insertion-deletion variants, indels) from 27 brains (50% of the entire cohort - Controls: n=6/14, AD: n=9/20, LB: n=12/20) and 173 total samples.

Eighteen mutations (48.7% of all variants) were present in only one brain region within an individual (Single Region Mutations, SRMs, Figure.6.2a&b, Figure.6.3a-d, Appendix 84 & 85). These SRMs were detected at a mean VAF of 0.84% (sd=0.005) (cerebellum, CB = 6/54 brains, 11.1%; entorhinal cortex, EC = 7/53 brains, 13.2%; frontal cortex, FC = 2/32 brains, 6.3%; medulla, Med = 3/24 brains, 12.5%), and were equally likely to occur in neurodegenerative disease (7/132,617bp) or cancer (11/152,519bp) (P = 0.64, Chi-squared with Yates correction). The majority of SRMs in brain were C>T substitutions (n=15/18 (83.3%), Figure. 6.3f) consistent with spontaneous deamination of 5-methyl-cytosine (Pfeifer 2006), as observed in single neurons (Lodato, Woodworth et al. 2015). Purine-purine transitions on the non-template strand were exclusively seen in case genes (n=4/7) (p=0.01 vs Control genes, Fisher's exact test) in-keeping with replication-transcription collisions (Sankar, Wastuwidyaningtyas et al. 2016), as seen in single cortical neurons (Lodato, Woodworth et al. 2015). However, given that the same rare mutations were detected in ≥0.5% of alleles, it is highly likely that they arose during development. The flanking 5' and 3' sequence of the SRMs were distinct from mutational 'signatures' described in cancers (Alexandrov, Nik-Zainal et al. 2013), suggestive of a different mechanism of mutagenesis (Figure. 6.3f, Appendix 86). Seven mutations occurred in neurodegenerative disease genes with a mean VAF of 0.82% (SD=0.003, range 0.47-1.56%). There was no difference in the proportion of SRMs in neurodegenerative disease genes between any disease group (AD 5/20; LB = 1/20; Control 1/14) (Figure. 6.3a-d).

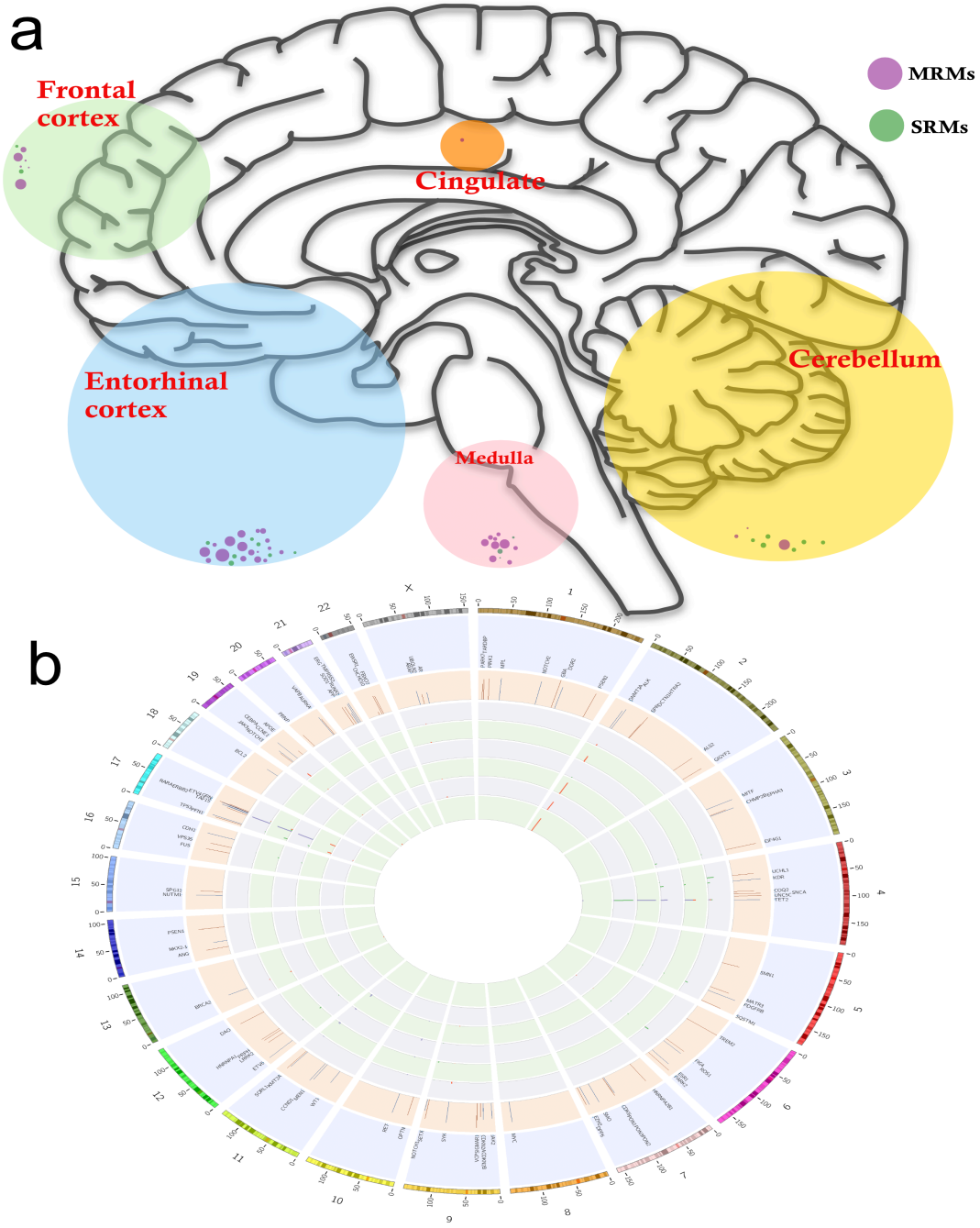


Figure 6.2. Validated detected mutations in 179 samples from 54 individuals. (a) Neuroanatomical origin of the brain samples sequenced. Large circle radii are proportionate to the number of DNA molecules sequenced in each region. Smaller circles represent the proportion of Single Region Mutations (SRMs, green) or Multiple Region Mutations (MRM, purple) detected within that region in the 54 brains. (b) Circos plot showing the detected variants. Outer to inner: genomic positions on the autosomes and X-chromosome; 102 genes sequenced; mean sequencing depth for each gene; Variant Allele Frequency (VAF) in the cerebellum; VAF in the entorhinal cortex; VAF in the frontal cortex; VAF in the medulla; VAF in the Cingulate; VAF in the blood. Neurodegenerative disease gene mutations shown in red, and cancer gene mutations in blue.

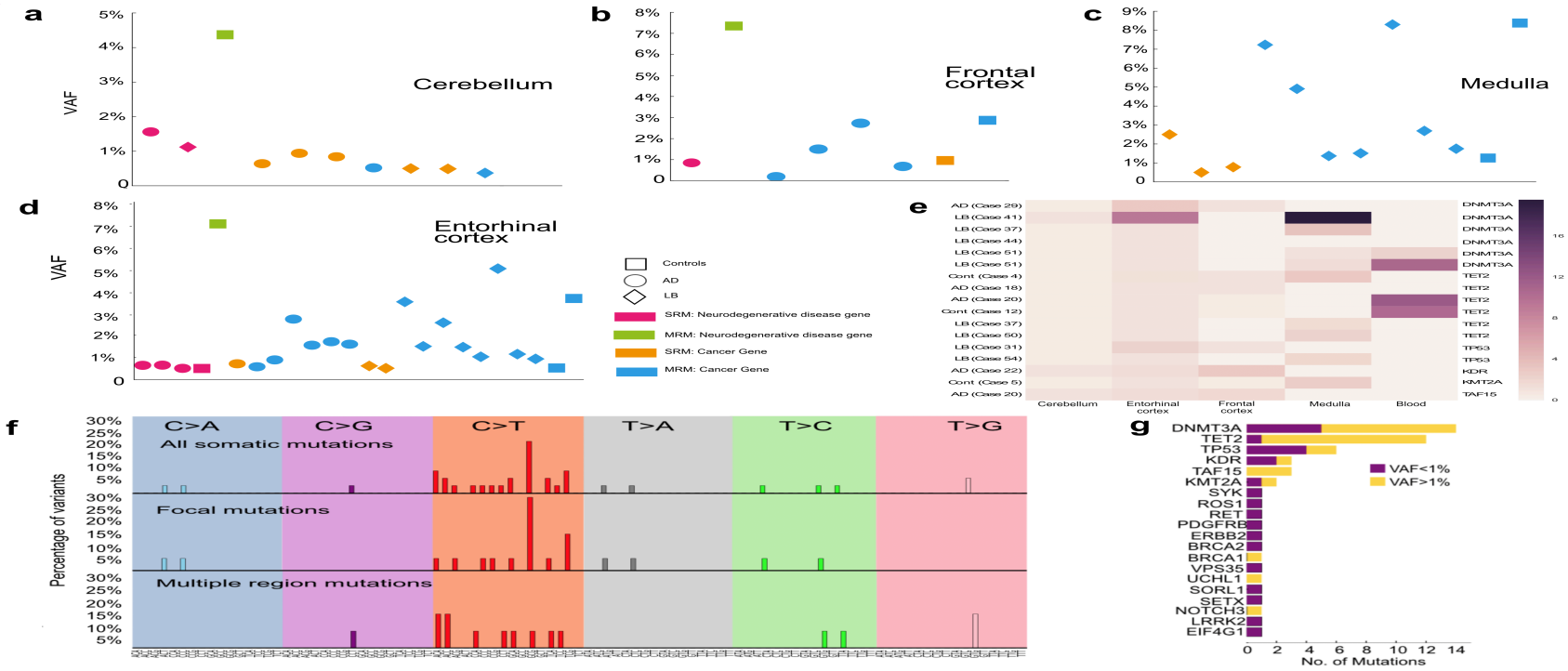


Figure 6.3. The regional distribution of mutations detected in 179 samples from 54 human brains. (a-d) All detected somatic mutations within each brain region. Single Regional Mutations (SRM) or Multiple Regional Mutations (MRM) are indicated by differing colours, and the disease phenotype indicated by differing symbols. The Variant Allele Frequency (VAF) for each mutation from the Accuracy and Content Enhanced (ACE) platform is shown for each case. Only a single mutation in the cingulate was detected and therefore is not shown. (e) Heat map showing the relative Variant Allele Frequencies (VAF) as defined as the ratio between the lowest VAF compared to the VAF for the indicated sample for MRMs, highlighting that the VAF was consistently higher in the Medulla when sampled. (f) The mutational signatures of each detected mutation in the study. The x-axis shows the 5' and 3' flanking base of each detected mutation, with the middle of the three alleles the reference allele that was mutated. The single base change for that allele is shown the column in which it is located (eg C>T etc), with each base mutation within a different column and depicted by a different colour. (g) Frequency of SRM and MRM mutations in specific genes seen in the 54 individuals.

Having determined the somatic mutation rate across the case and control panel (285,136 bases) in 179 brain regions (~611,285 cells based on barcode tagging data), we used approximate Bayesian computation (Figure. 6.4a & Appendix 87 – and all mathematical modelling work was performed by Dr Nick Jones and Juvid Aryaman at Imperial College London), was used to determine the somatic focal mutation rate in the human brain at 2.99 $x10^{-9}$ – 4.8 x 10^{-10} per base per cell division (95% Bayesian credible interval (BCI)) (Figure. 6.4b); a figure highly comparable to somatic mutation rates in mitotically active tissues (Figure. 6.4b) (Tomasetti, Vogelstein et al. 2013). Using the measured mutation rate, we modelled the frequency and size of brain regions harbouring known pathogenic mutations in neurodegenerative disease genes (Human Gene Mutation Database (Stenson, Mort et al. 2014)) using contrasting models of neurodevelopment. We found extremely similar estimates for the size and number of foci of cells harbouring pathogenic mutations predicted to occur within the brain under these models (Figure. 6.4c-e), with 10.76% (± 0.11%) of simulated individuals having one or more regions of $2.62x10^5$ spatially contiguous pathogenically mutated cells (Figure. 6.4f), and all individuals harbouring 75-481 regions (95% BCI) of 128 pathologically mutated cells (Figure. 6.4g).

Taken together, this shows that focal regions harbouring pathogenic point mutations are highly likely to be very common in the human brain. These regions have the potential to generate mutant proteins that form novel fibrillar structures, which can spread and cause different neurodegenerative diseases (Frost, Ollesch et al. 2009), or modify the clinical phenotype, depending on the original mutated gene.

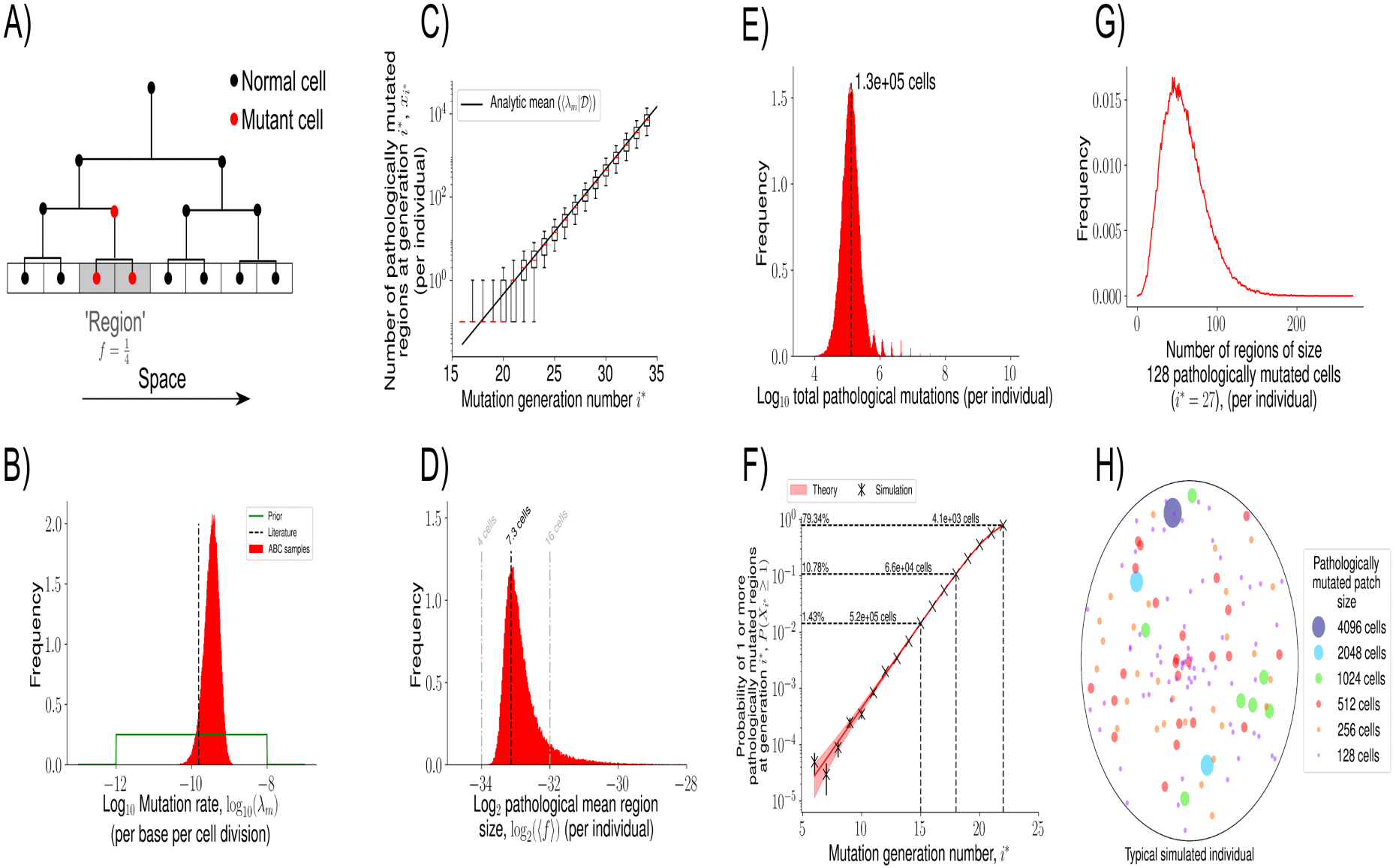

A)
Normal cell
Mutant cell
'Region'
$f = \frac{1}{4}$
Space
C)
Analytic mean $\langle \lambda_m | \mathcal{D} \rangle$
Number of pathologically mutated regions at generation i^*, x_{i^*} (per individual)
Mutation generation number i^*
E)
1.3e+05 cells
Frequency
Log_{10} total pathological mutations (per individual)
G)
Frequency
Number of regions of size 128 pathologically mutated cells ($i^* = 27$), (per individual)
B)
Prior
Literature
ABC samples
Frequency
Log_{10} Mutation rate, $\log_{10}(\lambda_m)$ (per base per cell division)
D)
4 cells
7.3 cells
16 cells
Frequency
Log_2 pathological mean region size, $\log_2(\langle f \rangle)$ (per individual)
F)
Theory
Simulation
79.34%
4.1e+03 cells
10.78%
6.6e+04 cells
1.43%
5.2e+05 cells
Probability of 1 or more pathologically mutated regions at generation i^*, $P(X_{i^*} \geq 1)$
Mutation generation number, i^*
H)
Pathologically mutated patch size
4096 cells
2048 cells
1024 cells
512 cells
256 cells
128 cells
Typical simulated individual

Figure 6.4. The frequency and distribution of somatic mutagenesis and pathogenic mutations in the human brain copy number variants seen in EXD3 and TRPM2. (a) Model of mutagenesis arising from neurodevelopment (see Supplementary Information, Supplementary Fig. 3). Red circle = cell containing a somatic mutation, black circles = cell with the wild-type allele. We model regional mutations as originating from neurodevelopment, where neurodevelopment consists of a simple branching process where a founder cell doubles exactly 36 times to form an adult brain of exactly 2^{36} neurons. Every base is assumed to have a constant probability of mutation during replication. When a mutation arises in the lineage, its mutant daughters occupy a fraction f of the adult brain which we refer to as a 'region'. (b) Prior (green solid) and approximate posterior (red bars) for the mutation rate using approximate Bayesian computation. We infer the mutation rate (λ_m) to be 1.28–7.9 $\times 10^{-10}$ per base per cell division (95% Bayesian credible interval (B.C.I.)). This is compatible with studies by Roach *et al.*(Rahbari, Wuster et al. 2016) (black dotted) and Tomasetti *et al.*(Tomasetti, Vogelstein et al. 2013) (black dashed). The value from Roach *et al.* was converted from per base pair per generation to per base per cell division by dividing by (2 × 36), being the number of strands per DNA and approximately the number of cell divisions during neurodevelopment respectively. (c–f) Using the approximate posterior samples for λ_m (see (b)) we simulated the neurodevelopment of individuals under our model (one sample for each individual). (c) Distribution of the number of mutated regions per individual, for each mutant generation number (i^*), in any of the case genes associated with pathology (298 bp/132617 bp). Boxplot whiskers show the 5th and 95th percentiles, median shown in red (when an individual has 0 mutations associated with generation i^*, the number of mutations is assigned to be 0.1, for representation on a log-scale). For $i^* \leq 19$, the majority of patients had zero regions seeded at these generations; for the prevalence in the population of these larger mutant regions, see (f). Displaying the analytic mean of Eq.(15) (Appendix 88) for a choice of λ_m as the mean of the approximate posterior distribution (black solid). (d) Distribution of the pathological mean relative region size across individuals. Showing the mode of the distribution in number of cells (black dashed) and the number of cells corresponding to two region sizes (grey dashed). (e) Distribution of total number of pathological mutations across individuals. Note that this is not necessarily equal to the number of mutated cells, as a single cell may harbour multiple mutations. Showing the mode of the distribution (black dashed). Multimodality is induced by the largest pathological region (see Supplementary Information, Supplementary Fig. 3). (f) Fraction of simulated individuals with at least 1 pathologically mutated region seeded at generation i^*. Showing the corresponding number of cells in the adult brain under the model as $2^{36} \cdot f$ where $f = 2^{-(i+2)}$ (Appendix 87 and Appendix 88). Comparison to theory shown in red, mean defined in Eq.(19) and standard deviation Eq.(20). (Error bars are a Bernoulli error model, see Appendix 88) (g) Frequency distribution over individuals of pathologically mutated regions of size 128 cells. (h) Visual representation of the size and frequency distribution of pathologically mutated regions (seeded between $22 \leq i^* \leq 27$) in an individual. Whole brain area (black circle) is not to scale with the mutated regions (coloured circles) (Appendix 88).

Seventeen mutations (43.6% of the total 39 detected variants) were present in more than one brain region, or in a paired blood sample and brain (Multiple Region Mutations, MRMs) (Figure. 6.2a, Appendix 84 & 85). These mutations had a significantly higher VAF than SRMs (3.67%, SD=0.04, p=0.0024, Figure. 6.3a-e). Only one of these variants occurred in a neurodegenerative disease gene (Case 6: p.R464R in *TAF15*, mean VAF 6.23%, SD=0.016) which was present in all 3 brain regions sampled from this single control individual (VAFs in Cerebellum: 4.37%, Entorhinal cortex:6.98%, Frontal cortex:7.35%). Based on these observations, mutations within neurodegenerative disease genes will be present diffusely across the brain in up to 9.77% of all humans (95% CI 0.33%-9.77%, Wilson score interval test). Sixteen MRMs (94%) occurred in cancer genes, with 15 (93.8%) known to be associated with myeloproliferative blood disorders (n=18 of the 53 genes in the cancer panel). This was greater than expected when compared to solid organ tumour or non-cancerous control genes (Appendix 78) (n=28 of 53 genes, P=7.45 x10^{-6}), raising the suspicion that specific MRMs were derived from the circulating blood cells. The two genes most frequently mutated in our study (*DNMT3A*, n=6; *TET2*, n= 6) account for the majority of age related clonal haematopoietic mutations (Jaiswal, Fontanillas et al. 2014, McKerrell, Park et al. 2015); and four of the MRMs (23.6% of MRMs, *DNMT3A* p.R882H, *DNMT3A* p.P700L, *DNMT3A* c.1667_splice, *TET2* c.3472_splice) involved known mutational hotspots (Jaiswal, Fontanillas et al. 2014) (Appendix 85), and given that such specific alleles were also detected in our study, together with the consistent spatial VAFs observed within the brain, strongly support a clonal haematopoetic aetiology for these mutations rather than an early developmental origin. In keeping with this, the VAF of the myeloproliferative gene mutations was always greater in available paired blood samples than in the brain (n=4, Appendix 84 & 85). However, the fold difference was surprisingly low (mean Blood:brain VAF ratio = 7.92 (range 2.10 to 11.98), suggesting that at least some of the rare clonal mutations were present in cells outside the

vasculature, most probably including migratory immune cells (Prinz and Priller 2017). Intriguingly, non-synonymous or frame-shift mutations in hematopoietic disorder genes were detected in 40% of LB brains (8/20), in contrast with controls (7%, 1/14; p=0.05, where the frequency was consistent with previous reports (Thal, Rub et al. 2002, Braak, Del Tredici et al. 2003, Busque, Patel et al. 2012, McKerrell, Park et al. 2015)).

Given the role of *DNMT3A* and *TET2* in regulating DNA methylation (Rasmussen, Jia et al. 2015, Yang, Rau et al. 2015), these findings provide an explanation for the concordant changes in DNA methylation seen in the blood and brains of PD patients who share Lewy body pathology (Masliah, Dumaop et al. 2013). The spatial distribution of *DNMT3A* and *TET2* variants mirrored the quantitative neuropathology, with the VAF in medulla 2.1-fold greater than entorhinal cortex (SD=0.69, p=0.0064). This could reflect regional weakness of the blood-brain barrier (BBB) seen in PD (Feuk, Carson et al. 2006, Monahan, Warren et al. 2008, Lubbe, Escott-Price et al. 2016) and other neurodegenerative disorders (Desai, Monahan et al. 2007). Given that clonal haematopoetic variants in *TET2* accelerate age-related atherosclerosis in mice (Ruderfer, Hamamsy et al. 2016), our findings raise the possibility that blood cell precursors harbouring somatic mutations translocate into the brain and contribute to the pathogenesis and clinical presentation of neurodegenerative diseases through cells derived from myeloid precursors (Genovese, Kahler et al. 2014).

6.5 Discussion

In conclusion, based on the observations of 173 human brain tissue samples, and ~ 611,000 cells, our findings indicate that the human brain is likely to contain many zones of cells harbouring somatic mutations in neurodegenerative disorders. If pathogenic, these mutations have the potential to cause neurodegenerative disease and modify the clinical course. This provides a potential explanation for common sporadic neurodegenerative diseases, which

currently affects ~10% of people in the developed world (Hurd, Martorell et al. 2013). In addition, the similar mutation rate seen in cancer genes (albeit predisposing to non-neurological forms of cancer) also suggest that somatic mutagenesis may predispose to other focal neurological abnormalities. It is conceivable that detecting these mutations during life will increase diagnostic precision, leading to new therapies, particularly if they involve targets amenable to pharmacological intervention within vulnerable neural circuits (Canter, Penney et al. 2016).

Chapter 7

The interaction of inherited and acquired mitochondrial DNA mutations with Lewy body pathology in a mouse model of Parkinson's disease

7.1 Aims

1. To generate mice with both inherited and acquired mitochondrial DNA heteroplasmy in a transgenic mouse model of Parkinson's disease in which human alpha-synuclein is over-expressed.
2. To perform thorough phenotypic assessment of all transgenic mice.
3. To determine whether the presence of either inherited or acquired mitochondrial DNA mutations exacerbate or improve the phenotype of mice over-expressing human alpha-synuclein.
4. To determine whether the presence of either inherited or acquired mitochondrial DNA mutations exacerbate or improve the neuuropathology of mice over-expressing human alpha-synuclein.
5. To begin to develop hypotheses by which mitochondrial DNA mutations may influence the development of Parkinson's disease-like pathology in mice.

7.2 Introduction

Parkinson's disease (PD) is the second most common neurodegenerative disorder, affecting approximately 1% of the elderly population in the Western world (Nussbaum and Ellis 2003). It is clinically characterized by a progressive movement disorder (Postuma, Berg et al. 2015) and pathologically by α-synuclein-positive neuronal inclusions (commonly referred to as Lewy bodies) and associated dopaminergic neuronal loss primarily within the substantia nigra pars compacta (Schulz-Schaeffer 2010).

Mitochondrial dysfunction is a common feature of Parkinson's disease, with the observation of mitochondrial biochemical abnormalities within the substantia nigra and frontal cortex of the brain, together with non-neuronal tissues of patients (Shoffner, Watts et al. 1991, Schapira 1994, Keeney, Xie et al. 2006, Navarro, Boveris et al. 2009). One mechanism hypothesised to predispose to these biochemical abnormalities is the development of somatic heteroplasmic mutations occurring with age within the 16.5Kb mitochondrial genome, which is present in tens to thousands of copies within every nucleated cell, and functions to encode key proteins within the mitochondrial respiratory chain (Sciacco, Bonilla et al. 1994). Significantly increased levels of somatic mitochondrial DNA (mtDNA) point mutations and deletions within the substantia nigra (Simon, Lin et al. 2004, Bender, Krishnan et al. 2006, Kraytsberg, Kudryavtseva et al. 2006, Coxhead, Kurzawa-Akanbi et al. 2016) and neocortex (Simon, Lin et al. 2004, Coxhead, Kurzawa-Akanbi et al. 2016) in patients with PD compared to age-matched controls have been observed by numerous authors, confirming that they are a common feature of the disease and suggesting that they may contribute to disease pathogenesis.

The potential that such heteroplasmic mutations may contribute to age-related neurodegenerative disorders has recently been given additional focus following observations

in mice showing that the inheritance of increased low level point mutations may predispose to premature aging (Ross, Stewart et al. 2013), thus raising the interesting possibility that inherited low level mtDNA mutations and not just somatically acquired mutations may predispose to aging and therefore, potentially, to age related neurodegenerative disorders.

However, patients with primary mitochondrial disorders, for example pathogenic mutations in mitochondrial DNA polymerase gamma (*POLG*) which generate higher levels of somatic mtDNA mutations within the substantia nigra and cortex do not develop Lewy Body accumulation (Tzoulis, Tran et al. 2013, Rajakulendran, Pitceathly et al. 2016). In addition, *in vitro* studies suggest that deletions may function to actually trigger neuroprotective compensatory mechanisms (Perier, Bender et al. 2013). Therefore, whether somatically acquired mtDNA mutations are a cause or consequence of Lewy body pathology remains unclear.

To investigate this, we generated a mouse colony with mice that have a predisposition to both inherited or acquire mtDNA mutations in conjunction with over expression of human α-synuclein to determine whether inherited or acquired low level variants exacerbates motor, cognitive, or pathological features of Parkinson's disease.

7.3 Methods

7.3.1 Mouse selection and breeding

Mice: Two Male C57Bl/6 *PolgA*$^{wt/mt}$ mice (Ross, Stewart et al. 2013) were obtained from Dr J Stewart and Prof N Larsson at The Max Planck Institute for Biology of Ageing, Cologne, Germany. These mice contain a mutation in the exonuclease domain of POLG that results in exonuclease deficiency (Figure 7.1). These mice were crossed with two C57Bl/6 wt females to generate initial breeding pairs (Figure 7.2). In parallel, two male C57Bl/6 tg(PDGFB-

SNCA) mice (Masliah, Rockenstein et al. 2000) (Figure 7.1) were purchased from QPS labs (Austria) and were also crossed with C57Bl/6 wt females to generate initial breeding pairs of that arm of the breeding programme. Subsequent offspring were bred as outlined in Figure 7.2.

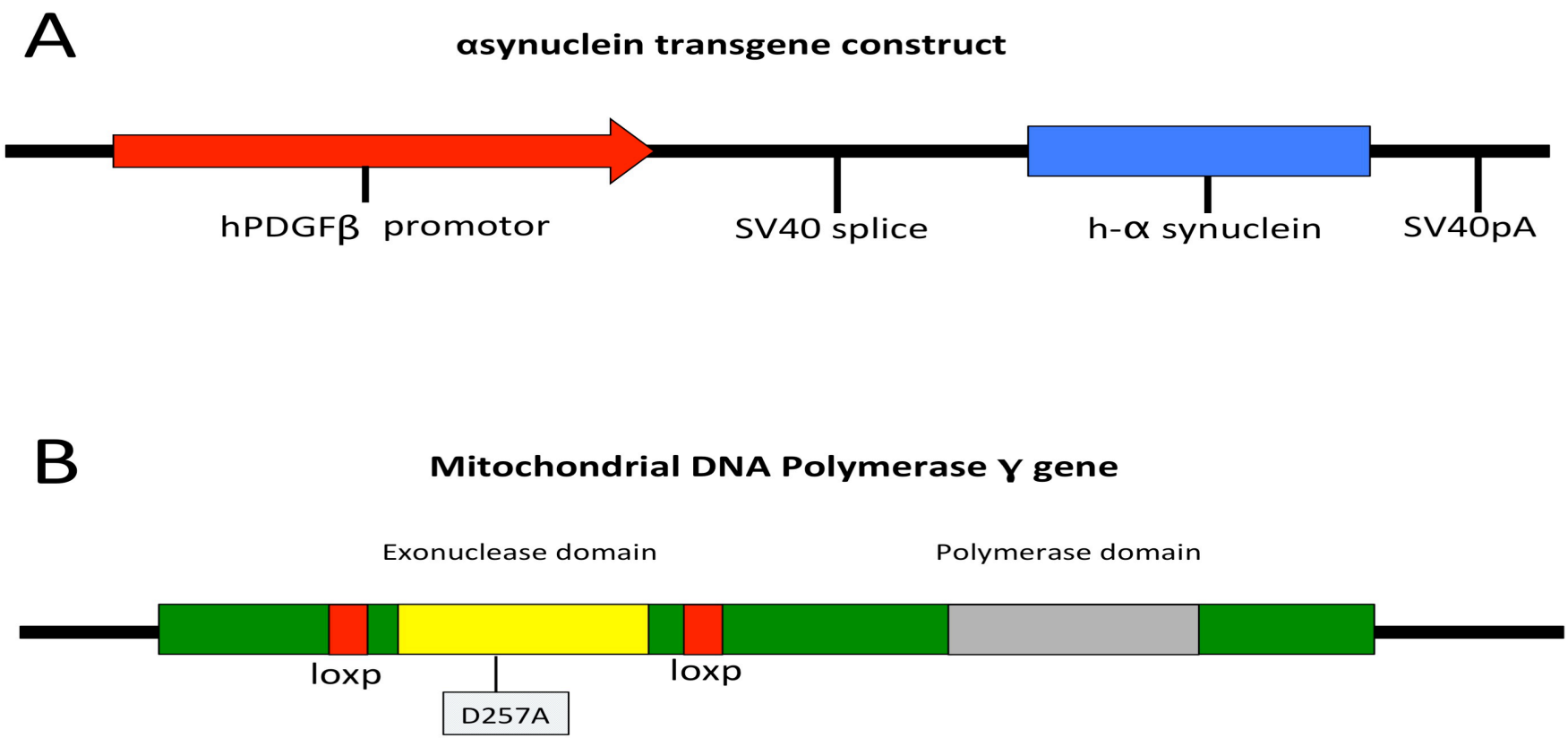


Figure 7.1. Gene constructs of both the mutant α-synuclein gene and mitochondrial DNA polymerase gamma (POLG). (A) A 1480 base pair fragment of human PDGF-b chain gene is situated upstream of a Not I-Sal I fragment sonsisting from 5' to 3' of an SV40 splice, 423bp of human cDNA encoding full length wild-type a-synuclein and SV40 sequence from the pCEP4 vector providing a polyadenylate signal as previously described(Masliah, Rockenstein et al. 2000). (B) A mutation within the gDNA of the murine POLG gene causing an amino acid change from aspartate to alanine within the exonuclease domain is present, flanked by loxp flanking sites either side of the exonuclease domain.

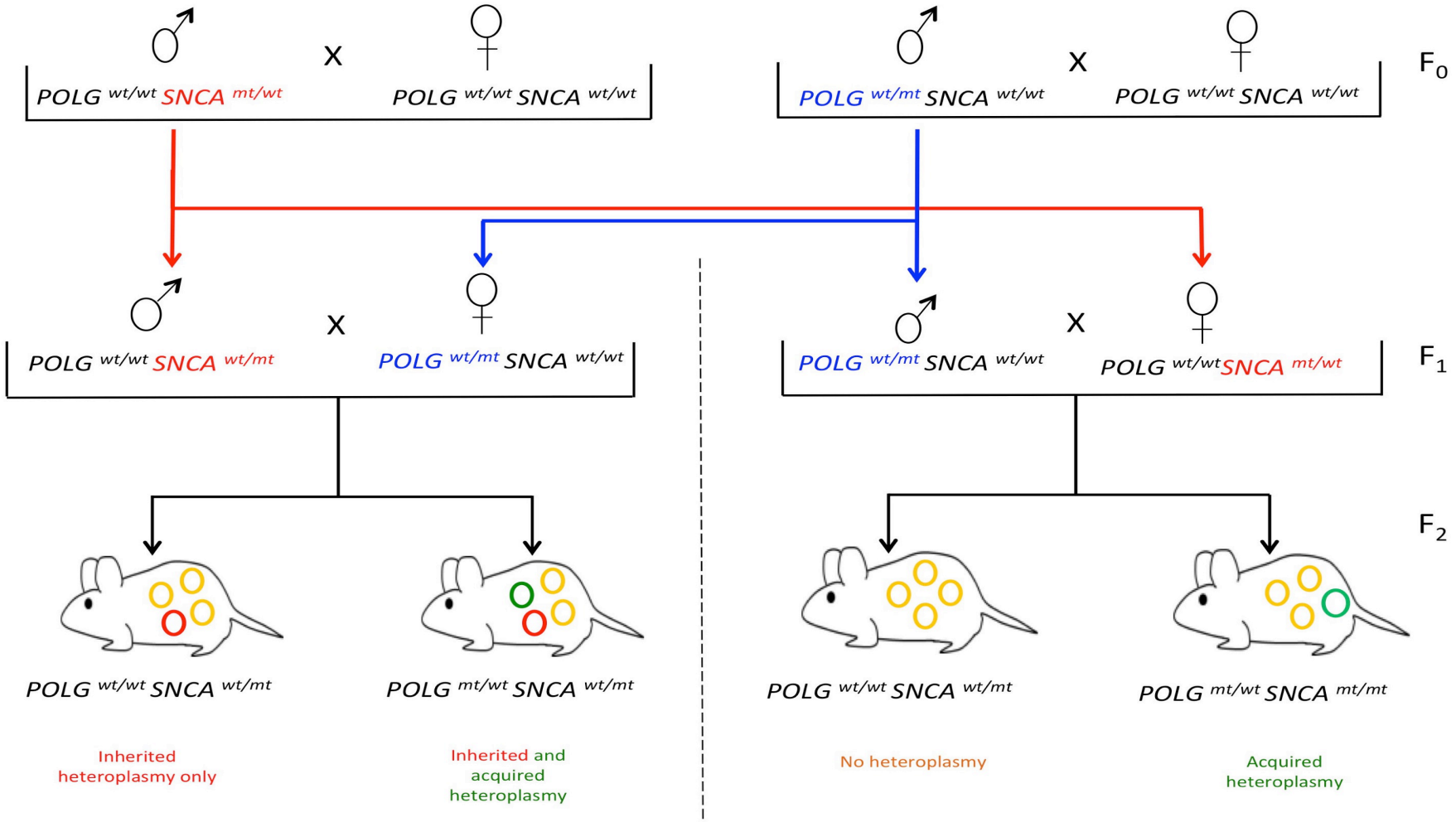


Figure 7.2. The mouse breeding programme. Initial breeding pairs of male POLG mutant and a-synuclein mutant mice were bred with female C57/bl6 littermates. Offspring with heterozygous mutations in the POLG and synuclein genes were then bred together in generation F1 to generate the final four genotypes of mice (F2) that inherit mtDNA mutations (left two genotypes) and those that do not inherit mtDNA mutations (right two genotypes)

7.3.2 DNA extraction and genotyping

DNA extraction: Mice were ear notched at 6 weeks of age, and DNA was extracted from ear clippings via a rapid lysis protocol consisting of mixing tissue with 100 µl of lysis buffer for 30 minutes at 95°C, before adding a further 100 µl of 40mmol tris(hydroxymethyl)aminomethane (TRIS, ThermoFisher Scientific, UK).

POLG genotyping: Was adapted from the method previously published by *Ross et al* (Ross, Stewart et al. 2013) and *Trifunovic et al* (Trifunovic, Wredenberg et al. 2004); Genotyping was performed in a 50 µl total volume reaction; 2µl of extracted DNA was mixed with 10µl of GoTaq buffer (Promega Ltd, UK), 5µl of PCR nucleotide mix, 1.26µl of forward and reverse primers (10pmol/µl) [F: CTTCGGAAGAGCAGTCGGGTG, R: GGGCTGCAAAGACTCCGAAGG], 0.35µl of GoTaq (DNA polymerase) (Promega Ltd, UK), 30ul H_2O. PCR protocol: 1x 94°C for 60 sec, 30 cycles of [94°C for 30 sec, 60°C for 30 sec, 72°C for 60 sec], 72°C for 3 minutes. 10µl of PCR product loaded onto a 2% agarose gel; 65V for 25 minutes. Genotype was determined by PCR fragment length; wt/wt: 520bp, wt/mt: 720 + 520bp, mt/mt: 720bp (Figure 7.3).

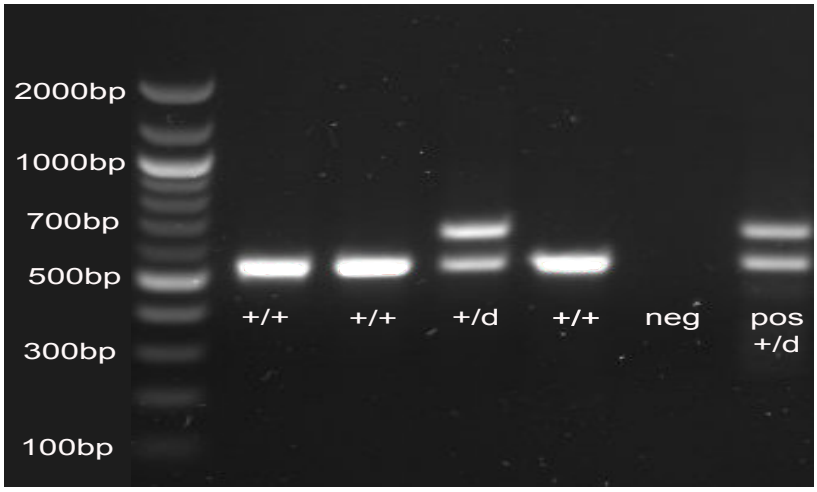


Figure 7.3. Agarose gel electrophoresis results for POLG genotyping. Results for four mice are shown together with positive and negative controls. Wild type mice (+/+) have a single ~ 520bp band whereas heterozygous POLG mutant mice (+/d) have a 520bp band plus at 700bp. An established POLG heterozygous founder male provided by Dr James Stewart (Max-Planck Institute, Germany) was used as a positive control.

***SNCA genotyping**:* Was adapted from that published by *Masliah et al* (Masliah, Rockenstein et al. 2000). Genotyping was performed in a 50µl total volume reaction; 2ul of extracted DNA was mixed with 10µl of GoTaq buffer (Promega Ltd, UK), 5µl of PCR nucleotide mix, 1.26µl of forward and reverse primers (10pmol/µl) [F:CCAGCGGGGCCGCTCTAGAACTAGTG, R: CCAAGGTTGTTAACTTGTTTATTGCAGC], 0.35µl of GoTaq (DNA polymerase) (Promega Ltd, UK), 30µl H_2O. PCR protocol: 1x 94°C for 60 sec, 30 cycles of [94°C for 30 sec, 55°C for 30 sec, 72°C for 30 sec], 72°C for 5 minutes.

10µl of PCR product loaded onto a 2% agarose gel; 65V for 25 minutes. Genotype was determined by PCR fragment of 800nt (Figure 7.4).

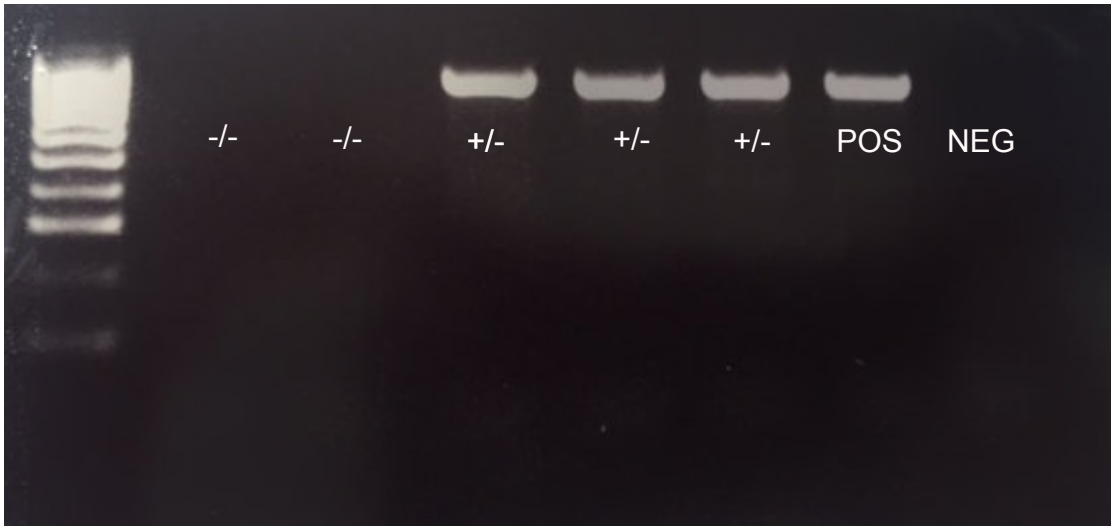


Figure 7.4. Agarose gel electrophoresis results for SNCA genotyping. Results for five mice are shown together with positive and negative controls. Mice containing the human SNCA gene are indicated by the presence of a band at ~800bp (+/-), and those that only contain the native mouse synuclein gene by the absence of any band (-/-). An established SNCA mutant founder male from the initial breeding pair was used as a positive control.

7.3.3 Phenotypic assessment of mice

7.3.3.1 Mouse handling and housing

Mice were handled and housed in accordance with the Home Office Code of Practice in 290x180x160mm cages in a 12hr:12hr light/dark cycle (active phase 08:00-20:00). The room was kept at a temperature range of 19-23°C with humidity at 50% (+/-10%). Both filtered water and CRM(P) feed (SDS, UK) were available to all mice at all times.

7.3.3.2 Weight

Mice were weighed every month from the age of three months onward using an animal weighing scale (Kent scientific, USA).

7.3.3.3 Rotarod

Animals were pre-trained on an automated 4-lane rotarod unit (Columbus Inst; Columbus,

Ohio) prior to the first assessment at 3 months of age. Pre-training involved one day of habituation with the equipment (10 minutes per mouse), followed by three days of practice runs on the rotarod (outlined subsequently). Motor coordination testing was performed using an accelerating rotarod protocol by measuring the latency to fall for each animal (Sedelis, Schwarting et al. 2001). Briefly, mice were transferred in their home cages to the testing room, and acclimatised for 30 minutes. Each mouse undertook 3 trials on the rotarod separated by 15 minute inter-trial intervals. The rotarod unit was set on an accelerating protocol from 4 to 40rpm in 300 seconds. The time taken for the mouse to fall was measured for each animal on each assessment. Mean latency times (seconds) and speed at time of fall (rpm) were recorded for each animal (Figure 7.5). This was performed every month from 3 to 14 months of age for each animal. Subsequent analysis was performed between age and genotype matched animals.

Figure 7.5. An image of mice being tested on the rotarod. The speed of the rod gradually increases until the mice are unable to continue to balance on the rod. They subsequently fall, automatically triggering the end of the timing period.

7.3.3.4 Mousetrapp

At 9,12 and 14 months of age, assessment of spontaneous fine motor behaviour was performed. Mice were placed for 5 minutes within an acrylic chamber (21cm (I) x 14 cm (w) x 15 cm (h)), with the base of chamber being the touch-screen sensitive portion of a Samsung Galaxy Tab E Quad Core tablet (Samsung, Suwon, South Korea). MouseTrapp software (Neurolytical, Ann Arbor, MI, USA) recording was commenced as the mouse is placed into the chamber (Mabrouk, Dripps et al. 2014). The relative position of the contact between the mouse paw and screen were recorded in a two dimensional X-Y grid format. These co-ordinates were converted to the number of steps, stride length, total distance, and position in the field at each time point either automatically or manually using R from X-Y coordinate data (Mabrouk, Dripps et al. 2014).

7.3.3.5 Open field test (OFT)

Were conducted in a 19cm x 35cm x 54cm open field plexiglass chamber at the same temperature and humidity as the standard housing conditions of the mice. A camera (Logitech HD Webcam C270, Logitech, USA) was suspended 85cm above the centre of the field, and data was recorded using QuickTime Player 7 on an Apple Macintosh OSX operating system at 25 frames per second (FPS). On day one to three of the experiment, mice were habituated to the open field for 10 minutes every 12 hours. On day four, the experiment was conducted, and prior to testing, mice were again acclimatized to the testing room for 10 minutes within their home cages.

For testing, mice were transferred from their home cages into the open field and recorded for 5 minutes. Afterwards the open field was cleaned with 70% ethanol and allowed to dry for 5 minutes. Subsequently, Tracker Video Analysis and Modelling Tool software (https://physlets.org/tracker/) was utilized to track the centre position of each mouse within

the field with an automated scaling factor relative to the size of the field. Mouse tracking was manually performed when automated tracking failed, with the centre position of each mouse recorded at 0.3 second intervals for 5 minutes. Subsequently, an X-Y plot of each mouse within the field was utilized to calculate both (a) total distance moved, and (b) centre-time (as determined by the proportion of time that each mouse spent within the X-Y co-ordinates of X: 13.5-40.5cm and Y: 8.75-26.25cm) (Figure 7.6).

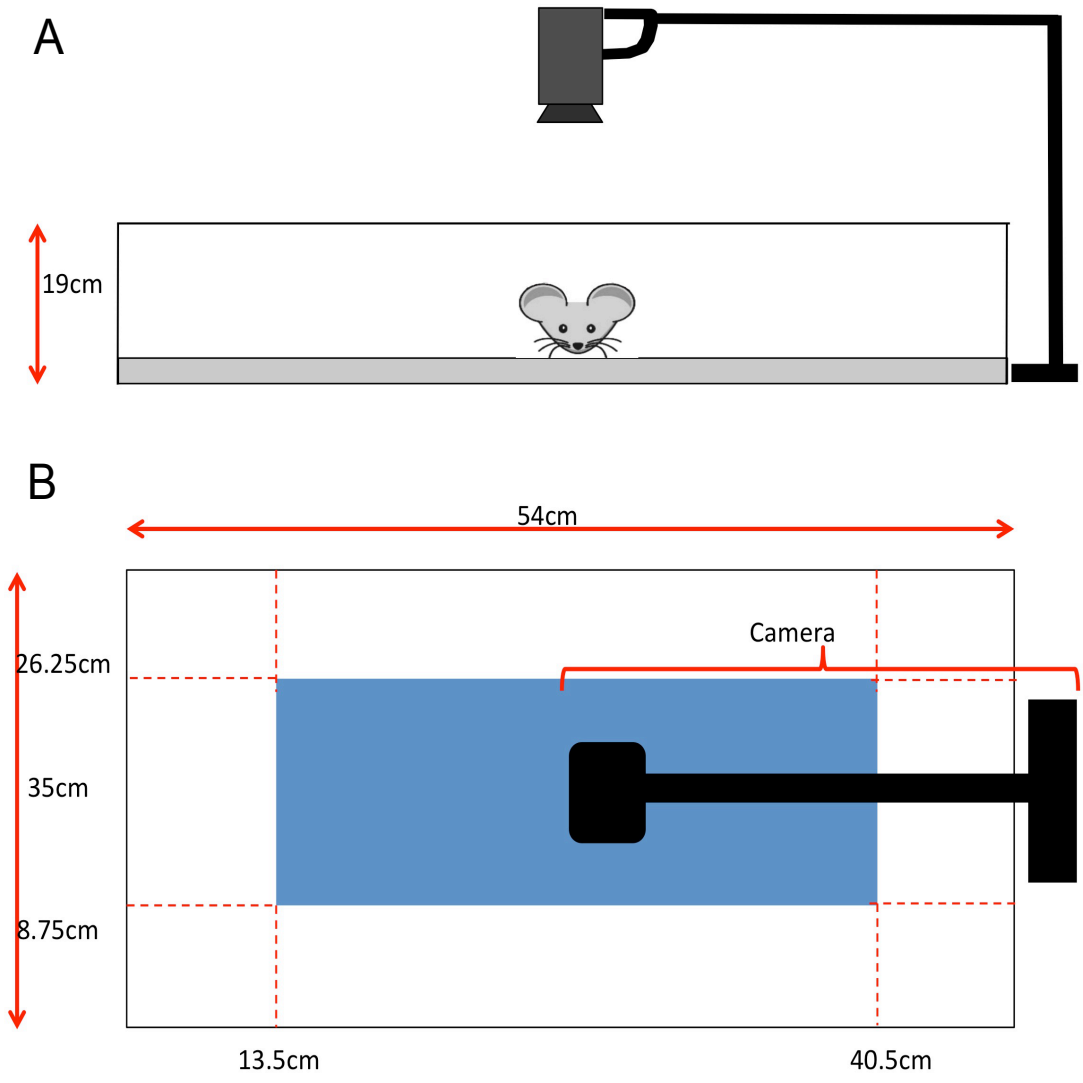


Figure 7.6. Open field testing apparatus and set-up. Results A: Side view, illustrating the camera placed centrally over the open field. B: Aerial schematic illustrating the size of both the total field and central field.

7.3.3.6 Novel object recognition test (NORT)

Were again conducted in a 19cm x 35cm x 54cm open field plexiglass chamber at the same temperature and humidity as the standard housing conditions of the mice. A camera (Logitech HD Webcam C270, Logitech, USA), was suspended 85cm above the centre of the field, and data was recorded using QuickTime Player 7 on an Apple Macintosh OSX operating system at 25 frames per second (FPS). On day one to three of the experiment, mice were habituated to the open field for 10 minutes every 12 hours.

On day four, mice were acclimatized to the testing room for 30 minutes within their home cages before 10 minutes of object exploration time was performed. The objects utilized in this exploration phase were two sand filled flasks (which have appropriate features as described by Ennaceur et al to act as objects of recognition (Ennaceur 2010) (Leger, Quiedeville et al. 2013)) (Figure 7.7) and were situated 5cm in from the walls of opposing corners of the cage. All experiments were conducted between 08:00am and 12:00pm (Walf and Frye 2007). A Perspex sheet was placed under each object with a 2cm radius around each object marked. Following the exploration phase, the field and objects were cleaned with 70% ethanol and allowed to dry for at least 5 minutes, before the mouse was placed back into it's home cage.

Four hours later, one of the objects utilised in the familiarization phase was replaced with a different object (Figure 7.7), with the position of the novel object randomized between the two corners each time. Again a Perspex sheet was placed under each object with an appropriate 2cm margin around each object. Mice were recorded for 5 minutes and viewed at a later time. During playback, mice were considered to have 'explored' the object whenever the mouse sniffed or touched the object (and in which the nose-object distance was less than 2 centimetres). Climbing onto the object, and chewing the object, together with standing within 2 centimetres with the nose pointed away from the object were not considered as exploratory

behaviour (Leger, Quiedeville et al. 2013). In addition, mice that spent less than 10 seconds exploring either object were a removed from the analysis. The total time spent exploring either the novel and old object were recorded manually using a stopwatch, and the discrimination index (DI) was subsequently calculated. This measure allows discrimination between the novel and familiar objects, i.e., it uses the difference in exploration time for familiar object, but then dividing this value by the total amount of exploration of the novel and familiar objects [$DI = (T_N - T_F)/(T_N + T_F)$], where T_N is the time spent exploring the novel object and TF the time exploring the familiar object. This result can range between +1 and −1, where a positive score indicates more time spent with the novel object, a negative score indicates more time spent with the familiar object, and a zero score indicates a null preference (Aggleton, Albasser et al. 2010, Antunes and Biala 2012).

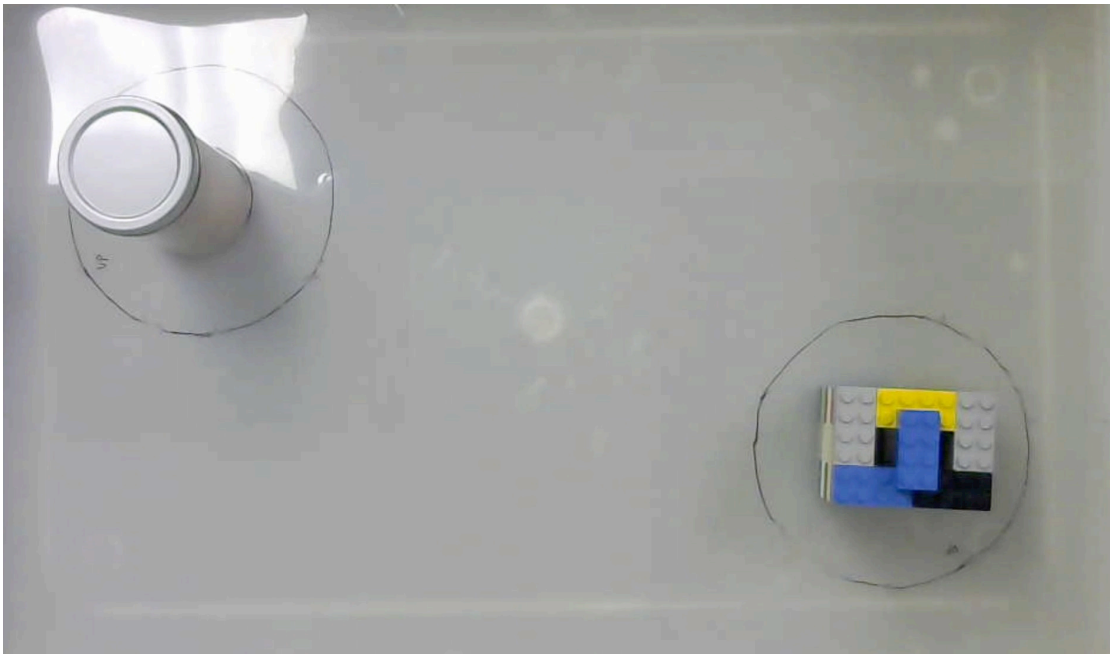


Figure 7.7. Novel object recognition test apparatus and set-up. The 'old' object (top left – sand filled container) is shown along with a 2cm margin to define the exploratory field, and bottom right the novel object (lego tower), also with a 2cm margin is shown. The position of the novel and old object were varied with each run.

7.3.3.7 Spectrophotometric assessment

These assessments performed after culling using a handheld R300 Minolta spectrophotometer (Konica, Singapore) which measures the Commission International d'Eclairage (CIE) L*a*b* system of colour which was established in 1976, with L* representing the lightness-darkness axis, a* the red-green axis, and b* the blue-yellow axis. Prior to the assessment of mice, the spectrophotometer was calibrated against a pure white background to ensure that L* values were greater than 95.

Subsequently, 4 mice from each genotype within the study were analysed with triplicate measurements from the hind-limb (Figure 7.8) using the spectrophotometer, with the L*, a* and b* average taken for each mouse, and ΔL*, Δa*, Δb* determined by subtracting a standard matt-black sample. Cohort differences were then determined by the ΔL*, Δa*, Δb* between groups together with the total colour difference between groups (ΔE*) as determined by the mean $\Delta E^* = [\Delta L^{*2} + \Delta a^{*2} + \Delta b^{*2}]1/2$.

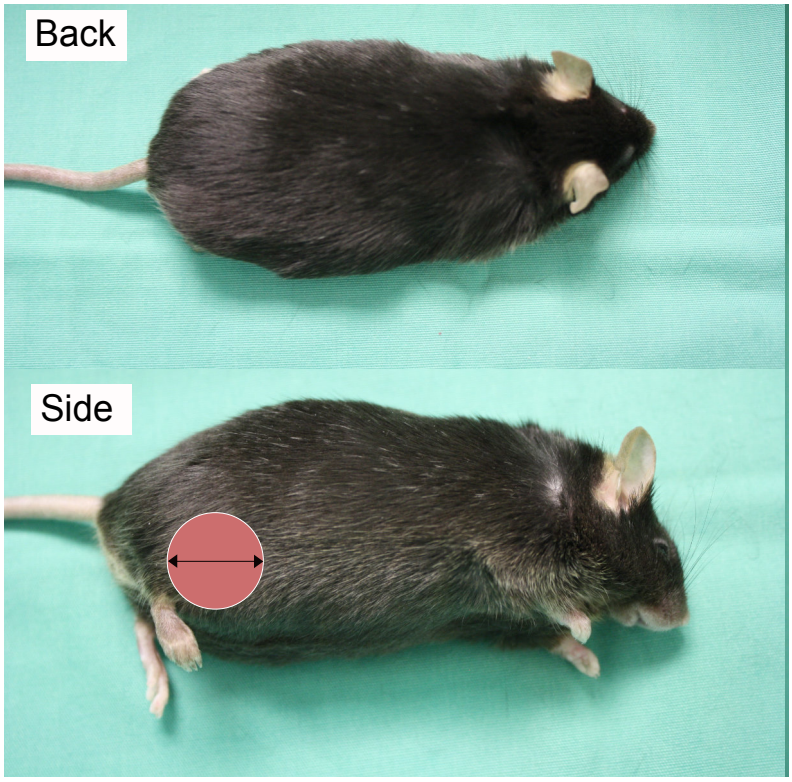


Figure 7.8. The areas on mice in which spectrophotometric readings of fur colour were taken. The lower half of the right lower limb was measured using a Minolta Spectrophotometer, with the mean of 3 readings taken and utilised for each measurement.

7.3.3.8 Photographic assessment

At death, mice were photographed to further analyse coat colour. Mice were positioned so that the back, side and abdomen could be photographed separately (Figure 7.8), and light was standardised and recorded by a light meter (Whitegoods, Light Meter App, USA), with photographs taken at both 682 and 982 lux using a Canon 1000D camera (Canon, Toyko, Japan). All photographs were taken using lens (Canon 18-55mm lens), with fixed manual settings; F5.6, ISO 400, 0.1 seconds), and at a standardised height of 1m above the bench (Figure 7.9).

Photographs were subsequently analysed with ImageJ, with the same areas of the body analysed as with the spectrophotometer (Figure 7.9).

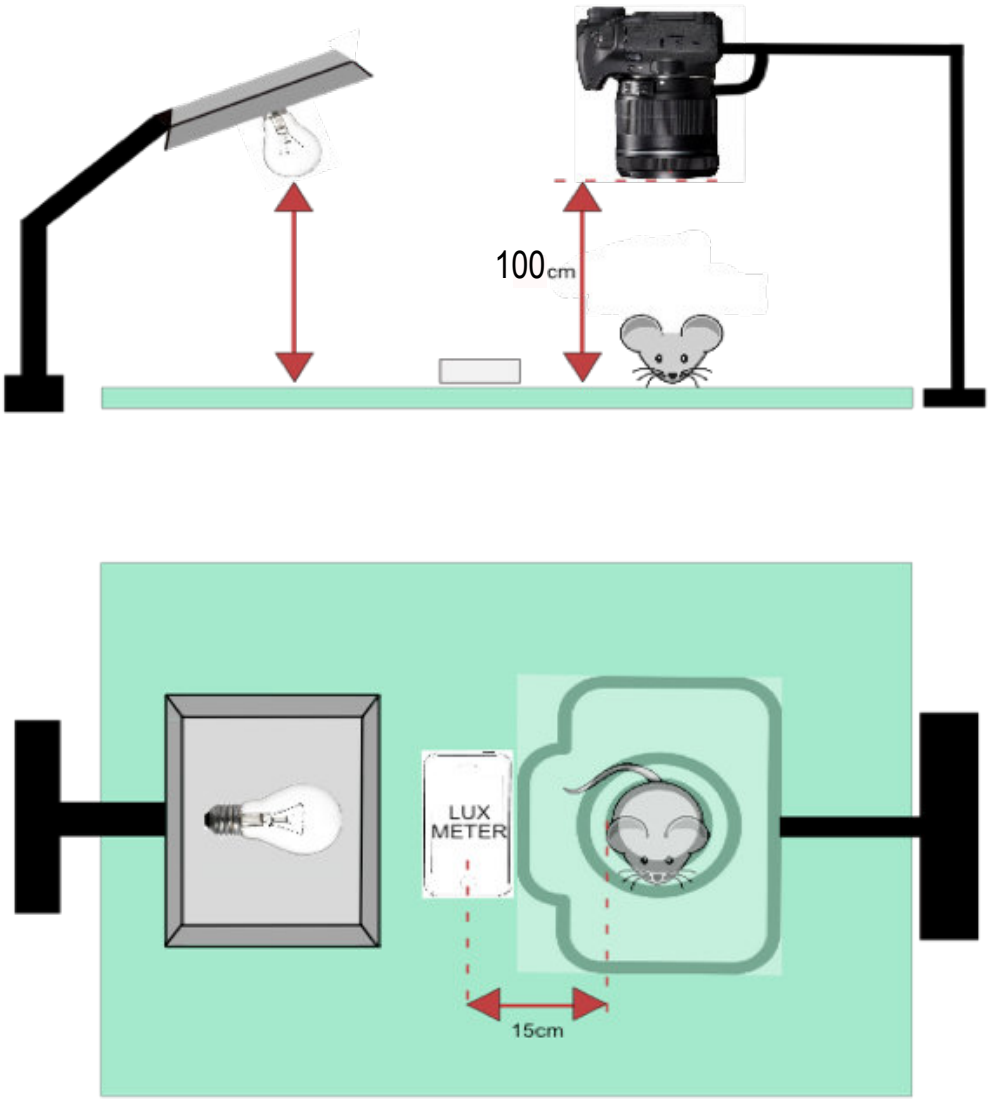


Figure 7.9. The photographic assessment of mice. Top – side view illustrating the consistent lighting source and position of the camera. Bottom – an aerial schematic of the relative positions of the light, lux meter and camera to the mice.

7.3.3.9 Culling

All mice were culled in accordance with home office guidelines (UK, Scientific Procedures Act, 1986), and under the study project licence, using isoflurane anaesthesia followed by

cervical dislocation. Mice were placed within a sealed Perspex box with 50% isoflurane air supply for ~ 2 minutes until the mice were rendered unconscious. Mice were then removed, and manual cervical dislocation was performed. Death was confirmed by a lack of spontaneous breathing or movement.

7.3.4 Molecular analysis

7.3.4.1 Post-mortem dissection

Following death, ~ 1ml of whole blood was extracted directly from a cardiac stab, and the heart, ovaries/testes, together with a sample of hindlimb striated muscle were all snap frozen in liquid nitrogen. The brain was also removed hemisected immediately, with the following structures from one hemisphere sub-dissected; cortex, cerebellum, striatum, hippocampus, substantia nigra, and remaining tissue, which were all also snap frozen in liquid nitrogen before freezing at -80°C.

7.3.4.2 Fixing tissue

The remaining cerebral hemisphere (whole) was placed into 0.4% buffered formalin for 48hrs before embedding in wax for histological examination. Brain tissue was subsequently sectioned coronally to enable sections of the hippocampus, substantia nigra and cortex to be identified.

7.3.4.3 DNA extraction

DNA from the cerebral cortex and hippocampus was extracted from frozen tissue using Qiagen's QIAamp DNA Mini Kit as per the manufacturers protocol (Qiagen, USA). Briefly, tissue was placed into a 1.5ml epindorf tube and 180µl of ATL buffer (Qiagen, USA) was added together with 20µl of proteinase K (Qiagen, USA), which were then mixed thoroughly by vortexing, and incubated using a heat block at 56°C for 3-6 hours until completely lysed.

Subsequently, 200µl of Buffer AL (Qiagen, USA) and 200ul of 100% ethanol were added and vortexed thoroughly. The solution was then pipetted into the DNeasy Mini Spin column n a 2ml collection tube and centrifuged at 8000rpm for one minute. The flowthrough was discarded and the column placed into a new collection tube and 500µl of buffer AW1 (Qiagen, USA) added before centrifugation at 8000rpm for one minute. This step was repeated, with the spin column again placed into a new collection tube, but with buffer AW2 (Qiagen, USA) added and centrifuged at 14000rpm for 3 minutes. Finally, 100µl of Buffer AE (Qiagen, USA) was added to the column, incubated at room temperature for 1 minute before centrifuging at 8000rpm for one minute.

7.3.4.4 DNA quantification

DNA quantity was determined using the Qubit DNA Standard Assay Kit (Thermo Fisher Scientific, USA) as per manufacturers instructions. Briefly, standard concentrations to generate a standard curve were performed by mixing Qubit dsDNA HS reagent 1:200 in Qubit dsDNA HS Buffer. Subsequently 10µl of Qubit Standard 1 and 2 were added to 190µl of the solution, left to mix for 5 minutes, and then read with a QuBit 3.0 Flurometer. Subsequently, 1µl of each DNA extraction was mixed with 199µl of the Qubit solution and the DNA concentration together with A260/280 ratio determined.

7.3.4.5 Long-range PCR

Whole mtDNA genome long-range PCR was performed in a 25µl total volume reaction; 1µl of extracted DNA was mixed with 5ul of Takara PrimeSTAR GXL buffer (Takara Bio Inc, Japan), 5µl of dNTPs, 0.5ul of forward and reverse primers (10pmol/µl) [F:CCCAGCTACTACCATCATTCAAGT, R:GAGAGTTTTATGGGTGTAATGCGG], 0.5µl of Takara PrimeSTAR (DNA polymerase) (Takara, Bio Inc, Japan) 8.5µl of H_2O. PCR protocol: 1x 94 °C for 60 sec, 30 cycles of [98 °C for 10 sec, 68 °C for 13 mins, 72 °C for

10min]. 10µl of PCR product loaded onto a 0.6% agarose gel; 65V for 40 minutes. The presence of deletions was determined by products shorter than 16.5kb in length (Figure 7.18).

7.3.4.6 Protein extraction

10-20mg of frozen tissue from each brain region (as indicated) were mechanically homogenized using a glass tissue homogenizer (Cole-Palmer, UK) in 100µl of 1 x Phosphate-buffered saline (PBS) (NaCl 137mmol/L, KCl 2.7 mmol/L, Na_2HPO_4 10 mmol/L, KH_2PO_4 1.8 mmol/L) at pH 7.4. To prevent excessive disruption of synuclein aggregates mechanical homogenization was performed gently for ~ 20 seconds. Subsequently the lysed tissue was centrifuged at 14000g for 15 minutes at 4°C, and the lysate pipetted out into a fresh epindorf and the pellet frozen.

Subsequently, protein concentration was determined using a Bradford Protein Assay. One part Bradford Dye Reagent (Bio-Rad Protein Assay Dye Reagent Concentrate, California USA) was diluted with 4 parts sterile water. Each protein lysate was diluted 1 in 10 and 1 in 30 with water, and then one microliter of each dilution was mixed with 199µl of the diluted dye reagent in triplicate. The results were compared against a standard curve that was generated by proportional dilutions of stock 1mg/µl of Bovine Serum Albumin (A6003, Sigma-Aldrich, USA) with sterile water to generate 0.05mg/µl – 1mg/µl of protein (Figure 7.10). Ten microliters of each of these standard solutions was mixed with 190µl of the diluted dye reagent and placed along with the experimental samples into a 96 well plate (Thermo Fisher Scientific) and left in the dark for 10 minutes. Subsequently the plate was placed in a UV spectrophotometer (Multiskan Ascent Plate Reader (MTX Lab Systems, FL, USA)), with the absorbance of each well determined at 595nm.

To determine protein concentrations from absorbance data, a standard curve as generated from the diluted BSA samples (Figure 7.10), and then this was compared against the result

from each sample from each original tissue sample. Subsequently, 100µg of total protein was normalised at 2-5µg/µl for each brain region, and then these samples utilized for further Western blotting and aggregation assays.

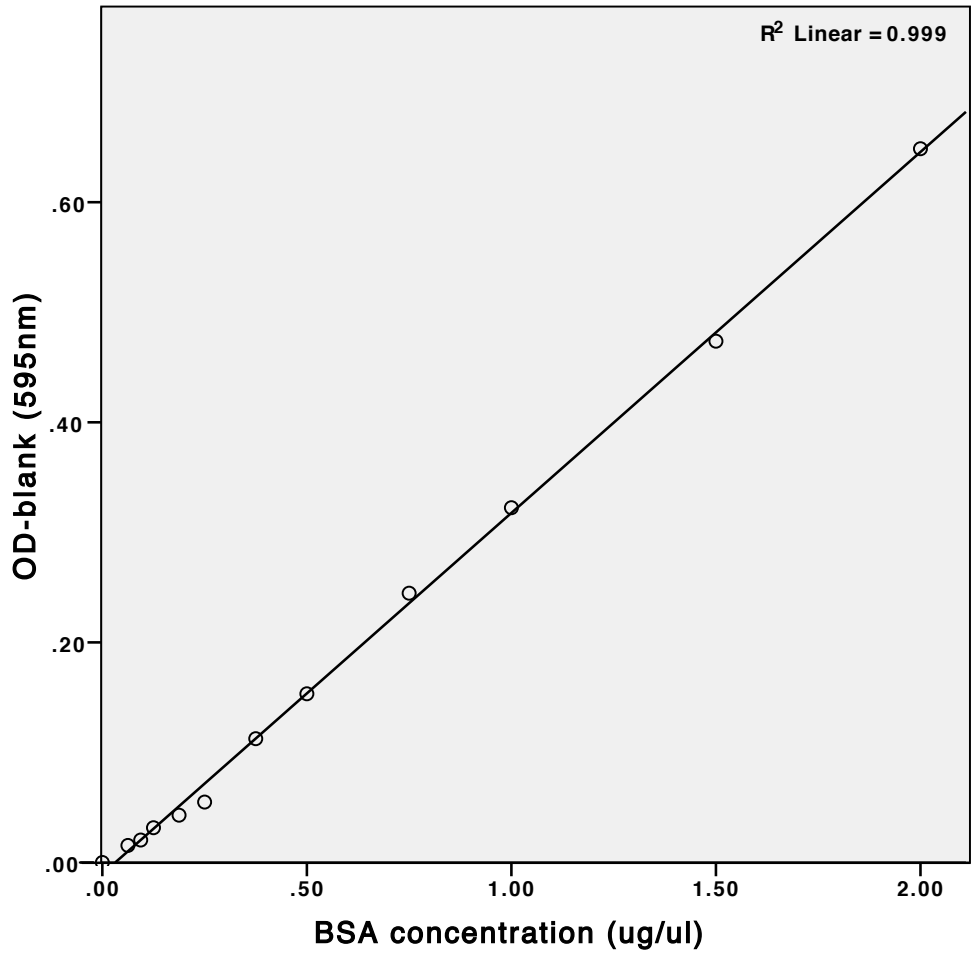


Figure 7.10. The Bradford reagent standard curves for Bovine Serum Albumin (BSA). Top – A regression trendline of normalised data was drawn and it's equation used to calculate protein concentration in tissue homogenates.

7.3.4.7 Western blotting of α-synuclein

On ice, 10mg of protein from the normalised lysate from each brain region in each individual were mixed with 2µl of 10x Sample Reducing Agent, and 5µl of 4x LDS Sample Buffer (both Novex®, Thermo Fisher Scientific), and the remainder of the 25µl volume made up with

sterile water. The samples were then boiled at 70°C for 5 minutes in a dry bath incubator. Samples were then vortexed for 10 seconds and centrifuged at 8000g for one minute.

Transfer to a polyvinylidene difluoride (PVDF) membrane was undertaken using the iBlot 2 Semi Dry Transfer System (Thermo Fisher Scientific). Each gel was removed from its case using a spatula, followed by soaking in Tris Buffer Saline with Tween 20 ® (TBST buffer, consisting of 20mM Tris hydrochloride pH 7 (Sigma-Aldrich), 0.5mM NaCl (VWR, Lutterworth, UK) and 0.1% Tween 20 (Sigma-Aldrich)) for 1 minute. IBlot 2 Transfer Stacks were used as instructed by the manufacturer. The supplied filter paper was soaked in dH_2O and the stack assembled as follows; stack-membrane-gel-filter paper-stack. The P0 protocol for mixed molecular weight was used for 7 minutes at a constant 1.3 A.

Initially following transfer, the membrane was directly soaked in 100ml of 4% paraformaldehyde (Sigma Aldrich, UK) with 0.01% glutaraldehyde (Sigma Aldrich, UK) in dH20 for 30 minutes to induce cross-linking and improve detection of endogenous α-synuclein(Lee and Kamitani 2011). The membrane was then washed with TBST buffer and then blocked with 5% (w/v) milk in TBST. Following this, the membrane was incubated with the relevant primary antibody at 4^oC under the conditions indicated in Table 7.1. The membrane was then washed three times for 5 minutes in TBST and then incubated with secondary antibody at a 1:2000 concentration for 2 hours at room temperature. After washing three times for 5 minutes in TBST, protein signal was detected by developing with Clarity Western ECL Western Blotting substrate (Bio-Rad) for 5 minutes in the dark. The membrane was then imaged using the Amersham Imager 600 (GE Healthcare). Densitometric analysis was performed using ImageQuant TL 8.1 software (GE Healthcare). Betactin, a constitutively expressed protein, was used to normalize the results and the protein of interest/Betactin ratio was calculated.

One antibody (mouse monoclonal α-synuclein antibody (Clone 42), BD Biosciences, UK) required the utilization of a fluorescent secondary antibody (Goat anti-mouse 800CW, Li-Cor) which was incubated in darkness for one hour at room temperature (1:10,000). Subsequently, this was analysed on the Odyssey CLx (Li-Cor), using Image Studio 4.0 (Li-Cor), and relative intensities of the protein of interest / Betactin ratio calculated.

7.3.4.8 α-synuclein aggregation assay

50µg of normalised protein lysate from each brain region was also utilized to look for the presence of protein aggregates. Each protein sample was diluted with 5mg/µl DNAse I (Sigma-Aldrich, UK), 1 MTRIS (tris(hydroxymethyl)aminomethane) (pH 8.0) and 100 mmol $MgCl_2$ to make a final concentration of the 50µg of extracted protein in 50µg/ml DNAse I, 50mmol TRIS, and 10mmol $MgCl_2$. This solution was then heated in a dry bath incubator at 37°C for one hour, before 10% SDS (Sodium dodecyl sulfate solution, Sigma-Aldrich) was added in a 1:1 volumetric ratio to the protein solution.

Each sample was then added to an individual well within a Minifold I manifold 96-well filtration/incubation unit. A 0.2micrometer cellular acetate membrane was placed below the wells together with filter paper and in sequential layers. Suction was applied using a vacuum pump until all of the lysate was drawn through.

The membrane was subsequently washed five times with TBST before blocking with 5% milk for one hour. Thereafter the membrane was incubated overnight with a mouse monoclonal α-synuclein antibody ((Clone 42), BD Biosciences, UK) at 1:500 concentration, before further washing 5 times with TBST for 5 minutes per wash. Thereafter, the membrane was incubated with a fluorescent secondary antibody (Goat anti-mouse 800CW, Li-Cor) which was incubated in darkness for one hour at room temperature (1:10,000). Subsequently, this was analysed on the Odyssey CLx (Li-Cor), using Image Studio 4.0 (Li-Cor).

Table 7.1. All antibodies used within the study. Variant All primary and relevant secondary antibodies are shown together with their manufacturer details and concentrations.

Primary antibody	Epitope	Manufacturer	Concentration	Secondary antibody	Manufacturer	Concentration
Clone 42 / α-synuclein IgG (610787)	All forms of synuclein	DB Biosciences, UK	1 in 2000	IRDye 800CW Goat anti-Mouse IgG (H+L)	LI-COR	1 in 10000
LB509 Synuclein IgG (ab27766)	Human synuclein	Abcam, UK	1 in 1000	P026002-2 Rabbit Anti-mouse HRP	Agilent	1 in 2000
Actin (a1978)	Actin	Sigma, UK	1 in 1000	IRDye 800CW Goat anti-Mouse IgG (H+L)	LI-COR	1 in 10000
				P026002-2 Rabbit Anti-mouse	Agilent	1 in 2000

7.3.4.9 MRI image acquisition

MRI images were taken immediately post mortem. Scans were collected on a 7T preclinical MRI system (Agilent, DirectDrive system). Animals were positioned in a cradle, and the head placed in a transmit-receive birdcage head coil (inner diameter 39mm, Rapid Biomedical GmbH) which was used for all imaging studies. Following scout image collection to confirm animal positioning in the scanner, magnet shimming to maximise field homogeneity over the brain and power calibration steps, the main series of axial and coronal T2 weighted imaging datasets were acquired. All scans were collected with the same scanning parameters (apart from slice orientation) as described below:

- TR/TE=4000/46ms,
- Fast spin echo (FSE) sequence, with 8 echoes and 11.5ms inter-echo spacing,
- Field of view 25.6x25.6mm,
- Matrix size 256x256 delivering a nominal in-plane resolution of 0.1mm,
- Slice thickness=1.0 mm, with 20 contiguous slices,
- Number of averages = 6
- Sweep width of 60kHz,
- Total scan time 12mins 56s per orientation.

7.3.4.10 MRI image analysis

MRI image analysis: Voxel Based Morphometric analysis was performed by Dr Steve Sewiak at the University of Cambridge using his own custom scripts. Comparisons were made between all combinations of all three genotypes imaged ($POLG^{wt/wt}$, $POLG^{wt/wt}$ +

heteroplasmy, and POLG$^{wt/mt}$+heteroplasmy) at the Familywise error correction ($p<0.05$) and at the uncorrected thresholds as stated.

7.4 Results

7.4.1 Phenotypic data

7.4.1.1 Morphological outcomes

In total, 58 mice were culled at 14 months of age: (1) POLG$^{wt/mt}$ SNCA$^{wt/mt}$ + inherited heteroplasmy: n=12 (M:6, F:6), (2) POLG$^{wt/wt}$ SNCA$^{wt/mt}$ + inherited heteroplasmy n=12 (M:7, F:6), POLG$^{wt/mt}$ SNCA$^{wt/mt}$ n=9 (M:7, F:2), POLG$^{wt/wt}$ SNCA$^{wt/mt}$ n=9 (M:4, F:5), Wild-type mice; n=11, (m=2, F=9). Given that all cohorts other than WT mice contained the SNCA mutant allele, the groups are subsequently referred to by their POLG genotype and susceptibility to having inherited heteroplasmy (Figure 7.2a, 7.11a).

We observed no premature deaths in any cohort, with all mice living to 14 months of age at which point they were culled according to home office guidelines. There was no difference in weight between any genotype of mice at any time point (Figure 7.11b), nor any difference in the mean rotarod speed or time between any genotype at any age (Fig 7.11c and d). There were no gross morphological abnormalities with any individual mice, and the only unexpected phenotypic feature was that of fur colour changes in isolated mice ranging from gross sections of white fur (n=2 mice) (Figure 7.12), to grey speckling of fur beyond that which would be expected in WT mice of an equivalent age (personal communication, animal husbandry technicians) (Figure 7.12). We also observed some evidence of over grooming of some animals (n=4), within 3 different genotypes (not shown).

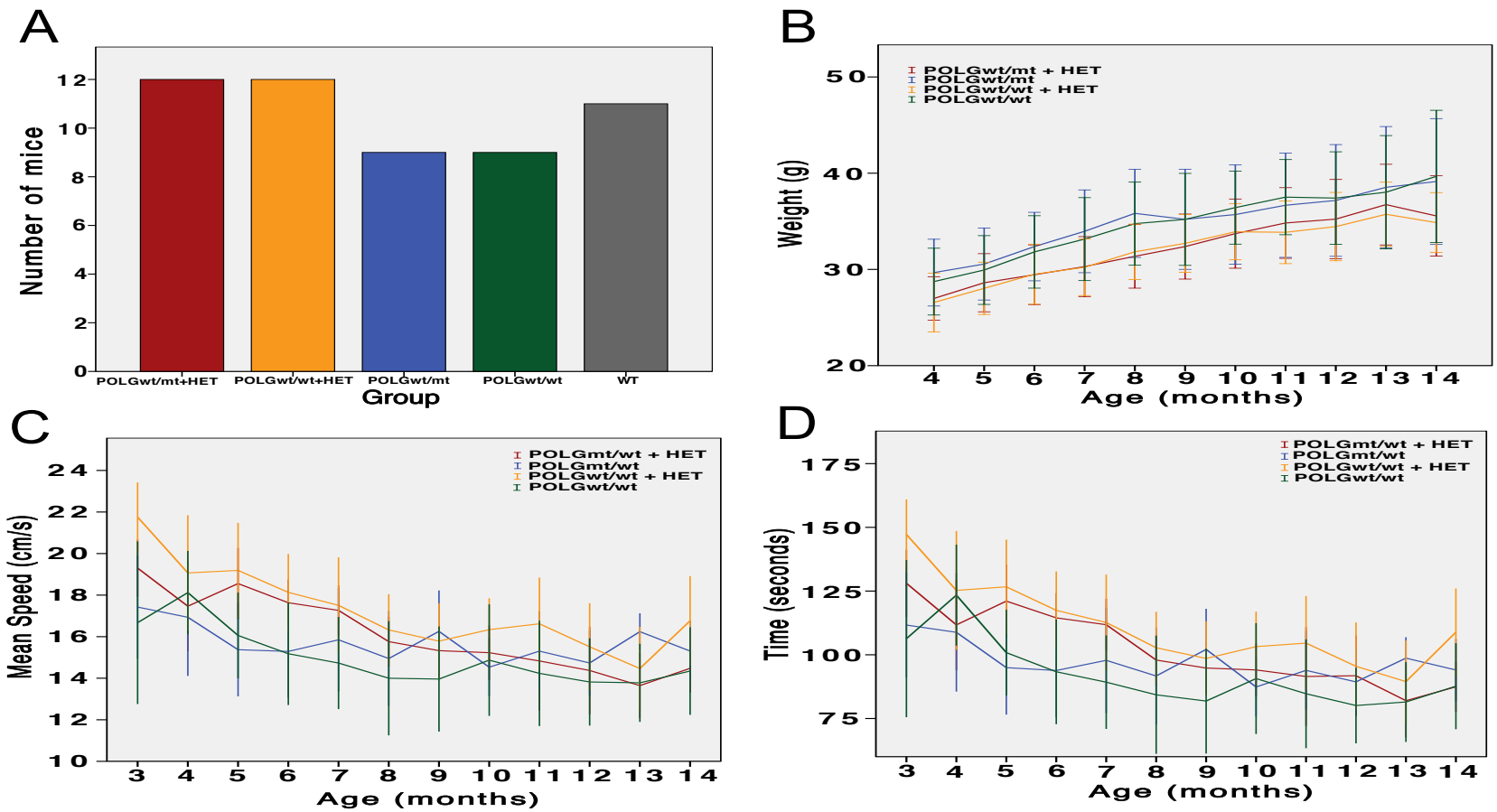


Figure 7.11. Breeding outcomes and the results of the rotarod assessments. IA: The number of mice from each genotype is shown. B: The mean and standard error of the mean (SEM) for the weight of each genotype of mice is shown. C: The mean speed and SEM from each genotype on the rotarod is shown. D: The mean time and SEM from each genotype are shown.

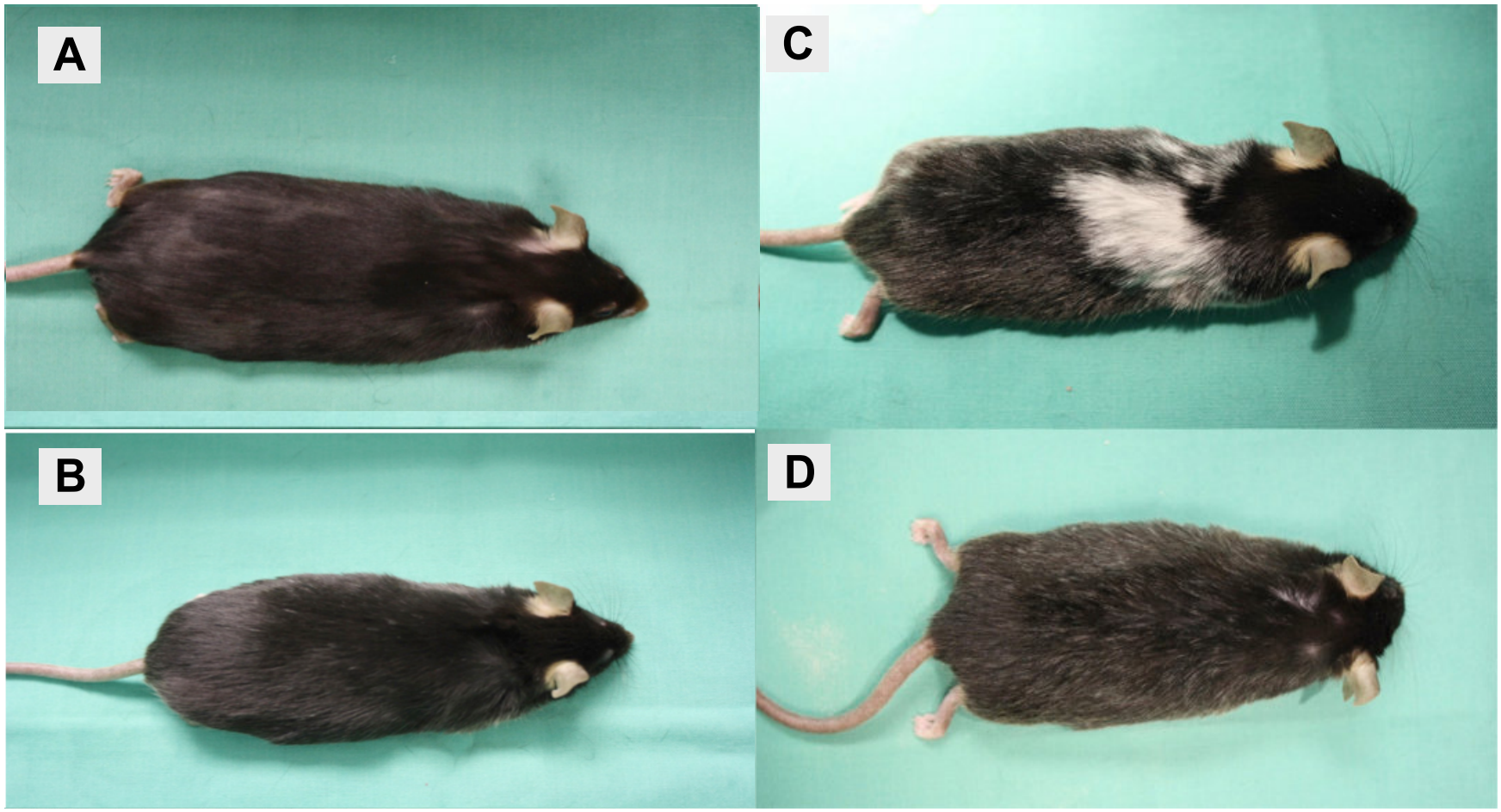


Figure 7.12. Fur colour of different genotypes of mice. Four male mice are shown. A: a WT C57bl6 male mouse with no mutant POLG or inherited heteroplasmy and no mutant SNCA allele (POLGwt/wt, SNCAwt/wt). B: Mutant SNCA allele only (POLGwt/wt/ SNCAwt/mt), C: Mutant POLG and mutant SNCA with no inherited heteroplasmy (POLGwt/mt, SNCAwt/mt). D: Mutant POLG and inherited heteroplasmy plus mutant SNCA (POLGwt/mt, SNCAwt/mt + heteroplasmy).

7.4.1.2 MouseTrapp data

All POLG genotypes carrying the SNCA transgene showed a reduction in total distance travelled, total number of touches and stride length compared to controls at 14 months of age (Total $P<0.05$, ANOVA and post-hoc testing) (Figure 7.13). However, we observed no difference between any POLG genotype in any of those parameters at any age (Figure 7.13).

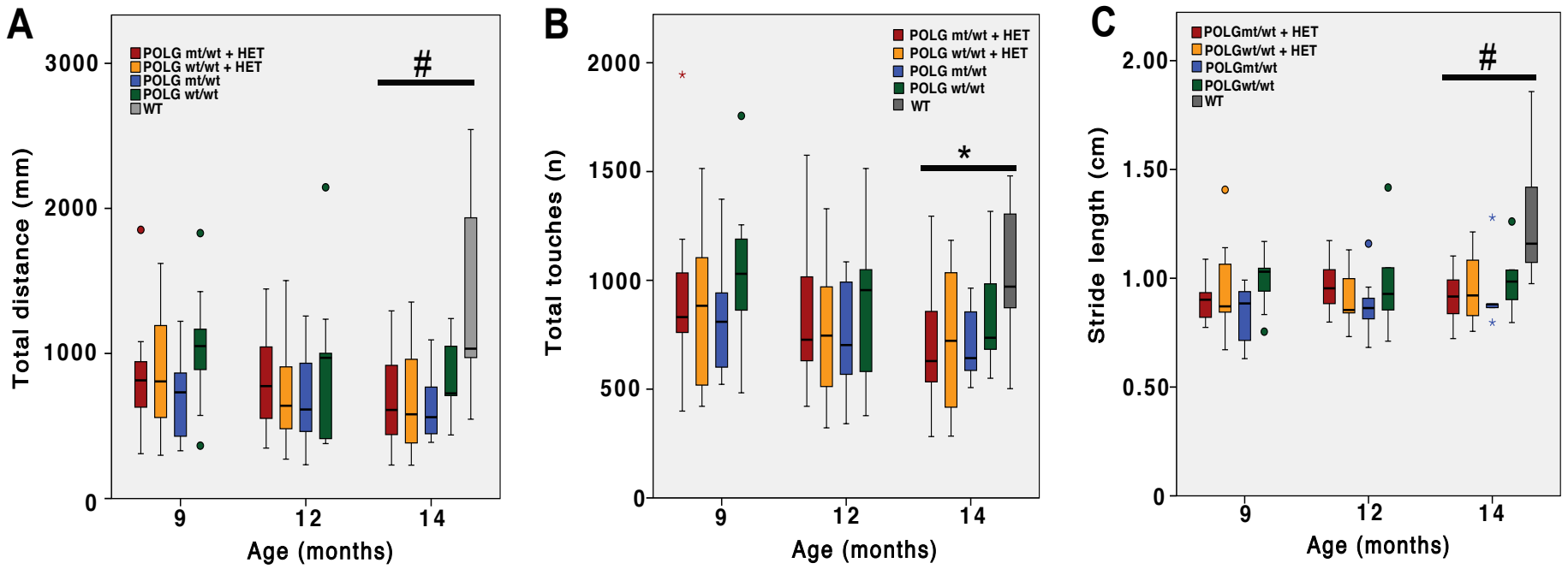


Figure 7.13. Results of the MouseTrapp experiments. Box-whisker plots show the median and inter quartile range (IQR) of teach genotype of mice at 9, 12 and 14 months of age for A: Total distance travelled (left panel), B: The total number of touches on the glass in 5 minutes, and C: the stride length (cm). Wild-type mice were only tested at 14 months of age. Significant differences between groups in a one-way ANOVA are shown at 14 months of age for all measures (#: $p<0.01$, *: $p<0.05$), with post-hoc testing results as shown in the text.

7.4.1.3 Open Field Testing and Novel Object Recognition

Open field testing showed significant differences between all four POLG genotypes and controls in both the total distance moved (p=0.013, ANOVA) and in central field times (p=0.002, ANOVA) at 14 months of age (Figure 7.14), with post-hoc testing revealing that both $POLG^{wt/mt}$ and $POLG^{wt/wt}$ with a susceptibility to have inherited heteroplasmy had significantly reduced total movement within the open field compared to controls (p=0.001, and p=0.024, post-hoc LSD). In addition, all SNCA mutant mice with any corresponsing POLG genotype or inherited heteroplasmy ($POLG^{wt/mt}$ + het, $POLG^{wt/wt}$ + het, and $POLG^{wt/mt}$) showed significantly reduced central field times compared to both controls ($p < 0.05$, post-hoc testing), and the $POLG^{wt/wt}$ genotype which had no inherited heteroplasmy ($p<0.05$, post-hoc testing) (Figure 7.14). Novel Object Recognition Testing (NORT) however showed no difference between any POLG genotype and controls (p=0.172, ANOVA).

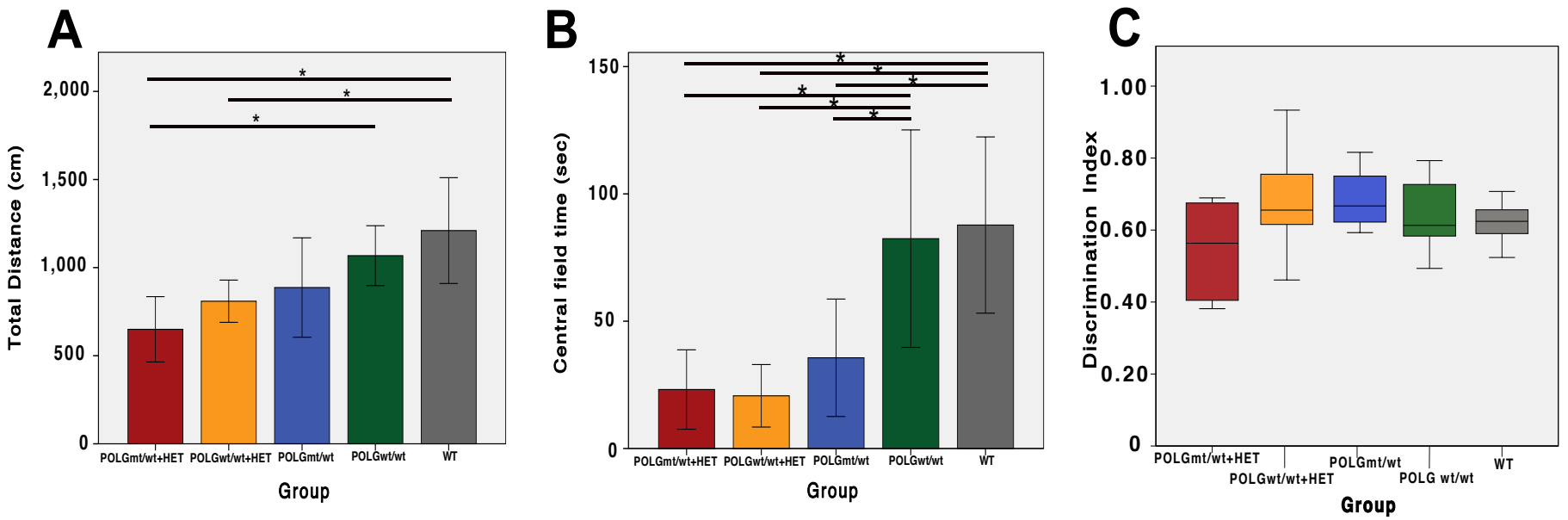


Figure 7.14. Results of open field testing (OFT). Panel A shows the mean and standard error of the mean (SEM) for the total distance travelled across the entire field for each genotype. Significant differences revealed with post-hoc least squared difference (LSD) testing after one-way ANOVA are shown (*; p=<0.05). B. The mean and standard error of the mean (SEM) for the total time in the centre of the open field are shown for each genotype. Significant differences revealed with post-hoc least squared difference (LSD) testing after one-way ANOVA are shown (*; p=<0.05). C: A box-whisker plot of the discrimination index of each genotype of mice in the novel object discrimination test. No significant differences were observed between groups.

7.4.1.4 Fur signal intensity

To try and determine any additional objective phenotypic features indicative of aging, we performed greyscale analysis of the fur colour of each genotype of mouse. This showed no difference in the mean signal intensity of fur on either the back or side of any POLG genotype compared to controls at both 629 and 982 lux (Figure 7.15 and 7.16). Some differences in the minimum and maximum signal intensities were observed between differing POLG genotypes at 982 lux, though any genotype rarely differed compared to controls (Figure 7.15). The same changes in minimum and maximum signal intensity were not reflected in the data at 629 lux.

We also did not observe any differences using the spectrophotometer in the colour of fur between groups (Figure 7.17).

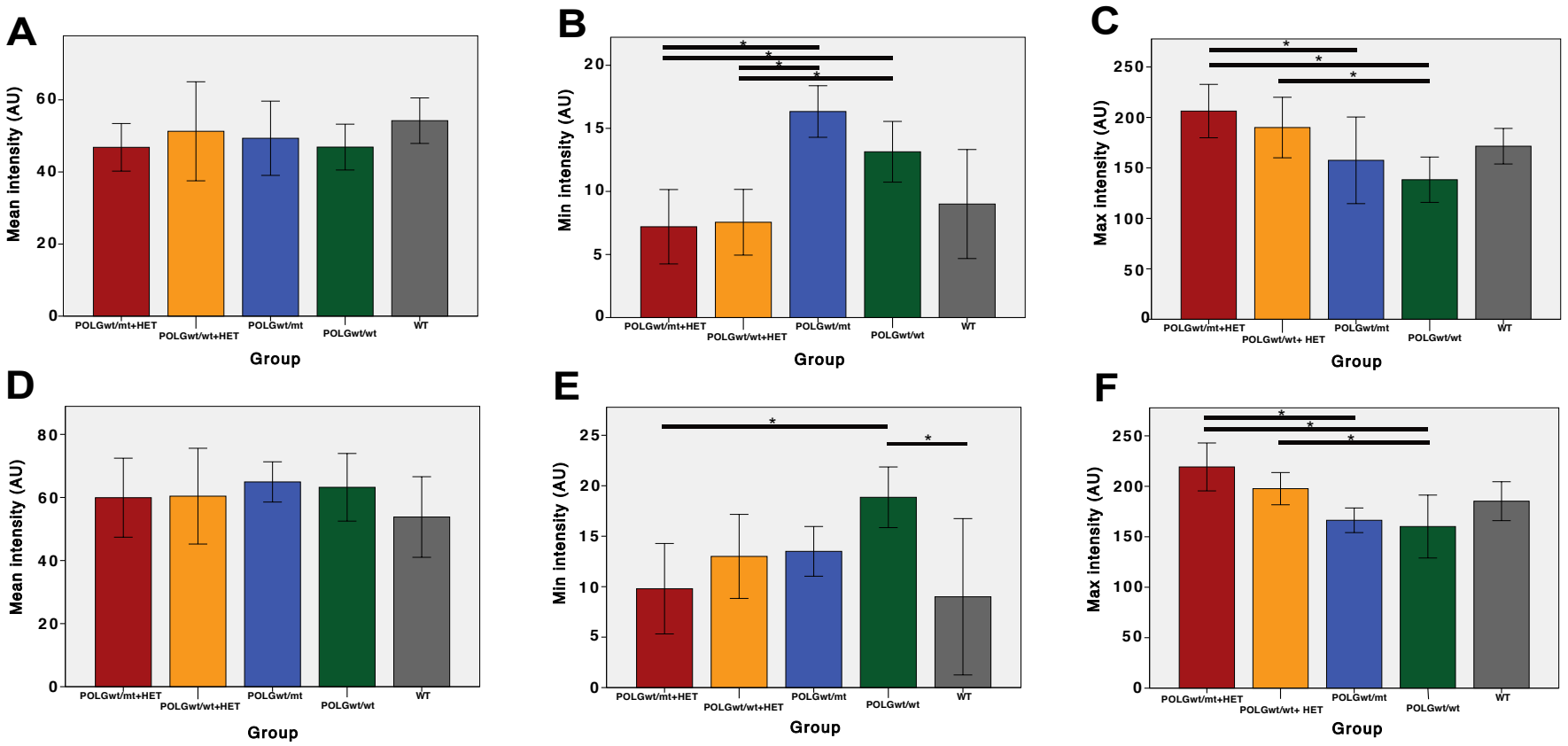


Figure 7.15. Results of fur colour image intensity at 982 lux. The mean, minimum and maximum signal intensity at 982 lux for all mice at death (14 months of age) are shown together with the standard error of the mean (SEM). A: shows the mean signal intensity from aerial photos, B: the minimum signal intensity from aerial photos and C the maximum intensity from aerial photos. D shows the mean intensity from side photos, E the minimum intensity of side photos and F the maximum intensity of side photos. Significant differences with post-hoc least squared difference (LSD) testing are shown when significant differences ($P<0.05$) were observed between groups with a one-way ANOVA. (*=$p<0.05$).

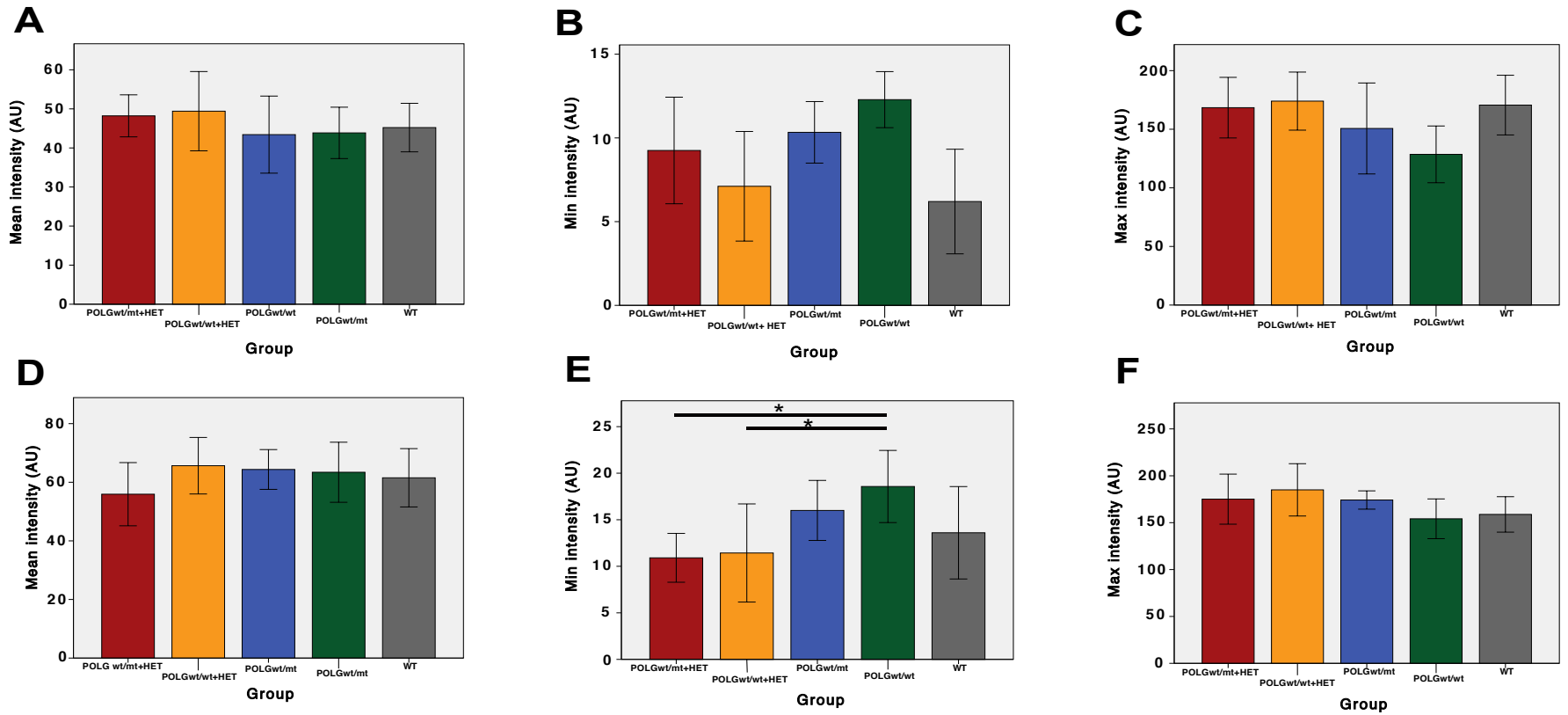


Figure 7.16. Results of fur colour image intensity at 629 lux. Panel A shows the mean and standard error of the mean (SEM) for the total distance travelled across the entire field for each genotype. Significant differences revealed with post-hoc least squared difference (LSD) testing after one-way ANOVA are shown (*; p=<0.05). B. The mean and standard error of the mean (SEM) for the total time in the centre of the open field are shown for each genotype. Significant differences revealed with post-hoc least squared difference (LSD) testing after one-way ANOVA are shown (*; p=<0.05). C: A box-whisker plot of the discrimination index of each genotype of mice in the novel object discrimination test. No significant differences were observed between groups.

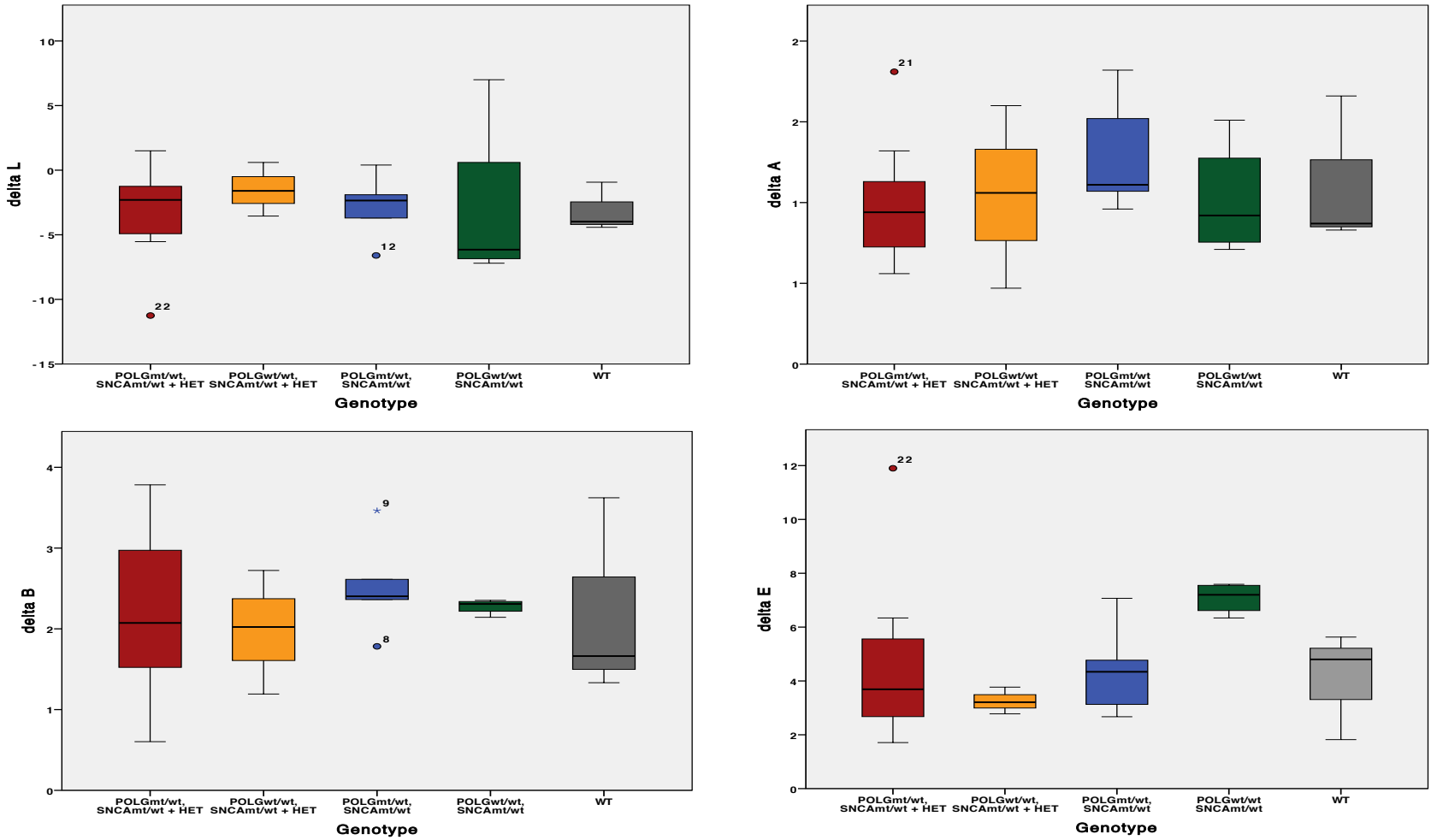


Figure 7.17. Results of spectrophotometric analysis. Each panel shows a box-plot of the values for each parameter of colour assessment (L, A, B and E) for each genotype. A one-way ANOVA for each of these 4 assessments was conducted etween all 5 genotypes and revealed no difference between them in each instance.

7.4.1.5 MRI Imaging

Three genotypes of mice ($POLG^{wt/mt}$ + heteroplasmy, $POLG^{wt/wt}$ + heteroplasmy, and $POLG^{wt/wt}$) were analysed using Voxel Based Morphometry (VBM) of segmented grey matter and white matter images. Using unpaired two-tailed t-tests and controlling for global differences in voxel intensity by including the overall mean of voxel intensity as a confounding covariate in the design matrix, together with gender and total brain volume, we observed significant differences in the $POLG^{wt/wt}$ cohort compared to the other two cohorts at the uncorrected level of $P < 0.001$. Increased T2 signal intensity was observed within both the primary motor cortex (Figure 7.17), and within both the anterior aspects of the dentate gyrus together with the subiculum and dorsal hippocampal commissures (Figure 7.17). There were no differences between the two genotypes with inherited heteroplasmy ($POLG^{wt/mt}$ + het and $POLG^{wt/wt}$ + het). The analysis of VBM images was performed by Dr Steve Sewiak (University of Cambridge).

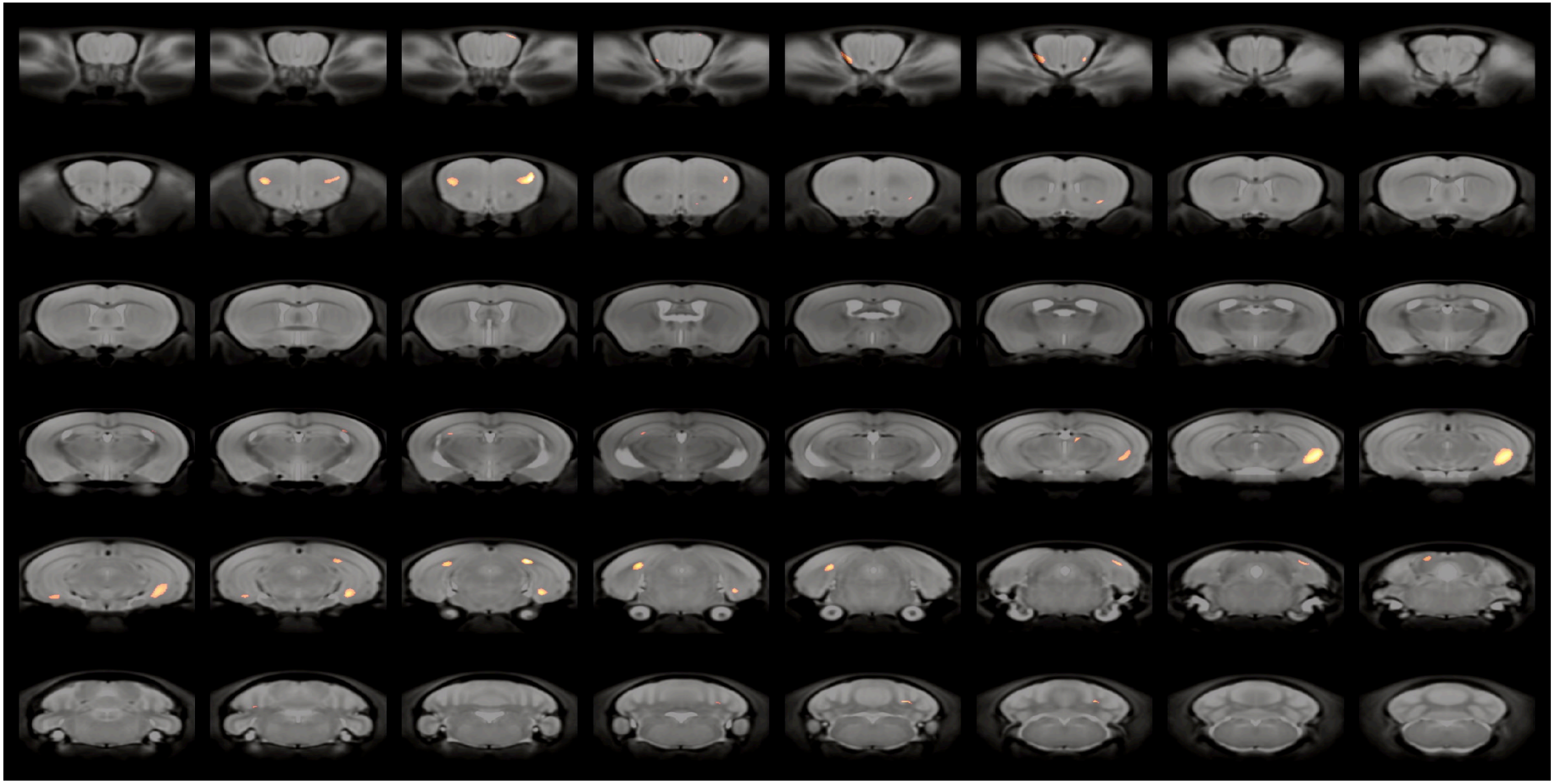


Figure 7.18. Results of voxel based morphometry (VBM). Panel A Three genotypes of mice were imaged: (1) POLGwt/wt, SNCAwt/mt, (2) POLGwt/mt, SNCAwt/mt, and (3) POLGwt/mt+heteroplasmy, SNCAwt/mt. Significant differences between the cohorts are shown using an uncorrected p-value of <0.001.

7.4.2 Molecular and histopathological data

7.4.2.1 Long-range PCR

We observed no evidence of mtDNA deletions within either the cortex or hippocampus in any POLG genotype (n=4 mice per cohort) or in control mice (n=6), with full-length 16.5kb PCR products seen in all samples (Figure 7.19).

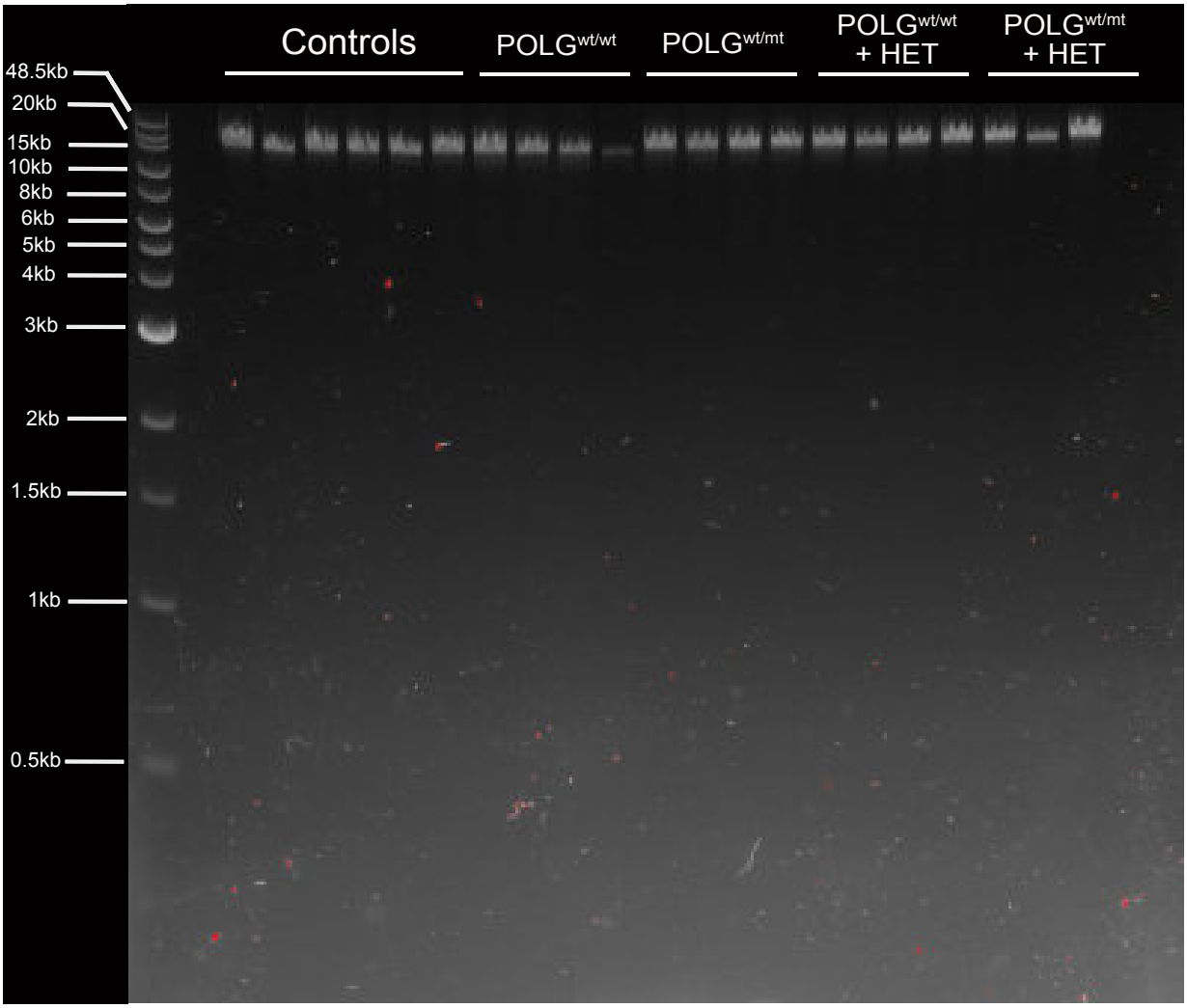


Figure 7.19. Results of long-range PCR of mitochondrial DNA. The five different genotypes of mice are shown, with six control samples and four from each other genotype. No bands smaller than 16.5kb were observed. Note that no positive control is shown due to the absence of available positive control samples in the lab from murine samples at the time of the experiment. Subsequent experiments by others in the group since the above experiment was performed have shown that the primers and protocol can detect multiple mtDNA deletions.

7.4.2.2 Western blotting

Across the entire cohort (22 animals; n=4 in each POLG genotype and n=6 controls), total α-synuclein expression was significantly higher in the hippocampus than the cortex (mean cortex: 0.40 (SD=0.14), mean hippocampus: 0.97 (SD=0.17) ($p=6.2x10^{-11}$, paired t-test)). These experiments were performed singularly for each of the 22 samples across those genotypes.

There was no difference in total α-synuclein expression levels between any POLG genotype and wild-type controls in the cortex ($POLG^{wt/mt}$ + heteroplasmy; mean = 0.54, SD = 0.22, $POLG^{wt/wt}$ + heteroplasmy; mean = 0.41, SD=0.032, $POLG^{wt/mt}$; mean = 0.36, SD = 0.057, $POLG^{wt/wt}$; mean = 0.41, SD = 0.10, WT; mean = 0.34, SD = 0.14) (p=0.236, ANOVA) (Figure 7.20).

In the hippocampus there was a strong trend towards differing levels of total α-synuclein expression between all genotypes and controls, though this just failed to reach statistical significance (p=0.052, ANOVA) ($POLG^{wt/mt}$ + heteroplasmy; mean = 0.89, SD = 0.35, $POLG^{wt/wt}$ + heteroplasmy; mean = 1.00, SD=0.282, $POLG^{wt/mt}$; mean = 1.08, SD = 0.11, $POLG^{wt/wt}$; mean = 1.16, SD = 0.055, WT; mean = 0.83, SD = 0.13) (Figure 7.20).

Human α-synuclein was only expressed within the mice carrying the PDGFB-SNCA mutation with no expression observed in control mice as expected (Figure 7.19), and within mice expressing human α-synuclein gene, the same trend was observed as with total synuclein, with significantly increased levels of expression within the hippocampus compared to the cortex (cortex: mean = 0.46 (SD=0.12), hippocampus, mean=0.75 (SD=0.18) $p=4.57x10^{-7}$, paired t-test).

Again there was no difference in the level of expression between all four POLG genotypes that possessed the α-synuclein mutations in either the cortex (p=0.11, ANOVA), or hippocampus (p=0.306, ANOVA) (Figure 7.20).

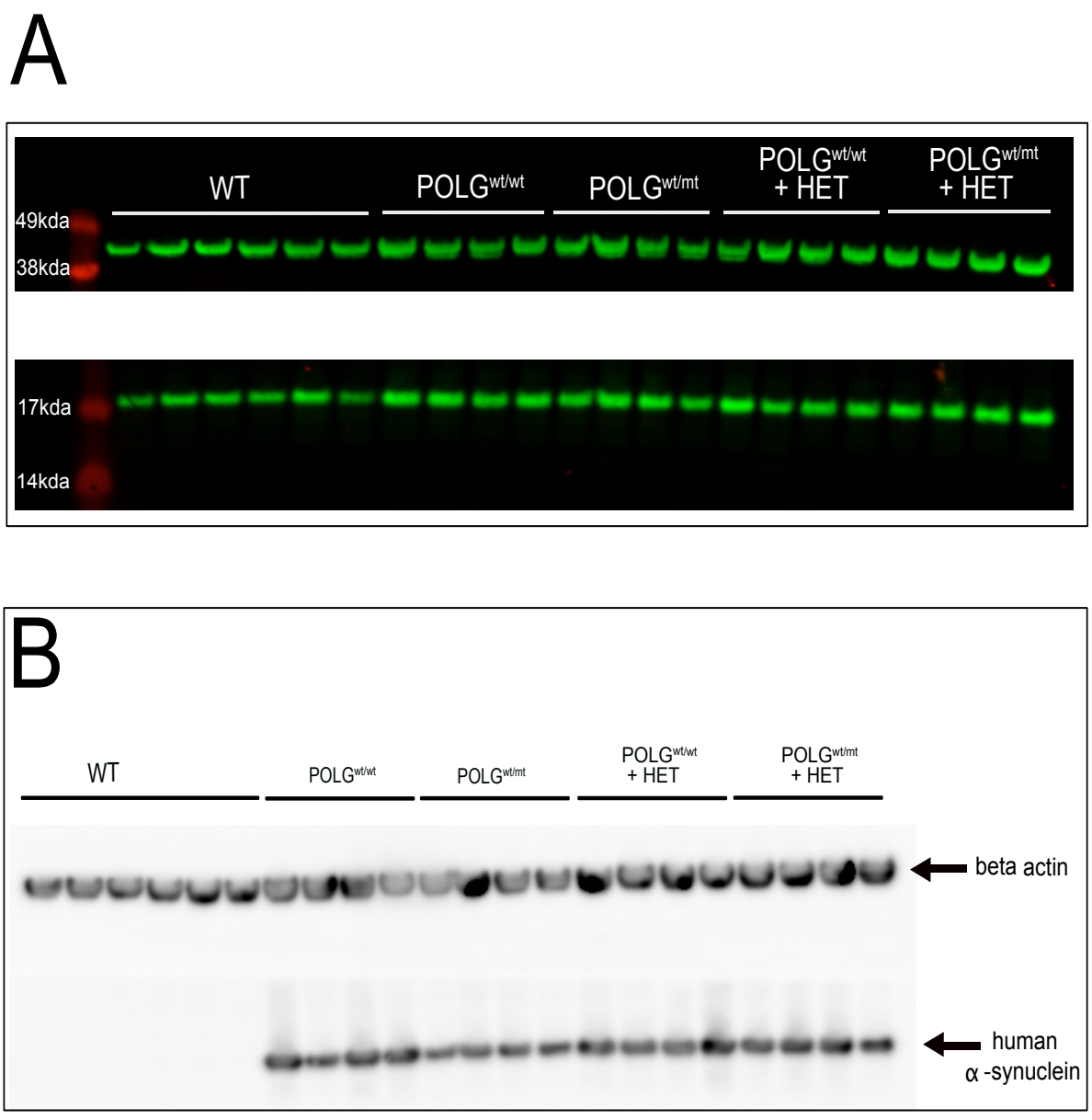


Figure 7.20. Western blots of both total a-synuclein (A) and human α -synuclein (B). The five different genotypes of mice are shown, with six control samples and four from each other genotype. A: The top band shows the beta actin control (42Kda), and total α-synuclein (17Kda). B: The top band again shows the loading control of beta actin, and the bottom band human a-synuclein. No expression of human a-synuclein could be seen in WT control mice (as expected)

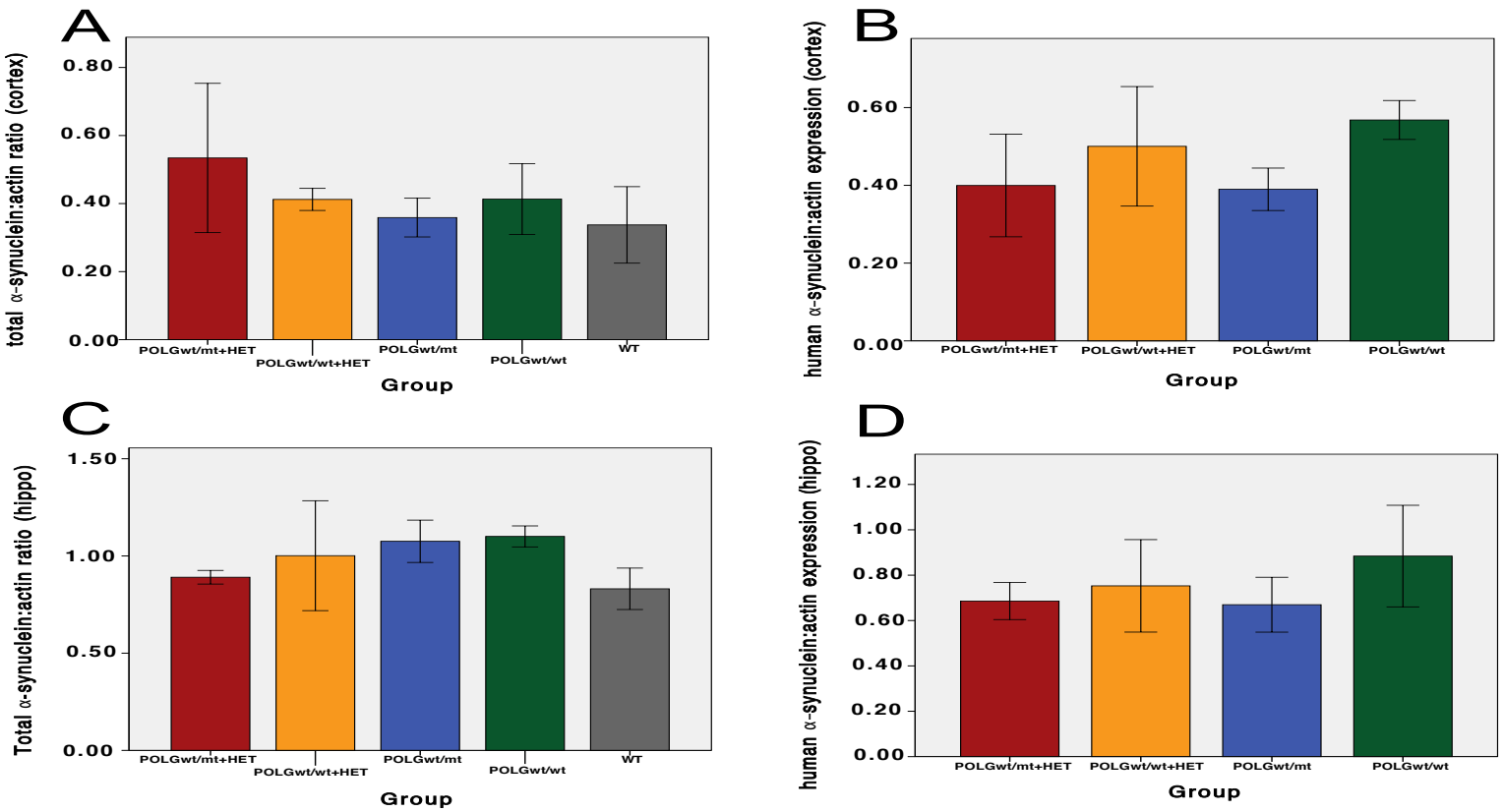


Figure 7.21. Results of α-synuclein expression between genotypes. All panels show the mean and standard error of the mean (SEM) of either human or total α-synuclein expression (as indicated) normalised against beta-actin expression. A: shows total α-synuclein expression in the cortex, B: shows human α-synuclein expression in the cortex (note that WT mice showed no expression and are thus not represented). C: shows total α -synuclein expression in the hippocampus, D: shows human α-synuclein expression in the hippocampus. No significant differences were observed. All experiments were single assessments of individual samples.

We subsequently grouped mice into those that have a predisposition to generate somatic mutations (heterozygous POLG mice ($POLG^{wt/mt}$) (n=8)) and mice that do not have a predisposition to generate somatic mtDNA mutations (homozygous wild-type POLG mice ($POLG^{wt/wt}$) (n=8), and compared these two groups against controls. This revealed a strong trend towards a significant difference in total α-synuclein expression within the hippocampus (p=0.051, ANOVA), with the highest levels observed in mice that do not generate somatic mutations ($POLG^{wt/wt;}$ mean = 1.40 (SD= 0.20), $POLG^{wt/mt}$ mice; mean = 1.24 (SD= 0.12), WT mice; mean = 0.83 (SD= 0.13)), though there was no trend towards significance within the cortex (p=0.339, ANOVA).

When comparing only human α-synuclein expression between those that have a predisposition to generate somatic mutations and those that do not, we observed no difference in expression between cohorts within the hippocampus (p=0.12, unpaired t-test), but did see increased expression in the cortex of mice that do not generate somatic mitochondrial DNA mutations (p=0.018, un-paired t-test) ($POLG^{wt/wt;}$ mean = 0.53 (SD= 0.11), $POLG^{wt/mt}$ mice; mean = 0.39 (SD= 0.09). This difference in expression may have been mediated by a shift towards lower expression of human α-synuclein compared to native mouse synuclein in the cortex of mice generating somatic mutations (p=0.086, unpaired t-test) (Figure 7.21).

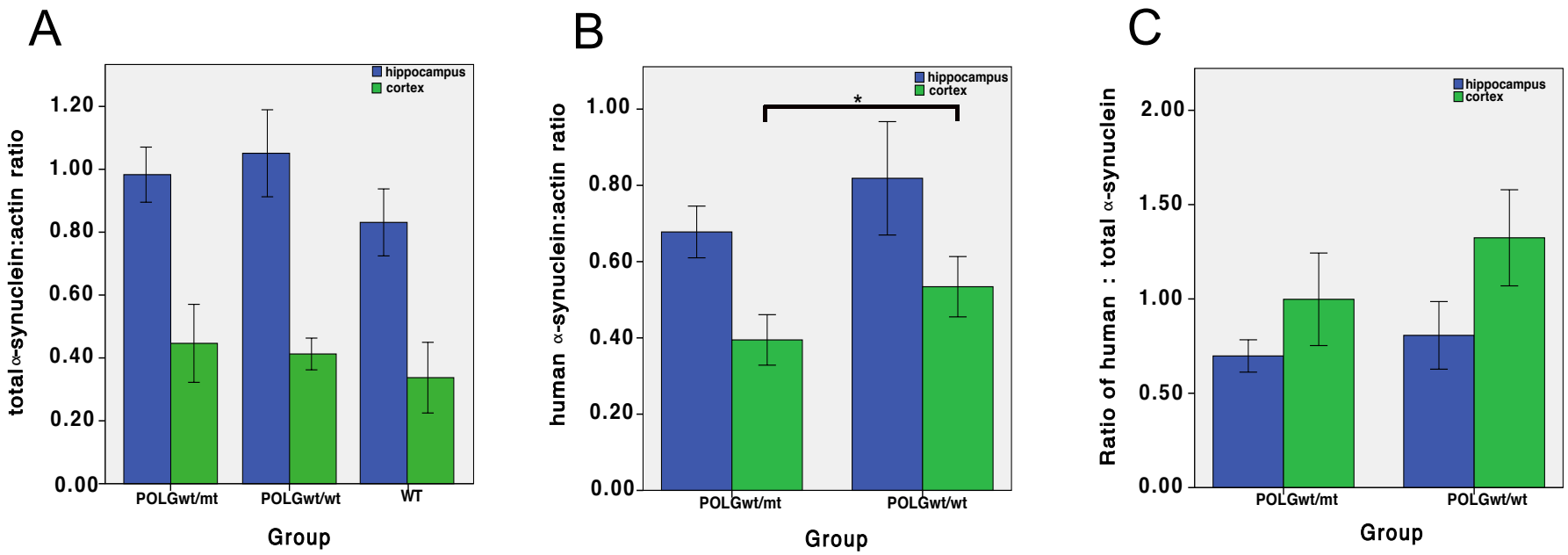


Figure 7.22. Results of α-synuclein expression between mice that generate heteroplasmy and those that do not. All panels show the mean and standard error of the mean (SEM) of either human or total α-synuclein expression (as indicated) normalised against beta-actin expression. A: shows total α-synuclein expression in the hippocampus (blue) and cortex (green), B: shows human α-synuclein expression in the cortex and hippocampus. C: shows the ratio of total α-synuclein expression to human α-synuclein expression in the hippocampus and cortex. Significant differences (unpaired t-test, $p<0.05$) are indicated by a *).

We also performed bivariate linear regression analysis to look for correlates between both human and total α-synuclein expression in the cortex and hippocampus with phenotypic metadata (OFT total distance, OFT central time, Novel Object Recognition Time and rotarod speed at 14 months of age). The only correlation that survived multiple testing (corrected p value = 0.0125) was the correlation between novel object exploration time and human α-synuclein expression in the hippocampus in mice that did not have any mitochondrial DNA heteroplasmy ($POLG^{wt/wt}$ ($p=0.007$, Pearson's product moment correlation coefficient, $R2 = 0.986$) (Figure 7.22 and 7.23).

7.4.2.3 Aggregation assay

Given that clinical synucleinopathies occur in the presence of synuclein aggregates (Spillantini, Schmidt et al. 1997), we aimed to determine whether we could detect evidence of any synuclein aggregates in any genotype of SNCA mutant mouse or in controls. Other than the positive control, we observed no evidence of α-synuclein aggregation within the cortex of any animal (Figure 7.24).

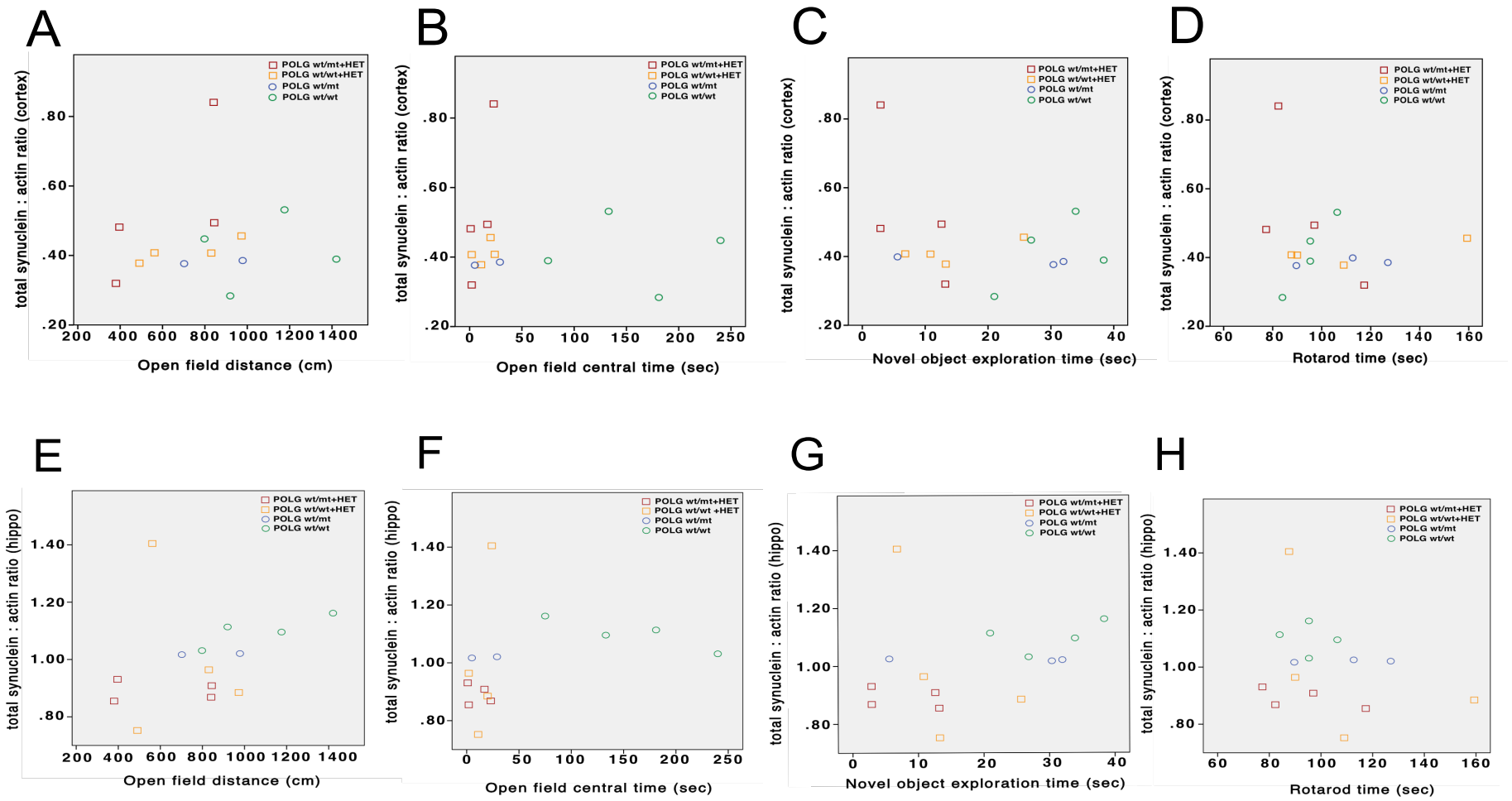


Figure 7.23. The correlation between total α-synuclein expression and clinical phenotypic measures of individual mice. All panels show the relevant phenotypic measure for individual mice on the x-axis and the expression of total expression on the Y-axis. The genotype of each mouse is indicated by a coloured symbol. Linear regression modelling was performed for each genotype for each measure and significant results (Pearson's product moment p-value < 0.05 at the corrected threshold) shown where appropriate.

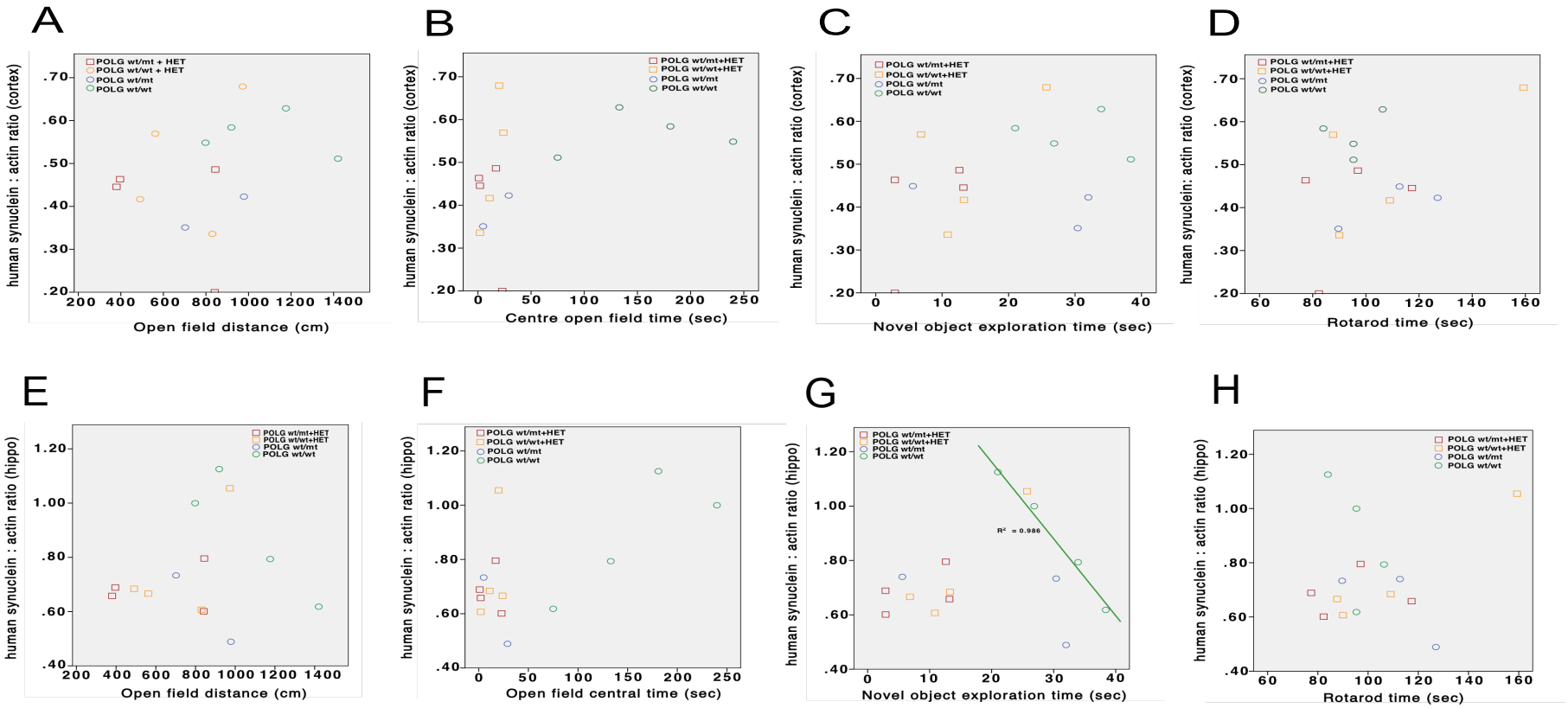


Figure 7.24. The correlation between human α-synuclein expression and clinical phenotypic measures of individual mice. All panels show the relevant phenotypic measure for individual mice on the x-axis and the expression of total expression on the Y-axis. The genotype of each mouse is indicated by a coloured symbol. Linear regression modelling was performed for each genotype for each measure and significant results (Pearson's product moment p-value < 0.05 at the corrected threshold) shown where appropriate.

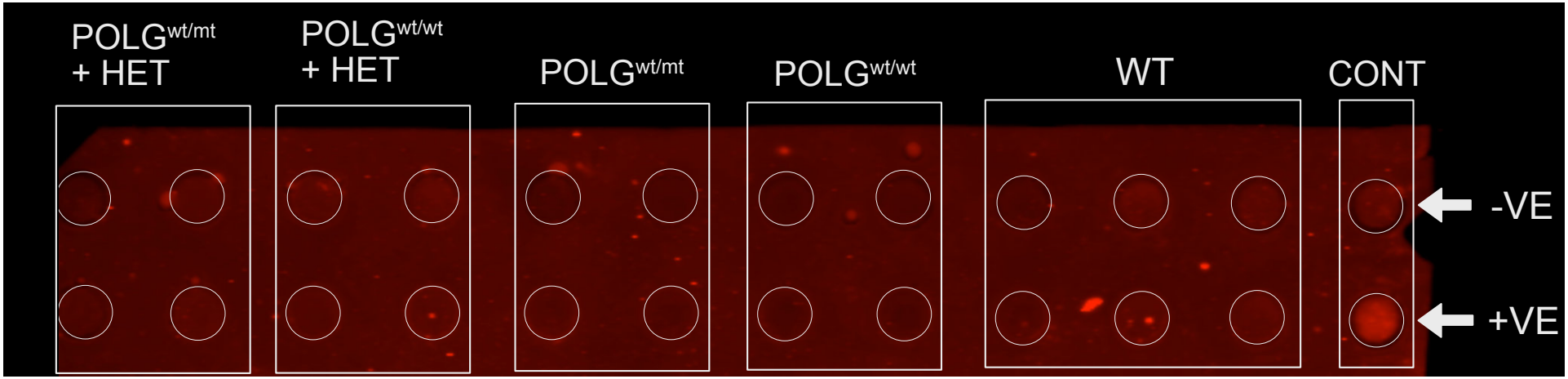


Figure 7.25. The results of α-synuclein aggregation assay. Each well (circle) represents protein extracted from the cortex of a single mouse, which are then clustered by genotype (indicated by boxes). A positive control shows the presence of α-synuclein on the membrane. No other samples showed evidence of any aggregated α-synuclein.

7.5 Discussion

7.5.1 Breeding programme

To try and model the interaction between low-level inherited mtDNA mutations and/or acquired somatic mutagenesis of mtDNA consistent with the potential paradigms of human aging, we decided to use the *POLG* D257 mutant mouse in the heterozygous state rather than in the homozygous state (which is otherwise known as the 'mutator mouse') (Trifunovic, Wredenberg et al. 2004). We utilized these genotypes for several reasons. Firstly, the lifespan of the mutator mouse is approximately 9 months of age (Trifunovic, Wredenberg et al. 2004, Kujoth, Hiona et al. 2005), and this is prior to the expected onset of phenotypic changes and neuropathology within the PDGFB-SNCA mouse (Rockenstein, Mallory et al. 2002, Chesselet and Richter 2011). Secondly, we used both the offspring of heterozygous females and heterozygous male mice as both the 'inherited' and 'acquired' genotypes respectively given that work from the Larsson lab revealed them to have a somatic mutation rate roughly equivalent to twice that of wild-type mice (Ameur, Stewart et al. 2011), which is more analogous to the relatively low level rises in heteroplasmy observed in the brain of humans with age (Wei, Keogh et al. 2017). Thirdly, working with homozygous POLG mutant 'mutator' mice often show less than Mendelian inheritance and poor fecundicity (personal correspondence), and in-keeping with the 3 Rs of animal testing (http://www.nc3rs.org.uk/ARRIVE) we wished to minimise the number of mice utilised in this study and ergo worked with heterozygous mice. Fourthly, we wished to utilize a mouse strain that expressed a native form of human α-synuclein that recapitulates the wild-type human protein conformation seen in adult humans with Parkinson's disease. Finally, our initial primary motor end point was that of a change in rotarod speed or time, and for which we had an 80% power to see a 20% difference in speed by that age based on previous data using the same *SNCA* mutant mouse model (Masliah, Rockenstein et al. 2000).

7.5.2 Phenotypic data

We observed no difference in the weight of mice between with *POLG* genotype or inherited heteroplasmy status. There did however appear to be a subjective trend towards a more similar weight within cohorts that inherited their *POLG* and *SNCA* alleles from the same maternal or paternal lineages (Figure 7.11), with mice born to *POLG* heterozygous mothers appearing to be lighter than those born to *POLG* heterozygous fathers. This difference was not statistically different at any age (un-paired t-test) (Figure 7.11), and in-keeping with general observations over the 14 month breeding programme, we conclude that there were no gross morphological or physical differences between any genotypes of mice with carrying the *SNCA* transgene.

The observation of highly atypical patches of white fur on two mice with inherited heteroplasmy (one $POLG^{wt/wt}$ and one $POLG^{wt/mt}$ (not shown)) prompted the investigation of whether there were significant differences in fur colour between cohorts. We did observe some differences in the maximum and minimum signal intensities at one level of illumination (982 Lux), detecting that mice with inherited heteroplasmy had both lower minimum signal intensities (in-keeping with a lighter colour), and higher maximum signal intensities (in-keeping with a darker colour). The possibility that fur colour may be a biomarker of genotype, and potentially therefore reflect the severity of the underlying neuropathology is intriguing, as lighter human hair colour is strongly associated with the risk of developing PD (Gao, Simon et al. 2009). In addition, skin melanoma risk is strongly associated with the development of Parkinson's disease and vice-versa (Liu, Gao et al. 2011). Whilst the mechanism between this association is unclear, there is putative biological rational to associate the two conditions given that dopamine, like melanin, is synthesized from the amino-acid tyrosine, and that impairment of this synthesis pathway may result in both impaired melanin in the skin and hair and also within the substantia nigra which are

strongly enriched with neuromelanin (Good, Olanow et al. 1992). However, the findings in our mice must be interpreted with caution given the lack of replication at 682 lux and the lack of significant difference to controls mean that these changes that suggest that the presence of inherited heteroplasmy may predispose to changes in fur pigmentation can only be considered preliminary, but further studies addressing the level of melanin within the fur from transgenic mice may be an interesting avenue of further exploration.

Our primary motor end-point for the study was a difference in speed and time spent on the rotarod, and we calculated that we had an 80% power to detect a 20% change in rotarod performance prior to the commencement of the study. Whilst we observed a general decrease in maximal speed and total time with age (as is expected) there was no difference in speed or time between any *POLG* genotype at any age. Whilst we therefore must conclude that there was no difference in rotarod ability between genotypes, several factors must be acknowledged with regard to this result. Firstly, the rotarod test is established as a measure of co-ordination, being most sensitive to detect cerebellar dysfunction rather than changes in strength or other motor features (Shiotsuki, Yoshimi et al. 2010), and cerebellar pathology is not a feature of the PDGFB-SNCA mouse (Chesselet and Richter 2011). In addition, the sensitivity of rotarod performance to detect motor deficits is strongly dependent on the individual task and protocol used (Pallier, Drew et al. 2009). Speculatively, the frequent testing of rotarod performance employed in this study (monthly) may have functioned as inadvertent 'training' thus improving improve rotarod performance (Scholz, Niibori et al. 2015), retaining functional capacity despite the presence of increased α-synuclein deposition (Scholz, Niibori et al. 2015).

Activity level in the open field test (OFT) (measured by total distance travelled) revealed that mice with a susceptibility to inherit heteroplasmy (irrespective of their POLG genotype)

travelled less far during the 5 minute period than control mice. The OFT is an extremely useful modality to assess locomotive impairment and general exploratory behaviour in mouse models of neuromuscular disorders (Raben, Nagaraju et al. 2000, Tatem, Quinn et al. 2014). It is also a reasonable surrogate of the six-minute walk test in humans (Grounds, Radley et al. 2008, Kobayashi, Rader et al. 2012) which has shown some correlation with the clinical severity of PD in humans (Falvo and Earhart 2009). Taken together, our data suggest that the inheritance of some baseline heteroplasmy exacerbates this locomotor impairment. Central field time also revealed that mice with inherited heteroplasmy and those that generate somatic heteroplasmy show reduced central field time compared to controls and PDGFB-SNCA mice without heteroplasmy. Central field time in the open field test is a reliable marker of anxiolysis (Prut and Belzung 2003), with reduced central field time indicative of increased levels of anxiety (Gould 2009). Taken together our data suggest that the presence of either inherited or acquired heteorplasmy exacerbate anxiolysis in PDGFB-SNCA transgenic mice.

Given that the OFT can be confounded by several factors such as circadian rhythm of the mice, intercurrent illness, and genetic background (Walsh 1976), we carefully controlled for these factors by performing all assessments at a standardised time of the day, and by ensuring that all POLG genotypes were created from both the maternal and paternal inheritance of each mutant allele and the same background C57bl/6 strain (Tatem, Quinn et al. 2014). Our results also suggest that unlike the human A53T mutant mouse model of PD in which mice showed less anxiety related behaviour compared to controls (Giasson, Duda et al. 2002), mtDNA mutations in combination with wild-type human α-synuclein may be a better model of PD related anxiety.

We did not observe any difference in the novel object recognition test (NORT) performance between the genotypes suggesting no difference in the short-term memory abilities of the different genotypes of mice (Antunes and Biala 2012). This task is highly dependent upon the performance of the perhinal cortex and to a lesser extent the hippocampus (Reger, Hovda et al. 2009, Goulart, de Lima et al. 2010, Warburton and Brown 2015). Whilst we detected a trend towards significant differences in both total and human α-synuclein expression within the hippocampus between genotypes (see below), these pathological changes were not reflected in the performance in the novel object recognition test, suggesting that α-synuclein deposition may not proportionally impair working memory in mice. However, it should be noted that whilst the OFT is a commonly used assessment to detect anxiety, previous studies of different transgenic α-synuclein mice have observed that the behavioural phenotype can only be captured in an elevated maze test and not an open field test (George, Mok et al. 2010), and other models have shown that significant hippocampal impairment only results in the presence of the aggregated form of protein (Hall, Yang et al. 2015), which was not present in our study (see below). In addition, our observation that in the absence of mtDNA mutations a strong negative correlation between novel object exploration time and hippocampal human α-synuclein expression was observed may also suggest that human α-synuclein deposition only impairs hippocampal function in the absence of mtDNA mutations. Therefore, mtDNA mutations may promote phenotypic heterogeneity in memory performance via an as yet unknown mechanism. This however is largely speculative and further work would be required to support such a hypothesis.

7.5.3 MRI imaging

The PDGFB-SNCA transgenic model of Parkinson's disease (Masliah, Rockenstein et al. 2000) is known to exhibit the highest level of expression of α-synuclein within the hippocampus, followed by the neocortex and olfactory cortex (Rockenstein, Mallory et al.

2002), with low levels in the cerebellum, substantia nigra and brainstem compared to other overexpression models of synuclein (Rockenstein, Mallory et al. 2002). The model is also known to reduce TH+ terminals in the striatum (Masliah, Rockenstein et al. 2000, Clark, Clore et al. 2010). VBM revealed a significant difference in signal in the hippocampus and primary motor area suggestive of increased volume in the $POLG^{wt/wt}$ group with no inherited heteroplasmy compared to the two cohorts with inherited heteroplasmy ($POLG^{wt/mt}$ and $POLG^{wt/wt}$). The suggestion that mice that are not exposed to mtDNA heteroplasmy have increased volumes of their hippocampi and primary motor cortex may suggest the mechanism by which the $POLG^{wt/wt}$ mice without inherited mtDNA mutations perform better in the open field test compared to other genotypes, although the relative importance and role of this region of the brain in the performance in such task is unknown.

However whilst most common interpretation of VBM results is the inference of volumetric changes, several factors that affect T2 signal intensity can be erroneously interpreted as changes in volume (given that they alter signal intensity) such as iron deposition, oedema, inflammation, necrosis and protein deposition (Schenck 1995, Deoni 2010). Additional factors such as subtle changes in the orientation and local field changes in image acquisition in the MRI scanner together with the segmentation and registering process can also create artefactual differences with VBM (Ashburner and Friston 2000, Jubault, Brambati et al. 2009), although we tried to mitigate against this by performing standardized scans using the same protocol in the same scanner.

Finally, VBM also suffers from a problem of multiple comparisons. To assess statistical significance, we used an adjusted p-value of 10^{-3} which is similar to approaches in human subjects (Rohrer, Ridgway et al. 2010). Our threshold is likely to limit the number of false negative results, but may increase the rate of false positive discovery (Ashburner and

Friston 2000). However, the clear anatomical correlation with the most prominent area of established neuropathology within the PDGFB-SNCA mouse model (Masliah, Rockenstein et al. 2000, Rockenstein, Mallory et al. 2002) and the absence of any significant differences between groups using opposite contrasts suggest that we can have a high degree of confidence in the results.

7.5.4 Molecular and pathological analysis

Long-range PCR observed no evidence of deletion formation within the brain. This is in contrast to previous studies which showed evidence of multiple mtDNA deletions within the neocortex and hippocampus but not in the cerebellum using both an overexposed southern blotting technique and LR-PCR in heterozygous *POLG* D257 mutant mice (Fuke, Kametani et al. 2014). These authors however also observed low level deletions within control mice, and the differences between our studies may be due to differences in the sensitivity of the techniques used to detect deletions, or due to the lox-p flanked D257 *POLG* mutation in our model compared to that used *Fuke* et al (Fuke, Kametani et al. 2014).

Surprisingly we found no difference in the absolute level of total α-synuclein expression in the cortex or hippocampus of any SNCA genotype compared to controls. However, as expected, human α-synuclein was only observed in the PDGFB-SNCA transgenic mice. Given that the PDGFB-SNCA retains it's native mouse α-synucelin gene (Masliah, Rockenstein et al. 2000), these data suggest that the expression of human α-synuclein significantly represses the expression of native mouse α-synuclein within the hippocampus and cortex, ostensibly replacing it with human α-synuclein. Further work will aim to determine the exact quantity of mouse α-synuclein expression that remains which is of interest given the propensity of a mix of α-synuclein from two species to reduce the

likelihood of aggregation (Fares, Maco et al. 2016, Luk, Covell et al. 2016) and the degree of pathology in mice (Cabin, Gispert-Sanchez et al. 2005).

Whilst we did not observe any difference in human α-synuclein or total α-synuclein levels between any of the 4 genotypes, we did observe significant differences when grouping mice by the presence or absence of the heterozygous mutant *POLG* allele. This revealed that mice with a heterozygous POLG mutation (and thus mice that have a predisposition to produce somatic heteroplasmy) showed a reduction in human α-synuclein expression in the cortex, with a trend towards increased expression in the hippocampus. These data are compelling given that the SNCA and POLG alleles within these cohorts were inherited from both parents, reducing the risk that background strains could be modifying this effect (Doetschman 2009). At present, our data is insufficient to explain this observation, but they suggest that impairment of mtDNA replication (rather than the presence of mtDNA burden *per se*) may reduce human alpha-synuclein expression in the cortex and therefore the acquisition of somatic mtDNA mutations may be protective within that brain region.

The mechanism mediating this is however unclear and warrants further study. However, a hypothesis may be generated from existing published data. For example, impairment of POLG activity in the mouse brain of homozygous *POLG* D257 carriers showed that of the top 20 differentially expressed proteins compared to WT mice, 15 were in the mitochondrial respiratory chain as perhaps may be expected (Hauser, Dillman et al. 2014). However, the top three non-respiratory chain proteins with differential expression were AP2A1, CRYM and GSTM1, and all three have been associated with α-synuclein metabolism or transport, or dopaminergic cellular impairment: AP2A1 (Adaptor-related protein complex 2, alpha 1) is a critical protein present in the wall of clathrin vesicles, and impairment triggers dystrophic changes in domapinergic axons (Cao, Wu et al. 2017). CRYM (crystalline) is

one of the most up-regulated genes induced by viral-vector mediated α-synuclein expression in the brain of rats (Qin, Buckley et al. 2016), and genetic variants in GSTM1 (Glutathione S-Transferase Mu1) have been associated with the development of PD (Ahmadi, Fredrikson et al. 2000), together with the common feature of reduced glutathione concentration in the post-mortem brain of patients (Sofic, Riederer et al. 1987). Whether dysregulation of the expression of these genes occurs in the context of a heterozygous POLG mutation too remains unclear and will be determined by future experiments, but may explain potentially differing levels of α-synuclein expression between genotypes.

Finally, we observed no aggregates of α-synuclein in any genotype or controls using our aggregation assay. Whilst small aggregates can be seen histologically in the brain of PDGFB-SNCA mice, whether either a sufficient amount, or large enough aggregates (Poehler, Xiang et al. 2014) are able to be detected by our assay is unknown. Further histological examination should help prove informative, and further mouse-crosses using the PDGFB-SNCA allele on a background strain that has removed the mouse endogenous murine α-synuclein gene may also provide valuable information given the ability of different strains of α-synuclein to reduce aggregation ability (Fares, Maco et al. 2016, Luk, Covell et al. 2016).

Chapter 8

Final discussion

Whilst the data derived in each chapter was discussed in turn, herein I provide an over-arching discussion of the whole body of work and it's relevance to the field of neurodegenerative disease research.

Chapter 2: The central dataset of this study was derived from tissue samples contained within four centres of the MRC Brain Bank (The University of Oxford, Kings College London, Newcastle University and Edinburgh University). Brain Bank donations nationally number only a few hundred per year but offer a unique resource for the study of neurological disorders and the central processes of neurodevelopment and degeneration. Whilst *in-vitro* studies and animal studies can recapitulate some components of disease phenotypes and pathology, the ability to observe and study neurological disorders within human tissue offers the unique possibility to study cellular specific responses to disease within their native tissue.

To date, the MRC Brain Bank curates clinical data together with relevant neuropathological data at all sites. While extensive data is collected, much is largely historical, with several clinical and pathological diagnoses and assessments having been made several decades ago, often having not been revisited to challenge or investigate a particular putative diagnosis. This means that several more recent neuropathological molecular techniques which may suggest a specific genetic aetiology have often not been undertaken. In addition, the life expectancy of the parents and grand parents of elderly individuals who died 10-20 years ago (and who are therefore present within the brain bank) was significantly shorter than the

current life expectancy of adults. This means that familial cases within the MRC brain bank may not have been able to be identified from a family history due to death prior to the onset of symptoms. They may therefore have been erroneously considered as sporadic in nature further reducing the putative rationale to look for genetically mediated forms of disease.

The work in Chapter 2 of this thesis used a multi-modal sequencing approach in order to detect single nucleotide variants, small exonic insertions, deletions, small and large copy number variations, the *C9orf72* hexanucleotide repeat expansion and the *APOE* genotype of 1511 cases within the MRC Brain Bank. This data is now available to download and is becoming incorporated and cross-referenced with the MRC Brain Bank metadata within their archive. This will result in the first genetically stratified brain tissue resource in the world (to our knowledge), and it is an extremely satisfying achievement to have produced this.

The data in Chapter 2 also identified at least 61 cases of monogenic disease, and 349 with risk factors across case and control cases. The thorough clinical assessment of variant pathogenicity and the curation of previously calculated statistical associations with disease for over 400 variants in 1500 individuals was a huge challenge, but now provides a framework for variant interpretation both within the brain bank and as a reference resource for the greater community. As new genes are identified as causing or contributing to neurodegenerative disorders, the variant interpretation data can be re-interrogated and added to at regular intervals. Such a resource has not to our knowledge been provided before. In particular, the identification of a spectrum of ages of control individuals carrying risk factor alleles is likely to help facilitate our understanding of the interaction of these risk variants with the process of aging, which remains the biggest risk factor for almost all neurodegenerative diseases.

To organize and co-ordinate sample selection and shipment, finalize appropriate metadata, liaise with all sites to ensure data homogeneity, ensure that all data storage was appropriately coded and anonymised, and ensure that all tissue and data was stored and accessed in accordance with all relevant guidelines was a huge challenge. This study also involved a lot of additional work that isn't directly reflected in the thesis, for example, organizing upgrades to the computational infrastructure and storage (involving procuring additional nodes for the computer cluster, data storage and programme licences). The co-ordination of vast quantities of resulting data from exome sequencing, SNP genotyping and *APOE* genotyping together with the *C9orf72* data was a huge logistical and technical challenge. Subsequent data quality control was extensive, and variant interpretation and cross-referencing was complex, together with the incorporation of this data into statistical models and tests to understand genotype-phenotype correlations and interactions.

This resource will however now enable tissue selection to be stratified by genetic background. This offers the possibility both to select individuals with rare variants and also exclude cases carrying rare variants from studies of ostensibly sporadic disease. In addition, the dataset also offers the possibility for researchers to cross-reference novel alleles to either confirm or refute pathogenicity in clinical cases; something often not possible from large control datasets of young individuals such as the Exome Aggregation Consortium (Lek, Karczewski et al. 2016). I have also received personal communication from at least two brain banks that they are now going to begin to collect additional non-CNS tissues (such as blood, muscle, skin, and possibly hair) in order to facilitate the ability to investigate the transcriptomic and metabolomic changes that result within the CNS and peripheral tissues as a consequence of the genetic variation we have now captured. This should significantly improve the utility of brain banks for the scientific community in the future.

Chapter 3: Following careful curation of variants in order to determine pathogenicity in Chapter 2, we proceeded to investigate whether combinations of variants may increase the risk of developing neurodegenerative disorders. However, it immediately became apparent when re-examining the data to identify individuals with two or more variants in an unbiased manner, that the majority of cases identified were those in whom we had already identified a monogenic case of disease. This observation led to the development of the work in Chapter 3. We firstly observed that the presence of two variants conferred a high sensitivity and specificity to detect individuals with a monogenic case of disease or a disease risk factor. Given that the majority of genes that cause neurological disorders are highly conserved (Samocha, Robinson et al. 2014, Lek, Karczewski et al. 2016), i.e. that they show little genetic variation within the population due to their propensity to cause disease and therefore are selected against by natural selection, it perhaps stands to reason that the presence of more than one variant therefore increases the chances that at least one of these variants causes or contributes to disease. Conversely, thinking of the problem from the alternative perspective (as presented in Chapter 3), monogenic individuals must have at least one variant that causes disease, and if they also have the same background rate of benign polymorphic variation as the rest of the population, then a reasonable proportion of monogenic individuals will have at least two variants. The odds of this scenario are significantly higher than of an individual having two benign variants. Through Chapter 3 we showed that this was indeed the case, with almost all oligogenic individuals having a single established pathogenic allele alongside a benign variant or variants. We are therefore not only the first to highlight this systematic bias that may have taken place in previous studies of oligogenic disease mechanisms, but we also confirmed this hypothesis by re-examining the data by *Cady et al* (Cady, Allred et al. 2015). In addition, we defined the sensitivity and specificity for an oligogenic status to suggest the presence of a genetically mediated case of

disease across a spectrum of neurodegenerative disorders and gene panels, which is entirely novel data.

Whilst not discussed extensively within the chapter, we are also aware of the potential for these findings to facilitate clinical diagnostics. For example, the presence of more than one variant within the panels we have used should immediately alert the clinician to the potential that said individual may have a monogenic case of disease or disease risk factor. Larger scale sequencing studies (such as Genomics England's 100,000 Genomes Project), will be able to further define the accuracy of this approach as they develop robust well phenotyped cohorts of common disorders. It is however important to bear in mind that gene panels that are extremely large and encompass multiple genes including those with a high degree of natural benign polymorphic variation (lower conservation) will be unlikely to observe the same enrichment of monogenic variants within oligogenic individuals. In such large panels, almost all individuals within the population will carry at least two variants, thus meaning that the sensitivity would be 100%, but the specificity extremely low. Therefore, the positive predictive and negative predictive values are important to define at the panel level within case and control individuals in order to determine their ability to suggest the presence of a highly penetrant allele or risk factor.

Chapter 4: During the course of my research Fellowship, the hypothesis that somatic variation could contribute to causing non-cancerous disorders began to gather significant momentum (Poduri, Evrony et al. 2012, Poduri, Evrony et al. 2013, Cai, Evrony et al. 2015). Next-generation sequencing paradigms employed to call somatic variants almost invariably involve a paired tissue of interest (normally cancer / tumour) and normal tissue (either macroscopically healthy tissue from the same organ, or blood). Our central study of brains from the MRC Brain Bank only utilized a single sample from each individual, which

was in most instances extracted from brain. This meant that detecting somatic mutations was extremely technically challenging from a single sample with no comparative 'normal tissue' to compare to. We therefore developed a novel strategy that built upon previous studies that arose during the course of my Fellowship (Genovese, Kahler et al. 2014). The first step involved the re-alignment and calling of the FASTQ files from the MRC dataset using a GATK pipeline as this was suggested to have the best ability to detect somatic alleles (as determined by those of intermediate allele frequencies inconsistent with either homozygous or heterozygous mutations) (Genovese, Kahler et al. 2014). This involved utilizing GATK Perl scripts that I had used during previous projects (Keogh, Pyle et al. 2015, Keogh, Steele et al. 2015) (but was performed by Dr Wei Wei) in order to generate new BAM files. Thereafter, we collated both reference BED files of known structural variation and incorporated those from our SNP genotyping data as performed in Chapter 2. Thereafter, we required a method to decipher between alleles at established positions of mis-mapping (in which variable variant allele frequencies are common) and true somatic mutations. To do this we used a programme that ostensibly tested the observed VAF against the distribution of variant allele frequencies from > 300 other individuals within the dataset (DeepSNV) (programming performed by Dr Wei Wei). Following this we confirmed the accuracy of this approach by performing validation using both amplicon and pyrosequencing.

The development of this pipeline was a significant undertaking, and led to the detection of 22 validated alleles. Given that these alleles were at a relatively high variant allele frequency, we suggest that they were likely to have arisen during early embryogenesis, especially given the distribution of allele frequencies observed consistent with their acquisition during the first or second cell divisions (12.5% and 25% VAF). The frequency of these mutations were also closely consistent with those recently described in a similar

experiment using whole-genome sequencing, giving us high confidence in our findings (Ju, Martincorena et al. 2017). However, with only single samples from each individual, we cannot confirm that these mutations were not focal, and present in only the region of the brain from which the DNA was extracted. None the less, our data still show that at last 1% of individuals have a high frequency protein coding mutation in at least one region within the brain, and this significantly changes our understanding of the genetic architecture of the human brain and the potential for somatic mutations to cause or contribute to disease susceptibility.

Whilst we did not observe any mutations in genes that are known to cause neurological disorders, the profile of the variants, irrespective of their aetiology, was consistent with a lack of the same selection pressures that are present within the germline which limit the development of *de novo* mutations in highly conserved genes (Samocha, Robinson et al. 2014, Lek, Karczewski et al. 2016). This suggests that somatic mutations in highly conserved genes are more likely to arise somatically within the brain than within the germline, and therefore, somatic mutations offer the potential to be an underappreciated cause of neurological disorders.

Further studies are of course required to adapt and improve the calling algorithm that we developed. Secondly, additional work should aim to confirm the potential utility of our approach to detect somatic variants in whole genome sequencing (WGS) which is likely to be the most frequently employed sequencing methodology in clinical practice for the next decade. In addition, determining the ability of peripheral blood samples to detect somatic mutations that are also present in the brain is of vital importance in order to define both the diagnostic utility of blood samples to determine neurological mosaicism, and also to confirm the origins of the somatic mutations during embryogenesis. An additional important

area that requires further investigation is the development of assays or experimental processes that can help model somatic mutations. Whilst we can suggest through either clinical, bioinformatic or *in vitro* work that mutations in the hetero or homozygous state may cause disease, the thresholds for the proportions of cells required containing these mutations to influence disease pathogenesis are unknown. Further models are required in order to test the likely effect of somatic mutations within the brain and warrant additional focus.

Chapter 5: In the fifth chapter of this thesis we investigated the potential that copy number variants within the genome may be associated with particular forms of neurodegenerative disorders. Previous experimental platforms have used SNP genotyping arrays with SNPs relatively evenly distributed across the genome, with the majority of SNPs present within non-coding regions. This has offered the ability to accurately detect large copy number gains with relatively high accuracy leading to the association of large CNVs with childhood developmental disorders and autism for example (Cook and Scherer 2008, Glessner, Wang et al. 2009, Coe, Witherspoon et al. 2014).

CNVs in neurodegenerative disorders have only been observed in genes in which coding regional SNVs also lead to disease (such as *APP, SNCA,* and *PARK2*). In Chapter 5, we utilized a SNP genotyping array that has the majority of SNPs within the exome, offering us a unique ability to detect small coding region CNVs in individuals with neurodegenerative disorders. We were also able to confirm the accuracy of this approach using comparative data derived from exome sequencing depth-based programmes. Remarkably we observed 3 copy number gains that were associated with CJD, and primarily sCJD. CNVs in *LAMA5* were most frequently observed, and CNVs in *EXD3* and *TRPM2* were also observed in a subset of these individuals. Remarkably none of these CNVs were observed in any other

individual in the study (either with any other neurodegenerative disease or as a control). This finding remains difficult to entirely explain. It appears statistically unlikely that several unrelated individuals (which was confirmed by identity by descent analysis) would all develop the same rare CNV combinations independently. We therefore suggest that a common large CNV encompassing all 3 genes arose somewhere in their distant common ancestry and over time this has been 'pruned' within some branches of that ostensibly large pedigree. The disease risk is likely to be mediated by the *LAMA5* gain (given that it is ubiquitously present in all individuals carrying the *EXD3* or *TRPM2* gain), but we observed no evidence that it promotes the development of any particular form of CJD, or modulates any parameter of disease phenotype. Further work is vital to determine the mechanism by which this gain may contribute to increasing disease risk, and whether whole genome sequencing (the primary genetic investigation for the next decade) can detect these CNVs accurately, thus offering the possibility to screen and identify patients.

Chapter 6: Following the observation of high variant allele frequency somatic mutations in Chapter 4, we performed additional experiments on new tissue samples designed to determine whether lower VAF somatic mutations (<10%) are present within the brain. In addition, we aimed to determine their spatiotemporal distribution, and whether mutational profiles may differ between diseased individuals and controls. Given the heterogeneity of data across the exome and the difficulty of calling even high variant allele frequency mutations (Chapter 4), together with the cost and likely methodological challenges of a pan-exome approach in multiple brain regions at very high depth, we discounted this approach. We therefore focused on sequencing established genes that cause neurodegenerative diseases to investigate the hypothesis that they may accumulate in regions of the brain in which high levels of neuropathology may occur, choosing to focusing our attention on Lewy

Body disorders and Alzheimer's disease for which we had access to high quality neuropathological data and samples available.

We first carefully designed spiked in control samples, performed in duplicate, to enable accurate data pertaining to the sensitivity and specificity of calling approaches. We subsequently performed a variety of calling approaches using both paired and un-paired callers to detect our relevant calling thresholds. These experiments were carefully designed and conducted providing high-quality data on detection thresholds that we had observed was often lacking in many deep-sequencing studies. By virtue of this considered approach, we subsequently were able to detect and validate 39 somatic variants within at least one brain region. Such high depth sequencing with such high fidelity calling has never been performed in any study of neurodegenerative disorders before and provides significant methodological advances together with biological findings to the field. Whilst it could perhaps be argued that the number of focal somatic mutations could have been estimated from DNA polymerase error rates, we confirm that somatic mutations at relatively high variant allele frequencies not only occur, but we can accurately define the levels to which they rise, their spatial distribution, and the profile of the mutations. In conjunction with Juvid Aryaman and Dr Nick Jones (mathematicians at Imperial College London) we were able to estimate the origin and likely prevalence throughout the brain, which, although based on crude models, provide a framework to understand the prevalence of such mutations within the brain.

Perhaps the most remarkable finding of this chapter was the frequent observation of mutations that are associated with clonal haematopoesis. The development of age-related clonal haematopoesis is an area of exciting research, and during the course of this thesis, several studies have both defined the prevalence and nature of these mutations in the blood

of the aging population, and were beginning to link them to age-related disorders such as stroke and myocardial infarction (Jaiswal, Fontanillas et al. 2014, Jaiswal, Natarajan et al. 2017). In addition, *in vivo* studies in mice have shown that transgenic mice with a deficiency of some of the genes that most commonly acquire somatic mutations in the blood exacerbate atherosclerosis (Fuster, MacLauchlan et al. 2017), suggesting that murine models may offer a robust experimental paradigm to understand the interaction between the central nervous system and age-related clonal haematopoetic variants.

Despite these recent studies, we are the first to our knowledge to describe the association between somatic variants seen commonly with clonal haematopoesis and Lewy body disorders. In addition, we are also the first to show that there may be a well conserved spatial profile of the accumulation of these somatic variants within the brain, with a predilection for higher levels to accrue within the medulla and to a lesser extent entorhinal cortex and frontal cortex.

Further work is however required to further understand the mechanism by which these age-related clonal variants may influence the aging brain. For example, are the detected mutations present in leukocytes that have translocated across the blood-brain barrier? Or, are they present only in microglia? If the latter is true, then at what stage of development did these mutations arise, and can it help clarify key questions about the development and turnover of glia within the brain (Reu, Khosravi et al. 2017)? To clarify the specific cell types in which these mutations are present will offer significant methodological challenges for the future, and will require significant improvements in our ability to purify different cell types from frozen tissue homogenate, for example using flow cytometry. Secondly, if these experiments were indeed to show that the mutations were present in translocated leukocytes, then trying to identify the physical location of cells carrying the mutation/s and their

proximity with respect to cells containing aberrant neuropathology would require even more advanced methodological techniques based most likely on sequencing in-situ technologies. However, the recent success of transgenic mouse models that showed that knock-down of commonly mutated genes within the aging blood can exacerbate atherosclerosis suggests that murine modelling may be an efficacious paradigm in the first instance or in parallel with such studies to determine the role of clonal haemaotpoetic variants in Lewy body pathology.

Chapter 7: The final chapter of the thesis also aimed to study the interaction between age-related clonal variants and Lewy Body disorders, though in this experiment it was the clonal acquisition of mitochondrial DNA mutations rather than haematopoetic variants in nuclear genes. Following a plethora of studies measuring mitochondrial DNA burden in individuals with a variety of neurodegenerative disorders (though primarily Lewy Body disorders) over the past 10 years (Keogh and Chinnery 2015), no experimental evidence has robustly determined whether their development is a cause of, or effect of, the deposition of Lewy bodies within the brain. We initially considered whether *in-vitro* cell studies could help address this question, however, the marked cell-type specific accumulation of both mitochondrial DNA mutations and α-synuclein meant that the heterogeneity of the propensity of cells to develop each of these pathologies was difficult to recapitulate *in vitro*. In addition, we were also keen to explore the behaviour of inherited low-level variants (occurring as a result of having a POLG mutant mother in our experimental paradigm) over a prolonged period of time (~14 months), which would have been impossible to conduct in cellular models *in vitro*. This model we felt is most analogous to human inheritance of mitochondrial DNA heteroplasmic variants, with a small number of low-level variants being transmitted between individuals, thereafter either clonally expanding or being removed from

different tissues and cell types (Elson, Samuels et al. 2001, Taylor, Barron et al. 2003, Payne, Wilson et al. 2013).

The crosses between these mice, and the work in Chapter 7, has enabled us to make several novel but preliminary observations about the role that mitochondrial DNA mutations may play in Lewy Body disorders. Firstly, we did not see any evidence that mitochondrial DNA mutations (either inherited or acquired) exacerbate rotarod performance, which was our primary motor endpoint. However, we did observe that mice without any mitochondrial DNA mutations showed much lower levels of anxiety (by virtue of the results of central field time), and also increased total activity levels within the open field. These data suggest that the presence of mitochondrial DNA mutations (whether inherited or acquired) may exacerbate some behavioural traits, in particular those associated with hippocampal function. These data are also supported by MRI volumetric data that revealed that mice with no mitochondrial DNA mutations had apparently greater hippocampal volumes than mice with heteroplasmic variants, suggesting that mitochondrial DNA mutations may promote cell loss or atrophy within this region of the brain.

In contrast to the behavioural and volumetric changes, the results of protein expression studies actually revealed a trend towards lower levels of α-synuclein expression in the cortex of mice that somatically produce mitochondrial DNA mutations. Taken together, one possible hypothesis resulting from this study is that mitochondrial DNA mutations in the presence of Lewy Body deposition result in cells becoming more susceptible to cell death or to the loss of their dendritic connections, worsening anxiety-like behaviour and mobility, and resulting in the apparent 'shrinking' of certain brain regions. Conversely, the accumulation of mtDNA mutations may also paradoxically slow the expression of α-synuclein, or improve it's degradation and metabolism.

Several further lines of study are required to further clarify the potential mechanisms underlying these observations. Firstly, a larger sample size of each genotype should have Western blotting performed for α-synuclein (human, mouse and all forms) in order to clarify potential sub-threshold associations seen in the data. Secondly, stereological analysis will determine whether indeed mitochondrial DNA mutations potentiate cell-death, thus explaining the poor performance in phenotypic testing and volumetric loss within brain regions. These findings would be of vital importance and would suggest that reducing the accumulation of these mutations with age may offer therapeutic utility in patients. Thirdly, further studies of the unfolded protein response (UPR) and the pathways involved in protein aggregation and degradation are vital in order to confirm whether mice with mtDNA mutations up regulate degradation pathways. This would enable us to clarify whether the reduction in observable α-synuclein could be due to an increased ability to clear the protein, or whether it is likely to simply reduce it's relative expression. Further work by future students should bring clarity to this area, and determine the exact role that mitochondrial DNA mutations may have in Lewy Body pathology.

Appendices

Appendix 1 All 80 cases included in the 'other' category within the study.

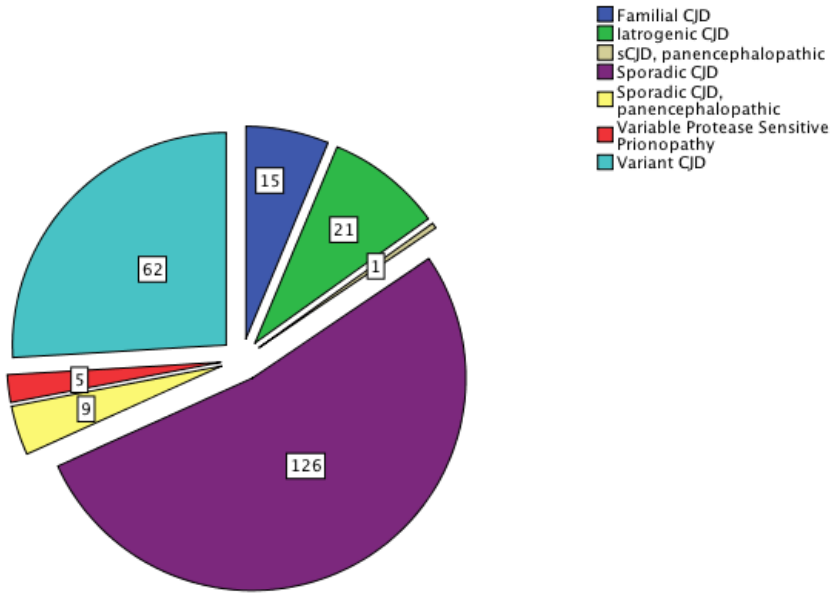
All 80 cases included in the 'other' category within the study. Number of cases together with age of onset, death, sex and family history of disease are recorded

	N	Age of onset (years)		Age of death (years)		Female	Male	FH of disease	
		Mean	SD	Mean	SD	N	N	Y	%
AD and FTD	1			73		0	1	0	0.0%
Adult onset gangliosidosis	1	65		71		0	1	0	0.0%
Agyrophilic Grain Disease	4	78		87.5	1.9	2	2	0	0.0%
Amyloid angiopathy	1			84		1	0	0	0.0%
Ataxia	1	85		85		1	0	1	100.0%
Atypical dementia not consistent with major classifications	3	55		77.3	11	1	2	0	0.0%
Atypical Substantia Nigra degeneration	2	79		83.5	2.1	1	1	0	0.0%
Atypical tauopathy	3	75	13.1	85.3	12.5	1	2	0	0.0%
Atypical TDP-43 deposition	1	76		83		1	0	0	0.0%
CADASIL	1	40		62		0	1	1	100.0%
Corticobasal Degeneration and concomitant AD	1			84		1	0	0	0.0%
Cerebello-olivary atrophy	1			75		1	0	0	0.0%
Cerebello-olivary degeneration	1	12		58		1	0	1	100.0%
Chorea-acanthocytosis	1			40		0	1	0	0.0%
Chronic encephalopathy	1			71		0	1	0	0.0%
Central Pontine Myelinolysis	1	64		67		0	1	0	0.0%
Chronic Traumatic Encephalopathy	1			69		0	1	0	0.0%
Demyelinating disorders	2			43.5	13.4	1	1	0	0.0%
Epilepsy	1			24		1	0	0	0.0%
Huntington Disease phenocopy	1	40		62		1	0	1	100.0%
Hepatic encephalopathy	1			52		1	0	0	0.0%
Kuf's disease	1	43		58		0	1	0	0.0%
Lewy Body Disease – no clinical features	2			90	0	2	0	0	0.0%
Learning difficulty and Epilepsy	1	10		54		0	1	0	0.0%

	N	Age of onset (years)		Age of death (years)		Female	Male	FH of disease	
Mild Cognitive Impairment	1			85		0	1	0	0.0%
MELAS	1	13		49		0	1	0	0.0%
Mitochondrial disease	1	0		3		0	1	0	0.0%
Mixed Alzheimer Disease and Lewy Body Disease	1	63		75		1	0	0	0.0%
Mixed Corticobasal degeneration and Dementia with Lewy Bodies	1	79		77		0	1	0	0.0%
Multiple Sclerosis	3	34.3	16.7	61.7	11.9	1	2	1	33.3%
Neocortical Lewy Body Disease – no clinical features	1			73		0	1	0	0.0%
Neuroaxonal dystrophy	1	14		16		1	0	0	0.0%
Neurofibribrillary tangle only dementia	3	80.3	8	87.7	9.1	0	3	0	0.0%
Normal Pressure Hydrocephalus	1	64		65		0	1	1	100.0%
Paraneoplastic encephalopathy	1			66		0	1	0	0.0%
Parkinson's Disease and Motor Neuron Disease	1					0	1	0	0.0%
Parkinson's disease related changes	2			84.5	3.5	0	2	0	0.0%
Primary Familial Basal Ganglia Calcification	2	75		76	2.8	2	0	0	0.0%
Possible Alzheimer Disease – clinical history unclear	1	62		82		1	0	1	100.0%
Possible paraneoplastic dementia	1	52		64		0	1	0	0.0%
Pre-clinical tauopathy	1			68		0	1	0	0.0%
Pre-symptomatic Dementia with Lewy Bodies	2	67		86.5	13.4	1	1	0	0.0%
Pre-symptomatic Frontotemporal dementia	1			79		1	0	0	0.0%
Probable Alzheimer Disease	4	84		86.3	6.1	2	2	0	0.0%
Spinocerebellar ataxia (SCA)	1			50		0	1	0	0.0%
SCA1	1			76		0	1	0	0.0%
SCA14	1			103		0	1	1	100.0%
SCA2	2			68	18.4	1	1	1	50.0%
SCA7	1			58		1	0	0	0.0%
Spinal muscular atrophy	1					1	0	0	0.0%
Superficial Siderosis	1	63		72		1	0	0	0.0%
Tauopathy	5	63	1.4	72.8	11.2	2	3	0	0.0%
Uncategorized dementia	1			95		1	0	0	0.0%
Unusual tauopathy	1			67		0	1	0	0.0%
Vascular disease / DLB	1	90		91		1	0	0	0.0%
Total	**80**	**57.4**	**25.3**	**72.4**	**18.3**	**35**	**45**	**9**	**11.3%**

Appendix 2 Stratification of CJD cases within the cohort

Numbers identify the number of cases of each sub-type of CJD across the entire cohort of CJD cases.



Appendix 3 Genes and inheritance patterns causing their relevant neurodegenerative disease.

Disease category; PD – Parkinson's disease, AD – Alzheimer's disease, FTD – Frontotemporal dementia, ALS – Amyotrophic lateral sclerosis. Inheritance; AD – Autosomal dominant, AR – Autosomal recessive, XLD – X-linked dominant, RF- Risk factor (dominant).

Gene	Disease	Inheritance
SNCA	PD	AD/RF
PARK2	PD	AR
PINK1	PD	AR
EIF4G1	PD	AD
GIGYF2	PD	AD/RF
HTRA2	PD	AD
UCHL	PD	AD
SPG11	PD	AR
VPS35	PD	AD
FBX07	PD	AR
APP	AD	AD
PSEN1	AD	AD
PSEN2	AD	AD
c9orf72	FTD / ALS	AD
GRN	FTD/AD	AD
CHCHD10	FTD	AD
TARDBP	FTD	AD
SOD1	ALS	AD/AR
FUS	ALS	AD
PFN1	ALS	AD
hnRNPA2B1	ALS	AD
hnRNPA1	ALS	AD
SETX	ALS	AR
VAPB	ALS	AD
OPTN	ALS	AR
VCP	ALS	AD
DAO	ALS	AD
ANG	ALS	AD
DCTN1	ALS	AD
PARK7	PD	AR
CHMP2B	FTD/ALS	AD
SQSTM1	FTD/ALS	AD/RF
PRPH	ALS	AD
DPP6	ALS	AR
MATR3	ALS	AD
MAPT	FTD	AD
ALS2	ALS	AR
SIGMAR1	ALS	AD
UBQLN2	FTD	XLD
NOTCH3	CADASIL	AD
PRNP	fCJD	AD
COQ2	MSA	AD/AR
GBA	DLB	RF
LRRK2	PD/DLB	RF
TREM2	AD	RF
SCARB2	DLB	RF
PON1	ALS	RF
PON3	ALS	RF
APOE	AD	RF

Appendix 4 Genes analysed for rare presentations within the cohort.

Heterozygous and homozygous rare variants (as described in methods) in all genes within the table were assessed, together with homozygous variants only for the patient with infantile mitochondrial disease.

Disease	**Genes analysed by phenotype**											
Neurodegeneration with Brain Iron Accumulation (NBIA)	C19orf12	FA2H	PANK2	PLA2G6	WDR45	PLA2	ATP13A2	COASY	CP	DCAF17	SCP	FTL
Infantile onset Mitochondrial disease	BCS1L	CLEC4GP1	CLEC4M	COX10	COX15	ECHS1	FBN3	FOXRED1	GYG2	IARS2	LRPPRC	NDUFA10
	NDUFA12	NDUFA2	NDUFA9	NDUFAF2	NDUFAF5	NDUFAF6	NDUFS2	NDUFS3	NDUFS4	NDUFS7	NDUFS8	PDHA1
	PDHB	PET100	PRR36	SDHA	SURF1	TACO1	TRAP1	TRAPPC5	TTC19	AARS2	ACAD8	ACAD9
	ACO1	ACO2	AGK	AGTR1	AIFM1	AMACR	ANGEL2	APOPT1	ATP5A1	ATP5E	ATPAF2	BCS1L
	BOLA3	C10orf2	C12orf65	CHKB	CHRNA4	CHRNA7	CHRNB2	CLEC4GP1	CLEC4M	COA5	COA6	COX10
	COX14	COX15	COX20	COX6B1	CPS1	CPT1A	CYC1	DGUOK	DLAT	DLD	DNA2	DNAH8
	DNM1L	EARS2	ECHS1	ECI1	ECSIT	ELAC2	EPHX1	FARS2	FASTKD2	FBN3	FOXRED1	FXN
	GAD1	GALNS	GFER	GFM1	GNPAT	GOLPH3	GPAM	GYG2	HFE	HMGCS2	HNF1A	HOGA1
	IARS2	IBA57	IVD	LARS2	LIPT1	LRPPRC	LYRM4	MCCC2	MFF	MGST3	MPC1	MPV17
	MRPL3	MRPL44	MRPS16	MRPS22	MRRF	NARS2	NDUFA1	NDUFA10	NDUFA11	NDUFA12	NDUFA13	NDUFA2
	NDUFA3	NDUFA4	NDUFA4L2	NDUFA5	NDUFA6	NDUFA7	NDUFA8	NDUFA9	NDUFAB1	NDUFAF1	NDUFAF2	NDUFAF3
	NDUFAF4	NDUFAF5	NDUFAF6	NDUFAF7	NDUFB1	NDUFB10	NDUFB11	NDUFB2	NDUFB3	NDUFB4	NDUFB5	NDUFB6
	NDUFB7	NDUFB8	NDUFB9	NDUFS1	NDUFS2	NDUFS3	NDUFS4	NDUFS5	NDUFS6	NDUFS7	NDUFS8	NDUFV1
	NDUFV2	NDUFV3	NFU1	NIPSNAP1	NIPSNAP3A	NPL	NUBPL	OPA1	OPA3	OXCT1	PARS2	PC
	PCK2	PDHA1	PDHB	PDHX	PET100	PHYH	PNPT1	POLG	POLG2	PPARG	PRR36	PTCD1
	PUS1	RMND1	RRM2B	SCO1	SCO2	SDHA	SDHAF1	SDHD	SFXN4	SLC25A4	SLC35G2	SUGCT
	SURF1	TACO1	TK2	TMEM70	TP53	TRAP1	TRAPPC2	TRAPPC5	TRMU	TSFM	TTC19	TTPA
	TUFM	TYMP	UQCC2	UQCC3	UQCR10	UQCR11	UQCRB	UQCRC1	UQCRC2	UQCRFS1	UQCRH	UQCRQ

Disease	**Genes analysed by phenotype**											
	WFS1	YARS2										
Primary Familial Basal Ganglia Calcification (PFBGC)	PDGFB	PDGFRB	SLC20A2	XPR1	SLC20A2							
Spinal Muscular Atrophy (SMA)	AARS	AR	ASAH1	ATP7A	BICD2	BSCL2	C1QB	CHMP2B	CKB	CKM	CKMT1A/CKMT1B	CKMT2
	DCTN1	DNAJB2	DYNC1H1	DYSF	ETV1	FBLN5	FBXO38	GARS	GOT1	GOT1L1	GOT2	GPT
	GPT2	HINT1	HSPB1	HSPB2	HSPB3	HSPB7	HSPB8	IGHMBP2	LMNA	MFN2	PLEKHG5	POLG
	REEP1	RRM1	RRM2	RRM2B	SETX	SLC52A3	SLC5A7	SMAD2	SMAD3	SMN1/SMN2	TARDBP	TRPV4
	UBA1	VAPB										
Huntington's disease phenocopy	JPH3	PRNP	TBP	XK	PARK2	VPS13A						
Chorea-acanthocytosis	XK	PARK2	JPH3	VPS13A								
Adult onset Gangliosidosis	CFTR	GLB1	GM2A	HEXA	HEXB	PSAP	S1PR3	SPHK1	UGCG			
Kuf's disease	CLN5	CLN6	CTSF	DNAJC5	GRN	PPT1						

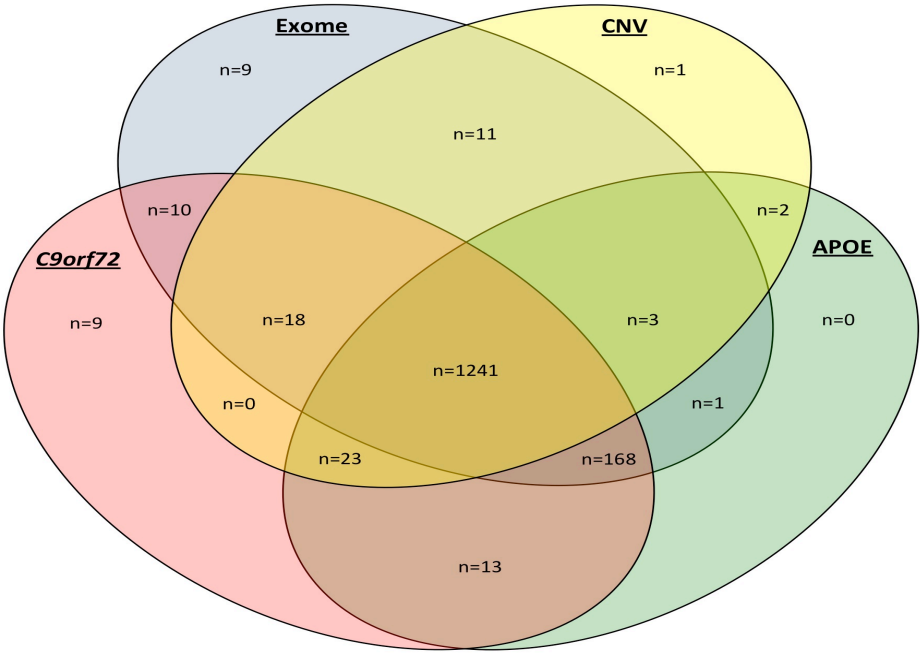
Appendix 5 Mean coverage of CCDS regions within genes.

Mean coverage of CCDS regions within genes implicated as risk factors or causing familial forms of neurodegenerative disease

Gene	Mean sequencing depth of each gene (n=1461 samples)	Standard deviation of sequencing depth	Percentage of samples with < 10 fold mean sequencing depth
ALS2	38.8	9.9	0
ANG	27.7	8.9	0.1
APOE	39.6	12.8	0
APP	44.1	11.4	0
C9orf72	38	9.9	0
CHCHD10	19.8	7.6	2.4
CHMP2B	50.3	16.4	0
COQ2	54.1	13.8	0
DAO	51.1	14.1	0
DCTN1	67.4	17	0
DPP6	60.2	15.1	0
EIF4G1	59	14.8	0
FBXO7	42.9	10.8	0
FUS	43.6	11.1	0
GIGYF2	49.8	12.8	0
GBA	42.4	11.5	0
GRN	73.6	20.8	0
HNRNPA1	31.6	8.3	0.1
HNRNPA2B1	72.3	19.4	0
HTRA2	39.9	10.8	0
LRRK2	43.7	12.5	0
MAPT	32.2	12	1.3
MATR3	44.4	12.2	0
Notch3	43.2	13.6	0
OPTN	65.1	16.7	0
PARK2	59.1	15	0
PARK7	35.1	9.9	0
PFN1	44.5	12.1	0
PINK1	52.3	13.9	0
PON1	44.1	11.9	0
PON3	42.5	11.5	0
PRNP	68.2	18.4	0
PRPH	73.2	20.2	0
PSEN1	48.7	12.9	0
PSEN2	57.9	15.9	0
SETX	43.4	10.9	0
SIGMAR1	44.4	12.4	0
SCARB2	47	10.9	0
SNCA	63.5	18.7	0
SOD1	54.3	14.2	0
SPG11	39.4	10.3	0
VAPB	50	7.4	0
SQSTM1	59.6	16.7	0
TARDBP	39	10.6	0
TREM2	55	16.3	0
UBQLN2	25.9	11.1	0.9
UCHL1	46	12.6	0
VCP	52.2	13.5	0
VPS35	38.1	10.8	0
Mean sequencing depth	**49.1**		**0.1**
Standard deviation	**13.8**		
Total percentage of samples in which mean was below 10 fold in any gene			**4.8**

Appendix 6 A Venn diagram of all the sequencing modalities utilized in this study

The total case cohort was 1511 cases. Overlapping ovals represent the number of samples covered by each of those modalities. In total, 1241 samples were sequenced on all 4 platforms.



Appendix 7 Pathogenic heterozygous mutations – MacArthur criteria.

All relevant data and references are provided to support pathogenicity. Minor Allele Frequencies (MAF) from the 1000Genomes (1000G), NHLBI 6500 ESP and ExAC databases are provided. Key: MND – Motor Neuron Disease, CJD – Creutzfelt-Jakob Disease, AD – Alzheimer's disease, CADASIL – Cerebral Autosomal-Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy. U – Unknown.

Variant		Genetic					Informatic		Experimental			Study cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
SOD1	p.I114T	Described within in several families with MND (Rosen, Siddique et al. 1993)	Segregation full (Rosen, Siddique et al. 1993)	NS	NS	NS	0.00003334	Located in exon 4. Pathogenic variants can lie throughout the polypeptide (Cleveland and Rothstein 2001)	Reduced axonal outgrowth and reduced cell survival (Karumbayaram, Kelly et al. 2009) and moderate aggregate formation (Prudencio, Hart et al. 2009)	Moderate degrees of protein aggregation seen in cell culture (Prudencio, Hart et al. 2009)	U	3	0
SOD1	p.D102N	Seen in several families with MND (Prudencio, Hart et al. 2009)	Segregates with disease (Prudencio, Hart et al. 2009)	NS	NS	NS	0.000002818	Located in exon 4. Pathogenic variants can lie throughout the polypeptide (Cleveland and Rothstein 2001)	Able to rapidly induce protein aggregates in comparison to WT or other mutant SOD1 variants (Ayers, Lelie et al. 2014)	Induces intracellular aggregates in vitro (Ayers, Lelie et al. 2014)	U	2	0
SOD1	p.E101G	Seen in several families with MND (Rosen et al. 1993)	Segregates with disease (Rosen et al. 1993)	NS	NS	NS		Located in exon 4. Pathogenic variants can lie throughout the polypeptide (Cleveland and Rothstein 2001)	Induces protein aggregation in vitro (Prudencio, Hart et al. 2009), and is also recognized by a monoclonal antibody to pathogenic variants (Fujisawa, Homma et al. 2012)	Induces intracellular aggregates in vitro (Ayers, Lelie et al. 2014)	U	1	0

Variant		Genetic					Informatic		Experimental			Study cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
PSEN1	p.R265H; p.R269H	Described in several families with AD (Gomez-Isla et al. 1997; Larner et al. 2007) (Gomez-Isla, Wasco et al. 1997, Larner, Ray et al. 2007)	Segregates with disease (Gomez-Isla, Wasco et al. 1997)	NS	NS	0	0.000025 12	Situated within TM-V domain in which pathogenic mutations are known to reside	Alters interaction with actin-binding proteins (Zhang, Han et al. 1998) and enhances deposition of A-bet x-42/43 (Gomez-Isla, Growdon et al. 1999)	Unknown	U	1	0
PSEN1	p.A246E; p.A242E	Described in several cases and families with AD (Sherrington, Rogaev et al. 1995)	Segregates with disease	NS	NS	NS	0.000034 51	Situated within TM-IV	Abolishes ER calcium leak(Nelson, Tu et al. 2007), increases BACE1 activity (Giliberto, Borghi et al. 2009) and elevates levels if the 42 residue beta-amyloid protein (Jankowsky, Fadale et al. 2004)	Transgenic mouse with the develops AD pathology (Borchelt, Ratovitski et al. 1997)	U	1	0
PSEN1	p.F237L; p.F233L	Seen in one sporadic case and one family previously (Sodeyama, Iwata et al. 2001, Janssen, Beck et al. 2003)	Segregated across 2 generations and 3 patients (Janssen, Beck et al. 2003)	NS	NS	NS	0.000007 762	Situated in TM-V domain in which several pathogenic mutations previously described	Unknown	Unknown	U	1	0
PSEN1	p.G202D; p.G206D	Seen in previous families segregating with disease (Wu, Cheng et al. 2011)	Segregates with disease (Wu, Cheng et al. 2011)	NS	NS	NS	7.30E-07	Situated in TM-IV domain in which several pathogenic mutations previously described	Increases AB42 production (Chen, Hsieh et al. 2014)	Unknown	U	1	0

Variant		Genetic					Informatic		Experimental			Study cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
PSEN1	p.S170F; p.S166F	Seen in several previous families with AD (Piccini, Zanusso et al. 2007) Piccini et al. 2007) and DLB (Snider, Norton et al. 2005)	Segregates with disease (Snider, Norton et al. 2005)	NS	NS	NS	0.000597	Within the TM-III domain	Increase AB42/40 ratio (Giliberto, Borghi et al. 2009)	Increase AB42/40 ratio (Giliberto, Borghi et al. 2009)	U	1	0
PSEN1	p.L153V; p.L149V	Seen previously in several cases and no controls (Raux, Gantier et al. 2000, Cornejo-Olivas, Yu et al. 2014)	Segregates with disease (Raux, Gantier et al. 2000)	NS	NS	NS	0.006138	Situated in TM-II domain in which several pathogenic mutations previously described	Unknown	Unknown	U	1	0
PSEN1	p.M135I; p.M139I	Seen previously in one family (Kim, Kim et al. 2010)	No data for segregation	NS	NS	NS	0.0001109	Situated in TM-II domain in which several pathogenic mutations previously described	Increases AB42 levels in vitro (Murayama, Tomita et al. 1999)	Unknown	U	1	0
PSEN1	p.P436S; p.P432S	Seen in previously in cases (Palmer, Beck et al. 1999)	Segregated with disease (Palmer, Beck et al. 1999)	NS	NS	NS	4.19E-07	Situated in TM-IX domain	Significant reduction in gamma secretase activity (Heilig, Xia et al. 2010)	Unknown	U	1	0
PRNP	p.E211Q	Previously seen in at least one family of gCJD (Peoc'h, Manivet et al. 2000)	Fully penetrant and segregates (Peoc'h, Manivet et al. 2000)	NS	NS	NS	0.0005309	Located within the third alpha-helix of PrP	Significantly alters primary structure and stability and aggregation propensity (Peoc'h, Levavasseur et al. 2012)	Yes	U	1	0

Variant		Genetic					Informatic		Experimental			Study cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
PRNP	p.E200K	Fully penetrant cause of gCJD (Spudich, Mastrianni et al. 1995)	Fully penetrant (Spudich, Mastrianni et al. 1995)	NS	NS	NS	0.0007031	Sits within known region of pathogenicity and results in a deamination of a CpG dinucleotide	Significantly alters protein confirmation consistently (Peoc'h, Levavasseur et al. 2012)	Fatal phenotype and recapitulation in mice (Friedman-Levi, Meiner et al. 2011)	U	7	0
NOTCH3	p.R153C	Described previously in several families (Mandellos, Limbitaki et al. 2005)	Segregates with disease (Mandellos, Limbitaki et al. 2005)	NS	NS	NS	0.0004887	Present in EGF domain and affects cysteine residue	Alters protein folding of NOTCH3 (Dichgans, Ludwig et al. 2000)	Unknown	U	1	0
HEXA	p.R499C	Previously described in several cases (Mules, Hayflick et al. 1992)((Tanaka, Hoang et al. 2003)	Segregated in previous studies (Mules, Hayflick et al. 1992)	NS	NS	NS	8.97E-03	Located in the α-subunit of the resulting in decreased solubility and aggregation (Paw, Moskowitz et al. 1990)	Unknown	Unknown	U	1	0
GRN	p.C139R	Seen in 2 cases and 0 controls (n=200) (Bernardi, Tomaino et al. 2009) and 3 cases and no controls (n=459) (Brouwers, Sleegers et al. 2008)	No data for segregation	NS	NS	NS	0.0001828	Has a destabilizing effect on the granulin-fold by disrupting one of the cysteine disulfide bridges	Reduced plasma granulin levels suggesting a partial loss of function (Finch, Baker et al. 2009)	Unknown	U	1	0

Variant		Genetic					Informatic		Experimental			Study cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
GRN	p.Q130fs*125	Cited in several papers (Baker, Mackenzie et al. 2006, Le Ber, van der Zee et al. 2007, Yu, Bird et al. 2010)	Segregates with disease in several reports	NS	NS	NS		Located in exon 5 and GrafF domain	Reduces plasma progranulin levels (Carecchio, Fenoglio et al. 2009)	Unknown	U	1	0
GRN	p.C31fs*35	Seen in at least 4 families (Beck, Rohrer et al. 2008)	Segregates with families (Beck, Rohrer et al. 2008)	NS	NS	NS	0.00134	Located in exon 2 and paraGran domain	Reduces gross neural connectivity and results in similar synaptic vesicle phenotype tp that seen in vivo (Tapia, Milnerwood et al. 2011)	Unknown	U	1	0
FUS	p.R517C; p.R521C; p.R520C	Variant at same position segregates with disease in several families (Yan, Deng et al. 2010)	Variant at same position segregates with disease in several families (Vance, Rogelj et al. 2009, Yan, Deng et al. 2010)	NS	NS	NS		Located at the C-terminal region of the protein	Alters ratio of soluble to insoluble FUS (Vance, Rogelj et al. 2009)	Induces axonal defects in vitro (Groen, Fumoto et al. 2013)	U	1	0
DNAJC5	p.L115R	Seen previously in numerous cases (Benitez, Alvarado et al. 2011, Cadieux-Dion, Andermann et al. 2013)	Has previously been shown to segregate in families	NS	NS	NS	1.38E-05	Occurs in the cysteine-string domain; post-translationally modified by extensive palmitoylation (Greaves, Lemonidis et al. 2012)	Occurs in the cysteine-string domain; post-translationally modified by extensive palmitoylation (Greaves, Lemonidis et al. 2012)	Forms mutant aggregates in vitro (Greaves, Lemonidis et al. 2012)	U	1	0

Appendix 8 Pathogenic heterozygous mutations – ACMG criteria.

The gene and protein alteration are shown together with whether there is evidence to support strong pathogenicity (PS), moderate pathogenicity (PM) or supporting pathogenicity (PP) according to ACMG 2015 criteria (Richards, Aziz et al. 2015). In addition, evidence supporting a benign nature of the variant is also provided (BS and BP). Final ACMG classification based on both the pre-study and post-study data are shown.

		Pathogenic																Benign												ACMG classification									
Variant		Very Strong	Strong				Moderate						Supporting					Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
SOD1	p.I114T		1		1		1	1					1	1	1																1					1				
SOD1	p.D102N		1		1		1	1					1	1	1																1					1				
SOD1	p.E101G		1		1		1	1					1	1	1																1					1				
PSEN1	p.R265H; p.R269H		1		1		1	1																							1					1				
PSEN1	p.A246E; p.A242E		1		1		1	1					1	1	1																1					1				
PSEN1	p.F237L; p.F233L		1				1	1	1																						1					1				
PSEN1	p.G202D; p.G206D		1				1	1						1	1																1					1				
PSEN1	p.S170F; p.S166F		1		1		1	1					1	1	1																1					1				

PSEN1	p.L153V; p.L149V		1				1	1					1	1	1																1					1					
PSEN1	p.M135I; p.M139I		1		1			1						1																	1					1					
PSEN1	p.P436S; p.P432S		1		1		1	1																							1					1					
PRNP	p.E211Q		1		1		1	1					1	1	1																1					1					
PRNP	p.E200K		1			1	1	1						1	1																1					1					
NOTCH 3	p.R153C	1	1				1	1					1	1																	1					1					
HEXA	p.R499C		1		1		1						1	1		1															1					1					
GRN	p.C139R				1	1		1						1	1																1					1					
GRN	p.Q130fs*125		1		1		1	1																							1					1					
GRN	p.C31fs*35		1				1	1						1	1																1					1					
FUS	p.R517C; p.R521C; p.R520C		1		1	1	1	1					1																		1					1					
DNAJC5	p.L115R		1		1		1						1	1		1															1					1					

Appendix 9 Clinical data - Pathogenic heterozygous mutations.

All relevant clinical data are shown for all variant ordered by ACMG pathogenicity category in addition to in-silico prediction by SIFT (Kumar, Henikoff et al. 2009) and PolyPhen2 (Adzhubei, Schmidt et al. 2010). Key: MND – Motor Neuron Disease, CJD – Creutzfelt-Jakob Disease, AD – Alzheimer's disease, CADASIL – Cerebral Autosomal-Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy, U – Unknown, NS – Not significant.

Case number	Sex	Age onset (years)	Age death (years)	Family History	Clinical Diagnosis	Neuropath Diagnosis	Chromosome	Position	Gene	Transcript	Protein Change	1000G MAF (%)	ESP6500 MAF (%)	ExAC MAF (%)	SIFT	PolyPhen2
1	Female		73	N	MND	MND	21	33039672	SOD1	c.341T>C	p.I114T	NS	NS	NS	Damaging	Probably Damaging
2	Female		71	Y	MND	MND	21	33039672	SOD1	c.341T>C	p.I114T	NS	NS	NS	Damaging	Probably Damaging
3	Unknown		78	N	MND	MND	21	33039672	SOD1	c.341T>C	p.I114T	NS	NS	NS	Damaging	Probably Damaging
4	Female		43	N	MND	MND	21	33039635	SOD1	c.304G>A	p.D102N	NS	NS	NS	Damaging	Probably Damaging
5	Female		42	Y	MND	MND	21	33039635	SOD1	c.304G>A	p.D102N	NS	NS	NS	Damaging	Probably Damaging
6	Female	36	48	N	MND	MND	21	33039633	SOD1	c.302A>G	p.E101G	NS	NS	NS	Tolerated	Benign
7	Female	60	72	N	AD	AD	14	73664775	PSEN1	c.806G>A; c.794G>A	p.R265H; p.R269H	NS	NS	0	Damaging	Probably Damaging
8	Female	65	77	N	AD	AD	14	73659540	PSEN1	c.725C>A; c.737C>A	p.A246E; p.A242E	NS	NS	NS	Damaging	Possibly Damaging
9	Male		38	N	CJD	AD	14	73653589	PSEN1	c.509C>T; c.497C>T	p.S170F; p.S166F	NS	NS	NS	Damaging	Probably Damaging
10	Female		49	Y	Dementia	AD	14	73640392	PSEN1	c.457C>G; c.445C>G	p.L153V; p.L149V	NS	NS	NS	Damaging	Probably Damaging
11	Male		47	Y	Dementia	AD	14	73640352	PSEN1	c.417G>T; c.405G>T	p.M135I; p.M139I	NS	NS	NS	Tolerated	Benign
12	Male		52	Y	Dementia	AD	14	73685899	PSEN1	c.1294C>T; c.1306C>T	p.P436S; p.P432S	NS	NS	NS	Damaging	Probably Damaging

Case number	Sex	Age onset (years)	Age death (years)	Family History	Clinical Diagnosis	Neuropath Diagnosis	Chromosome	Position	Gene	Transcript	Protein Change	1000G MAF (%)	ESP6500 MAF (%)	ExAC MAF (%)	SIFT	PolyPhen2
13	Female		49	N	AD	AD	14	73659512	PSEN1	c.697T>C; c.709T>C	p.F237L; p.F233L	NS	NS	NS	Tolerated	Possibly Damaging
14	Male		43	N	AD	AD	14	73659420	PSEN1	c.605G>A; c.617G>A	p.G202D; p.G206D	NS	NS	NS	Damaging	Probably Damaging
15	Male	53	53	N	CJD	CJD	20	4680497	PRNP	c.631G>C; c.*320G>C	p.E211Q	NS	NS	NS	Tolerated	Possibly Damaging
16	Female	53	53	N	CJD	CJD	20	4680464	PRNP	c.598G>A; c.*287G>A	p.E200K	NS	NS	NS	Damaging	Probably Damaging
17	Female	42	42	Y	CJD	CJD	20	4680464	PRNP	c.598G>A; c.*287G>A	p.E200K	NS	NS	NS	Damaging	Probably Damaging
18	Female	55	56	Y	CJD	CJD	20	4680464	PRNP	c.598G>A; c.*287G>A	p.E200K	NS	NS	NS	Damaging	Probably Damaging
19	Female	54	55	N	CJD	CJD	20	4680464	PRNP	c.598G>A; c.*287G>A	p.E200K	NS	NS	NS	Damaging	Probably Damaging
20	Male	66	66	N	CJD	CJD	20	4680464	PRNP	c.598G>A; c.*287G>A	p.E200K	NS	NS	NS	Damaging	Probably Damaging
21	Male	51	51	Y	CJD	CJD	20	4680464	PRNP	c.598G>A; c.*287G>A	p.E200K	NS	NS	NS	Damaging	Probably Damaging
22	Male	60	61	N	CJD	CJD	20	4680464	PRNP	c.598G>A; c.*287G>A	p.E200K	NS	NS	NS	Damaging	Probably Damaging
23	Male	40	62	Y	CADASIL	CADASIL	19	15302993	NOTCH3	c.457C>T	p.R153C	NS	NS	NS	Damaging	Possibly Damaging
24	Male	65	71	N	Psychiatric	Adult onset gangliosidosis	15	72637818	HEXA		p.R499C	NS	NS	NS	Damaging	Probably Damaging
25	Female	56	57	N	CJD	CJD	17	42427661	GRN	c.415T>C	p.C139R	NS	NS	0.02	Damaging	Probably Damaging
26	Female		72	N	Dementia	FTD	17	42427632	GRN	c.386_389delGTCA	p.Q130fs*125	NS	NS	NS		
27	Female	55	60	Y	AD	FTD	17	42426621	GRN	c.86_89dupCCTG	p.C31fs*35	NS	NS	NS		
28	Female		61	Y	MND	MND	16	31202739	FUS	c.1561C>T; c.1558C>T; c.1549C>T	p.R517C; p.R521C; p.R520C	NS	NS	NS	Damaging	Benign
29	Male	43	58	N	Kuf's disease	Kuf's disease	20	62562226	DNAJC5		p.L115R	NS	NS	NS	Tolerated	Benign

Appendix 10 Likely Pathogenic heterozygous mutations – MacArthur criteria.

All relevant data and references are provided to support pathogenicity. Minor Allele Frequencies (MAF) from the 1000Genomes (1000G), NHLBI 6500 ESP and ExAC databases are provided. Key: MND – Motor Neuron Disease, CJD – Creutzfelt-Jakob Disease, AD – Alzheimer's disease, CADASIL – Cerebral Autosomal-Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy. U – Unknown.

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
SQSTM1	p.A427T; p.A343T	Not previously reported	Unknown	NS	NS	0	0.0001062	Situated within the UBA domain in which pathogenic mutations have previously been identified.	Mutations at codon 425 result in loss of polyubiquitin-binding and increased activation of NF-KB (Hocking, Lucas et al. 2004)	Unknown	U	1	0
PSEN2	p.G70W	Novel	Unknown	NS	NS	NS	0.002951	Located in N-terminal domain	Unknown	Unknown	U	1	0
PSEN2	p.L204I	Novel	Unknown	NS	NS	NS	0.005284	Located in the 4th TM domain	Unknown	Unknown	U	1	0
PSEN2	p.V101M	Novel	Unknown	NS	0.01	NS	8.22E-07	Located in TM-I domain in which several pathogenic mutations described	Unknown	Unknown	U	1	0
PSEN2	p.D439A; p.D438A	Described in several cases of disease, but also seen in low-frequency in controls (Lleo, Blesa et al. 2002, Sassi, Guerreiro et al. 2014)	Possible incomplete penetrance (Sassi, Guerreiro et al. 2014)	NS	0.02	0	0.000006053	Outside TM domain, but within the C-terminal region which is involved in X-terminal trafficking	Unknown	Doesn't alter Ab40/42 ratio (Walker, Martinez et al. 2005)	U	2	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
PSEN1	p.I110N	Novel	Unknown	NS	NS	NS	0.000025	Located in exon 4 - HL-I domain	Unknown	Unknown	U	1	0
HNRNPA1	p.G316R; p.G264R	Novel	Unknown	NS	NS	NS	0.000008091	Situated 42 bases away from a previously described variant (Kim, Kim et al. 2013)	Unknown	Unknown	U	1	0
GRN	P.V452Wfs *39	Described previously in a single case (Chen-Plotkin, Martinez-Lage et al. 2011)	Unknown	NS	NS	NS	N/A	Situated in exon 11, one amino acid away from another previously described possible pathogenic mutation (Brouwers, Sleegers et al. 2008)	Unknown	Unknown	U	1	0
CHMP2B	p.R69Q; p.R28Q	Described previously in a case of PMA (van Blitterswijk, Vlam et al. 2012)	Unknown	NS	0.02	0	0.000003802	Located in well conserved N-terminal region	Unknown	Unknown	U	1	0
APP	p.M698I; p.M612I;	Novel	Unknown	NS	NS	NS	0.000001854	Sits within TM-I region in which pathogenic variants have previously been described.	Unknown	Unknown	U	1	0

Appendix 11 Likely Pathogenic heterozygous mutations – ACMG criteria.

The gene and protein alteration are shown together with whether there is evidence to support strong pathogenicity (PS), moderate pathogenicity (PM) or supporting pathogenicity (PP) according to ACMG 2015 criteria (Richards, Aziz et al. 2015). In addition, evidence supporting a benign nature of the variant is also provided (BS and BP). Final ACMG classification based on both the pre-study and post-study data are shown.

		Pathogenic															Benign													ACMG classification									
Variant		Very Strong	Strong				Moderate						Supporting					Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
SQSTM1	p.A427T; p.A343T							1	1						1	1																1					1				
PSEN2	p.G70W								1						1	1																1					1				
PSEN2	p.L204I							1	1						1	1																1					1				
PSEN2	p.V101M						1	1							1	1	1															1					1				
PSEN2	p.D439A; p.D438A			1										1	1		1				1											1					1				
PSEN1	p.I110N						1	1						1	1																	1					1				
HNRNPA1	p.G316R; p.G264R							1	1						1	1																1					1				
GRN	P.V452Wfs*39			1				1	1						1																	1					1				
CHMP2B	p.R69Q; p.R28Q		1					1							1	1	1	1														1					1				
APP	p.M698I; p.M612I;							1	1						1	1																1					1				

Appendix 12 Clinical data – Likely Pathogenic heterozygous mutations.

Case number	Sex	Age onset	Age death	FH	Clinical Dx	Neuropath Dx	Chromosome	Position	Gene	Transcript	Protein Change	1000G MAF (%)	ESP6500 MAF (%)	ExAC MAF (%)	SIFT	PolyPhen2
30	Male	55	61	N	MND	MND	5	179263549	SQSTM1	c.1027G>A; c.1279G>A	p.A427T; p.A343T	NS	NS	0	Damaging	Probably Damaging
31	Female	81	82	N	Stroke with dementia	Vascular disease / AD	1	227071472	PSEN2	c.208G>T	p.G70W	NS	NS	NS	Damaging	Benign
32	Female		70	N	Dementia	AD	1	227076573	PSEN2	c.610C>A	p.L204I	NS	NS	NS	Tolerated	Benign
33	Female		93	N	AD	AD	1	227071565	PSEN2	c.301G>A	p.V101M	NS	0.01	NS	Damaging	Probably Damaging
34	Female		81	N	AD	AD	1	227083249	PSEN2	c.1313A>C; c.1316A>C	p.D439A; p.D438A	NS	0.02	0	Damaging	Possibly Damaging
35	Male		86	N	Control	High BRAAK (3/4)	1	227083249	PSEN2	c.1313A>C; c.1316A>C	p.D439A; p.D438A	NS	0.02	0	Damaging	Possibly Damaging
36	Female	38	44	Y	AD	AD	14	73640276	PSEN1	c.341T>A	p.I110N	NS	NS	NS	Damaging	Probably Damaging
37	Male		62	N	MND	MND	12	54677634	HNRNPA1;	c.790G>C; c.946G>C	p.G316R; p.G264R	NS	NS	NS	Tolerated	Probably Damaging
38	Male	65	69	N	Dementia	FTD	17	42429557	GRN	c.1354delG	p.V452fs*39	NS	NS	NS		
39	Female	64	71	N	Dementia	FTD	3	87294943	CHMP2B	c.206G>A; c.83G>A	p.R69Q; p.R28Q	NS	0.02	0	Damaging	Probably Damaging
40	Female	71	77	N	AD	AD	21	27264079	APP	c.1773G>T; c.1836G>T;	p.M698I; p.M612I;	NS	NS	NS	Damaging	Possibly Damaging

Appendix 13 Likely Benign mutations – MacArthur criteria.

All relevant data and references are provided to support pathogenicity. Minor Allele Frequencies (MAF) from the 1000Genomes (1000G), NHLBI 6500 ESP and ExAC databases are provided. Key: MND – Motor Neuron Disease, CJD – Creutzfelt-Jakob Disease, AD – Alzheimer's disease, CADASIL – Cerebral Autosomal-Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy. U – Unknown.

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
PSEN2	p.S30F	Novel	U	NS	NS	0	0.0003041	Located in N-terminal domain	Unknown	Unknown	Unknown	1	0
PSEN2	p.G37V	Novel	U	NS	NS	NS		Located in N-terminal domain	Unknown	Unknown	Unknown	1	2
PSEN2	p.R62C	Same amino acid change seen in 9% of Africans suggesting it is unlikely to be pathogenic (Guerreiro, Baquero et al. 2010), however they also have an earlier age of disease onset suggesting they may be disease modifiers (Cruchaga, Haller et al. 2012)	U	NS	0.01	0.02		Located in N-terminal domain	Unknown	Unknown	Unknown	2	0
PSEN2	p.V64F	Novel	U	NS	NS	NS		Located in N-terminal domain	Unknown	Unknown	Unknown	1	0
PSEN2	p.C65F	Novel	U	NS	NS	NS		Located in N-terminal domain	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
PSEN2	p.R110H	Not previously associated with disease	U	NS	NS	0	0.00005508	Located in the topological domain prior to a transmembrane region	Unknown	Unknown	Unknown	1	0
PSEN2	p.S130L	Previously described in cases, but also seen in controls (Sassi, Guerreiro et al. 2014)	U	0.1	0.1	0.06	0.00006714	Located in the topological domain prior to a transmembrane region	Does not alter AB 40/42 levels (Walker, Martinez et al. 2005)	Unknown	Unknown	5	1
DCTN1	p.V1065I	Not previously seen	U	0.06	NS	0.02		Outside major functional domains	Unknown	Unknown	Unknown	1	0
DCTN1	p.V1074M	Not previously associated with disease	U	NS	NS	0		Outside major functional domains	Unknown	Unknown	Unknown	2	0
DCTN1	p.E1067*	Not previously associated with disease	U	NS	0.01	0		Within the second coiled coil domain	Unknown	Unknown	See previous	1	0
DCTN1	p.R1029Q	Seen in relatively high frequency in international reference databases	U	0.04	0.23	0.1	0.000001679	Within the second coiled coil domain	Mutant protein is not ubiquitinated, but does alter cellular morphology causing thin dynatic filaments (Munch, Sedlmeier et al. 2004)	Mutant protein causes thin dynatic filaments (Munch, Sedlmeier et al. 2004)	Unknown	9	2

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
DCTN1	p.E1026G	Not previously associated with disease	U	NS	0.02	0.01	0.00001702	Within the second coiled coil domain	Unknown	Unknown	Unknown	1	0
DCTN1	p.K851R	Not previously associated with disease	U	NS	NS	0	0.00001884	Outside major functional domains	Unknown	Unknown	Unknown	2	0
DCTN1	p.A803G	Not previously associated with disease	U	NS	0.01	NS	0.00000171	Outside major functional domains	Unknown	Unknown	Unknown	1	0
DCTN1	p.R778W	Detected previously in cases of ALS (Liu, Li et al. 2014) however it has also been seen in affected and unaffected relatives (Munch, Sedlmeier et al. 2004) with a high incidence in population databases	N	NS	0.08	0.02	0.0003228	Outside major functional domains	Does not alter cellular morphology in vitro (Stockmann, Meyer-Ohlendorf et al. 2013)	No (Stockmann, Meyer-Ohlendorf et al. 2013)	Unknown	1	1
DCTN1	p.V697L	Novel	U	NS	NS	NS	0.000001449	Outside major functional domains	Unknown	Unknown	Unknown	1	0
DCTN1	p.S661F	Novel	U	NS	NS	NS	0.00005598	Outside major functional domains	Unknown	Unknown	Unknown	1	0
DCTN1	p.A653V	Novel	U	NS	NS	NS	0.000001832	Outside major functional domains	Unknown	Unknown	Unknown	1	0
DCTN1	p.K617N	Novel	U	NS	NS	NS	0.008128	Outside major functional domains	Unknown	Unknown	Unknown	1	0
DCTN1	p.V599I	Not previously associated with disease	U	NS	0	0		Outside major functional domains	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
DCTN1	p.P512S	Novel	U	NS	NS	NS	0.000113	Within the first coiled coil domain	Unknown	Unknown	Unknown	1	0
DCTN1	p.R465C	Not previously associated with disease	U	NS	NS	0	0.0000618	Within the first coiled coil domain	Unknown	Unknown	Unknown	0	1
DCTN1	p.A354T	Novel	U	NS	NS	NS	0.00006745	Within the first coiled coil domain	Unknown	Unknown	Unknown	1	0
DCTN1	p.R218Q	Novel	U	NS	NS	NS	0.00000113	Within the first coiled coil domain	Unknown	Unknown	Unknown	1	0
HTRA2	p.R173H	Novel	U	NS	NS	NS		Within serine protease domain	Unknown	Unknown	Unknown	1	0
HTRA2	p.P120S	Novel	U	NS	NS	NS	0.00887	Within transmembrane domain	Unknown	Unknown	Unknown	1	0
HTRA2	p.N106H	Not previously associated with disease	U	NS	NS	0	0.00004	Within transmembrane domain	Unknown	Unknown	Unknown	1	0
HTRA2	p.V60L	Not previously associated with disease	U	0.04	0.13	0.13		Between N-terminal mitochondrial targeting domain and transmembrane domain	Unknown	Unknown	Unknown	2	2
HTRA2	p.A32S	Not previously associated with disease	U	NS	0.01	0		Within N-terminal mitochondrial targeting sequence	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
HTRA2	p.L72P	Seen previously in 2 cases of PD	U	0.16	0.3	0.27		Between N-terminal mitochondrial targeting domain and transmembrane domain	Unknown	Unknown	Unknown	2	1
HTRA2	p.T76N	Novel	U	NS	NS	NS		Between N-terminal mitochondrial targeting domain and transmembrane domain	Unknown	Unknown	Unknown	0	1
HTRA2	p.N98S	Novel	U	NS	NS	NS		Between N-terminal mitochondrial targeting domain and transmembrane domain	Unknown	Unknown	Unknown	1	0
GIGYF2	p.R168C	Not previously associated with disease	U	NS	NS	0		Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	0
GIGYF2	p.G208A	Not previously associated with disease	U	NS	NS	0	0.001679	Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	0
GIGYF2	p.I228M	Novel	U	NS	NS	NS		Mis-sense variant. No convincing specific region for pathogenicity in	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
								literature					
GIGYF2	p.G239V	Novel	U	NS	NS	NS	0.000001225	Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	0
GIGYF2	p.D257E	Novel	U	NS	NS	NS		Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	0
GIGYF2	p.D371E	Not previously associated with disease	U	NS	0.13	0.05		Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	1
GIGYF2	p.S396P	Novel	U	NS	NS	NS		Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	0
GIGYF2	p.Q439H	Not previously associated with disease	U	NS	0.02	0.01	0.0008433	Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	2	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
GIGYF2	p.N451T	Seen in 3/498 of cases of PD (Lautier, Goldwurm et al. 2008), but 0 of 727 in another together with one control (Bras, Simon-Sanchez et al. 2009).	U	0.02	0.14	0.04		Mis-sense variant. No convincing specific region for pathogenicity in literature	Does not cause dopaminergic cell loss in Zebra fish (Guella, Pistocchi et al. 2011)	Unknown	Unknown	1	0
GIGYF2	p.T702K	Novel	U	NS	NS	NS	0.000009863	Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	0
GIGYF2	p.R816C	Not previously associated with disease	U	NS	NS	0		Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	0
GIGYF2	p.R852Q	Not previously associated with disease	U	NS	0.01	0.04	0.0008147	Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	0
GIGYF2	p.R912Q	Not previously associated with disease	U	NS	0.01	0	0.0003565	Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	0	1

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
GIGYF2	p.R982Q	Not previously associated with disease	U	0.06	0.14	0.06	0.007834	Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	0
GIGYF2	p.S1056C	Not previously associated with disease	U	NS	0.19	0.08	0.00001469	Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	4	0
GIGYF2	p.E1169D	Novel	U	NS	NS	NS		Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	0
GIGYF2	p.H1171R	Previously seen in a young onset PD case (Ghani, Lang et al. 2015), but also seen frequently in controls.	U	0.08	0.22	0.16	0.00002178	Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	4	0
GIGYF2	p.K1210M	Novel	U	NS	NS	NS	0.00001901	Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	0
GIGYF2	p.L1230P	Not previously associated with disease	U	NS	NS	0.02		Mis-sense variant. No convincing specific region for pathogenicity in	Does not cause dopaminergic cell loss in Zebra fish (Guella,	Unknown	Unknown	54	6

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
								literature	Pistocchi et al. 2011)				
GIGYF2	p.Q1207P	Novel	U	NS	NS	NS		Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	3	2
GIGYF2	p.Q1223_ Q1226del	Novel	U	NS	NS	NS		Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	28	1
CHMP2B	p.R19Q	Not previously associated with disease	U	0.02	NS	0.01	0.006152	In N-terminal domain	Unknown	Unknown	Unknown	1	0
CHMP2B	p.R32Q	Not previously associated with disease	U	NS	0.01	0	0.0001837	Mis-sense variant in well conserved region	Unknown	Unknown	Unknown	1	0
CHMP2B	p.A33V	Novel	U	NS	NS	NS	0.00000456	Mis-sense variant in well conserved region	Unknown	Unknown	Unknown	1	0
EIF4G1	Splice-site alteration (1)	Novel	U	NS	NS	NS		Results in splice site loss	Unknown	Unknown	Unknown	1	0
EIF4G1	p.A63D	Novel	U	NS	NS	NS		Mis-sense variant. Functional effects unknown	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
EIF4G1	Splice-site alteration (2)	Not previously associated with disease	U	NS	0.01	0		Causes alternate splicing	Unknown	Unknown	Unknown	0	1
EIF4G1	p.P100L	Not previously associated with disease	U	NS	NS	0		Mis-sense variant. Functional effects unknown	Unknown	Unknown	Unknown	1	0
EIF4G1	p.I293T	Novel	U	NS	NS	NS		Mis-sense variant. Functional effects unknown	Unknown	Unknown	Unknown	1	0
EIF4G1	p.A265S	Not previously associated with disease	U	NS	0.02	0.01		Mis-sense variant. Functional effects unknown	Unknown	Unknown	Unknown	1	0
EIF4G1	p.G276_A278del	Novel	U	NS	NS	NS		Mis-sense variant. Functional effects unknown	Unknown	Unknown	Unknown	67	10
EIF4G1	p.P399S	Not previously associated with disease	U	0.02	0.06	0.08		Mis-sense variant. Functional effects unknown	Unknown	Unknown	Unknown	4	0
EIF4G1	p.P539L	Not previously associated with disease	U	0.04	NS	0.01		Mis-sense variant. Functional effects unknown	Unknown	Unknown	Unknown	0	1
EIF4G1	p.A510P	Not previously associated with disease	U	0.12	0.19	0.1	0.004416	Mis-sense variant. Functional effects unknown	Unknown	Unknown	Unknown	1	1

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
EIF4G1	p.S489L	Novel	U	NS	NS	NS	0.004074	Mis-sense variant. Functional effects unknown	Unknown	Unknown	Unknown	1	0
EIF4G1	p.G490C	Seen in 2 of 251 cases and 0/231 controls (Lesage, Condroyer et al. 2012) and in 2 patients and 0/4000 controls (Chartier-Harlin, Dachsel et al. 2011)	U	NS	NS	0.01	0.001901	Mis-sense variant. Functional effects unknown	Unknown	Unknown	Unknown	2	0
EIF4G1	p.P703T	Novel	U	NS	NS	NS	0.0004786	Some transcripts alter EIF4E binding site	Unknown	Unknown	Unknown	1	0
EIF4G1	p.G534A	Seen in 1 of 975 cases and 0 of 1014 controls (Schulte, Mollenhauer et al. 2012)	U	NS	0.01	0	0.000005572	Mis-sense variant. Functional effects unknown. May alter eIF3/EIF4A binding	Unknown	Unknown	Unknown	2	0
EIF4G1	p.A553P	Not previously associated with disease	U	NS	0.15	0.07		Mis-sense variant. Functional effects unknown. May alter eIF3/EIF4A binding	Unknown	Unknown	Unknown	1	1
EIF4G1	p.I642V	Not previously associated with disease	U	0.02	0.05	0.04	0.000006295	Mis-sense variant. Functional effects unknown. May alter eIF3/EIF4A binding	Unknown	Unknown	Unknown	2	1
EIF4G1	p.T634S	Not previously associated with disease	U	NS	0.02	0	0.0002178	Mis-sense variant. Functional effects unknown. May alter eIF3/EIF4A binding	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
EIF4G1	p.E936*	Novel	U	NS	NS	NS	6.82E-07	Stop-codon insertion	Unknown	Unknown	Unknown	1	0
EIF4G1	p.R1074C	Not previously associated with disease	U	NS	0	0	0.000003873	Mis-sense variant. Functional effects unknown. May alter eIF3/EIF4A binding	Unknown	Unknown	Unknown	1	0
EIF4G1	p.R1099C	Novel	U	NS	NS	0.01	0.0004064	Mis-sense variant in well conserved region	Unknown	Unknown	Unknown	2	0
EIF4G1	p.L1312V	Novel	U	NS	NS	0		Mis-sense variant in well conserved region	Unknown	Unknown	Unknown	1	0
EIF4G1	p.M1336V	Not previously associated with disease	U	0.1	0.02	0.02	0.00001702	Mis-sense variant in well conserved region	Unknown	Unknown	Unknown	2	0
EIF4G1	p.I1493V	Novel	U	NS	NS	NS		Mis-sense variant in well conserved region	Unknown	Unknown	Unknown	1	0
EIF4G1	Splice-site alteration (3)	Novel	U	0	0	0	0.002133	Causes alternate splicing	Unknown	Unknown	Unknown	1	0
EIF4G1	p.A1549S	Novel	U	NS	NS	NS		Mis-sense variant in well conserved region	Unknown	Unknown	Unknown	1	0
COQ2	p.N417S	Novel	U	NS	NS	0.01		Mis-sense variant in well conserved region	Unknown	Unknown	Unknown	0	1
COQ2	p.Y353C	Novel	U	NS	NS	0	0.0005848	Mis-sense variant in well conserved region	Unknown	Unknown	Unknown	0	1

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
COQ2	p.A201T	Not previously associated with disease	U	NS	0	0	0.0001995	Situated in the (PT_UbiA_1) domain.	Unknown	Unknown	Unknown	1	0
COQ2	p.R197H	Only previously seen in childhood onset q10 deficiency (Lopez, Quinzii et al. 2010)and CoQ2 nephropathy (Diomedi-Camassei, Di Giandomenico et al. 2007)	N	NS	NS	0	0.000001175	Located in a critical and well-established functional domain (PT_UbiA_1)	Unknown	Unknown	Unknown	1	0
COQ2	p.S107T	Not previously associated with disease	U	0.02	NS	0.12	0.00007907	Mis-sense variant. No convincing specific region for pathogenicity in literature for heterozygous cases	Unknown	Unknown	Unknown	2	1
COQ2	p.Q105H	Not previously associated with disease	U	NS	NS	0.05		Mis-sense variant. No convincing specific region for pathogenicity in literature for heterozygous cases	Unknown	Unknown	Unknown	0	1
COQ2	p.P96S	Not previously associated with disease	U	NS	NS	0		Mis-sense variant. No convincing specific region for pathogenicity in literature for heterozygous cases	Unknown	Unknown	Unknown	3	1

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
COQ2	p.S54W	Not previously associated with disease	U	NS	0.01	0.14		Mis-sense variant. No convincing specific region for pathogenicity in literature for heterozygous cases	Unknown	Unknown	Unknown	1	0
COQ2	p.G53R	Novel	U	NS	NS	NS		Well conserved region	Unknown	Unknown	Unknown	1	0
MATR3	p.Y214C	Novel	U	NS	NS	NS		Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	0	1
MATR3	p.E262*	Novel	U	NS	NS	NS	0.0001377	Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	0
MATR3	p.V159M	Novel	U	NS	NS	NS	0.00005321	Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	0
MATR3	p.A762T	Not previously associated with disease	U	NS	0.06	0.02		Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
MATR3	p.R503C	Not previously associated with disease	U	NS	NS	0.01		Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	3	1
SQSTM1	p.S24G	Novel	U	NS	NS	NS		Mis-sense variant in PB1 domain	Unknown	Unknown	Unknown	0	1
SQSTM1	p.A33V	Not previously associated with disease	U	NS	0.08	0.12		Mis-sense variant in PB1 domain	Unknown	Unknown	Unknown	2	2
SQSTM1	p.P111L	Not previously associated with disease	U	NS	0.01	0		Mis-sense variant in PB1 domain	Unknown	Unknown	Unknown	1	0
SQSTM1	p.A33V	Previously described in 1 case of FALS and 2 cases of SALS (Fecto, Yan et al. 2011) but also seen in controls in relatively high frequency	U	0.08	0.12	0.15		Mis-sense variant in SH2 binding domain	Unknown	Unknown	Unknown	2	1
SQSTM1	p.P118S	Not previously associated with disease	U	NS	0.01	0.02	0.00006531	Mis-sense variant in PB1 domain	Unknown	Unknown	Unknown	4	0
SQSTM1	p.P50L	Not previously associated with disease	U	NS	0.01	0	0.000001426	Mis-sense variant in PB1 domain	Unknown	Unknown	Unknown	1	0
SQSTM1	p.V150G	Novel	U	NS	NS	NS	0.001365	Sits in TRAF6 binding domain	Unknown	Unknown	Unknown	1	0
SQSTM1	p.K238E	Seen in both cases of ALS and controls in equal measure (van der Zee, Van	U	0.24	0.35	0.24		Sits in TRAF6 binding domain	Unknown	Unknown	Unknown	17	1

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
		Langenhove et al. 2014)											
SQSTM1	p.G168V	Novel	U	NS	NS	NS	0.000005383	Sits in TRAF6 binding domain	Unknown	Unknown	Unknown	0	1
SQSTM1	p.P392L	Seen in 1 of 187 cases of FTD and 0 of 539 controls (Le Ber, Camuzat et al. 2013).Also detected in 0.97% of cases of ALS (Kwok, Morris et al. 2013), Known RF for Paget's disease of bone.	Y – single study (van der Zee, Van Langenhove et al. 2014)	0.24	0.21	0.09	0.000001854	Located in the Ubiquitin-associated domain	Osteoclasts harbouring this mutation show sensitivity to NF-kb (Kurihara, Hiruma et al. 2007)	Unknown	Unknown	9	0
HNRNPA2B1	p.G323del	Novel	U	NS	NS	NS	0.000001191	This sits within the nuclear targeting sequence! No previous pathogenic mutations described in this region	Unknown	Unknown	Unknown	1	0
C9orf72	p.M332V	Not previously associated with disease	U	NS	0	0	0.00001549	Unknown functional domain	Unknown	Unknown	Unknown	2	1
C9orf72	p.M318V	Novel	U	NS	NS	0	0.00001549	Unknown functional domain	Unknown	Unknown	Unknown	1	0
C9orf72	p.I138V	Novel	U	NS	NS	NS	0.0004102	Unknown functional domain	Unknown	Unknown	Unknown	1	0
C9orf72	p.T49R	Seen previously in a single case of MND (Couthouis, Raphael et al. 2014) but also	U	0.02	0.02	0.02	0.005458	Unknown functional domain	Unknown	Unknown	Unknown	3	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
		seen in control databases.											
VCP	p.N616fs*12	Novel	U	NS	NS	NS		Contained within the D2 a/b region of which impaired function is unknown	Unknown	Unknown	Unknown	27	4
VCP	p.N616fs*63	Novel	U	NS	NS	NS		Contained within the D2 a/b region of which impaired function is unknown	Unknown	Unknown	Unknown	3	0
VCP	p.D501E	Novel	U	NS	NS	NS	0.006902	Contained with the ATPase binding cassette region	Unknown	Unknown	Unknown	1	0
VCP	p.R453Q	Novel	U	NS	NS	NS	0.00008551	Contained within the helical structural domain of the protein	Unknown	Unknown	Unknown	0	2
VCP	p.R323L	Novel	U	NS	NS	NS	3.77E-07	Contained within the AAA2 region	Unknown	Unknown	Unknown	1	0
VCP	p.I27V	Previously reported in 2 cases of IBMPFD(Mehta, Khare et al. 2013). Also seen in 1/768 cases of PD and 2 of 716 controls (aged in their 5th and 8th decades) (Majounie, Traynor et al. 2012)	U	0.1	0.03	0.05	0.0004227	well-established functional domain (MopB_CT_1)	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
VCP	p.G9V	Novel	U	NS	NS	NS		Contained within the N-terminal region	Unknown	Unknown	Unknown	1	0
HNRNPA1	p.F216L	Novel	U	NS	NS	NS		Situated in close proximity to previously described pathogenic mutations	Unknown	Unknown	Unknown	1	0
DAO	p.R2S	Novel	U	NS	NS	NS		Situated in nucleotide binding domain. Functional impairment unclear	Unknown	Unknown	Unknown	1	1
DAO	p.A8E	Novel	U	NS	NS	NS	0.000004853	Situated in nucleotide binding domain. Functional impairment unclear	Unknown	Unknown	Unknown	1	0
DAO	p.Q63E	Not previously associated with disease	U	NS	0.02	0.01		Functional impairment unclear	Unknown	Unknown	Unknown	1	1
DAO	p.P103P	Not previously associated with disease	U	NS	0.02	0.01		Alters splicing	Unknown	Unknown	Unknown	1	0
DAO	p.R115W	Not previously associated with disease	U	NS	0.01	0.01		Situated in well conserved region	Unknown	Unknown	Unknown	1	0
DAO	p.Y144H	Not previously associated with disease	U	0.3	0	0.08	0.00003133	Situated in well conserved region	Unknown	Unknown	Unknown	2	0
DAO	p.W260*	Novel	U	NS	NS	NS	0.00002851	Involved in enzymatic activity (amino-acids Tyr228-His307)	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
ANG	p.G17V	Novel	U	NS	NS	NS		Situated in n-terminal region	Unknown	Unknown	Unknown	1	0
ANG	p.I70V	Seen in equal numbers of cases and controls (Gellera, Colombrita et al. 2008)	U	0.02	0.06	0.06		Located in a critical and well-established functional domain (RNase_A_canonical)	Reduces ribonucleolytic activity (Crabtree, Thiyagarajan et al. 2007)	Unknown	Unknown	2	1
ANG	p.K78E	Previously described in isolated sporadic cases (Fernandez-Santiago, Hoenig et al. 2009, Kirby, Highley et al. 2013) but also seen in reference databases.	U	NS	0.01	0.02	0.004477	Located in the second helix of the mature ANG protein	Predicted to lose nuclear-translocational activity (Padhi, Vasaikar et al. 2013)	Unknown	Unknown	0	1
ANG	p.N83K	Novel	U	NS	NS	NS		Located in the second helix of the mature ANG protein	Unknown	Unknown	Unknown	1	0
FUS	p.S57del	Seen in 1 of 200 patients and none of 475 controls (Belzil, Daoud et al. 2011)	U	NS	NS	NS	0.0002742	Located in the QGSY-Region. Additional variants in these region are unlikely to be pathogenic (Cruts, Theuns et al. 2012)	Unknown	Unknown	Does rescue the phenotype (Kabashi, Bercier et al. 2011)	3	0
FUS	p.R379H	Not previously associated with disease	U	NS	NS	0	0.0004592	Located in the RRM region	Unknown	Unknown	Unknown	1	0
FUS	p.R483H	Not previously associated with disease	U	0.02	NS	0		Located in the RRM region	Unknown	Unknown	Unknown	0	1

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
VPS35	p.V476A	Not previously associated with disease	U	NS	0.01	0	0.001374	Located in region which interacts with SLC11A2	Unknown	Unknown	Unknown	1	0
VPS35	p.K296E	Novel	U	NS	NS	NS	0.000009795	Located in a critical and well-established functional domain	Unknown	Unknown	Unknown	1	0
VPS35	p.M57L	Not previously associated with disease	U	NS	NS	0	0.00001406	Located in region which interacts with SLC11A2	Unknown	Unknown	Unknown	1	0
VPS35	p.M30K	Unknown	U	NS	0	0.01	0.00001687	Unknown	Unknown	Unknown	Unknown	1	0
PFN1	p.V103F	Novel	U	NS	NS	NS		Located in a region which involved in a beta strand in the secondary protein	Unknown	Unknown	Unknown	1	0
PFN1	Splice site alteration	Novel	U	NS	NS	NS	0.002735	Splice site altering variant. This is highly likely to result in a frameshift mutation (MaxEntScore decrease 100%)	Unknown	Unknown	Unknown	1	0
PFN1	p.P29L	Novel	U	NS	NS	0	0.003573	Mis-sense variant between two beta-strands in the secondary structure	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
GRN	p.T18M	Not previously associated with disease	U	0.02	0.02	0		Situated in exon 2 - ParaGran region. Mutation at codon 19 is non-pathogenic (Gass, Cannon et al. 2006)	Unknown	Unknown	Unknown	2	0
GRN	p.P50T	Novel	U	NS	NS	NS		Sits between Paragranulin and Granulin-1 domains	Unknown	Unknown	Unknown	1	0
GRN	p.M286T	Novel	U	NS	NS	NS		Mis-sense variant in the Granulin-4 domain	Unknown	Unknown	Unknown	1	0
GRN	p.E287D	Seen previously in a single patient (Gass, Cannon et al. 2006)	U	NS	0.01	0		Mis-sense variant in the Granulin-4 domain	Unknown	Unknown	Unknown	1	0
GRN	p.C404F	Not seen previously	U	NS	NS	NS	0.000005176	Located in GranC region. Unclear whether mutations in this region are pathogenic	Unknown	Unknown	Unknown	1	0
GRN	p.R418Q	Seen in control databases and also seen in controls (Gass, Cannon et al. 2006) (Le Ber, van der Zee et al. 2007)	U	NS	0.02	0.02		Located in Exon 11 - InterCD region	Unknown	Unknown	Unknown	2	0
GRN	p.V514M	Seen previously in a familial AD case (Cruchaga, Haller et al. 2012)	U	NS	0.05	0		Sits between Granulin-6 and Granulin-7 domains	Unknown	Unknown	Unknown	2	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
GRN	p.A588S	Not previously associated with disease	U	NS	NS	0.05		Located after Granulin-7 domain	Unknown	Unknown	Unknown	30	7
MAPT	p.S501I	Novel	U	NS	NS	NS	0.000236	Located outside the tubulin binding domain	Unknown	Unknown	Unknown	0	1
MAPT	p.A210T	Previously described in a patient with FTLD (King, Al-Sarraj et al. 2013), but also seen in control databases.	U	0.02	0.14	0.08		Located outside the tubulin binding domain	Unknown	Unknown	Unknown	4	2
NOTCH3	p.T2270M	Not previously associated with disease	U	NS	0	0.01	0.00003214	Located outside an EGF domain and doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.P2222L	Novel	U	NS	NS	NS	0.000415	Located outside an EGF domain and doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.P2209L	Not previously associated with disease	U	0.04	NS	0.07		Located outside an EGF domain and doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.A2190V	Not previously associated with disease	U	0.02	NS	0.02		Located outside an EGF domain and doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
NOTCH3	p.R2150S	Novel	U	NS	NS	NS	0.0004178	Located outside an EGF domain and doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.P2115L	Novel	U	NS	NS	NS	0.0005559	Located outside an EGF domain and doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.V2021M	Not previously associated with disease	U	0.04	NS	0.02	0.000006808	Not present in EGF and doesn't alter cysteine residue	Unknown	Unknown	Unknown	2	0
NOTCH3	p.A1947V	Not previously associated with disease	U	NS	NS	0.01	0.000003381	Located outside an EGF domain and doesn't affect cysteine residue	Unknown	Unknown	Unknown	0	1
NOTCH3	p.V1762M	Described in a single case with a CADASIL like phenotype (Bersano, Ranieri et al. 2012)	U	NS	NS	0.02		Not present in EGF and doesn't alter cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.S1420R	Novel	U	NS	NS	NS		Located outside an EGF domain and doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.P1408S	Novel	U	NS	NS	NS		Located outside an EGF domain and doesn't affect cysteine residue	Unknown	Unknown	Unknown	0	1

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
NOTCH3	p.C1405F	Novel	U	NS	NS	NS	0.00000471	Not present in EGF but does alter cysteine residue				0	1
NOTCH3	p.E1404D	Novel	U	NS	NS	NS		Located outside an EGF domain and doesn't affect cysteine residue	Unknown	Unknown	Unknown	0	1
NOTCH3	p.D1398Y	Novel	U	NS	NS	NS	0.001803	Situated outside an EGF domain and doesn't affect a cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.R1285P	Novel	U	NS	NS	NS		Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.G1269V	Novel	U	NS	NS	0		Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.Q1253P	Novel	U	NS	NS	0.01		Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.R1242H	Not previously associated with disease	U	0.04	0.02	0		Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	0	1

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
NOTCH3	p.C1222G	Not previously associated with disease	U	NS	0.05	0.01	0.001972	Situated in EGF and affects cysteine but not seen in association with vascular disease in this cohort	Unknown	Unknown	Unknown	1	0
NOTCH3	p.A1213D	Novel	U	NS	NS	NS		Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.M1107I	Not previously associated with disease	U	0.02	NS	0		Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.C997*	Novel	U	NS	NS	NS		Situated in EGF and affects cysteine but not seen in association with vascular disease in this cohort	Unknown	Unknown	Unknown	1	0
NOTCH3	p.S991R	Novel	U	NS	NS	NS		Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	2	0
NOTCH3	p.H981Y	Previously seen in healthy cases (Ross, Soto-Ortolaza et al. 2013)	U	NS	NS	0		Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.S947R	Novel	U	NS	NS	NS		In EGF domain but doesn't affect cysteine	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
								residue					
NOTCH3	p.P913L	Not previously associated with disease	U	NS	0.01	0		Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.S893N	Novel	U	NS	NS	NS	0.000001622	In EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.G878R	Novel	U	NS	NS	NS	0.000001622	Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	0	1
NOTCH3	p.G869A	Not previously associated with disease	U	NS	NS	0		Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.A815D	Novel	U	NS	NS	NS		Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.P761L	Not previously associated with disease	U	NS	0	0		Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	0	1

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
NOTCH3	p.V644D	Not previously associated with disease	U	0.02	0.09	0.07	0.00003396	Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	3	1
NOTCH3	p.R544H	Novel	U	NS	NS	0		Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	1
NOTCH3	p.A534S	Novel	U	NS	NS	NS	0.001	Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	1
NOTCH3	p.R532L	Novel	U	NS	NS	NS	0.005888	Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.G490A	Not previously associated with disease	U	NS	0.01	0	0.000002547	Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.P426L	Not previously associated with disease	U	NS	NS	0	0.0002198	In EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.E403D	Novel	U	NS	NS	NS		In EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
NOTCH3	p.L295R	Not previously associated with disease	U	0.02	0.07	0.02		Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	2	0
NOTCH3	p.G248A	Not previously associated with disease	U	0.02	NS	0.01	0.0007345	In EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.H170R	Not previously associated with disease	U	0.14	0.15	0.19	0.00002818	Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	7	3
NOTCH3	p.R169H	Not previously associated with disease	U	NS	NS	0		Present in EGF domain but does not affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.R113Q	Not previously associated with disease	U	NS	0.07	0.08		Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	1
NOTCH3	p.A102V	Novel	U	NS	NS	NS	0.004592	Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0
NOTCH3	p.V97A	Novel	U	NS	NS	NS	0.0009817	Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
NOTCH3	p.S95R	Not previously associated with disease	U	NS	NS	0	0.00004688	Located within an EGF domain but doesn't affect cysteine residue	Unknown	Unknown	Unknown	0	1
PRNP	p.R48H	Novel	U	NS	NS	NS	0.0009016	Sits in N-terminal region, no pathogenic mutations have been seen in this region before (Mead 2006, Beck, Poulter et al. 2010)	Unknown	Unknown	Unknown	1	0
PRNP	p.G54S	Seen in healthy controls (Beck, Poulter et al. 2010, Forbes, Goodwin et al. 2014)	U	NS	NS	0.07	0.0009016	Sits in N-terminal region, no pathogenic mutations have been seen in this region before (Mead 2006, Beck, Poulter et al. 2010)	Unknown	Unknown	Unknown	0	1
PRNP	p.W26L	No point mutations within this region have been described as being pathogenic (Mead 2006)	U	NS	NS	NS		Present in N-terminal domain	Unknown	Unknown	Unknown	1	0
PRNP	p.R164M	Novel	U	NS	NS	NS		Present in major coding region	Unknown	Unknown	Unknown	1	0
PRNP	p.K194R	Novel	U	NS	NS	NS	0.002223	Present in major coding region	Unknown	Unknown	Unknown	0	1

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
VAPB	p.S160del	Seen in equal frequency between cases and controls(Landers, Leclerc et al. 2008), but seen in one fALS case and none of > 300 controls (Kabashi, El Oussini et al. 2013)	U	NS	NS	NS		Sits within exon 5, but lies outside the MSP domain	Alters the structure of the protein but retains normal subcellular localisation (Landers, Leclerc et al. 2008)	Mutant protein does not rescue the phenotype (Landers, Leclerc et al. 2008)	No	11	0
VAPB	p.M170I	High frequency in population databases	U	0.1	0.16	0.14	6.75E-07	Located near MSP domain	Unknown	Unknown	Unknown	5	1
VAPB	p.F231L	Novel	U	NS	NS	NS	0.0000271	Located in the transmembrane domain	Unknown	Unknown	Unknown	0	1
APP	p.I665V	Not previously associated with disease	U	NS	NS	0	0.0006095	Sits within TM-I	Unknown	Unknown	Unknown	1	1
APP	p.T553M	Not previously associated with disease	U	NS	NS	0	0.002723	Not within TM domain	Unknown	Unknown	Unknown	0	1
APP	p.T525M	Not previously associated with disease	U	NS	NS	0.01	0.0001466	Not within TM domain	Unknown	Unknown	Unknown	0	1
APP	p.E489K	Not previously associated with disease	U	0.06	0.1	0.13	0.0001466	Not within TM domain	Unknown	Unknown	Unknown	1	0
APP	p.C285F	Novel	U	NS	NS	NS	0.000004207	Not within TM domain	Unknown	Unknown	Unknown	1	0
APP	p.M338V	Not previously associated	U	0	0.01	0		Not within TM	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
		with disease						domain					
APP	p.R328W	Not previously associated with disease	U	NS	NS	0.02	0.004592	Not within TM domain	Unknown	Unknown	Unknown	1	0
APP	p.R250L	Not previously associated with disease	U	0	0.01	0	0.000003899	Not within TM domain	Unknown	Unknown	Unknown	1	0
APP	p.A295S	Novel	U	NS	NS	NS	0.000003899	Not within TM domain	Unknown	Unknown	Unknown	1	0
APP	p.S198P	Not previously associated with disease	U	NS	0.07	0.05	0.001042	Not within TM domain	Unknown	Unknown	Unknown	0	1
SOD1	p.Q16R	Novel	U	NS	NS	NS		Sits in N-terminal region and in which several pathogenic mutations have been described	Unknown	Unknown	Unknown	1	0
CHCHD10	p.Y135H	Not previously associated with disease	U	NS	0.08	0.03		Sits distal to the CHCH domain	Unknown	Unknown	Unknown	2	0
CHCHD10	p.P80L	Not previously associated with disease	U	NS	NS	0.03	0.00001442	Sits within the non-structured domain between the hydrophobic helix ad CHCH domain	Unknown	Unknown	Unknown	1	0
CHCHD10	p.M48K	Novel	U	NS	NS	NS	0.00317	Sits within the hydrophobic helix in which previous pathogenic mutations	Unknown	Unknown	Unknown	0	1

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
								have been described					
CHCHD10	p.P34S	Described previously in 2 cases of FTD-ALS (Chaussenot, Le Ber et al. 2014)	U	0.04	NS	0.11		Sits within the nonstructured N-terminal region, and the biological relevance is unknown	Unknown	Unknown	Unknown	12	2
UBQLN2	p.A28P	Novel	U	NS	NS	NS		Sits within a nonstructured domain	Unknown	Unknown	Unknown	1	0
UBQLN2	p.S140R	Novel	U	NS	NS	NS		Sits within a nonstructured domain	Unknown	Unknown	Unknown	1	0
UBQLN2	p.S301P	Novel	U	NS	NS	NS		Sits within a nonstructured domain	Unknown	Unknown	Unknown	1	0
UBQLN2	p.M563I	Novel	U	NS	NS	NS	0.00005875	Sits within a nonstructured domain	Unknown	Unknown	Unknown	1	0

Appendix 14 Likely Benign mutations – ACMG criteria.

The gene and protein alteration are shown together with whether there is evidence to support strong pathogenicity (PS), moderate pathogenicity (PM) or supporting pathogenicity (PP) according to ACMG 2015 criteria (Richards, Aziz et al. 2015). In addition, evidence supporting a benign nature of the variant is also provided (BS and BP). Final ACMG classification based on both the pre-study and post-study data are shown.

		Pathogenic																Benign													ACMG classification									
Variant		Very Strong	Strong				Moderate						Supporting					Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
PSEN2	p.S30F							1						1	1																				1					1	
PSEN2	p.G37V							1						1	1																				1					1	
PSEN2	p.R62C																		1											1					1					1	
PSEN2	p.V64F							1																											1					1	
PSEN2	p.C65F							1						1	1																				1					1	
PSEN2	p.R110H							1						1	1																				1					1	
PSEN2	p.S130L							1						1																					1					1	
DCTN1	p.V1065I							1																				1							1					1	
DCTN1	p.V1074M							1																				1							1					1	
DCTN1	p.E1067*														1																				1					1	
DCTN1	p.R1029Q														1																				1					1	
DCTN1	p.E1026G																																		1					1	
DCTN1	p.K851R							1																											1					1	
DCTN1	p.A803G																																		1					1	
DCTN1	p.R778W																		1	1																1					1	
DCTN1	p.V697L							1																											1					1	
DCTN1	p.S661F							1																											1					1	
DCTN1	p.A653V							1																											1					1	

		Pathogenic																Benign													ACMG classification									
Variant		Very Strong	Strong				Moderate						Supporting					Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
DCTN1	p.K617N							1																											1					1	
DCTN1	p.V599I																																		1					1	
DCTN1	p.P512S							1																											1					1	
DCTN1	p.R465C							1							1																				1					1	
DCTN1	p.A354T							1							1																				1					1	
DCTN1	p.R218Q							1							1																				1					1	
HTRA2	p.R173H																																		1					1	
HTRA2	p.P120S															1																			1					1	
HTRA2	p.N106H															1																			1					1	
HTRA2	p.V60L																		1																	1				1	
HTRA2	p.A32S																																		1					1	
HTRA2	p.L72P																		1																	1				1	
HTRA2	p.T76N								1																										1					1	
HTRA2	p.N98S								1																										1					1	
GIGYF2	p.R168C							1							1																				1					1	
GIGYF2	p.G208A							1							1																				1					1	
GIGYF2	p.I228M							1							1																				1					1	
GIGYF2	p.G239V							1							1																				1					1	
GIGYF2	p.D257E							1																											1					1	
GIGYF2	p.D371E																		1									1							1				1	
GIGYF2	p.S396P							1																				1							1					1	
GIGYF2	p.Q439H																																		1					1	
GIGYF2	p.N451T																																		1				1	
GIGYF2	p.T702K							1																											1					1	
GIGYF2	p.R816C							1							1																				1					1	

		Pathogenic																	Benign												ACMG classification									
Variant		Very Strong	Strong				Moderate						Supporting						Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
GIGYF2	p.R852Q																											1								1				1	
GIGYF2	p.R912Q																											1								1				1	
GIGYF2	p.R982Q																		1									1							1				1	
GIGYF2	p.S1056C																		1									1							1				1	
GIGYF2	p.E1169D							1																												1				1	
GIGYF2	p.H1171R																		1									1							1				1	
GIGYF2	p.K1210M							1							1																				1				1	
GIGYF2	p.L1230P																											1								1				1	
GIGYF2	p.Q1207P																											1								1				1	
GIGYF2	p.Q1223_Q1226del																																			1				1	
CHMP2B	p.R19Q							1							1																					1				1	
CHMP2B	p.R32Q														1																					1				1	
CHMP2B	p.A33V							1																				1								1				1	
EIF4G1	Splice-site alteration (1)							1																												1				1	
EIF4G1	p.A63D							1																												1				1	
EIF4G1	Splice-site alteration (2)																																			1				1	
EIF4G1	p.P100L							1																												1				1	
EIF4G1	p.I293T							1																				1								1				1	
EIF4G1	p.A265S																											1								1				1	

		Pathogenic																Benign													ACMG classification									
Variant		Very Strong	Strong				Moderate						Supporting					Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
EIF4G1	p.G276_A278del							1																											1					1	
EIF4G1	p.P399S																		1									1							1					1	
EIF4G1	p.P539L																											1							1					1	
EIF4G1	p.A510P																		1									1							1					1	
EIF4G1	p.S489L							1																				1							1					1	
EIF4G1	p.G490C							1						1	1													1							1					1	
EIF4G1	p.P703T							1																				1							1					1	
EIF4G1	p.G534A													1	1																				1					1	
EIF4G1	p.A553P																		1									1							1					1	
EIF4G1	p.I642V																		1									1							1					1	
EIF4G1	p.T634S																											1							1					1	
EIF4G1	p.E936*							1		1				1	1																				1					1	
EIF4G1	p.R1074C													1	1																				1					1	
EIF4G1	p.R1099C							1						1	1																				1					1	
EIF4G1	p.L1312V							1																											1					1	
EIF4G1	p.M1336V													1	1																				1					1	
EIF4G1	p.I1493V							1																				1							1					1	
EIF4G1	Splice-site alteration (3)							1																											1					1	
EIF4G1	p.A1549S							1						1	1																				1					1	
COQ2	p.N417S							1																											1					1	
COQ2	p.Y353C							1																											1					1	

		Pathogenic																Benign													ACMG classification									
Variant		Very Strong	Strong				Moderate						Supporting					Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
COQ2	p.A201T													1	1																				1					1	
COQ2	p.R197H													1	1																				1					1	
COQ2	p.S107T							1																											1					1	
COQ2	p.Q105H							1																											1					1	
COQ2	p.P96S							1																											1					1	
COQ2	p.S54W													1	1																				1					1	
COQ2	p.G53R													1	1																				1					1	
MATR3	p.Y214C							1																											1					1	
MATR3	p.E262*							1		1					1																				1					1	
MATR3	p.V159M							1																											1					1	
MATR3	p.A762T																		1									1								1				1	
MATR3	p.R503C																																		1					1	
SQSTM1	p.S24G																																		1					1	
SQSTM1	p.A33V																		1									1								1				1	
SQSTM1	p.P111L																											1								1				1	
SQSTM1	p.A33V																		1									1								1				1	
SQSTM1	p.P118S																																		1					1	
SQSTM1	p.P50L																																		1					1	
SQSTM1	p.V150G							1						1																					1					1	
SQSTM1	p.K238E																		1									1								1				1	
SQSTM1	p.G168V							1							1																				1					1	
SQSTM1	p.P392L						1		1				1	1																					1					1	
HNRNPA2B1	p.G323del							1		1					1																				1					1	
C9orf72	p.M332V																																		1					1	
C9orf72	p.M318V							1																											1					1	

		Pathogenic																Benign													ACMG classification									
Variant		Very Strong	Strong				Moderate						Supporting					Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
C9orf72	p.I138V							1																											1					1	
C9orf72	p.T49R																																		1					1	
VCP	p.N616fs*12							1		1				1	1																				1					1	
VCP	p.N616fs*63							1		1				1	1																				1					1	
VCP	p.D501E							1																				1							1					1	
VCP	p.R453Q							1																				1							1					1	
VCP	p.R323L							1																											1					1	
VCP	p.I27V														1																				1					1	
VCP	p.G9V							1																				1							1					1	
HNRNPA1	p.F216L						1	1						1																					1					1	
DAO	p.R2S				1																														1					1	
DAO	p.A8E							1						1	1																				1					1	
DAO	p.Q63E																											1							1					1	
DAO	p.P103P																																		1					1	
DAO	p.R115W													1	1																				1					1	
DAO	p.Y144H													1	1																				1					1	
DAO	p.W260*						1							1	1																				1					1	
ANG	p.G17V							1																				1							1					1	
ANG	p.I70V																											1							1					1	
ANG	p.K78E																																		1					1	
ANG	p.N83K							1																				1							1					1	
FUS	p.S57del							1												1															1					1	

		Pathogenic																Benign												ACMG classification									
Variant		Very Strong	Strong				Moderate						Supporting					Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
FUS	p.R379H													1														1							1					1	
FUS	p.R483H																											1							1					1	
VPS35	p.V476A																											1							1					1	
VPS35	p.K296E							1						1																					1					1	
VPS35	p.M57L							1																				1							1					1	
VPS35	p.M30K													1																					1					1	
PFN1	p.V103F							1						1																					1					1	
PFN1	Splice site alteration							1		1																									1					1	
PFN1	p.P29L							1																											1					1	
GRN	p.T18M							1																											1					1	
GRN	p.P50T							1																											1					1	
GRN	p.M286T							1																				1								1				1	
GRN	p.E287D																																		1					1	
GRN	p.C404F							1						1	1																				1					1	
GRN	p.R418Q																	1																	1					1	
GRN	p.V514M																	1										1								1				1	
GRN	p.A588S							1																											1					1	
MAPT	p.S501I							1																				1							1					1	
MAPT	p.A210T																	1	1									1								1				1	
NOTCH3	p.T2270M							1																				1								1				1	
NOTCH3	p.P2222L							1																				1								1				1	
NOTCH3	p.P2209L							1																				1								1				1	
NOTCH3	p.A2190V							1										1										1								1				1	
NOTCH3	p.R2150S							1																				1								1				1	

		Pathogenic																Benign												ACMG classification									
Variant		Very Strong	Strong				Moderate						Supporting					Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
NOTCH3	p.P2115L							1																				1												
NOTCH3	p.V2021M							1										1										1							1					
NOTCH3	p.A1947V							1																				1												
NOTCH3	p.V1762M							1										1										1							1					
NOTCH3	p.S1420R							1																				1												
NOTCH3	p.P1408S							1																				1												
NOTCH3	p.C1405F							1																				1								1				
NOTCH3	p.E1404D							1																				1												
NOTCH3	p.D1398Y							1																				1							1					
NOTCH3	p.R1285P							1																				1												
NOTCH3	p.G1269V							1																				1												
NOTCH3	p.Q1253P							1										1										1							1					
NOTCH3	p.R1242H																											1												
NOTCH3	p.C1222G						1	1						1														1								1				
NOTCH3	p.A1213D							1																				1												
NOTCH3	p.M1107I							1																				1												
NOTCH3	p.C997*	1					1	1						1																						1				
NOTCH3	p.S991R							1																				1												
NOTCH3	p.H981Y							1																				1												
NOTCH3	p.S947R							1																				1							1					
NOTCH3	p.P913L																											1												
NOTCH3	p.S893N							1																				1							1					
NOTCH3	p.G878R							1																				1												
NOTCH3	p.G869A							1																				1							1					
NOTCH3	p.A815D							1																				1												

		Pathogenic																Benign												ACMG classification									
Variant		Very Strong	Strong				Moderate						Supporting					Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
NOTCH3	p.P761L																											1												
NOTCH3	p.V644D																	1										1												
NOTCH3	p.R544H							1																				1												
NOTCH3	p.A534S							1																				1												
NOTCH3	p.R532L							1																				1												
NOTCH3	p.G490A																											1												
NOTCH3	p.P426L							1						1														1							1						
NOTCH3	p.E403D							1																				1							1						
NOTCH3	p.L295R																	1										1												
NOTCH3	p.G248A							1																				1							1						
NOTCH3	p.H170R																	1										1												
NOTCH3	p.R169H							1																				1								1					
NOTCH3	p.R113Q																		1																1						
NOTCH3	p.A102V																																								
NOTCH3	p.V97A																																								
NOTCH3	p.S95R																																								
PRNP	p.R48H							1						1	1																					1						
PRNP	p.G54S							1																											1						
PRNP	p.W26L							1						1	1																					1						
PRNP	p.R164M							1						1	1																					1						
PRNP	p.K194R							1																												1						
VAPB	p.S160del																	1		1															1						
VAPB	p.M170I																	1	1																1						
VAPB	p.F231L							1							1																					1						
APP	p.I665V						1	1																												1						

		Pathogenic																Benign												ACMG classification									
Variant		Very Strong	Strong				Moderate						Supporting					Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
APP	p.T553M							1																				1								1				1	
APP	p.T525M							1																												1				1	
APP	p.E489K																	1	1									1							1				1	
APP	p.C285F							1																												1				1	
APP	p.M338V																											1								1				1	
APP	p.R328W							1						1	1																					1				1	
APP	p.R250L							1						1	1																					1				1	
APP	p.A295S																																			1				1	
APP	p.S198P																																			1				1	
SOD1	p.Q16R							1						1																						1				1	
CHCHD10	p.Y135H																	1	1									1							1				1	
CHCHD10	p.P80L							1																				1								1				1	
CHCHD10	p.M48K							1																												1				1	
CHCHD10	p.P34S							1																				1								1				1	
UBQLN2	p.A28P							1																				1								1				1	
UBQLN2	p.S140R							1																				1								1				1	
UBQLN2	p.S301P							1																												1				1	
UBQLN2	p.M563I							1																												1				.	

Appendix 15 Clinical data – Likely Benign variants.

Gene	Protein alteration	Chromosome	Position	1000G MAF (%)	ESP6500 MAF (%)	Exac MAF (%)	SIFT	PolyPhen2	ACMG	Cases with variant	Controls with variant		AD	CBD	CJD	DLB	FTD-ALS	HD	MSA	Other	PD	PSP	VD	VD/AD	VD/control		Control	Control (High Braak)	Young control
PSEN2	p.S30F	1	227069697	NS	NS	0	D	B	Likely Benign	1	0				1														
PSEN2	p.G37V	1	227069718	NS	NS	NS	T	B	Likely Benign	1	2			1													2		
PSEN2	p.R62C	1	227071448	NS	0.01	0.02	T	PoD	Likely Benign	2	0						2												
PSEN2	p.V64F	1	227071454	NS	NS	NS	T	B	Likely Benign	1	0																		1
PSEN2	p.C65F	1	227071458	NS	NS	NS	D	B	Likely Benign	1	0										1								
PSEN2	p.R110H	1	227071593	NS	NS	0	D	B	Likely Benign	1	0												1						
PSEN2	p.S130L	1	227073271	0.1	0.1	0.06	D	PoD	Likely Benign	5	1		2		2						1						1		
DCTN1	p.V1065I	2	74589791	0.06	NS	0.02	T	B	Likely Benign	1	0		1																
DCTN1	p.V1074M	2	74590525	NS	NS	0	T	B	Likely Benign	2	0						1			1									
DCTN1	p.E1067*	2	74590753	NS	0.01	0			Likely Benign	1	0		1																
DCTN1	p.R1029Q	2	74592252	0.04	0.23	0.1	T	PoD	Likely Benign	9	2		3		2		1			1			2				2		
DCTN1	p.E1026G	2	74592261	NS	0.02	0.01	T	B	Likely Benign	1	0				1														
DCTN1	p.K851R	2	74592717	NS	NS	0	T	B	Likely Benign	2	0				1													1	
DCTN1	p.A803G	2	74593947	NS	0.01	NS	T	B	Likely Benign	1	0											1							
DCTN1	p.R778W	2	74594023	NS	0.08	0.02	D	PoD	Likely Benign	1	1		1														1		
DCTN1	p.V697L	2	74594858	NS	NS	NS	T	B	Likely Benign	1	0		1																
DCTN1	p.S661F	2	74594914	NS	NS	NS	D	B	Likely Benign	1	0									1									
DCTN1	p.A653V	2	74595155	NS	NS	NS	T	B	Likely Benign	1	0		1																
DCTN1	p.K617N	2	74595241	NS	NS	NS	T	B	Likely Benign	1	0					1													
DCTN1	p.V599I	2	74595893	NS	0	0		B	Likely Benign	1	0														1				
DCTN1	p.P512S	2	74596281	NS	NS	NS	T	B	Likely Benign	1	0		1																
DCTN1	p.R465C	2	74596507	NS	NS	0	D	PoD	Likely Benign	0	1																1		
DCTN1	p.A354T	2	74597660	NS	NS	NS	D	B	Likely Benign	1	0		1																
DCTN1	p.R218Q	2	74598236	NS	NS	NS	D	PrD	Likely Benign	1	0		1																
HTRA2	p.R173H	2	74755884	NS	NS	NS	T	B	Likely Benign	1	0						1												

Gene	Protein alteration	Chromosome	Position	1000G MAF (%)	ESP6500 MAF (%)	Exac MAF (%)	SIFT	PolyPhen2	ACMG	Cases with variant	Controls with variant		AD	CBD	CJD	DLB	FTD-ALS	HD	MSA	Other	PD	PSP	VD	VD/AD	VD/control		Control	Control (High Braak)	Young control
HTRA2	p.P120S	2	74756044	NS	NS	NS	D	PoD	Likely Benign	1	0		1																
HTRA2	p.N106H	2	74756282	NS	NS	0	D	PrD	Likely Benign	1	0																	1	
HTRA2	p.V60L	2	74756499	0.04	0.13	0.13	T	B	Likely Benign	2	2				1					1							2		
HTRA2	p.A32S	2	74756583	NS	0.01	0	T	B	Likely Benign	1	0		1																
HTRA2	p.L72P	2	74757348	0.16	0.3	0.27	D	B	Likely Benign	2	1		1		1												1		
HTRA2	p.T76N	2	74757360	NS	NS	NS	T	B	Likely Benign	0	1																1		
HTRA2	p.N98S	2	74757426	NS	NS	NS	T	B	Likely Benign	1	0				1														
GIGYF2	p.R168C	2	233626116	NS	NS	0	D	B	Likely Benign	1	0																	1	
GIGYF2	p.G208A	2	233651884	NS	NS	0	T	B	Likely Benign	1	0			1															
GIGYF2	p.I228M	2	233651945	NS	NS	NS	D	PoD	Likely Benign	1	0						1												
GIGYF2	p.G239V	2	233655411	NS	NS	NS	D	PoD	Likely Benign	1	0		1																
GIGYF2	p.D257E	2	233655484	NS	NS	NS		B	Likely Benign	1	0		1																
GIGYF2	p.D371E	2	233655834	NS	0.13	0.05	T	B	Likely Benign	1	1													1			1		
GIGYF2	p.S396P	2	233655997	NS	0	NS	T	B	Likely Benign	1	0				1														
GIGYF2	p.Q439H	2	233659492	NS	0.02	0.01	T	PoD	Likely Benign	2	0						1			1									
GIGYF2	p.N451T	2	233659545	0.02	0.14	0.04	T	B	Likely Benign	1	0				1														
GIGYF2	p.T702K	2	233677199	NS	NS	NS	D	B	Likely Benign	1	0		1																
GIGYF2	p.R816C	2	233684549	NS	NS	0	D	PoD	Likely Benign	1	0													1					
GIGYF2	p.R852Q	2	233697592	NS	0.01	0.04	T	B	Likely Benign	1	0						1												
GIGYF2	p.R912Q	2	233697709	NS	0.01	0	T	B	Likely Benign	0	1																1		
GIGYF2	p.R982Q;	2	233704674	0.06	0.14	0.06	T	B	Likely Benign	1	0						1												
GIGYF2	p.S1056C	2	233709083	NS	0.19	0.08	T	PoD	Likely Benign	4	0		1		1		2												
GIGYF2	p.E1169D	2	233712104	NS	NS	NS	T	PrD	Likely Benign	1	0		1																
GIGYF2	p.H1171R	2	233712109	0.08	0.22	0.16	T	PrD	Likely Benign	4	0		3							1									
GIGYF2	p.K1210M	2	233712163	NS	NS	NS	D	PrD	Likely Benign	1	0																	1	
GIGYF2	p.L1230P	2	233712223	NS	NS	0.02	T	B	Likely Benign	54	6		17	2	6	2	11			6	1		3		1		6		5
GIGYF2	p.Q1207P	2	233712235	NS	NS	NS	T	B	Likely Benign	3	2						1						1		1		2		

Gene	Protein alteration	Chromosome	Position	1000G MAF (%)	ESP6500 MAF (%)	Exac MAF (%)	SIFT	PolyPhen2	ACMG	Cases with variant	Controls with variant		AD	CBD	CJD	DLB	FTD-ALS	HD	MSA	Other	PD	PSP	VD	VD/AD	VD/control		Control	Control (High Braak)	Young control
GIGYF2	p.Q1223_Q1226del	2	233712264	NS	NS	NS			Likely Benign	28	1		8		5	2	6			1			2	2			1	1	1
CHMP2B	p.R19Q	3	87289870	0.02	NS	0.01	D	B	Likely Benign	1	0														1				
CHMP2B	p.R32Q	3	87289909	NS	0.01	0	D	B	Likely Benign	1	0				1														
CHMP2B	p.A33V	3	87289912	NS	NS	NS	T	B	Likely Benign	1	0																	1	
EIF4G1	Splice-site alteration (1)	3	184033272	NS	NS	NS			Likely Benign	1	0		1																
EIF4G1	p.A63D	3	184035149	NS	NS	NS	T	PrD	Likely Benign	1	0								1										
EIF4G1	Splice-site alteration (2)	3	184038412	NS	0.01	0			Likely Benign	0	1																1		
EIF4G1	p.P100L	3	184039163	NS	NS	0	T	PrD	Likely Benign	1	0						1												
EIF4G1	p.I293T	3	184039250	NS	NS	NS		B	Likely Benign	1	0							1											
EIF4G1	p.A265S	3	184039426	NS	0.02	0.01	T	B	Likely Benign	1	0									1									
EIF4G1	p.G276_A278del	3	184039785	NS	NS	NS			Likely Benign	67	10		15		7	7	11	1		4	2		3	1	3		10	1	12
EIF4G1	p.P399S	3	184039828	0.02	0.06	0.08		B	Likely Benign	4	0		1		1	1	1												
EIF4G1	p.P539L	3	184040339	0.04	NS	0.01	T	B	Likely Benign	0	1																1		
EIF4G1	p.A510P	3	184040371	0.12	0.19	0.1	T	B	Likely Benign	1	1												1				1		
EIF4G1	p.S489L	3	184040450	NS	NS	NS	T	B	Likely Benign	1	0				1														
EIF4G1	p.G490C	3	184040997	NS	NS	0.01	T	PoD	Likely Benign	2	0				2														
EIF4G1	p.P703T	3	184041027	NS	NS	NS	T	B	Likely Benign	1	0				1														
EIF4G1	p.G534A	3	184041200	NS	0.01	0	T	PoD	Likely Benign	2	0		1		1														
EIF4G1	p.A553P	3	184041256	NS	0.15	0.07	T	B	Likely Benign	1	1										1						1		
EIF4G1	p.I642V	3	184041709	0.02	0.05	0.04	D	B	Likely Benign	2	1									1							1		1
EIF4G1	p.T634S	3	184042001	NS	0.02	0	T	B	Likely Benign	1	0												1						
EIF4G1	p.E936*	3	184042831	NS	NS	NS			Likely Benign	1	0		1																
EIF4G1	p.R1074C	3	184044683	NS	0	0	D	PrD	Likely Benign	1	0									1									
EIF4G1	p.R1099C	3	184044758	NS	NS	0.01	D	PrD	Likely Benign	2	0		1				1												

Gene	Protein alteration	Chromosome	Position	1000G MAF (%)	ESP6500 MAF (%)	Exac MAF (%)	SIFT	PolyPhen2	ACMG	Cases with variant	Controls with variant		AD	CBD	CJD	DLB	FTD-ALS	HD	MSA	Other	PD	PSP	VD	VD/AD	VD/control		Control	Control (High Braak)	Young control
EIF4G1	p.L1312V	3	184045750	NS	NS	0	T	B	Likely Benign	1	0		1																
EIF4G1	p.M1336V	3	184046450	0.1	0.02	0.02	D	PoD	Likely Benign	2	0					1	1												
EIF4G1	p.I1493V	3	184049594	NS	NS	NS	T	B	Likely Benign	1	0									1									
EIF4G1	Splice-site alteration (3)	3	184052513	0	0	0			Likely Benign	1	0						1												
EIF4G1	p.A1549S	3	184052538	NS	NS	NS	D	PoD	Likely Benign	1	0				1														
COQ2	p.N417S	4	84185368	NS	NS	0.01			Likely Benign	0	1																1		
COQ2	p.Y353C	4	84188782	NS	NS	0			Likely Benign	0	1																1		
COQ2	p.A201T	4	84194740	NS	0	0			Likely Benign	1	0						1												
COQ2	p.R197H	4	84194751	NS	NS	0			Likely Benign	1	0				1														
COQ2	p.S107T	4	84205748	0.02	NS	0.12	T		Likely Benign	2	1		1										1				1		
COQ2	p.Q105H	4	84205753	NS	NS	0.05	T		Likely Benign	0	1																1		
COQ2	p.P96S	4	84205782	NS	NS	0	T		Likely Benign	3	1		1		1						1						1		
COQ2	p.S54W	4	84205907	NS	0.01	0.14			Likely Benign	1	0				1														
COQ2	p.G53R	4	84205911	NS	NS	NS			Likely Benign	1	0						1												
MATR3	p.Y214C	5	138643745	NS	NS	NS	D	PrD	Likely Benign	0	1																1		
MATR3	p.E262*	5	138643888	NS	NS	NS			Likely Benign	1	0		1																
MATR3	p.V159M	5	138655077	NS	NS	NS	T	PrD	Likely Benign	1	0				1														
MATR3	p.A762T	5	138661264	NS	0.06	0.02	T	B	Likely Benign	1	0		1																
MATR3	p.R503C	5	138665061	NS	NS	0.01	D	B	Likely Benign	3	1		1		1		1										1		
SQSTM1	p.S24G	5	179248006	NS	NS	NS	D	B	Likely Benign	0	1																1		
SQSTM1	p.A33V	5	179248034	NS	0.08	0.12		B	Likely Benign	2	2		1														2		1
SQSTM1	p.P111L	5	179250888	NS	0.01	0	T	B	Likely Benign	1	0					1													
SQSTM1	p.A33V	5	179250906	0.08	0.12	0.15	T	B	Likely Benign	2	1				1								1				1		
SQSTM1	p.P118S	5	179250908	NS	0.01	0.02	D	B	Likely Benign	4	0				2		1												1
SQSTM1	p.P50L	5	179250957	NS	0.01	0	D	B	Likely Benign	1	0		1																
SQSTM1	p.V150G	5	179252173	NS	NS	NS	D	PrD	Likely Benign	1	0		1																
SQSTM1	p.K238E	5	179252184	0.24	0.35	0.24	D	B	Likely Benign	17	1		4		1		7					2	1				1	1	1

Gene	Protein alteration	Chromosome	Position	1000G MAF (%)	ESP6500 MAF (%)	Exac MAF (%)	SIFT	PolyPhen2	ACMG	Cases with variant	Controls with variant		AD	CBD	CJD	DLB	FTD-ALS	HD	MSA	Other	PD	PSP	VD	VD/AD	VD/control		Control	Control (High Braak)	Young control
SQSTM1	p.G168V	5	179260032	0	0	NS	D	PoD	Likely Benign	0	1																		
SQSTM1	p.P392L	5	179263445	0.24	0.21	0.09	T	B	Likely Benign	9	0		1				4						1					1	2
HNRNPA2B1	p.G323del	7	26232906	NS	NS	NS			Likely Benign	1	0		1																
C9orf72	p.M332V	9	27556656	NS	0	0	D	B	Likely Benign	2	1									1							1		1
C9orf72	p.M318V	9	27556698	NS	NS	0	T	B	Likely Benign	1	0												1						
C9orf72	p.I138V	9	27566707	NS	NS	NS	T	B	Likely Benign	1	0									1									
C9orf72	p.T49R	9	27566973	0.02	0.02	0.02	T	B	Likely Benign	3	0		1			1	1												
VCP	p.N616fs*12	9	35059655	NS	NS	NS			Likely Benign	27	4		9		4	1	2			2	1	1	2	2	1		4	2	
VCP	p.N616fs*63	9	35059655	NS	NS	NS			Likely Benign	3	0				3														
VCP	p.D501E	9	35060502	NS	NS	NS		B	Likely Benign	1	0		1																
VCP	p.R453Q	9	35061013	NS	NS	NS		B	Likely Benign	0	2																2		
VCP	p.R323L	9	35062113	NS	NS	NS		PoD	Likely Benign	1	0																		1
VCP	p.I27V	9	35068298	0.1	0.03	0.05		B	Likely Benign	1	0				1														
VCP	p.G9V	9	35068351	NS	NS	NS		B	Likely Benign	1	0				1														
HNRNPA1	p.F216L	12	54676421	NS	NS	NS	T	B	Likely Benign	1	0		1																
DAO	p.R2S	12	109278786	NS	NS	NS	D	B	Likely Benign	1	1																1		
DAO	p.A8E	12	109278805	NS	NS	NS	D	PrD	Likely Benign	1	0														1				
DAO	p.Q63E	12	109278969	NS	0.02	0.01	T	B	Likely Benign	1	1				1												1		
DAO	p.P103P	12	109281340	NS	0.02	0.01			Likely Benign	1	0				1														
DAO	p.R115W	12	109283278	NS	0.01	0.01	D	PrD	Likely Benign	1	0											1							
DAO	p.Y144H	12	109284027	0.3	0	0.08	D	PrD	Likely Benign	2	0		1				1												
DAO	p.W260*	12	109292538	NS	NS	NS			Likely Benign	1	0										1								
ANG	p.G17V	14	21161773	NS	NS	NS	T	B	Likely Benign	1	0												1						
ANG	p.I70V	14	21161931	0.02	0.06	0.06	T	B	Likely Benign	2	1		1				1										1		
ANG	p.K78E	14	21161955	NS	0.01	0.02	T	B	Likely Benign	0	1																1		
ANG	p.N83K	14	21161972	NS	NS	NS	T	B	Likely Benign	1	0					1													
FUS	p.S57del	16	31193964	NS	NS	NS			Likely Benign	3	0				1								1						1

Gene	Protein alteration	Chromosome	Position	1000G MAF (%)	ESP6500 MAF (%)	Exac MAF (%)	SIFT	PolyPhen2	ACMG	Cases with variant	Controls with variant		AD	CBD	CJD	DLB	FTD-ALS	HD	MSA	Other	PD	PSP	VD	VD/AD	VD/control		Control	Control (High Braak)	Young control
FUS	p.R379H; p.R382H; p.R383H	16	31201442	NS	NS	0	T	B	Likely Benign	1	0				1														
FUS	p.R483H; p.R486H; p.R487H	16	31202350	0.02	NS	0	T	B	Likely Benign	0	1																1		
VPS35	p.V476A	16	46705714	NS	0.01	0	T	B	Likely Benign	1	0		1																
VPS35	p.K296E	16	46710523	NS	NS	NS	D	PrD	Likely Benign	1	0						1												
VPS35	p.M57L	16	46716021	NS	NS	0	T	B	Likely Benign	1	0		1																
VPS35	p.M30K	16	46717433	NS	0	0.01	D	PoD	Likely Benign	1	0		1																
PFN1	p.V103F	17	4849941	NS	NS	NS	D	PrD	Likely Benign	1	0									1									
PFN1	Splice site alteration	17	4850116	NS	NS	NS			Likely Benign	1	0				1														
PFN1	p.P29L	17	4851604	NS	NS	NS	D	PoD	Likely Benign	1	0				1														
GRN	p.T18M	17	42426585	0.02	0.02	0	D	PoD	Likely Benign	2	0					1													1
GRN	p.P50T	17	42426803	NS	NS	NS	D	B	Likely Benign	1	0					1													
GRN	p.M286T	17	42428752	NS	NS	NS	T	B	Likely Benign	1	0												1						
GRN	p.E287D	17	42428756	NS	0.01	0	T	PoD	Likely Benign	1	0					1													
GRN	p.C404F	17	42429414	NS	NS	NS	D	PrD	Likely Benign	1	0						1												
GRN	p.R418Q	17	42429456	NS	0.02	0.02	T	B	Likely Benign	2	0				1								1						
GRN	p.V514M	17	42429835	NS	0.05	0	T	B	Likely Benign	2	0				2														
GRN	p.A588S	17	42430146	NS	NS	0.05	D	B	Likely Benign	30	7		1 1	1		1	1 1		2	3							7	1	
MAPT	p.S501I	17	44068947	NS	NS	NS	T	B	Likely Benign	0	1																1		
MAPT	p.A210T	17	44073923	0.02	0.14	0.08	T	B	Likely Benign	4	2		2	1			1										2		
NOTCH3	p.T2270M	19	15271630	NS	0	0.01	T	PoD	Likely Benign	1	0												1						
NOTCH3	p.P2222L	19	15271774	NS	NS	NS	T	B	Likely Benign	1	0		1																

Gene	Protein alteration	Chromosome	Position	1000G MAF (%)	ESP6500 MAF (%)	Exac MAF (%)	SIFT	PolyPhen2	ACMG	Cases with variant	Controls with variant		AD	CBD	CJD	DLB	FTD-ALS	HD	MSA	Other	PD	PSP	VD	VD/AD	VD/control		Control	Control (High Braak)	Young control
NOTCH3	p.P2209L	19	15271813	0.04	NS	0.07	T	B	Likely Benign	1	0																		1
NOTCH3	p.A2190V	19	15271870	0.02	NS	0.02	T	B	Likely Benign	1	0																		1
NOTCH3	p.R2150S	19	15271991	NS	NS	NS	T	B	Likely Benign	1	0						1												
NOTCH3	p.P2115L	19	15272095	NS	NS	NS	T	B	Likely Benign	1	0						1												
NOTCH3	p.V2021M	19	15272378	0.04	NS	0.02	D	PrD	Likely Benign	2	0				2														
NOTCH3	p.A1947V	19	15273349	NS	NS	0.01	D	PrD	Likely Benign	0	1																1		
NOTCH3	p.V1762M	19	15278138	NS	NS	0.02	T	B	Likely Benign	1	0				1														
NOTCH3	p.S1420R	19	15288479	NS	NS	NS		B	Likely Benign	1	0										1								
NOTCH3	p.P1408S	19	15288517	NS	NS	NS		B	Likely Benign	0	1																1		
NOTCH3	p.C1405F	19	15288525	NS	NS	NS	D	PrD	Likely Benign	0	1																1		
NOTCH3	p.E1404D	19	15288527	NS	NS	NS		B	Likely Benign	0	1																1		
NOTCH3	p.D1398Y	19	15288547	NS	NS	NS		PrD	Likely Benign	1	0										1								
NOTCH3	p.R1285P	19	15288885	NS	NS	NS	T	B	Likely Benign	1	0						1												
NOTCH3	p.G1269V	19	15289665	NS	NS	0	T	B	Likely Benign	1	0								1										
NOTCH3	p.Q1253P	19	15289713	NS	NS	0.01		PoD	Likely Benign	1	0						1												
NOTCH3	p.R1242H	19	15289746	0.04	0.02	0		B	Likely Benign	0	1																1		
NOTCH3	p.C1222G	19	15289890	NS	0.05	0.01		PrD	Likely Benign	1	0					1													
NOTCH3	p.A1213D	19	15289916	NS	NS	NS		B	Likely Benign	1	0				1														
NOTCH3	p.M1107I	19	15290889	0.02	NS	0		B	Likely Benign	1	0		1																
NOTCH3	p.C997*	19	15291775	NS	NS	NS		N/A	Likely Benign	1	0										1								
NOTCH3	p.S991R	19	15291793	NS	NS	NS		B	Likely Benign	2	0									1									1
NOTCH3	p.H981Y	19	15291825	NS	NS	0		B	Likely Benign	1	0																		1
NOTCH3	p.S947R	19	15291925	NS	NS	NS		PoD	Likely Benign	1	0				1														
NOTCH3	p.P913L	19	15292441	NS	0.01	0		B	Likely Benign	1	0						1												
NOTCH3	p.S893N	19	15292501	NS	NS	NS		PoD	Likely Benign	1	0																		1
NOTCH3	p.G878R	19	15292547	NS	NS	NS		PrD	Likely Benign	0	1																1		
NOTCH3	p.G869A	19	15292573	NS	NS	0		PoD	Likely Benign	1	0		1																
NOTCH3	p.A815D	19	15295228	NS	NS	NS	D	B	Likely Benign	1	0		1																

Gene	Protein alteration	Chromosome	Position	1000G MAF (%)	ESP6500 MAF (%)	Exac MAF (%)	SIFT	PolyPhen2	ACMG	Cases with variant	Controls with variant		AD	CBD	CJD	DLB	FTD-ALS	HD	MSA	Other	PD	PSP	VD	VD/AD	VD/control		Control	Control (High Braak)	Young control
NOTCH3	p.P761L	19	15296082	NS	0	0	T	B	Likely Benign	0	1																1		
NOTCH3	p.V644D	19	15297709	0.02	0.09	0.07	T	PrD	Likely Benign	3	1				2	1											1		
NOTCH3	p.R544H	19	15298125	NS	NS	0	T	B	Likely Benign	1	1						1										1		
NOTCH3	p.A534S	19	15298698	NS	NS	NS	T	PrD	Likely Benign	1	1														1		1		
NOTCH3	p.R532L	19	15298703	NS	NS	NS	T	B	Likely Benign	1	0				1														
NOTCH3	p.G490A	19	15299069	NS	0.01	0	T	PoD	Likely Benign	1	0					1													
NOTCH3	p.P426L	19	15299901	NS	NS	0	D	PrD	Likely Benign	1	0				1														
NOTCH3	p.E403D	19	15299969	NS	NS	NS	D	PrD	Likely Benign	1	0								1										
NOTCH3	p.L295R	19	15302387	0.02	0.07	0.02	T	B	Likely Benign	2	0				1													1	
NOTCH3	p.G248A	19	15302615	0.02	NS	0.01	D	PrD	Likely Benign	1	0																		1
NOTCH3	p.H170R	19	15302941	0.14	0.15	0.19	D	PoD	Likely Benign	7	3			1	1		3			1							3		1
NOTCH3	p.R169H	19	15302944	NS	NS	0	T	B	Likely Benign	1	0												1						
NOTCH3	p.R113Q	19	15303190	NS	0.07	0.08	T	B	Likely Benign	1	1				1												1		
NOTCH3	p.A102V	19	15303223	NS	NS	NS	T	B	Likely Benign	1	0																		1
NOTCH3	p.V97A	19	15303238	NS	NS	NS	T	B	Likely Benign	1	0									1									
NOTCH3	p.S95R	19	15303245	NS	NS	0	T	PrD	Likely Benign	0	1																1		
PRNP	p.R48H	20	4680009	NS	NS	NS	D	PoD	Likely Benign	1	0		1																
PRNP	p.G54S	20	4680026	NS	NS	0.07	T	B	Likely Benign	0	1																1		
PRNP	p.W26L	20	4680032	NS	NS	NS	D	PoD	Likely Benign	1	0		1																
PRNP	p.R164M	20	4680357	NS	NS	NS	D	PrD	Likely Benign	1	0						1												
PRNP	p.K194R	20	4680447	NS	NS	NS	T	PrD	Likely Benign	0	1																1		
VAPB	p.S160del	20	57016044	NS	NS	NS			Likely Benign	11	0		3		1	1	3			1		1	1						
VAPB	p.M170I	20	57016076	0.1	0.16	0.14	T	B	Likely Benign	5	1		3				2										1		
VAPB	p.F231L	20	57019250	NS	NS	NS		PrD	Likely Benign	0	1																		
APP	p.I665V	21	27264141	NS	NS	0	T	B	Likely Benign	1	1									1							1		
APP	p.T553M	21	27269961	NS	NS	0	T	B	Likely Benign	0	1																1		
APP	p.T525M	21	27284163	NS	NS	0.01	T	PrD	Likely Benign	0	1																1		
APP	p.E489K	21	27284167	0.06	0.1	0.13	T	PoD	Likely Benign	1	0						1												

Gene	Protein alteration	Chromosome	Position	1000G MAF (%)	ESP6500 MAF (%)	Exac MAF (%)	SIFT	PolyPhen2	ACMG	Cases with variant	Controls with variant		AD	CBD	CJD	DLB	FTD-ALS	HD	MSA	Other	PD	PSP	VD	VD/AD	VD/control		Control	Control (High Braak)	Young control
APP	p.C285F;	21	27372341	NS	NS	NS	T	PrD	Likley Benign	1	0												1						
APP	p.M338V	21	27372351	0	0.01	0	T	B	Likley Benign	1	0								1										
APP	p.R328W	21	27372381	NS	NS	0.02	D	PrD	Likley Benign	1	0				1														
APP	p.R250L	21	27372446	0	0.01	0	D	PrD	Likley Benign	1	0						1												
APP	p.A295S	21	27372480	NS	NS	NS	T	PrD	Likley Benign	1	0						1												
APP	p.S198P	21	27423386	NS	0.07	0.05	T	PoD	Likley Benign	0	1																1		
SOD1	p.Q16R	21	33032129	NS	NS	NS	T	B	Likley Benign	1	0												1						
CHCHD10	p.Y135H	22	24108321	NS	0.08	0.03	T	B	Likley Benign	2	0				1						1								
CHCHD10	p.P80L	22	24109583	NS	NS	0.03	T	B	Likley Benign	1	0				1														
CHCHD10	p.M48K	22	24109679	NS	NS	NS	D	PrD	Likley Benign	0	1																1		
CHCHD10	p.P34S	22	24109722	0.04	NS	0.11	T	B	Likley Benign	12	2		3				5					1	2				2		1
UBQLN2	p.A28P	X	56590388	NS	NS	NS	T	B	Likley Benign	1	0									1									
UBQLN2	p.S140R	X	56590726	NS	NS	NS	T	B	Likley Benign	1	0							1											
UBQLN2	p.S301P	X	56591207	NS	NS	NS	T	PoD	Likley Benign	1	0		1																
UBQLN2	p.M563I	X	56591995	NS	NS	NS	T	PoD	Likley Benign	1	0					1													

Appendix 16 Benign variants – MacArthur criteria.

All relevant data and references are provided to support pathogenicity. Minor Allele Frequencies (MAF) from the 1000Genomes (1000G), NHLBI 6500 ESP and ExAC databases are provided. Key: MND – Motor Neuron Disease, CJD – Creutzfelt-Jakob Disease, AD – Alzheimer's disease, CADASIL – Cerebral Autosomal-Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy. U – Unknown.

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
PSEN2	p.R71W	Seen in both cases and controls, but does confer and earlier disease onset (Cruchaga et al. 2012)	U	0.14	0.37	0.34		Located in N-terminal domain	Unknown	Unknown	Unknown	8	3
DCTN1	p.T1224I	Previously associated with a single case of ALS (Munch et al. 2004) However, the MAF has subsequently been shown to be high in population databases	U	0.1	0.45	0.3	0.0003436	Located within the coiled coil domain	Does not result in abnormal cellular morphology in vitro (Stockmann et al. 2013)	Does not result in abnormal cellular morphology in vitro (Stockmann et al. 2013)	Unknown	18	4
GIGYF2	p.L1230Q	Not previously associated with disease	U	NS	NS	0		Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	38	5
GRN	p.A324T	Seen in several cases in healthy subjects. No evidence acts as a risk factor (Cruchaga et al. 2012)	U	0.1	0.14	0.08		Mis-sense variant in the Granulin-4 domain	Unknown	Unknown	Unknown	4	1

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
GRN	p.R433W	Seen relatively commonly in control databases	U	0.16	0.28	0.43		Mis-sense variant between Granulin 5 and Granulin-6 domain	Does not alter GRN plasma levels (Finch et al. 2009)	Unknown	Unknown	11	1
MAPT	p.S427F	Seen in 1 of 436 familial cases of AD, and none of 1346 controls (Cruchaga et al. 2012)	U	0.1	0.21	0.15	0.001259	Present in the inter CD region of the gene	Unknown	Unknown	Unknown	3	0
VAPB	p.D130E	Seen in cases and controls and is also present in reference databases (Landers et al. 2008)	U	0.1	0.1	0.14		Located in the topological domain	Unknown	Unknown	Unknown	1	0

Appendix 17 Benign variants – ACMG criteria.

The gene and protein alteration are shown together with whether there is evidence to support strong pathogenicity (PS), moderate pathogenicity (PM) or supporting pathogenicity (PP) according to ACMG 2015 criteria (Richards, Aziz et al. 2015). In addition, evidence supporting a benign nature of the variant is also provided (BS and BP). Final ACMG classification based on both the pre-study and post-study data are shown.

		Pathogenic																Benign													ACMG classification									
Variant		Very Strong	Strong				Moderate						Supporting					Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
PSEN2	p.R71W																	1	1																1					1		
DCTN1	p.T1224I																	1	1															1					1		
GIGYF2	p.L1230Q																										1	1						1					1		
GRN	p.A324T																	1	1															1					1		
GRN	p.R433W																	1	1	1														1					1		
MAPT	p.S427F																	1	1															1					1		
VAPB	p.D130E																	1	1															1					1		

Appendix 18 Clinical data – Benign variants.

Gene	Protein alteration	Chromosome	Position	1000G MAF (%)	ESP6500 MAF (%)	Exac MAF (%)	SIFT	PolyPhen2	ACMG	Cases with variant	Controls with variant		AD	CBD	CJD	DLB	FTD-ALS	HD	MSA	Other	PD	PSP	VD	VD/AD	VD/control		Control	Control (High Braak)
PSEN2	p.R71W	1	227071475	0.14	0.37	0.34	D	B	Benign	8	3		2		2					1			1				3	
DCTN1	p.T1224I	2	74588717	0.1	0.45	0.3	T	B	Benign	18	4		6	1	3	2	1	1		1		1					4	
GIGYF2	p.L1230Q;	2	233712223	NS	NS	0	T	B	Benign	38	5		9		11	3	8				2	1	2				5	
GRN	p.A324T	17	42428954	0.1	0.14	0.08	T	B	Benign	4	1					2				1				1			1	
GRN	p.R433W	17	42429500	0.16	0.28	0.43	D	B	Benign	11	1		5		2		1			1	1	1					1	
MAPT	p.S427F	17	44067341	0.1	0.21	0.15	D	PrD	Benign	3	0				1		1											
VAPB	p.D130E	20	57014075	0.1	0.1	0.14	T	B	Benign	1	0								1									

Appendix 19 Pathogenicity questioned – MacArthur criteria.

All relevant data and references are provided to support pathogenicity. Minor Allele Frequencies (MAF) from the 1000Genomes (1000G), NHLBI 6500 ESP and ExAC databases are provided. Key: MND – Motor Neuron Disease, CJD – Creutzfelt-Jakob Disease, AD – Alzheimer's disease, CADASIL – Cerebral Autosomal-Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy. U – Unknown.

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
SNCA	p.H50Q	Previously described in several cases of PD (Appel-Cresswell, Vilarino-Guell et al. 2013) (Proukakis, Dudzik et al. 2013)	Unknown	NS	0	0		Disrupts the protein's amphipathic helix; and located between other pathogenic point mutations (Appel-Cresswell, Vilarino-Guell et al. 2013)	Unknown	Unknown	U	1	1
NOTCH3	p.C835fs*25	Not previously seen	No data for segregation	NS	NS	NS	0.000002158	In EGF domain and affects cysteine residue	Unknown	Unknown	U	1	
CHMP2B	p.I29V	Seen previously in ALS and FTD, but also in a control (Cox, Ferraiuolo et al. 2010)	Unknown	NS	0.02	0.01		Positioned within two conserved regions, the snf-7 and coiled coil domain	Significantly increases vacuolation (Cox, Ferraiuolo et al. 2010)	Unknown	U	0	1
SOD1	p.D91A	Incomplete penetrance well described (Al-Chalabi, Andersen et al. 1998, Khoris, Moulard et al. 2000)	Incomplete dominant inheritance (Al-Chalabi, Andersen et al. 1998)	0.04	0.08	0.11		The mutation destabilise the apoSOD1 monomer (Bystrom, Andersen et al. 2010)Pathogenic mutations in SOD1 exist throughout the gene.	Neurons harboring the variant are more susceptibility to oxidative stress (An, Lee et al. 2008)	Mild phenotype in mice (Jonsson, Graffmo et al. 2006)	U	2	2

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
SNCA	p.P117T	Not previously reported	Unknown	NS	0.01	0	0.004436	Located in the C-terminal acidic tail.	Unknown	Unknown	U	1	0
PSEN2	p.P123T	Novel	Unknown	NS	NS	NS	0.000001321	Located in HL-I domain, but one amino acid away from a known pathogenic variant (Finckh, Muller-Thomsen et al. 2000)	Variant p.Thr122Pro alters Ab40/42 ratio (Walker, Martinez et al. 2005)	Unknown	U	1	0
PSEN1	p.D40del; p.D36del	Seen in a single case previously (Nygaard, Lippa et al. 2014)	Unknown	NS	NS	NS		Located in N-terminal region	Unknown	Unknown	U	1	0
FUS	p.H513N; p.H516N; p.H517N	Novel	Unknown	NS	NS	NS	0.00006792	Located in the C-terminal region. Several other pathogenic mutations located within the region (Vance, Rogelj et al. 2009)	A variant in the same region (R521H) significantly alters function through a toxic gain of function (Kabashi, Bercier et al. 2011)	Variants at codon 521 cause phenotype (Kabashi, Bercier et al. 2011)	The variants at codon 521 do not rescue the phenotype (Kabashi, Bercier et al. 2011)	0	1
DAO	p.R199W	Segregated with ALS in a single family (Mitchell, Paul et al. 2010)	Segregated within family (Mitchell, Paul et al. 2010)	NS	0.06	0.03	0.002432	The mutation lies close to the FAD binding site and a region controlling enzymatic activity (Mitchell, Paul et al. 2010)	Co-cultured neurons with astrocytes expressing R199W induced cell death and neuronal cell lines decreased cell viability (Mitchell, Paul et	Ubiquitinated cell aggregates seen (Mitchell, Paul et al. 2010)	U	3	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
									al. 2010)				
ANG; RNASE4	p.K41I	Seen in 5 of 1037 ALS cases (van Blitterswijk, van Es et al. 2012)	Segregated in one case report (van Es, Diekstra et al. 2009)	0.04	0.26	0.14		Located in a critical and well-established functional domain (RNase_A_canonical)	Induces a complete loss of ANG function (Wu, Yu et al. 2007)	Alters neuron morphology (Thiyagarajan, Ferguson et al. 2012)	U	2	1
ANG; RNASE4	p.P136L	Seen in a single case previously (Wu, Yu et al. 2007)	Unknown	NS	0.01	0		Located in a critical and well-established functional domain (RNase_A_canonical)	Almost entirely abolishes ANG function and reduces nuclear translocation (Wu, Yu et al. 2007, Padhi, Kumar et al. 2012)	Unknown	U	1	0

Appendix 20 Pathogenicity questioned – ACMG criteria.

The gene and protein alteration are shown together with whether there is evidence to support strong pathogenicity (PS), moderate pathogenicity (PM) or supporting pathogenicity (PP) according to ACMG 2015 criteria (Richards, Aziz et al. 2015). In addition, evidence supporting a benign nature of the variant is also provided (BS and BP). Final ACMG classification based on both the pre-study and post-study data are shown.

		Pathogenic																Benign												ACMG classification									
Variant		Very Strong	Strong				Moderate						Supporting					Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
SNCA	p.H50Q		1		1		1																									1									1
NOTCH3	p.C835fs*25	1					1	1																								1									1
CHMP2B	p.I29V		1		1		1																									1									1
SOD1	p.D91A				1		1							1	1	1																1									1
SNCA	p.P117T					1		1							1																		1									1
PSEN2	p.P123T								1	1					1	1	1																1									1
PSEN1	p.D40del; p.D36del								1						1	1	1																1									1
FUS	p.H513N; p.H516N; p.H517N								1			1			1	1																	1									1
DAO	p.R199W					1								1	1	1																	1									1
ANG; RNASE4	p.K41I			1		1																											1									1
ANG; RNASE4	p.P136L					1			1							1																	1									1

Appendix 21 Clinical data – Pathogenicity questioned.

Case number	Sex	Age onset	Age death	FH	Clinical Dx	Neuropath Dx	Chromosome	Position	Gene	Transcript	Protein Change	1000G MAF (%)	ESP6500 MAF (%)	ExAC MAF (%)	SIFT	PolyPhen2
47	Female	54	58	N	AD	AD	4	90749307	SNCA	c.150T>G	p.H50Q	NS	0	0	Tolerated	Benign
48	Male		65	N	Control	Consistent with aging	4	90749307	SNCA	c.150T>G	p.H50Q	NS	0	0	Tolerated	Benign
49	Male	59	65	N	MND	MND	19	15295168	NOTCH3	c.2504delG	p.C835fs*25	NS	NS	NS		
50	Female		100	N	Control	Vascular disease	3	87289899	CHMP2B	c.85A>G; c.4-4965A>G	p.I29V	NS	0.02	0.01	Tolerated	Benign
51	Male		39	N	Control	Consistent with aging	21	33039603	SOD1	c.272A>C	p.D91A	0.04	0.08	0.11	Tolerated	Benign
52	Unknown	15	52	N	MS	MS	21	33039603	SOD1	c.272A>C	p.D91A	0.04	0.08	0.11	Tolerated	Benign
53	Male		87	N	Control	Possible early PD changes	21	33039603	SOD1	c.272A>C	p.D91A	0.04	0.08	0.11	Tolerated	Benign
54	Female	61	75	N	PD	PDD	21	33039603	SOD1	c.272A>C	p.D91A	0.04	0.08	0.11	Tolerated	Benign
55	Female		47	N	MND	MND	4	90650386	SNCA	c.349C>A; c.307-2575C>A	p.P117T	NS	0.01	0		Benign
56	Female	58	65	N	Dementia	FTD	1	227073249	PSEN2	c.367C>A	p.P123T	NS	NS	NS	Damaging	Possibly Damaging
57	Female		55	N	MND	MND	14	73637534	PSEN1	c.105_107delCGA; c.117_119delCGA	p.D40del; p.D36del	NS	NS	NS		
58	Male	59	65	N	MND	MND	12	109288126	DAO	c.595C>T	p.R199W	NS	0.06	0.03	Damaging	Probably Damaging
59	Female	87	93	N	Dementia	AD	12	109288126	DAO	c.595C>T	p.R199W	NS	0.06	0.03	Damaging	Probably Damaging
60	Male	58	63	N	PD	PD	12	109288126	DAO	c.595C>T	p.R199W	NS	0.06	0.03	Damaging	Probably Damaging
61	Male	41	41	N	CJD	CJD	14	21161845	ANG; RNASE4	c.122A>T	p.K41I	0.04	0.26	0.14	Tolerated	Benign

Case number	Sex	Age onset	Age death	FH	Clinical Dx	Neuropath Dx	Chromosome	Position	Gene	Transcript	Protein Change	1000G MAF (%)	ESP6500 MAF (%)	ExAC MAF (%)	SIFT	PolyPhen2
62	Male		32	N	Control	Consistent with aging	14	21161845	ANG; RNASE4	c.122A>T	p.K41I	0.04	0.26	0.14	Tolerated	Benign
63	Male		74	N	DLB	DLB	14	21161845	ANG; RNASE4	c.122A>T	p.K41I	0.04	0.26	0.14	Tolerated	Benign
64	Female		88	N	Control	High BRAAK	14	21162130	ANG; RNASE4	c.122A>T	p.P136L	NS	0.01	0	Damaging	Probably Damaging
65	Female		46	N	Control	Consistent with aging	16	31202727	FUS	c.1546C>A; c.1549C>A; c.1537C>A	p.H513N; p.H516N; p.H517N	NS	NS	NS	Damaging	Probably Damaging

Appendix 22 Uncertain Significance – MacArthur criteria.

All relevant data and references are provided to support pathogenicity. Minor Allele Frequencies (MAF) from the 1000Genomes (1000G), NHLBI 6500 ESP and ExAC databases are provided. Key: MND – Motor Neuron Disease, CJD – Creutzfelt-Jakob Disease, AD – Alzheimer's disease, CADASIL – Cerebral Autosomal-Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy. U – Unknown.

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
PSEN2	p.L192I	Novel	U	NS	NS	NS		Located in the topological domain prior to the third transmembrane region	Unknown	Unknown	Unknown	2	0
PSEN2	p.R435Q	Not previously described	U	NS	0.01	0	0.002972	Outside TM domain, but within the C-terminal region which is involved in X-terminal trafficking	Unknown	Unknown	Unknown	1	0
DCTN1	p.E895D	Novel	U	NS	NS	NS		Outside major functional domains	Unknown	Unknown	Unknown	1	0
DCTN1	p.T502M	Novel	U	NS	NS	NS	0.00004375	Within the first coiled coil domain	Unknown	Unknown	Unknown	1	0
DCTN1	p.T138I	Novel	U	NS	NS	NS	0.0006577	Outside major functional domains	Unknown	Unknown	Unknown	1	0
HTRA2	p.L43V	Novel	U	NS	NS	NS		Between N-terminal mitochondrial targeting domain and transmembrane domain	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
HTRA2	p.S15W	Not previously associated with disease	U	NS	NS	0.02	0.00042	Within N-terminal mitochondrial targeting sequence	Unknown	Unknown	Unknown	1	0
GIGYF2	p.Q922P	Novel	U	NS	NS	NS	0.00007178	Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	0
CHMP2B	p.S153L	Seen in a single case of FTD (Ghanim, Guillot-Noel et al. 2010) and also seen in 1 of 750 controls (van Blitterswijk, Vlam et al. 2012)	U	NS	0.02	0	0.000005272	Located in the c-terminal domain, and alters protein stability (van Blitterswijk, Vlam et al. 2012)	Located in the c-terminal domain, and alters protein stability (van Blitterswijk, Vlam et al. 2012)	Unknown	Unknown	1	0
EIF4G1	p.R945Q	Novel	U	NS	NS	NS	0.001589	Mis-sense variant. Functional effects unknown. May alter eIF3/EIF4A binding	Unknown	Unknown	Unknown	1	0
EIF4G1	p.P1075L	Novel	U	NS	NS	NS	0.000003873	Mis-sense variant in well conserved region	Unknown	Unknown	Unknown	1	0
EIF4G1	p.M1356T	Not previously associated with disease	U	NS	0.08	0.04	0.0005649	Mis-sense variant in well conserved region	Unknown	Unknown	Unknown	1	1
MATR3	p.D187E	Not previously associated with disease	U	NS	0.01	0	0.007112	Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
MATR3	p.A745V	Not previously associated with disease	U	NS	NS	0		Mis-sense variant. No convincing specific region for pathogenicity in literature	Unknown	Unknown	Unknown	1	0
SQSTM1	p.P29S	Not previously associated with disease	U	NS	NS	0.02		Mis-sense variant in PB1 domain	Unknown	Unknown	Unknown	1	0
SQSTM1	p.R55C	Not previously associated with disease	U	NS	NS	0	0.006209	Situated in the ZZ_ADA2 domain	Unknown	Unknown	Unknown	1	0
SQSTM1	p.R161W	Not previously associated with disease	U	NS	NS	0		Located in a critical and well-established functional domain (ZZ_ADA2)	Unknown	Unknown	Unknown	1	0
SQSTM1	p.R209C	Novel	U	NS	NS	NS		Mis-sense variant in no specific domain	Unknown	Unknown	Unknown	1	0
SQSTM1	p.P228L	Seen in 1/546 cases and 0/724 controls (Fecto, Yan et al. 2011)	U	NS	0.05	0.01	0.008017	Sits in TRAF6 binding domain	Unknown	Unknown	Unknown	1	0
C9orf72	p.D419N	Novel	U	NS	NS	NS	0.000004325	Unknown functional domain	Unknown	Unknown	Unknown	1	0
C9orf72	p.I413M	Not previously associated with disease	U	NS	NS	0		Unknown functional domain	Unknown	Unknown	Unknown	1	0
C9orf72	p.D347Y	Novel	U	NS	NS	NS	0.000001552	Unknown functional domain	Unknown	Unknown	Unknown	1	0
C9orf72	p.N207del	Not previously associated with disease	U	NS	NS	0		Amino acid 207 is known to have a polymorphism p.N207S	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
C9orf72	p.I17M	Novel	U	NS	NS	NS		Unknown	Unknown	Unknown	Unknown	1	0
VCP	p.R113S	Novel	U	NS	NS	NS	0.00006839	Contained within the AAA2 region and exon 4 in which pathogenic mutations previously described	Unknown	Unknown	Unknown	1	0
DAO	p.R2H	Not previously associated with disease	U	0	0.02	0.01		Situated in nucleotide binding domain. Functional impairment unclear	Unknown	Unknown	Unknown	1	0
DAO	p.F90V	Not previously associated with disease	U	NS	0.03	0.03		Functional impairment unclear	Unknown	Unknown	Unknown	1	0
DAO	p.P103L	Not previously associated with disease	U	0.06	0	0.04	0.000006281	Functional impairment unclear	Unknown	Unknown	Unknown	1	1
DAO	p.R274G	Novel	U	NS	NS	NS		Situated in well conserved region	Unknown	Unknown	Unknown	1	0
PSEN1	p.R27R	Novel	U	NS	NS	NS		Splice-site altering variant. Predicted by MaxEntScan to result in an in-frame insertion	Unknown	Unknown	Unknown	1	0
FUS	p.P18S	Seen in 2 of 454 cases and 0 of 450 controls(Belzil, Daoud et al. 2011), however has a relatively high freq in international databases.	U	NS	0.02	0	0.005662	Located in well conserved N-terminal region	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
FUS	p.S135N	Seen in 1 of 168 SALS patients (Rademakers, Stewart et al. 2010)	U	0.2	0.01	0.03	0.00002265	Located in the QGSY-Region. Additional variants in these region are not pathogenic (Cruts, Theuns et al. 2012)	Unknown	Unknown	Unknown	1	0
FUS	Splice-site alteration	Novel	U	NS	NS	NS	0.006966	Splice site altering variant. This is highly likely to result in exon skipping (MaxEntScore decrease 100%)	Unknown	Unknown	Unknown	1	0
VPS35	p.N384S	Not previously associated with disease	U	NS	NS	0	0.000007962	Mis-sense variant. Regional function unclear	Unknown	Unknown	Unknown	1	0
PFN1	p.E117G	Seen in 3/1090 patients with ALS and 3/7560 controls (p=0.03) (Wu, Fallini et al. 2012).This variant was also seen in 0.11% of 13089 controls and 0.25% of 5188 cases (p=0.036) with an odds ratio of 2.44 for ALS (Fratta, Charnock et al. 2014)	U	NS	NS	NS	0.003724	Located close to actin binding domain	Doesn't reduce bound actin (Wu, Fallini et al. 2012)	Moderate levels of protein aggregation (Wu, Fallini et al. 2012)	Partly reduces axonal outgrowth in vitro (Wu, Fallini et al. 2012)	5	0
GRN	p.V8M	Novel	U	NS	NS	0		Mis-sense variant. Regional function unclear	Unknown	Unknown	Unknown	1	0

Variant		Genetic					Informatic		Experimental			Number of cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
GRN	Splice site alteration	Seen in one previous case causing a corticobasal syndrome (Masellis, Momeni et al. 2006) ,also seen in very low freq in controls	U	NS	0.01	0	0.0002594	Alters splicing	Induces haploinsufficiency (Masellis, Momeni et al. 2006)	Unknown	Unknown	1	0
GRN	p.R298H	Seen previously in a single case of FTLD-U (Yu, Bird et al. 2010)	U	NS	NS	0.01		Mis-sense variant in the Granulin-4 domain	Unknown	Unknown	Unknown	1	0
MAPT	p.A2S	Novel	U	NS	NS	NS		Located after Granulin-7 domain	Unknown	Unknown	Unknown	1	0
MAPT	p.P184L	Novel	U	NS	NS	NS		Located outside the tubulin binding domain	Unknown	Unknown	Unknown	1	0
PRNP	p.D167G	Seen in 1 case of CJD before (Bishop, Pennington et al. 2009)	U	NS	NS	NS		Any disruption in coding domain may affect function	Unknown	Unknown	Unknown	1	0
APP	p.V173F	Novel	U	NS	NS	NS	0.000002056	Not within TM domain	Unknown	Unknown	Unknown	1	0
APP	p.V97I	Not previously associated with disease	U	NS	NS	0	0.00002864	Not within TM domain	Unknown	Unknown	Unknown	1	0
CHCHD10	p.A13T	Novel	U	NS	NS	NS	0.001633	Sits within the N-terminal domain	Unknown	Unknown	Unknown	1	0
UBQLN2	p.S222G	Novel	U	NS	NS	NS		Sits within a nonstructured domain	Unknown	Unknown	Unknown	1	0

Appendix 23 Uncertain significance – ACMG criteria.

The gene and protein alteration are shown together with whether there is evidence to support strong pathogenicity (PS), moderate pathogenicity (PM) or supporting pathogenicity (PP) according to ACMG 2015 criteria (Richards, Aziz et al. 2015). In addition, evidence supporting a benign nature of the variant is also provided (BS and BP). Final ACMG classification based on both the pre-study and post-study data are shown.

		Pathogenic																Benign												ACMG classification									
Variant		Very Strong	Strong				Moderate						Supporting					Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
PSEN2	p.L192I							1						1	1																				1					1
PSEN2	p.R435Q													1																					1					1
DCTN1	p.E895D							1																											1					1
DCTN1	p.T502M							1																											1					1
DCTN1	p.T138I																																		1					1
HTRA2	p.L43V								1																										1					1
HTRA2	p.S15W								1							1																			1					1
GIGYF2	p.Q922P							1																											1					1
CHMP2B	p.S153L						1							1																					1					1
EIF4G1	p.R945Q							1																											1					1
EIF4G1	p.P1075L							1																											1					1
EIF4G1	p.M1356T																		1																1					1
MATR3	p.D187E																																		1					1
MATR3	p.A745V							1																											1					1
SQSTM1	p.P29S							1																				1								1					1
SQSTM1	p.R55C							1						1																					1					1
SQSTM1	p.R161W							1						1																					1					1
SQSTM1	p.R209C													1																					1					1

Variant		Pathogenic																Benign												ACMG classification									
		Very Strong	Strong				Moderate						Supporting					Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
SQSTM1	p.P228L							1						1																					1					1
C9orf72	p.D419N							1																											1					1
C9orf72	p.I413M							1																											1					1
C9orf72	p.D347Y							1																											1					1
C9orf72	p.N207del							1																											1					1
C9orf72	p.I17M							1							1																				1					1
VCP	p.R113S						1	1						1	1																				1					1
DAO	p.R2H																																		1					1
DAO	p.F90V																																		1					1
DAO	p.P103L																																		1					1
DAO	p.R274G							1																				1							1					1
PSEN1	p.R27R																																		1					1
FUS	p.P18S													1														1							1					1
FUS	p.S135N													1														1							1					1
FUS	Splice-site alteration													1														1							1					1
VPS35	p.N384S							1																				1							1					1
PFN1	p.E117G					1	1							1																					1					1
GRN	p.V8M							1																											1					1
GRN	Splice site alteration							1						1																					1					1
GRN	p.R298H							1																											1					1
MAPT	p.A2S							1																				1							1					1
MAPT	p.P184L							1							1																				1					1
PRNP	p.D167G							1						1	1																				1					1
APP	p.V173F																																		1					1

		Pathogenic																Benign												ACMG classification									
Variant		Very Strong	Strong				Moderate						Supporting					Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
APP	p.V97I																																			1					1
CHCHD10	p.A13T							1																				1								1					1
UBQLN2	p.S222G							1																				1								1					1

Appendix 24 Clinical data – variants of uncertain significance.

Gene	Protein alteration	Chromosome	Position	1000G MAF (%)	ESP6500 MAF (%)	Exac MAF (%)	SIFT	PolyPhen2	ACMG	Cases with variant	Controls with variant		AD	CBD	CJD	DLB	FTD-ALS	HD	MSA	Other	PD	PSP	VD	VD/AD	VD/control		Control	Control (High Braak)	Young control
PSEN2	p.L192I	1	227076537	NS	NS	NS	D	B	Uncertain Significance	2	0					1												1	
PSEN2	p.R435Q	1	227083237	NS	0.01	0	T	B	Uncertain Significance	1	0																		1
DCTN1	p.E895D	2	74593386	NS	NS	NS	T	B	Uncertain Significance	1	0																		1
DCTN1	p.T502M	2	74596485	NS	NS	NS	T	B	Uncertain Significance	1	0						1												
DCTN1	p.T138I; p.T11I; p.T145I	2	74600074	NS	NS	NS	T	B	Uncertain Significance	1	0						1												
HTRA2	p.L43V	2	74756550	NS	NS	NS	T	B	Uncertain Significance	1	0										1								
HTRA2	p.S15W	2	74756713	NS	NS	0.02	T	PrD	Uncertain Significance	1	0										1								
GIGYF2	p.Q922P	2	233697739	0	0	NS	T	PrD	Uncertain Significance	1	0					1													
CHMP2B	p.S153L	3	87302911	NS	0.02	0	T	B	Uncertain Significance	1	0						1												
EIF4G1	p.R945Q	3	184043401	NS	NS	NS	T	B	Uncertain Significance	1	0					1													

Gene	Protein alteration	Chromosome	Position	1000G MAF (%)	ESP6500 MAF (%)	Exac MAF (%)	SIFT	PolyPhen2	ACMG	Cases with variant	Controls with variant		AD	CBD	CJD	DLB	FTD-ALS	HD	MSA	Other	PD	PSP	VD	VD/AD	VD/control		Control	Control (High Braak)	Young control
EIF4G1	p.P1075L	3	184044687	NS	NS	NS	T	B	Uncertain Significance	1	0					1													
EIF4G1	p.M1356T	3	184046529	NS	0.08	0.04	D	PrD	Uncertain Significance	1	1										1						1		
MATR3	p.D187E	5	138643665	NS	0.01	0	T	PoD	Uncertain Significance	1	0																		1
MATR3	p.A745V	5	138661214	NS	NS	0	T	B	Uncertain Significance	1	0						1												
SQSTM1	p.P29S	5	179248021	NS	NS	0.02		B	Uncertain Significance	1	0						1												
SQSTM1	p.R55C	5	179250971	NS	NS	0	D	PrD	Uncertain Significance	1	0						1												
SQSTM1	p.R161W	5	179251037	NS	NS	0	D	PoD	Uncertain Significance	1	0																		1
SQSTM1	p.R209C	5	179251275	NS	NS	NS	D	B	Uncertain Significance	1	0						1												
SQSTM1	p.P228L	5	179252155	NS	0.05	0.01	D	B	Uncertain Significance	1	0						1												
C9orf72	p.D419N	9	27548559	NS	NS	NS	D	B	Uncertain Significance	1	0																		1
C9orf72	p.I413M	9	27548575	NS	NS	0	D	B	Uncertain Significance	1	0						1												

Gene	Protein alteration	Chromosome	Position	1000G MAF (%)	ESP6500 MAF (%)	Exac MAF (%)	SIFT	PolyPhen2	ACMG	Cases with variant	Controls with variant		AD	CBD	CJD	DLB	FTD-ALS	HD	MSA	Other	PD	PSP	VD	VD/AD	VD/control		Control	Control (High Braak)	Young control
C9orf72	p.D347Y	9	27556611	NS	NS	NS		B	Uncertain Significance	1	0						1												
C9orf72	p.N207del	9	27561627	NS	NS	0			Uncertain Significance	1	0						1												
C9orf72	p.I17M	9	27567068	NS	NS	NS		PoD	Uncertain Significance	1	0						1												
VCP	p.R113S	9	35066780	NS	NS	NS		PoD	Uncertain Significance	1	0																		1
DAO	p.R2H	12	109278787	0	0.02	0.01	D	B	Uncertain Significance	1	0						1												
DAO	p.F90V	12	109281299	NS	0.03	0.03	T	PoD	Uncertain Significance	1	0						1												
DAO	p.P103L	12	109281339	0.06	0	0.04	T	PoD	Uncertain Significance	1	1						1										1		
DAO	p.R274G	12	109293159	NS	NS	NS	T	B	Uncertain Significance	1	0						1												
PSEN1	p.R27R	14	73614808	NS	NS	NS			Uncertain Significance	1	0																		1
FUS	p.P18S	16	31193847	NS	0.02	0		B	Uncertain Significance	1	0						1												
FUS	p.S135N; p.S134N	16	31195598	0.2	0.01	0.03	T	B	Uncertain Significance	1	0						1												

Gene	Protein alteration	Chromosome	Position	1000G MAF (%)	ESP6500 MAF (%)	Exac MAF (%)	SIFT	PolyPhen2	ACMG	Cases with variant	Controls with variant		AD	CBD	CJD	DLB	FTD-ALS	HD	MSA	Other	PD	PSP	VD	VD/AD	VD/control		Control	Control (High Braak)	Young control
FUS	Splice-site alteration	16	31195718	NS	NS	NS			Uncertain Significance	1	0						1												
VPS35	p.N384S	16	46708235	NS	NS	0	T	B	Uncertain Significance	1	0																		1
PFN1	p.E117G	17	4849267	NS	NS	NS	D	B	Uncertain Significance	5	0		1		1		1			1									1
GRN	p.V8M	17	42426554	NS	NS	0	T	B	Uncertain Significance	1	0											1							
GRN	Splice site alteration	17	42428169	NS	0.01	0			Uncertain Significance	1	0		1																
GRN	p.R298H	17	42428788	NS	NS	0.01	T	PrD	Uncertain Significance	1	0																		1
MAPT	p.A2S	17	44039707	NS	NS	NS		B	Uncertain Significance	1	0																	1	
MAPT	p.P184L	17	44073846	NS	NS	NS	D	PrD	Uncertain Significance	1	0		1																
PRNP	p.D167G	20	4680366	NS	NS	NS	D	PoD	Uncertain Significance	1	0				1														
APP	p.V173F	21	27423356	NS	NS	NS	T	PrD	Uncertain Significance	1	0																	1	
APP	p.V97I	21	27425563	NS	NS	0	T	B	Uncertain Significance	1	0		1																

Gene	Protein alteration	Chromosome	Position	1000G MAF (%)	ESP6500 MAF (%)	Exac MAF (%)	SIFT	PolyPhen2	ACMG	Cases with variant	Controls with variant		AD	CBD	CJD	DLB	FTD-ALS	HD	MSA	Other	PD	PSP	VD	VD/AD	VD/control		Control	Control (High Braak)	Young control
CHCHD10	p.A13T	22	24110025	NS	NS	NS	T	B	Uncertain Significance	1	0						1												
UBQLN2	p.S222G	X	56590970	NS	NS	NS	T	B	Uncertain Significance	1	0											1							

Appendix 25 Pathogenic recessive mutations - MacArthur criteria.

All relevant data and references are provided to support pathogenicity. Minor Allele Frequencies (MAF) from the 1000Genomes (1000G), NHLBI 6500 ESP and ExAC databases are provided. Key: MND – Motor Neuron Disease, CJD – Creutzfelt-Jakob Disease, AD – Alzheimer's disease, CADASIL – Cerebral Autosomal-Dominant Arteriopathy with Subcortical Infarcts and Leukoencephalopathy. U – Unknown.

Variant		Genetic					Informatic		Experimental			Study cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
PARK2	p.P437L; p.P409L; p.P288L	Previously described in an early onset case of PD (Mellick, Siebert et al. 2009)	N/A	0.04	0.24	0.15	1.13E-04	Situated in the RING 2 domain of the protein. RING2 forms a hydrophobic interface with the UPD, burying the catalytic Cys431, which is part of a conserved catalytic triad (Sriram, Li et al. 2005)	Almost completely abolishes the binding ability of Parkin (Sriram, Li et al. 2005)	Unknown	U	1	0
OPTN	p.R217*	Novel	N/A	NS	NS	NS	N/A	Premature stop codon. This is prior to the coiled coil 2 domain necessary for the binding to ubiquitin and ubiquitin receptor-interacting protein (Maruyama, Morino et al. 2010)	The previously described homozygous nonsense mutation (p.Q398X) showed no ability to inhibit NF-kB (Maruyama, Morino et al. 2010)	Unknown	U	1	0
SOD1	p.D91A	Seen in several pedigrees in the homozygous state (Andersen, Nilsson et al. 1995)	Fully penetrant	0.04	0.05	0.11		The mutation destabilises the apoSOD1 monomer (Bystrom, Andersen et al. 2010). Pathogenic mutations in SOD1 exist throughout the gene.	Increases susceptibility to oxidative stress and impairment of respiratory complexes (Ferri, Cozzolino et al. 2006)	Mild similar phenotype in mice (Jonsson, Graffmo et al. 2006)	U	1	0

Variant		Genetic					Informatic		Experimental			Study cases	
Gene	Protein alteration	Association (cases vs controls)	Segregation	1000G MAF (%)	6500 ESP MAF (%)	ExAC MAF (%)	Conservation	Location (appropriate)	Gene disruption	Phenotype recapitulation	Rescue	Number of cases	Number of controls
VPS13A	p.L1841*; p.L1802*	Novel	Unknown	NS	NS	NS	1.53E-05	No specific functional domain known in association with disease	Unknown	Unknown	U	1	0
VPS13A	p.W2347fs*36; p.W2308fs*36	Novel	Unknown	NS	NS	NS	2.17E-06	No specific functional domain known in association with disease	Unknown	Unknown	U		
PARK2	p.G430D; p.G281D; p.G402D	Pathogenic in homozygous state (Mellick, Siebert et al. 2009)	N/A	0	0.02	0.01	1.13E-04	Located in exon 12 of the PARK2 gene	NF-kB activation are reduced by PD-linked parkin pathogenic mutations(Henn, Bouman et al. 2007), and alters mitochondrial translocation and clearance (Geisler, Holmstrom et al. 2010)	Unknown	U	1	0
PARK2	p.R275W; p.R126W; p.R247W	Pathogenic in compound heterozygous cases (Klein, Djarmati et al. 2005)	N/A	0.04	0.2	0.21	6.41E-03	Located in exon 7	Alters mitochondrial translocation and clearance (Geisler, Holmstrom et al. 2010)and alters parkin localization (Cookson, Lockhart et al. 2003)	Causes cytoplasmic inclusions (Cookson, Lockhart et al. 2003)	U		

Appendix 26 Pathogenic recessive mutations – ACMG criteria.

The gene and protein alteration are shown together with whether there is evidence to support strong pathogenicity (PS), moderate pathogenicity (PM) or supporting pathogenicity (PP) according to ACMG 2015 criteria (Richards, Aziz et al. 2015). In addition, evidence supporting a benign nature of the variant is also provided (BS and BP). Final ACMG classification based on both the pre-study and post-study data are shown.

		Pathogenic															Benign												ACMG classification									
Variant		Very strong	Strong				Moderate						Supporting					Strong							Supporting							Pre-assessment					Post-assessment				
Gene	Protein alteration	PSV1	PS1	PS2	PS3	PS4	PM1	PM2	PM3	PM4	PM5	PM6	PP1	PP2	PP3	PP4	PP5	BS1	BS2	BS3	BS4	BS5	BS6	BS7	BP1	BP2	BP3	BP4	BP5	BP6	BP7	P	LP	B	LB	US	P	LP	B	LB	US
PARK2	p.P437L; p.P409L; p.P288L		1		1										1																1					1				
OPTN	p.R217*	1						1		1																					1					1				
SOD1	p.D91A			1		1		1					1	1	1																1					1				
VPS13A	p.L1841*; p.L1802*	1					1								1	1															1					1				
VPS13A	p.W2347fs*36; p.W2308fs*36	1					1			1					1	1															1					1				
PARK2	p.G430D; p.G281D; p.G402D		1		1		1																								1					1				
PARK2	p.R275W; p.R126W; p.R247W		1		1		1																								1					1				

Appendix 27 Clinical data – Pathogenic recessive mutations.

Case number	Sex	Age onset	Age death	Family History	Clinical Dx	Neuropath Dx	Chromosome	Position	Gene	Transcript	Protein Change	1000G MAF (%)	ESP6500 MAF (%)	ExAC MAF (%)	SIFT	PolyPhen2
41	Female		32	N	Control	Appearances consistent with aging	6	161771219	PARK2	c.1310C>T; c.863C>T; c.1226C>T	p.P437L; p.P409L; p.P288L	0.04	0.24	0.15	Damaging	Probably Damaging
42	Male		54	N	MND	FTD-ALS	10	13160910	OPTN	c.649A>T	p.R217*	NS	NS	NS	Stop-gain	Stop-gain
43	Male	55	65	N	MSA	Unclear: Atypical Multisystem degeneration	21	33039603	SOD1	c.272A>C	p.D91A	0.04	0.05	0.11	Tolerated	Benign
44	Male		40	N	Chorea-acanthocytosis	Chorea-acanthocytosis	9	79936191	VPS13A		p.L1841*; p.L1802*	NS	NS	NS	Stop-gain	Stop-gain
							9	79959075	VPS13A		p.W2347fs*36; p.W2308fs*36	NS	NS	NS	Frameshift	Stop-gain
45	Female		69	N	PD	PD	6	161771240	PARK2	c.842G>A; c.1289G>A; c.1205G>A	p.G430D; p.G281D; p.G402D	0	0.02	0.01	Damaging	Probably Damaging
							6	162206852	PARK2	c.823C>T; c.376C>T; c.739C>T	p.R275W; p.R126W; p.R247W	0.04	0.2	0.21	Probably Damaging	Probably Damaging

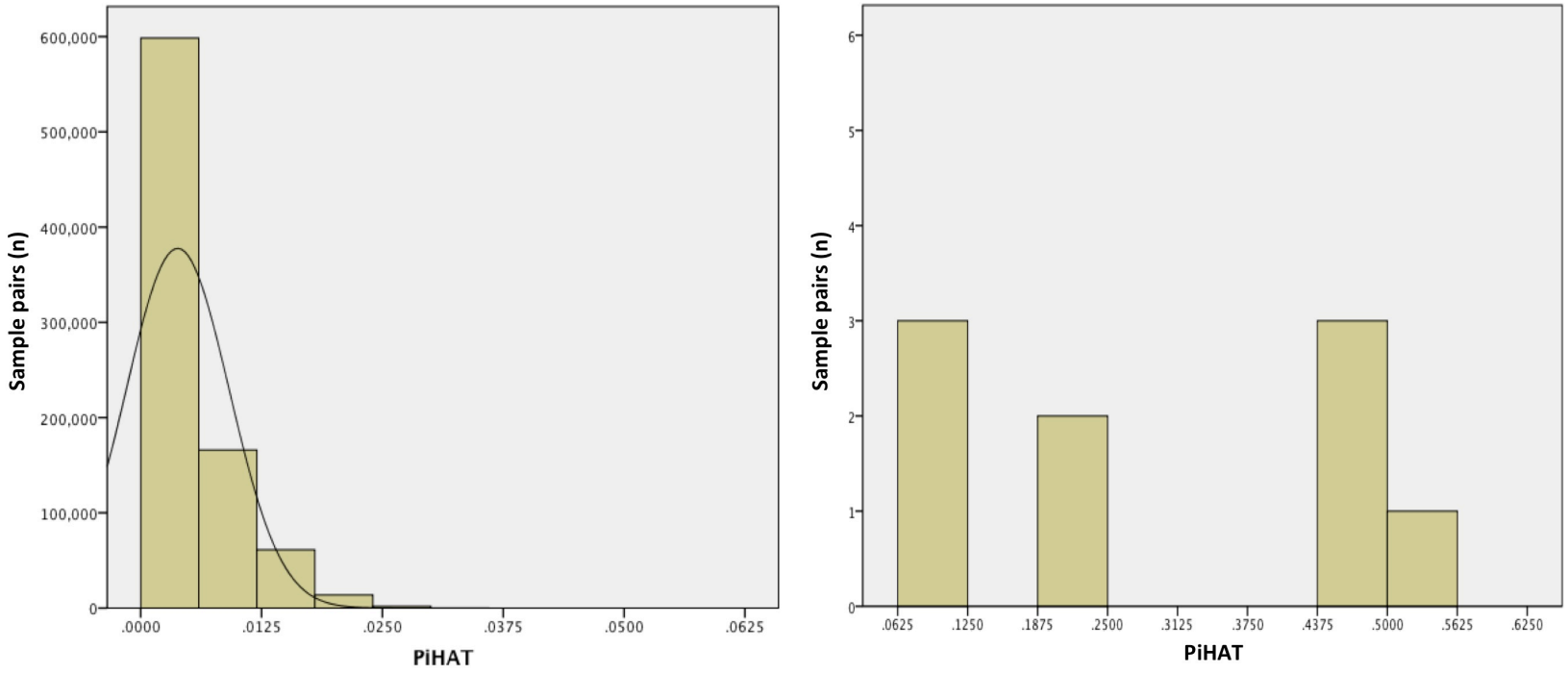
Appendix 28 Cases testing positive for the *C9orf72* hexanucleotide repeat expansion.

Key; FH – Family History. MND – Motor Neuron Disease, FTD – Frontotemporal Dementia, FTD-ALS – Frontotemporal dementia and Amyotrophic Lateral Sclerosis.

Case number	Sex	Age onset (years/SD)	Age death (years/SD)	FH of disease	Clinical Diagnosis	Neuropathological diagnosis
46	Female		63	N	MND	MND
47	Female		43	N	MND	MND
48	Male		71	N	MND	FTD-ALS
49	Female	73	75	N	Vascular dementia	FTD
50	Male		63	N	Parkinsonism	FTD
51	Female	52	56	N	Dementia - unspecified	FTD
52	Male	63	64	N	MND	MND
53	Male		56	N	Dementia - unspecified	FTD
54	Male	57	60	N	FTD-ALS	FTD-ALS
55	Female	49	50	Y	MND	MND
56	Male		67	N	MND	MND
57	Male		60	Y	FTD-ALS	FTD-ALS
58	Male		55	Y	MND	FTD-ALS
59	Male		62	Y	FTD-ALS	FTD-ALS
60	Male	54	57	N	MND	MND
Mean (SD)	**Male (75%)**	**58 (8.8)**	**60.1 (7.9)**	**Y (40%)**		

Appendix 29 Identity by descent analysis for the entire cohort.

Subfigure (left) shows cases with cryptic ancestry suggesting between 1st-4th degree relatedness. Right shows PiHAT scores of 0.625, 0.125, 0.25 and 0.5 were considered as 4th, 3rd, 2nd and 1st degree relatives respectively. A description of the 9 pairs with cryptic ancestry can be seen in Supplemental Table X



Appendix 30 Clinical and pathological features of inter-related individuals.

A summary of the clinical and pathological diagnosis is provided for each case together with the Pi-Hat score (see methods), and the subsequent inferred degree of relatedness based on that score. The results of the diagnostic outcome in relation to this information is provided in the results column.

Case 1:			**Case 2:**			**PI-HAT**	**Inferred relatedness**	**Results**
Age of onset (years)	**Age death (years)**	**Diagnosis**	**Age onset (years)**	**Age death (years)**	**Diagnosis**			
N/A	58	SCA7	N/A	50	SCA - molecular diagnosis unknown	0.5637	1st degree relative	SCA7 expansion confirmed in case 2
N/A	43	MND	N/A	42	MND	0.4797	1st degree relative	SOD1 variant (p.D102N) confirmed in both
N/A	68	AD	N/A	64	AD	0.2234	2nd degree	Relationship confirmed through clinical note review. No familial allele or RF identified
N/A	89	Control (High Braak) – Braak stage 4	N/A	90	Other; Lewy body disease (extensive neocortical Lewy bodies and parenchymal amyloid)	0.4827	1st degree relative	No risk factor or familial cause identified
N/A	82	Control (BNE-stage 1)	N/A	83	DLB	0.4689	1st degree relative	No risk factor or familial cause identified
N/A	24	Other: Epilepsy	N/A	45	Sporadic CJD	0.2395	2nd degree	No risk factor or familial cause identified
N/A	85	Control (Braak < 2)	N/A	39	MND	0.0823	3-4th degree	No risk factor or familial cause identified
N/A	89	PSP	N/A	77	Sporadic CJD	0.0673	4th degree	No risk factor or familial cause identified
N/A	83	AD	N/A	50	MND	0.0659	4th degree	No risk factor or familial cause identified

Appendix 31 Cases with rare copy number variants in established disease-associated genes.

Rare copy number variants (CNVs, present in <1% of reference databases) and spanning all genes previously associated with neurodegenerative disorders are shown. Clinical details (diagnosis, age of disease onset, age of death) of each patient together with the start and end positions of the CNV, copy number change, and associated disease gene within the CNV

Case	Diagnosis	Age onset	Age death	Gene	Chromosome	Start position	End position	width (bases)	Copy number change
Case 61	Alzheimer's disease		59	*APP*	21	26982980	28168090	1185111	3
Case 62	Alzheimer's disease	-	79	*PRPH*	12	49676010	49689404	13395	3
Case 63	Alzheimer's disease	-	85	*PRPH*	12	49676010	49689404	13395	3
Case 64	Alzheimer's disease	58	69	*PRPH*	12	49676010	49689404	13395	3
Case 65	Alzheimer's disease	-	77	*PRPH*	12	49676010	49689404	13395	3
Case 66	Alzheimer's disease	83	92	*PRPH*	12	49676010	49689404	13395	3
Case 67	Alzheimer's disease	-	92	*PRPH*	12	49676010	49691567	15558	3
Case 68	Control	-	60	*PRPH*	12	49676010	49689404	13395	3
Case 69	Atypical tauopathy with AD features	64	80	*PRPH*	12	49676010	49689404	13395	3
Case 70	Motor Neuron Disease	-	83	*PRPH*	12	49676010	49691225	15216	3
Case 71	Alzheimer's disease	-	82	*DAO*	12	109294301	109309971	15671	3
Case 72	Control (Braak 1)		74	*APOE*	19	45412040	45414451	2412	3
Case 73	Sporadic CJD	73	73	*SPG11*	15	44890903	44967847	76944	3
Case 74	Alzheimer's disease	-	72	*SCARB2*	4	77033590	77105700	72110	3
Case 75	Psychiatric disorder (possible NBIA at post-mortem)	18	85	*PARK2*	6	161815583	161859655	44073	1
Case 76	Motor Neuron Disease	-	53	*PARK2*	6	161815583	161859655	44073	1
Case 77	Sporadic CJD	72	73	*PARK2*	6	161817168	161980513	163346	1
Case 78	Frontotemporal Dementia	-	65	*PARK2*	6	162426517	162447612	21096	1
Case 79	Alzheimer's disease	-	58	*LRRK2*	12	40745181	40757533	12353	1

Appendix 32 Genetic variants in established disease risk-factor genes.

Variants with a MAF <3% in either the 1000G or ESP6500 were selected. Statistical association through a case-burden test was performed at the gene level using SKAT-O for genes known to act as risk factors for each disease. Subsequent analysis at the variant level was performed using Chi-squared testing. Odds ratios were calculated where Chi-squared were p-values were significant ($p<0.05$). Variants were defined as risk factors where either previous large cohort sequencing studies or metanalysis provided statistical evidence of an association with disease, or where it was present within this study (identified by 'Y' in final column).

Disease	Gene	(SKAT-0)	Cases (%)	Cont (%)	cDNA pos	AA change	1000G (MAF) %	ESP 6500 (MAF) %	Case (n)	Cont (n)	X2	OR	95% CI	Association	Study	RF
AD	TREM2	p=0.017	20 (7%)	8 (3%)	c.668C>T	p.T223I	0.14	0.05	0	1	0.48		-	Seen in a single AD case previously	(Cuyvers, Bettens et al. 2014)	
								0	1	0	1.0	-	-	None		
					c.407G>A	p.R136W		0.02	0	1	0.48	-	-	Seen in a single case previously	(Jin, Benitez et al. 2014)	
					c.292C>T	p.R98W	0.02	0.01	1	0	1.0	-	-	Strong association with AD ([CI], 2.77 - 9.16); P = 9.0×10−9]	(Guerreiro, Wojtas et al. 2013)	Y
					c.259G>A	p.D87N	0.06	0.09	0	1	0.48	-	-	Not reported		
					c.185G>A	p.R62H	0.5	0.78	12	5	0.14	2.2	0.72-8.3	Confers a significantly increased risk of disease (P = 2.36 × 10^{-4}, OR = 2.36 (1.47–3.80) in 2082 cases and 1648 controls.	(Jin, Benitez et al. 2014)	Y
					c.140G>A	p.R47H	0.2	0.18	7	0	0.015	-	-	Strong association with AD (p = 2.93×10(-17))	(Cuyvers, Bettens et al. 2014)	Y
	APOE	p=0.17	5 (1%)	2 (0%)	c.137T>C	p.L46P	0.08	0.12	3	1	0.63	2.7	0.2-	Significant risk of AD	(Kamboh, Aston et al.	Y

Disease	Gene	(SKAT-0)	Cases (%)	Cont (%)	cDNA pos	AA change	1000G (MAF) %	ESP 6500 (MAF) %	Case (n)	Cont (n)	X2	OR	95% CI	Association	Study	RF
													145	(13.2 (3.12-55.9)	1999)	
					c.761T>A	p.V254E		0.08	1	0	1.0	-	-	Not reported		
					c.805C>G	p.R269G			0	1	0.48	-	-	Novel		
					c.895G>C	p.E299Q		0	1	0	1.0	-	-	Novel		
DLB	GBA	0.0075	8 (13%)	10 (4%)	c.1301T>C;	p.L483P;	0.34	0.03	3	0	0.0068	-	-	Significant risk factor for PD (11.68 (5.23,26.06)	(Chen, Li et al. 2014)	Y
					c.962C>T;	p.T321M	0.18	0.52	3	2	0.05	6.6	0.73-80	Seen in 9% of PD patients and only 0.5% of controls previously, with an OR . 6.97, 95% CI 0.93–52.02; (P . 0.03)	(Seto-Salvia, Pagonabarraga et al. 2012) (Lesage, Condroyer et al. 2011)	Y
					c.946G>A;	p.E365K;	0.5	0.89	1	7	1	0.61	0.01-4.8	OR, 2.97; 95% CI: 1.3-6.4 for PD	(Duran, Mencacci et al. 2013)	Y
					c.665C>A;	p.A271D	0	0	1	0	0.19	-	-	Novel variant		
					c.-146-596C>G	p.S24R	0	0	0	1	1	-	--	Novel variant		
	APOE	0.759	0 (0%)	2 (0%)	c.137T>C;	p.L46P;	0.08	0.12	0	1	1	-	-	Significant risk of AD (OR 3.2 (CI; 3.12-55.9)	(Kamboh, Aston et al. 1999)	Y

Disease	Gene	(SKAT-0)	Cases (%)	Cont (%)	cDNA pos	AA change	1000G (MAF) %	ESP 6500 (MAF) %	Case (n)	Cont (n)	X2	OR	95% CI	Association	Study	RF
					c.883C>G	p.R295G	0	0	0	1	1	-	-	Not reported		
	TREM2	0.0084	7 (12%)	8 (3%)	c.483-101C>T	p.T223I	0.14	0.05	0	1	1	-	-	Seen in a single AD case previously	(Cuyvers, Bettens et al. 2014)	
					c.407G>A	p.R136Q	0	0.02	0	1	1	-	-	Seen in a single case previously	(Jin, Benitez et al. 2014)	
					c.259G>A	p.D87N	0.06	0.09	0	1	1	-	-	Not reported		
					c.185G>A	p.R62H	0.5	0.78	7	5	0.0024	3.2	1.7-27	Not reported in association with DLB, but reported with AD (see above)	(Jin, Benitez et al. 2014)	Y
	LRRK2	0.821	8 (13%)	46 (19%)	c.368C>G	p.T123R	0	0	1	0	0.19	-	-	Novel variant		
					c.518A>G	p.N173S	0	0	0	1	1	-	-	Seen in 1 PD case	(Pihlstrom, Rengmark et al. 2014)	
					c.1000G>A	p.E334K	0.14	0.07	0	1	1	-	-	Significance unknown	(Ross, Soto-Ortolaza et al. 2011)	
					c.1095C>A	p.H365Q	0	0	0	1	1	-	-	Novel variant		
					c.2189T>C	p.L730P	0	0	0	1	1	-	-	Novel variant		
					c.2611A>G	p.K871E	0	0	0	1	1	-	-	Significance unknown	(Ross, Soto-Ortolaza et al. 2011)	
					c.4165G>A	p.V1389I	0	0.02	0	1	1	-	-	Not reported		
					c.4541G>A	p.R1514Q	0.16	0.52	0	2	1	-	-	OR 1.13 (0.85, 1.49)	(Ross, Soto-Ortolaza et al. 2011)	

Disease	Gene	(SKAT-0)	Cases (%)	Cont (%)	cDNA pos	AA change	1000G (MAF) %	ESP 6500 (MAF) %	Case (n)	Cont (n)	X2	OR	95% CI	Association	Study	RF
					c.4624C>T	p.P1542S	1.32	2.34	2	17	0.55	0.5	0.05-2.1	OR 0.90 (0.77, 1.06)	(Ross, Soto-Ortolaza et al. 2011)	
					c.4894G>C	p.E1632Q	0	0	1	0	0.20	-	-	Novel		
					c.4937T>C	p.M1646T	0.48	1.13	2	8	1	1.06	0.11-5.5	OR 1.43 (1.15, 1.78) (p=0.0012)	(Ross, Soto-Ortolaza et al. 2011)	Y
					c.5183G>T	p.R1728L	0	0.01	0	1	1	-	-	Not reported		
					c.6185_6189delTACTC	p.L2063*	0	0	0	1	1	-	-	Novel		
					c.6241A>G	p.N2081D	0.98	1.35	2	10	1	0.86	0.09-4.1			
					c.6902T>C	p.M2301T	0	0	0	1	1	-	-	Novel		
					c.7151G>T	p.C2384F	0	0	0	1	1	-	-	Novel		
					c.7430G>A	p.R2477Q	0.02	0.02	0	1	1	-	-	Novel		
	SNCA	0.58	1 (0%)	1 (0%)	c.150T>G	p.H50Q	0	0.01	0	1	1	-	-	Previously described as familial cause of disease		
	SCARB2	0.077	2 (0%)	4 (1%)	c.475A>G;	p.M159V	0.38	0.86	0	3	1	-	-	Not reported		
					; c.445G>A	p.V149M	0.1	0.25	0	1	1	-	-	Not reported		

Disease	Gene	(SKAT-0)	Cases (%)	Cont (%)	cDNA pos	AA change	1000G (MAF) %	ESP 6500 (MAF) %	Case (n)	Cont (n)	X2	OR	95% CI	Association	Study	RF
FTD-ALS	SQSTM1	0.0076	29 (13%)	16 (7%)	c.70A>G	p.S24G	0	0	0	1	1.00	-	-	Novel		
					c.85C>T	p.P29S	0	0	1	0	0.47	-	-	Novel		
					c.98C>T	p.A33V	0	0.05	0	2	0.50	-	-	Seen in equal proportion of cases and controls	(van der Zee, Van Langenhove et al. 2014)	
					c.350C>T;	p.A117V	0.08	0.08	0	1	1.00	-	-	Seen in equal proportion of cases and controls	(van der Zee, Van Langenhove et al. 2014)	
					c.100C>T;	p.P34S	0	0.01	1	0	0.47	-	-	Seen in equal proportion of cases and controls	(van der Zee, Van Langenhove et al. 2014)	
					c.163C>T	p.R55C	0	0	1	0	0.47	-	-	Novel		
					c.373C>T	p.R125C	0	0	1	0	0.47	-	-	Novel		
					c.431C>T	p.P144L	0	0.04	1	0	0.47	-	-	Previously reported in 2 cases	(van der Zee, Van Langenhove et al. 2014)	
					c.460A>G	p.K154E	0.24	0.26	6	1	0.06	6.84	0.81-314	Seen in 23 of 2203 patients and 14 of 3899 cases	(van der Zee, Van Langenhove et al. 2014)	
					c.503G>T;	p.G168V	0	0	0	1	1.00	-	-	Novel		
					c.570G>C	p.E190D	0.9	1.87	12	8	0.26	1.76	0.62-4.9	Seen in 131 of 2203 cases and 79 of 3899 controls	(van der Zee, Van Langenhove et al. 2014)	

Disease	Gene	(SKAT-0)	Cases (%)	Cont (%)	cDNA pos	AA change	1000G (MAF) %	ESP 6500 (MAF) %	Case (n)	Cont (n)	X2	OR	95% CI	Association	Study	RF
					c.709C>T	p.R237C	0.9	1.1	1	2	1.00	0.56	0.01-10.7	Seen in 4 of 2203 cases and 1 of 3899 controls	(van der Zee, Van Langenhove et al. 2014)	
					c.1175C>T;	p.P392L	0.24	0.15	4	0	0.05	-	-	Seen in 18 of 2203 cases and 11 of 3899 controls	(van der Zee, Van Langenhove et al. 2014)	
					c.1279G>A	p.A427T	0	0	1	0	0.47	-	-	Novel – considered likely pathogenic in this study		
	CHMP2B	0.1136	2(0)	0(0)	c.83G>A	p.R28Q	0	0.02	1	0	0.47	-	-	Seen previously in one PLS case	[83]	
					c.458C>T	p.S153L	0	0.02	1	0	0.47	-	-	Seen in cases and controls previously	[83]	
	PON1	0.41	5 (2%)	5 (2%)	c.977C>G	p.T326R	0	0	1	0	0.47	-	-	Novel		
					c.602C>T	p.A201V	0.08	0.2	2	3	1.00	0.75	0.06-6.5	Seen in 7 of 1444 cases and 3 of 1159 controls	(Ticozzi, LeClerc et al. 2010)	
					c.55A>G	p.N19D	0.1	0.16	2	2	1.00	1.1	0.08-15	Seen in 8 of 1444 cases and 3/1159 controls	[142]	
	PON2	0.23	4 (1%)	4 (1%)	c.896C>A	p.P299H	0	0	0	1	1.00	-	-	Novel		
					c.304A>G	p.I102V	0	0	1	0	0.47	-	-	Novel		
					c.286delA	p.R96fs*5	0	0	2	1	0.61	2.24	0.12-132	Seen in 7 of 1444 cases and 4/1159 controls	(Ticozzi, LeClerc et al. 2010)	
					c.248G>A	p.G83E	0.02	0	1	0	0.47	-	-	Novel		
	PON3	0.4	3 (1%)	3 (1%)	c.971G>A	p.G324D	0.16	0.23	2	0	0.22	-	-	Seen in 3 of 1444 cases	142]	

Disease	Gene	(SKAT-0)	Cases (%)	Cont (%)	cDNA pos	AA change	1000G (MAF) %	ESP 6500 (MAF) %	Case (n)	Cont (n)	X2	OR	95% CI	Association	Study	RF
														and 3/1159 controls		
					c.436G>A	p.E146K	0.28	0.35	0	1	1.00	-	-	Not previously reported		
					c.94C>T	p.R32*	0.12	0.17	1	2	1.00	0.56	0.01-11	Not previously reported		
	GRN	0.19	13 (5)	10 (4%)	c.970G>A	p.A324T	0.1	0.14	0	1	1.00	-	-	Not associated with AD	(Sassi, Guerreiro et al. 2014)	
					c.1211G>T	p.C404F	0	0	1	0	0.47	-	-	Possibly pathogenic		
					c.1297C>T	p.R433W	0.16	0.22	1	1	1.00	1.11	0.01-87	Previously reported in a single patient	(Guerreiro, Washecka et al. 2010)	
					c.1354delG	p.V452fs *39	0	0	1	0	0.47	-	-	Considered to be a likely pathogenic variant		
					c.1544G>C	p.G515A	0.54	1.01	0	1	1.00	-	-	Suggested previously to be pathogenic	(Cruchaga, Haller et al. 2012)	
					c.1762G>T	p.A588S	0	0	10	7	0.46	1.66	0.54-5.8			
	PFN1	0.24	1 (0)	0 (0)	c.350_351delAAinsGT	p.E117G	0	0	1	0	0.47	-	-	Confers an Odds ratio of disease of 2.44	(Fratta, Charnock et al. 2014)	Y

Appendix 33 APOE genotypes of all cohorts.

Genotype was inferred from rs729358 and rs7412 SNVs. In total 1451 cases were successfully genotyped. The total number of cases for each genotype in each disease group and the percentage of that disease group are given (n,%). Associations between disease cohort and *APOE* genotype were performed for AD, FTD-ALS, and DLB vs controls (Chi-squared testing). **p<0.0001, *P<0.001

Disease Cohort	2/2 (n (%))	2/3 (n (%))	2/4 (n (%))	3/3 (n (%))	3/4 (n (%))	4/4 (n (%))	Total cohort in study	Total with APOE genotyping
AD	0 (0.0)	11 (4.0)	9 (3.3)	85 (31.3)	120 (44.1) **	47 (17.3)	289	272
CBD	0 (0.0)	1 (7.1)	0 (0)	9 (64.3)	4 (28.6)	0 (0)	14	14
CJD	3 (1.3)	28 (12.2)	2 (0.9)	130 (56.5)	64 (27.8)	3 (1.3)	239	230
Control	3 (0.8)	42 (11.7)	12 (3.3)	213 (59.3)	83 (23.1)	6 (1.7)	369	359
Control (High Braak)	0 (0.0)	3 (8.3)	0 (0)	24 (66.7)	9 (25)	0 (0)	38	36
DLB	0 (0.0)	1 (1.8)	1 (1.8)	24 (43.6)	25 (45.5) *	4 (7.3) *	58	55
FTD-ALS	2 (0.8)	18 (7.3)	10 (4.1)	140 (57.1)	68 (27.8)	7 (2.9)	251	245
HD	0 (0.0)	3 (42.9)	1 (14.3)	3 (42.9)	0 (0)	0 (0)	7	7
MSA	0 (0.0)	0 (0.0)	0 (0)	5 (50)	5 (50)	0 (0)	10	10
Other	1 (1.4)	13 (17.8)	2 (2.7)	35 (47.9)	18 (24.7)	4 (5.5)	80	73
PD	0 (0.0)	4 (10.5)	0 (0)	17 (44.7)	16 (42.1)	1 (2.6)	39	38
PSP	0 (0.0)	2 (11.8)	0 (0)	15 (88.2)	0 (0)	0 (0)	17	17
Vascular disease	1 (1.6)	10 (15.9)	1 (1.6)	31 (49.2)	20 (31.7)	0 (0)	65	63
Vascular disease / AD	0 (0.0)	0 (0.0)	0 (0)	5 (35.7)	7 (50)	2 (14.3)	17	14
Vascular disease / Control	0 (0.0)	2 (11.1)	1 (5.6)	14 (77.8)	1 (5.6)	0 (0)	18	18
Total	10	138	39	750	440	74	1511	1451

Appendix 34 Case-burden test results for all risk factor genes across all diseases compared to all controls (n=380).

Number of variants (V) together with the number (n) and percentage of each cohort possessing a variant in each gene are shown for each disease cohort. Associations significant at the uncorrected threshold are shaded grey. Key – AD; Alzheimer's disease, FTD-ALS; Frontotemporal dementia – Amyotrophic Lateral Sclerosis, CJD; Creutzfeldt-Jakob Disease, DLB; Dementia with Lewy Bodies.

	AD vs All controls						FTD-ALS vs All controls						CJD vs All controls						DLB vs All controls					
	V	Cases (n)	% case	Cont (n)	% cont	SKAT 0 (p-value)	V	Cases (n)	% case	Cont (n)	% cont	SKAT0 (p-value)	V	Cases (n)	% case	Cont (n)	% cont	SKAT 0 (p-value)	V	Cases (n)	% case	Cont (n)	% cont	SKAT 0 (p-value)
GBA	7	14	5.2	15	3.9	0.1678	9	12	5.5	15	3.9	0.2475	7	5	2.3	6	1.6	0.1621	7	8	13.8	15	3.9	0.0047
GRN	8	16	6.0	12	3.2	0.0329	8	13	5.9	12	3.2	0.0624	8	6	2.7	11	2.9	0.4927	8	7	12.1	12	3.2	0.0050
LRRK2	32	44	16.5	75	19.7	0.8938	30	43	19.5	77	20.3	0.6196	20	18	8.2	36	9.5	0.6779	24	8	13.8	77	20.3	0.8559
PFN1	1	1	0.4	1	0.3	0.3797	1	1	0.5	1	0.3	0.3686	3	3	1.4	1	0.3	0.0862	1	0	0.0	1	0.3	0.5430
PON1	2	4	1.5	6	1.6	0.5260	3	5	2.3	6	1.6	0.2752	2	6	2.7	6	1.6	0.2070	2	0	0.0	6	1.6	0.7953
PON2	4	2	0.7	2	0.5	0.4316	4	4	1.8	2	0.5	0.0788	2	0	0.0	2	0.5	0.7891	3	1	1.7	2	0.5	0.2040
PON3	4	3	1.1	6	1.6	0.6645	5	4	1.8	6	1.6	0.4570	6	3	1.4	6	1.6	0.5489	4	1	1.7	6	1.6	0.4654
SCARB2	5	7	2.6	8	2.1	0.3089	8	7	3.2	8	2.1	0.2254	4	6	2.7	8	2.1	0.2544	4	0	0.0	8	2.1	0.7891
SNCA	1	0	0.0	1	0.3	0.6918	2	1	0.5	1	0.3	0.4226	1	0	0.0	1	0.3	0.6779	1	0	0.0	1	0.3	0.5674
SQSTM1	12	19	7.1	24	6.3	0.3646	15	29	13.2	24	6.3	0.0055	8	4	1.8	11	2.9	0.8417	11	4	6.9	24	6.3	0.4856
TREM2	9	18	6.7	13	3.4	0.0343	7	7	3.2	13	3.4	0.6474	8	5	2.3	7	1.8	0.4879	7	7	12.1	13	3.4	0.0114

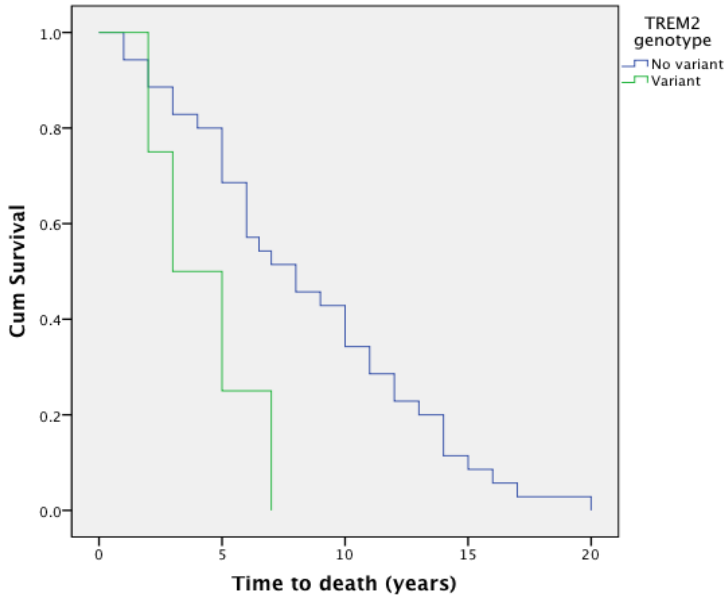
Appendix 35 Genetic variants in GRN observed in DLB patients.

All *GRN* variants (MAF < 3%) seen in DLB cases and controls. Variant position (GRCh37) together with the amino-acid change are provided and the number of cases and controls from each cohort (DLB n=58 and Controls n=380) are shown. *In-silico* predictions using SIFT and PolyPhen2 are shown together with Minor Allele frequency (MAF) in the 1000 Genomes database (1000G) and NHLBI ESP 6500 database. P-values are calculated by Chi-squared test.

Chromosome	Position	Reference allele	Variant allele	Protein alteration	Cases (n)	Controls (n)	SIFT	PolyPhen-2	1000G MAF (%)	NHLBI ESP 6500 (%)	P-value
17	42426585	C	T	p.T18M	1	1	Damaging	Possibly Damaging	0.02	0.02	0.25
17	42426803	C	A	p.P50T	1	0	Damaging	Benign		0	0.13
17	42428756	G	C	p.E287D	1	0	Tolerated	Possibly		0.01	0.13
17	42428788	G	A	p.R298H	0	1	Tolerated	Damaging		0	1.00
17	42428954	G	A	p.A324T	2	1	Tolerated	Benign	0.1	0.14	0.05
17	42429500	C	T	p.R433W	0	1	Damaging	Benign	0.16	0.28	1.00
17	42429839	G	C	p.G515A	1	1	Tolerated	Benign	0.54	0.01	0.25
17	42430146	G	T	p.A588S	1	7	Damaging	Benign		0	1.00

Appendix 36 Genotype-clinical and genotype-pathological correlates of disease.

A Kaplan Meier curve of time to death for Alzheimer disease cases with *TREM2* variants (green line) and no *TREM2* vari ant (blue line) (p=0.024, Mantel-Cox test).



Appendix 37 Clinical and demographic data for all cohorts in the oligogenic study.

Age of disease onset and death (years) together with the presence of a family history (FH) of disease are shown. Key: FTD-ALS – Frontotemporal dementia – amyotrophic lateral sclerosis, AD – Alzheimer's disease, DLB – Dementia with Lewy Bodies, PD – Parkinson's disease.

Phenotype	Number of cases	Male (number)	Female (number)	Mean age onset	Mean age death	Number with FH
Control	362	232 (64.1)	130 (35.9)	N/A	63.3 (18.8)	N/A
AD	277	131 (47.3)	146 (52.7)	65.4 (10.2)	77.7 (11.7)	11
FTD-ALS	244	143 (58.6)	101 (41.4)	59.4 (11.8)	64.6 (11.7)	14
DLB	58	36 (62.1)	22 (37.9)	66.7 (8.4)	76.7 (7.0)	2
PD	39	28 (71.8)	11 (28.2)	59.9 (10.9)	72.3 (9.2)	0

Appendix 38 Genes included in relevant disease gene panels.

Genes in which variants were assessed for pathogenicity are shown within each relevant cohort. 'Y' indicates that variants within the gene were included within the relevant panel. Key (inheritance): AD – autosomal dominant, AR – autosomal recessive, RF – risk factor.

Gene	Disease	Inheritance	AD panel	PD-DLB panel	Full FTD- ALS panel (n=28 genes)	Medium FTD-ALS panel (n=12 genes)	Small FTD-ALS panel (n=5 genes
SNCA	PD	AD/RF		Y			
PARK2	PD	AR		Y			
PINK1	PD	AR		Y			
EIF4G1	PD	AD		Y			
GIGYF2	PD	AD/RF		Y			
HTRA2	PD/AD	AD	Y	Y			
UCHL	PD	AD		Y			
SPG11	PD	AR		Y			
VPS35	PD	AD		Y			
FBX07	PD	AR		Y			
APP	AD	AD	Y	Y			
PSEN1	AD	AD	Y	Y			
PSEN2	AD	AD	Y	Y			
c9orf72	FTD / ALS	AD			Y		Y
GRN	FTD/AD	AD	Y	Y	Y		
CHCHD10	FTD	AD			Y		
TARDBP	FTD	AD			Y	Y	Y
SOD1	ALS	AD/AR			Y	Y	Y
FUS	ALS	AD			Y	Y	Y
PFN1	ALS	AD			Y	Y	
hnRNPA2B1	ALS	AD			Y		
hnRNPA1	ALS	AD			Y		
SETX	ALS	AR			Y	Y	
VAPB	ALS	AD			Y	Y	
OPTN	ALS	AR			Y	Y	
VCP	ALS	AD			Y	Y	
DAO	ALS	AD			Y	Y	
ANG	ALS	AD			Y	Y	Y
DCTN1	ALS	AD			Y	Y	
PARK7	PD	AR					
CHMP2B	FTD/ALS	AD			Y		
SQSTM1	FTD/ALS	AD/RF			Y	Y	
PRPH	ALS	AR			Y		
DPP6	ALS	AR			Y		
MATR3	ALS	AD			Y		
MAPT	FTD/AD	AD	Y	Y	Y		
ALS2	ALS	AR			Y		
SIGMAR1	ALS	AD			Y		
UBQLN2	FTD	XLD			Y		
NOTCH3	CADASIL	AD					
PRNP	fCJD	AD					
COQ2	MSA	AD/AR			Y		
GBA	DLB	RF		Y			
LRRK2	PD/DLB	RF		Y			
TREM2	AD	RF	Y	Y			
SCARB2	DLB	RF		Y			
PON1	ALS	RF			Y		
PON3	ALS	RF			Y		
APOE	AD	RF	Y	Y			

Appendix 39 Frequency of oligogenic variation by disease cohort and gene panel.

A table of the number and frequency of cases with >1 variant in each cohort and in each panel in the study. The proportion of cases with >1 variant in cases compared to controls were first tested before re-testing was performed after the removal of cases that harbour pathogenic variants, likely pathogenic variants, or known disease risk factors.

Panel	Total disease cohort	Disease cases with >1 variant	% of cases >1 variant	Control cohort size	Control cases with >1 variant	Percentage of controls with >1 variant	Fisher's test (cases with >1 variant vs controls)	Fisher's test (cases with >1 variant vs controls) after monogenic or RG cases removed
>1 variant: full FTD-ALS panel (MAF 5%)	244	48	19.67	362	48	13.26	**0.04**	0.45
>1 variant: full FTD-ALS panel (MAF 5%)	244	43	17.62	362	48	13.26	0.164	0.45
>1 variant: full FTD-ALS panel (MAF 1%)	244	19	7.79	362	26	7.18	0.875	0.14
>1 variant: full FTD-ALS panel (MAF 1%)	244	15	6.15	362	26	7.18	0.742	0.14
>1 variant: medium FTD-ALS panel (MAF 5%)	244	15	6.15	362	15	4.14	0.258	0.50
>1 variant: medium FTD-ALS panel (MAF 5%)	244	11	4.51	362	15	4.14	0.839	0.82
>1 variant: medium FTD-ALS panel (MAF 1%)	244	7	2.87	362	8	2.21	0.61	0.34
>1 variant: medium FTD-ALS panel (MAF 1%)	244	4	1.64	362	8	2.21	0.77	0.34
>1 variant: AD panel (MAF 5%)	277	8	2.89	362	10	2.76	1	0.057
>1 variant: AD panel (MAF 1%)	277	6	2.17	362	8	2.21	1	0.16
>1 variant: PD-DLB panel (MAF 5%)	97	39	40.21	362	92	25.41	**0.0002**	0.70
>1 variant: PD-DLB panel (MAF 1%)	97	23	23.71	362	37	10.22	**0.0011**	0.363

Appendix 40 All variants comprising oligogenic FTD-ALS cases at 1% MAF.

All variants in FTD-ALS cases with >1 variant in the full gene panel (28 genes) at a threshold of 1% MAF. The age of onset, death and disease duration is shown where available, together with all variant data. Key – ACMG – American College of Medical Genetics, B- Benign, LB – Likely Benign, UC – Unclassified, US – Uncertain Significance, RF – Risk Factor, LP – Likely Pathogenic, P – Pathogenic, T – Tolerated, D - Deleterious.

Case	Age onset	Age death	Disease duration	FH of disease	Monogenic?	Chromosome	Position	Reference Allele	Sample Allele	Variation Type	Gene Symbol	Protein Variant	Variant interpretation	Translation Impact	SIFT Function Prediction	SIFT Score	PolyPhen-2 Function Prediction	CADD Score	Conservation phyloP p-value	ExAC Frequency
1		71		N	Y	7	94944735	A	G	SNV	PON1	p.L90P	UC	missense	D	0	PrD	29.6	0.00003258	0.028
						17	42430146	G	T	SNV	GRN	p.A588S	LB	missense	D	0.04	B	13.92		0.046
						9	*C9orf72*						P							
2		74		N	N	7	94928347	G	C	SNV	PON1	p.T326R	UC	missense	T	1	B	10.21		0.001
						14	21161931	A	G	SNV	ANG	p.I70V	LB	missense	T	0.3	B			0.061
3		83		N	N	5	138665061	C	T	SNV	MATR3	p.R553C;	LB	missense			B	23.1		0.011
						12	49689404	G	T	SNV	PRPH	p.D141Y	UC	missense	D	0.01	PoD	27.4	0.0004159	0.247
4		87		N	N	2	74598723	T	C	SNV	DCTN1	p.I159V;	UC	missense	T	1	B		0.005445	0.799
						9	35059655		T	Ins	VCP	p.N616fs*12	LB	frameshift						
						17	42429414	G	T	SNV	GRN	p.C404F	LB	missense	D	0	PrD	32	0.00000517	
						17	44067341	C	T	SNV	MAPT	p.S427F	B	missense	D	0.05	PrD	28.5	0.001259	0.146
5	73	75	2	N	Y	9	135224754	T	C	SNV	SETX	p.Y21C	UC	missense	D	0	PrD	24	0.0004831	0.001
						9	*C9orf72*						P							
6	41	77	36	N	N	9	35059655		T	Ins	VCP	p.N616fs*12	LB	frameshift						
						20	57016044	TCT		Del	VAPB	p.S160del	LB	in-frame						
7	52	56	4	N	Y	7	94953733	T	C	SNV	PON1	p.N19D	UC	missense	T	0.51	B			0.159
						9	*C9orf72*						P							
8	55	60	5	Y	Y	7	95024007	G	A	SNV	PON3	p.R32*	UC	stop gain				36		0.142
						17	42426621		CCTG	Ins	GRN	p.C31fs*35	P	frameshift					0.00134	
9	57	60	3	N	Y	20	57016044	TCT		Del	VAPB	p.S160del	LB	in-frame						
						9	*C9orf72*						P							

Case	Age onset	Age death	Disease duration	FH of disease	Monogenic?	Chromosome	Position	Reference Allele	Sample Allele	Variation Type	Gene Symbol	Protein Variant	Variant interpretation	Translation Impact	SIFT Function Prediction	SIFT Score	PolyPhen-2 Function Prediction	CADD Score	Conservation phyloP p-value	ExAC Frequency
10	53	63	10	N	0	5	179250971	C	T	SNV	SQSTM1	p.R55C;	US	missense	D	0	PrD	32	0.006209	0.003
						7	94937419	G	A	SNV	PON1	p.A201V	UC	missense	T	0.47	B	23.3		0.158
11	49	50	1	Y	Y	9	135140020	A	G	SNV	SETX	p.I2547T	UC	missense	T	0.66	B			0.342
						9	*C9orf72*						P							
12	74	82	8	N	N	9	34635620	C	A	SNV	SIGMAR1	p.A155S	UC	missense				22.7		
						9	135140020	A	G	SNV	SETX	p.I2547T	UC	missense	T	0.66	B			0.342
13		71		Y	Y	2	202625862	C	A	SNV	ALS2	p.R285S	UC	missense			B			
						21	33039672	T	C	SNV	SOD1	p.I114T	P	missense	D	0.01	PrD	26.3	0.00003334	
14		55		Y	Y	5	179252184	A	G	SNV	SQSTM1	p.K154E	LB	missense	T	0.11	B	24.4		0.242
						9	135224757	C	T	SNV	SETX	p.R20H	UC	missense	T	0.15	B			0.906
						9	*C9orf72*						P							
15		62		N	Y	7	95024007	G	A	SNV	PON3	p.R32*	UC	stop gain				36		0.142
						9	135224757	C	T	SNV	SETX	p.R20H	UC	missense	T	0.15	B			0.906
16		78		N	Y	12	54677634	G	C	SNV	HNRNPA1	p.G316R	LP	missense	D	0.02	PrD	23.3	0.00000809	
						2	74592252	C	T	SNV	DCTN1	p.R1042Q;	LB	missense	T	0.31	PoD	26.3	0.00000167	0.099
17		53		N	N	21	33039672	T	C	SNV	SOD1	p.I114T	P	missense	D	0.01	PrD	26.3	0.00003334	
						9	34635679	G	A	SNV	SIGMAR1	p.R188W;	UC	missense	D	0.01	PrD	34		0.778
18		65		Y	N	9	135202325	A	C	SNV	SETX	p.C1554G	UC	missense	D	0.03	PrD	21.6	0.000526	0.584
						7	94953733	T	C	SNV	PON1	p.N19D	UC	missense	T	0.51	B			0.159
19		73		N	Y	17	42430146	G	T	SNV	GRN	p.A588S	LB	missense	D	0.04	B	13.92		0.046
						12	49689404	G	T	SNV	PRPH	p.D141Y	P	missense	D	0.01	PoD	27.4	0.0004159	0.247

Appendix 41 All variants comprising oligogenic PD-DLB cases at 1% MAF.

All variants in FTD-ALS cases with >1 variant in the full gene panel (14 genes) at a threshold of 1% MAF. The age of onset, death and disease duration is shown where available, together with all variant data. Key – ACMG – American College of Medical Genetics, B- Benign, LB – Likely Benign, UC – Unclassified, US – Uncertain Significance, RF – Risk Factor, LP – Likely Pathogenic, P – Pathogenic, T – Tolerated, D - Deleterious.

Case	Sub-phenotype	Age onset	Age death	Disease duration	FH of disease	McKeith Criteria	Monogenic ?	APOE genotype	Chromosome	Position	Reference Allele	Sample Allele	Variation Type	Gene Symbol	Protein Variant	ACMG Variant interpretation	Translation Impact	SIFT Function Prediction	SIFT Score	PolyPhen-2 Function Prediction	CADD Score	Conservation phyloP p-value	ExAC Frequency
1	DLB		79			Neo		3/3	3	184046450	A	G	SNV	EIF4G1	p.M1336V	LB	missense	D	0.01	PoD	26.3	0.00001702	0.022
									17	42430146	G	T	SNV	GRN	p.A588S	LB	missense	D	0.04	B	13.92		0.046
2	DLB		81			Limbic		3/4	15	44865000	T	C	SNV	SPG11	p.N2075S	UC	missense	T	0.65	B			0.259
									17	42428756	G	C	SNV	GRN	p.E287D	LB	missense	T	0.32	PoD	23.9		0.003
3	DLB	86	88	2		Neo		3/4	2	74759825	G	A	SNV	HTRA2	p.G399S	UC	missense	D	0.02	PrD	24.1	0.00052	0.437
									3	184044687	C	T	SNV	EIF4G1	p.P1075L	US	missense	T	0.14	B	26.6	0.000003873	
4	PD		80					3/3	3	184046529	T	C	SNV	EIF4G1	p.M1355T	US	missense	D	0	PrD	26.8	0.0005649	0.04
									15	44949354	C	T	SNV	SPG11	p.V270I	UC	missense	T	0.09	PoD	23.6	0.0001435	0.609
5	DLB	73	78	5				4/4	2	233712223	T	A	SNV	GIGYF2	p.L1230Q	LB	missense	T	0.4	B			0.002
									15	44907562	T	C	SNV	SPG11	p.K1013E	UC	missense	T	0.21	B	15.76		0.993
6	DLB	75	77	2				3/3	2	233712223	T	A	SNV	GIGYF2	p.L1230Q	LB	missense	T	0.4	B			0.002
									6	41129207	C	T	SNV	TREM2	p.R62H	RF	missense			B	11.11		0.826
7	PD	60	75	15		Neo		3/4	1	227073271	C	T	SNV	PSEN2	p.S130L	LB	missense	D	0.02	PoD	31	0.00006714	0.064
									3	184041256	G	C	SNV	EIF4G1	p.A717P	LB	missense	T	0.3	B	15.94		0.074
									6	41129207	C	T	SNV	TREM2	p.R62H	RF	missense			B	11.11		0.826
8	PD	68	77	9		Neo		3/4	1	155206167	C	T	SNV	GBA	p.E278K	RF	missense	T	0.88	B	17.33		0.979
									6	162683724	G	T	SNV	PARK2	p.A82E	UC	missense	T	1	B			0.472
									19	45411110	T	C	SNV	APOE	p.L46P	RF	missense	T	0.07	PoD	11.43		0.242
9	DLB	57	69	12		Limbic		3/3	12	40713856	G	C	SNV	LRRK2	p.E1632Q	UC	missense			PrD	24.3	0.000002056	

Case	Sub-phenotype	Age onset	Age death	Disease duration	FH of disease	McKeith Criteria	Monogenic ?	APOE genotype	Chromosome	Position	Reference Allele	Sample Allele	Variation Type	Gene Symbol	Protein Variant	ACMG Variant interpretation	Translation Impact	SIFT Function Prediction	SIFT Score	PolyPhen-2 Function Prediction	CADD Score	Conservation phyloP p-value	ExAC Frequency
									22	32894483	G		Del	FBXO7	p.R399fs	UC	frameshif					0.00003048	
10	DLB	65	70	5				3/3	6	41129207	C	T	SNV	TREM2	p.R62H	RF	missense			B	11.11		0.826
									6	161771219	G	A	SNV	PARK2	p.P409L	PR	missense	D	0.01	PrD	27.7	0.0001127	0.15
									12	40713899	T	C	SNV	LRRK2	p.M1646T	RF	missense			B	17.91	0.00001683	0.916
11	DLB	52	68	16		Neo		3/4	3	184043401	G	A	SNV	EIF4G1	p.R1039Q	US	missense	T	0.24	B	24.1	0.001589	
									12	40713899	T	C	SNV	LRRK2	p.M1646T	RF	missense			B	17.91	0.00001683	0.916
12	DLB	62	72	10		Neo		3/3	1	20960395	C	A	SNV	PINK1	p.S118R	UC	missense	T	0.32	B	16.26		
									1	155206037	G	A	SNV	GBA	p.T321M	RF	missense	T	0.11	B	22.2		0.657
									2	233712223	T	C	SNV	GIGYF2	p.L1230P	LB	missense	T	0.23	B			0.021
									15	44949354	C	T	SNV	SPG11	p.V270I	UC	missense	T	0.09	PoD	23.6	0.0001435	0.609
13	PD	58	63	5		Limbic		3/4	6	41129252	C	T	SNV	TREM2	p.R47H	RF	missense			PrD	33		0.206
									12	40745375	G	T	SNV	LRRK2	p.C2139F	UC	missense			PrD	33	0.000004074	
14	DLB	62	75	13		Limbic		3/3	1	155206167	C	T	SNV	GBA	p.E278K	RF	missense	T	0.88	B	17.33		0.979
									17	42429839	G	C	SNV	GRN	p.G515A	UC	missense	T	0.56	B			0.268
15	PD		76			Bstem		3/4	6	161771219	G	A	SNV	PARK2	p.P409L	PR	missense	D	0.01	PrD	27.7	0.0001127	0.15
									12	40713899	T	C	SNV	LRRK2	p.M1646T	RF	missense			B	17.91	0.00001683	0.916
16	DLB	65	66	1				3/4	1	155205043	A	G	SNV	GBA	p.L434P	RF	missense	D	0.04	PoD	24.8	0.0002443	0.31
									17	42426585	C	T	SNV	GRN	p.T18M	LB	missense	T	0.08	PoD	25.8		0.003
17	PD	66	69	3		Limbic		2/3	1	155205043	A	G	SNV	GBA	p.L434P	RF	missense	D	0.04	PoD	24.8	0.0002443	0.31
									12	40677726	G	T	SNV	LRRK2	p.S764I	UC	missense			B	16.12		
18	PD	63	70	7		Limbic		3/4	6	162206852	G	A	SNV	PARK2	p.R275W	PR	missense	D	0	PrD	34	0.006412	0.206
									19	45412358	C	G	SNV	APOE	p.R269G	UC	missense	D	0.01	B	25.6		0.042
19	PD	40	69	29		Limbic		3/4	6	41129100	G	A	SNV	TREM2	p.R98W	RF	missense			PoD	25.2		0.007
									6	161771240	C	T	SNV	PARK2	p.G430D	PR	missense	D	0	PrD	33	0.0001127	0.011
									6	162206852	G	A	SNV	PARK2	p.R275W	PR	missense	D	0	PrD	34	0.006412	0.206
20	PD	54	64	10		Neo		3/4	2	233712223	T	C	SNV	GIGYF2	p.L1230P	LB	missense	T	0.23	B			0.021
									12	40713899	T	C	SNV	LRRK2	p.M1646T	RF	missense			B	17.91	0.00001683	0.916
21	DLB		94			Neo		3/4	1	227076537	C	A	SNV	PSEN2	p.L192I	US	missense	T	0.15	B	25.4		

Case	Sub-phenotype	Age onset	Age death	Disease duration	FH of disease	McKeith Criteria	Monogenic ?	APOE genotype	Chromosome	Position	Reference Allele	Sample Allele	Variation Type	Gene Symbol	Protein Variant	ACMG Variant interpretation	Translation Impact	SIFT Function Prediction	SIFT Score	PolyPhen-2 Function Prediction	CADD Score	Conservation phyloP p-value	ExAC Frequency
									6	41129207	C	T	SNV	TREM2	p.R62H	RF	missense			B	11.11		0.826
22	DLB		80			Neo		3/3	3	184039828	C	T	SNV	EIF4G1	p.P399S	LB	missense	T	1	B			0.079
									6	41129207	C	T	SNV	TREM2	p.R62H	RF	missense			B	11.11		0.826
23	DLB		74			Neo		3/3	1	155205043	A	G	SNV	GBA	p.L434P	RF	missense	D	0.04	PoD	24.8	0.0002443	0.31
									2	74759825	G	A	SNV	HTRA2	p.G399S	UC	missense	D	0.02	PrD	24.1	0.00052	0.437

Appendix 42 The enrichment of monogenic alleles within oligogenic cases.

A table highlighting the enrichment of cases harbouring established monogenic or risk factor cases within cases that harbour >1 variant.

Panel	C9orf72 (inc/ex)	Cohort (n)	Number of monogenic or RF carriers	Oligogenic cases (>1 variant)	Number of carriers within oligogenic cases	Fisher's test (p-value)
>1 variant: full FTD-ALS panel (MAF 5%)	inc	244	33	48	**17**	0.0001
>1 variant: full FTD-ALS panel (MAF 5%)	ex	244	33	43	**12**	0.0054
>1 variant: full FTD-ALS panel (MAF 1%)	inc	244	33	19	**11**	0.0001
>1 variant: full FTD-ALS panel (MAF 1%)	ex	244	33	15	**7**	0.0013
>1 variant: medium FTD-ALS panel (MAF 5%)	inc	244	33	15	**9**	0.0001
>1 variant: medium FTD-ALS panel (MAF 5%)	ex	244	33	11	4	0.0461
>1 variant: medium FTD-ALS panel (MAF 1%)	inc	244	33	7	5	0.0006
>1 variant: medium FTD-ALS panel (MAF 1%)	ex	244	33	4	2	0.0895
>1 variant: AD panel (MAF 5%)		277	36	8	7	0.0001
>1 variant: AD panel (MAF 1%)		277	36	6	6	0.0001
>1 variant: PD-DLB panel (MAF 5%)		97	16	39	12	0.004
>1 variant: PD-DLB panel (MAF 1%)		97	16	23	10	0.0003

Appendix 43 The ability of oligogenic variation to suggest a monogenic disorder.

The sensitivity, specificity, Positive Likelihood Ratio (PLR), Negative Likelihood Ratio (NLR), Positive Predicative Value (PPV), and Negative Predicative Value (NPV) that an affected individual would have a highly penetrant allele, or risk factor for their disease upon the observation of >1 variant in the relevant panel at the relevant Minor Allele Frequency (MAF%) as indicated.

	>1 variant: full FTD-ALS panel (5% MAF)		**>1 variant: full FTD-ALS panel (1% MAF)**		**>1 variant: medium FTD-ALS panel (5% MAF)**		**>1 variant: medium FTD-ALS panel (1% MAF)**		**>1 variant in AD panel (5% MAF)**	**>1 variant in AD panel (1% MAF)**	**>1 variant in PD-DLB panel (5% MAF)**	**>1 variant in PD-DLB panel (1% MAF)**
	inc *C9orf72*	**ex *C9orf72***	**inc *C9orf72***	**ex *C9orf72***	**inc *C9orf72***	**ex *C9orf72***	**inc *C9orf72***	**ex *C9orf72***	**N/A**			
Sensitivity (%)	51.50	36.36	33.33	21.21	27.27	12.12	15.15	6.06	19.44	16.67	75.00	62.50
95% Sensitivity CI	33.50-69.20	20.40-54.88	17.96-51.83	8.98-38.91	13.30-45.52	3.40-28.20	5.11-31.90	0.74-20.23	8.19-36.02	6.37-32.81	47.62-92.73	35.43-84.80
Specificity (%)	85.30	86.64	96.21	96.21	97.16	96.68	99.05	99.05	99.59	100.00	66.67	83.95
95% specificity CI	79.80-89.80	81.57-90.74	92.67-98.35	92.67-98.35	93.91-98.95	93.28-98.66	96.62-99.89	96.62-99.89	97.71-99.99	98.48-100	55.32-76.76	74.12-91.17
PLR	3.50	2.72	8.79	5.59	9.59	3.65	15.98	6.39	46.86		2.25	3.89
95% PLR CI	2.2-5.6	1.56-4.75	3.82-20.23	2.17-14.40	3.65-25.19	1.13-11.80	3.23-79.04	0.93-43.85	5.94-369.83		1.48-3.42	2.08-7.28
NLR	0.57	0.73	0.69	0.82	0.75	0.91	0.86	0.95	0.81	0.83	0.38	0.45
95% NLR CI	0.40-0.81	0.56-0.96	0.54-0.88	0.68-0.98	0.61-0.92	0.80-1.03	0.74-0.99	0.87-1.04	0.69-0.95	0.72-0.96	0.16-0.89	0.24-0.87
PPV (%)	35.42	27.91	57.89	46.67	60.00	36.36	71.34	50.00	87.50	100.00	30.77	43.48
95% PPV CI	22.16-50.54	15.33-43.67	33.50-79.75	21.27-73.41	32.39-83.66	10.93-69.21	29.04-96.33	6.76-93.24	47.35-99.68	54.07-100	17.02-47.57	23.19-65.51
NPV (%)	91.84	90.54	90.22	88.65	89.52	87.55	88.19	87.08	89.22	88.93	93.10	91.89
95% NPV CI	87.08-95.26	85.90-94.05	85.57-93.77	83.81-92.45	84.81-93.17	82.62-91.50	83.38-92.00	82.17-91.05	84.89-92.66	84.57-92.41	83.27-98.09	83.18-96.67

Appendix 44 Enrichment of genetic variation in monogenic cases.

Number of controls with any variant across the relevant disease panel at the relevant threshold. A Fisher's test was performed to see if cases containing a known monogenic or risk factor variant within that panel were more likely that controls to also have an additional non-pathogenic variant.

Panel	Controls (n=362)	Controls with ≥1 variant. N (%)	Total number of monogenic cases	Monogenic cases with an extra non-pathogenic variant	Percentage of monogenic cases with a non-pathogenic variant	Fisher's test (p-value)
>1 variant: full FTD-ALS panel (MAF 5%) inc C9orf72	362	174 (48.07)	33	12	36.36	0.21
>1 variant: full FTD-ALS panel (MAF 5%)	362	174 (48.07)	19	9	47.37	0.12
>1 variant: full FTD-ALS panel (MAF 1%) inc C9orf72	362	117 (32.32)	33	9	27.27	0.68
>1 variant: full FTD-ALS genes (MAF 1%)	362	117 (32.32)	19	7	36.84	0.80
> 1 variant: AD panel (MAF 5%)	362	78 (21.55)	36	7	19.44	1.00
> 1 variant: AD panel (MAF 1%)	362	50 (13.81)	36	6	16.67	0.62
> 1 variant: PD-DLB genes (MAF 5%)	362	229 (63.26)	16	12	75.00	0.43
> 1 variant: PD-DLB genes (MAF 1%)	362	158 (43.64)	16	10	62.50	0.20

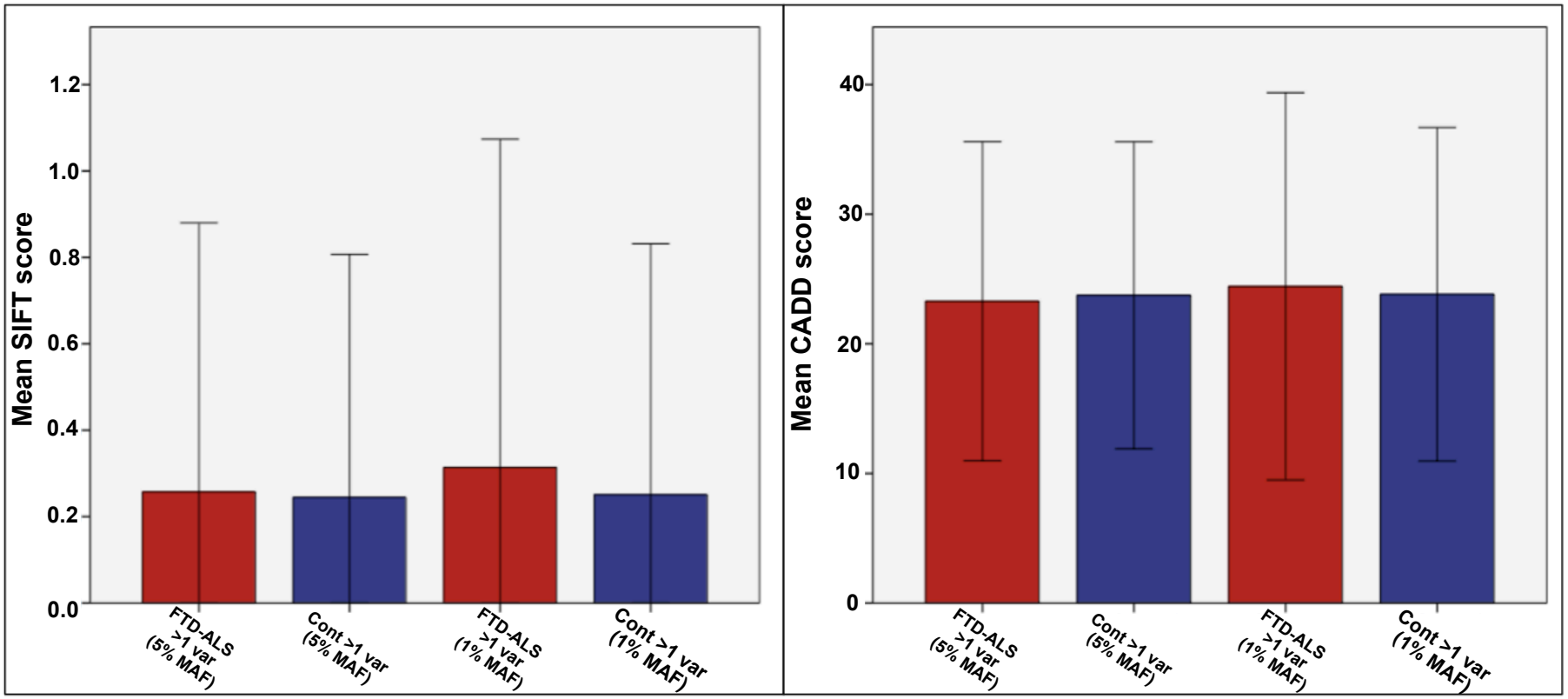
Appendix 45 Clinical data for oligogenic cases after the removal of monogenic cases.

Mean (SD) age of disease onset, death and duration for all cases in their relevant cohort after the removal of individuals with known highly penetrant alleles or disease risk factors. A longer disease duration was observed in cases of FTD-ALS with >1 variant compared to those with ≤ 1 variant.

	Age of onset (years)			Age of death (years)			Disease duration (years)		
	>1 variant	**≤1 variant**	**p-value**	**>1 variant**	**≤1 variant**	**p-value**	**>1 variant**	**≤1 variant**	**p-value**
FTD-ALS (MAF 5%) (mean, SD)	55.40 (11.60)	61.2 (12.10)	0.17	65.5 (13.30)	65.32 (11.60)	0.94	10.70 (10.80)	5.90 (4.30)	0.020
FTD-ALS (MAF 1%) (mean, SD)	56.00 (16.70)	60.5 (12.00)	0.54	73 (11.70)	65.04 (11.80)	0.062	18.00 (15.60)	6.30 (4.80)	0.00060
PD-DLB (MAF 5%) (mean, SD)	62.97 (11.99)	64.18 (9.26)	0.989	73.77 (10.49)	75.05 (7.29)	0.54	7.87 (4.22)	9.62 (5.73)	0.282
PD-DLB (MAF 1%) (mean, SD)	66.1 (9.92)	63.30 (10.22)	0.432	75.54 (7.18)	74.64 (7.75)	0.97	7.56 (4.30)	9.32 (5.49)	0.432

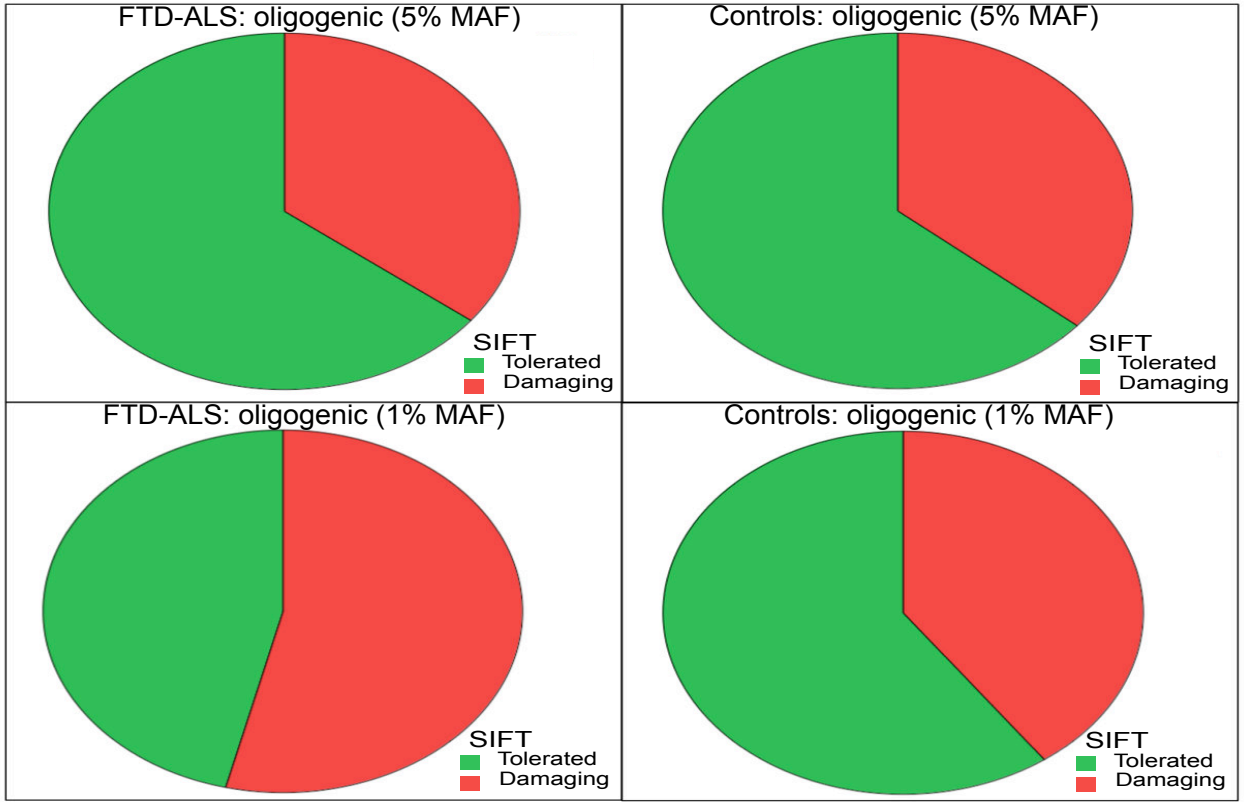
Appendix 46 Mean SIFT and CADD scores for FTD-ALS cases and controls.

Mean SIFT and CADD *in-silico* pathogenicity scores for variants in FTD-ALS cases (n=211) and controls (n=362) who have >1 variant, and of which neither variant was deemed to be either a pathogenic, or likely pathogenic variant based on ACMG criteria, nor an established risk factor for disease. There were no significant differences between cases or controls in either criteria at the 5% MAF or 1% MAF threshold (p>0.05, un-paired t-test). Error bars indicate standard deviation from mean.



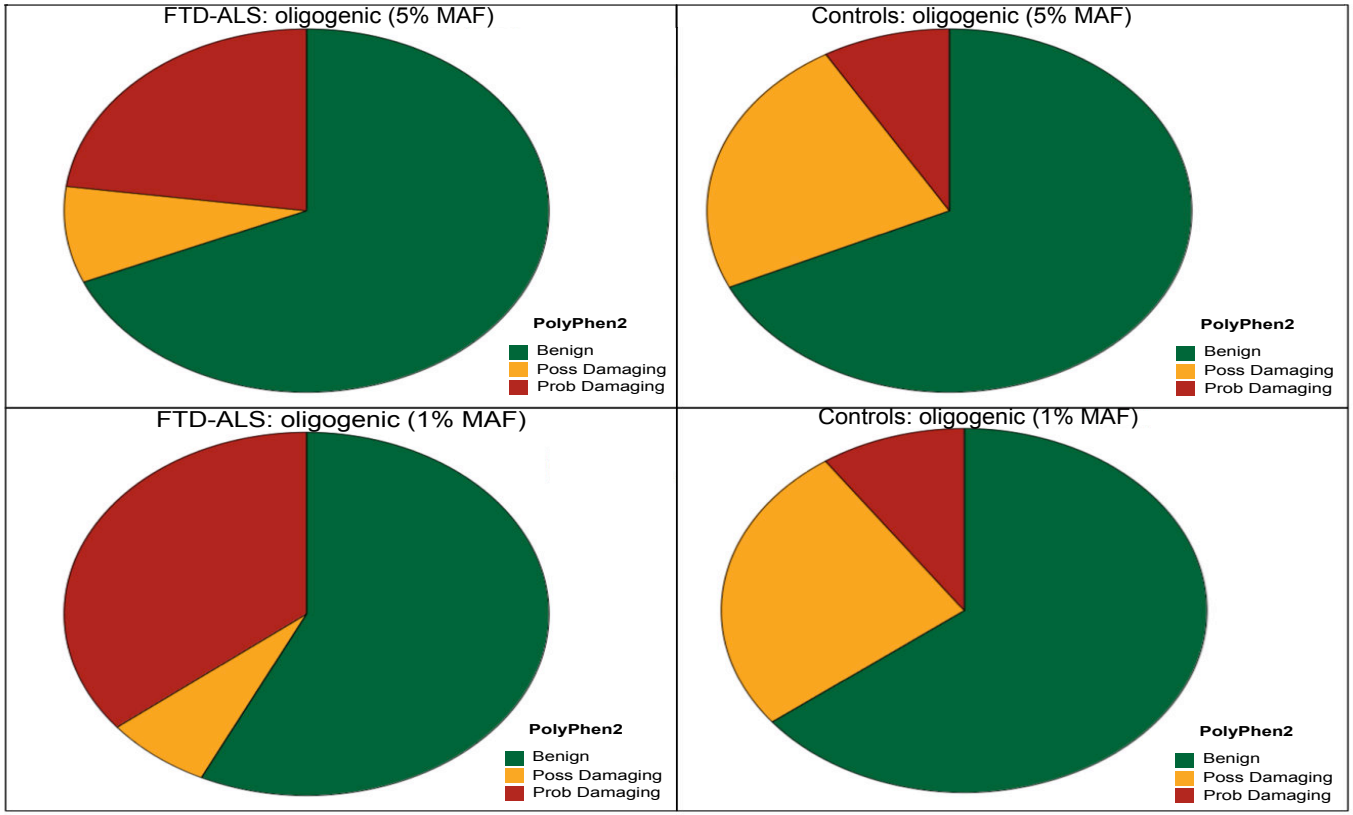
Appendix 47 The proportion of SIFT predicted damaging variants in FTD-ALS cases and controls.

The proportion of predicted damaging and tolerant SIFT *in-silico* pathogenicity scores for variants in FTD-ALS cases (n=211) and controls (n=362) who have >1 variant, and of which neither variant was deemed to be either a pathogenic, or likely pathogenic variant based on ACMG criteria, nor an established risk factor for disease. There were no significant differences between cases or controls in either criteria at the 5% MAF or 1% MAF threshold (P.0.05, Fisher's exact test).



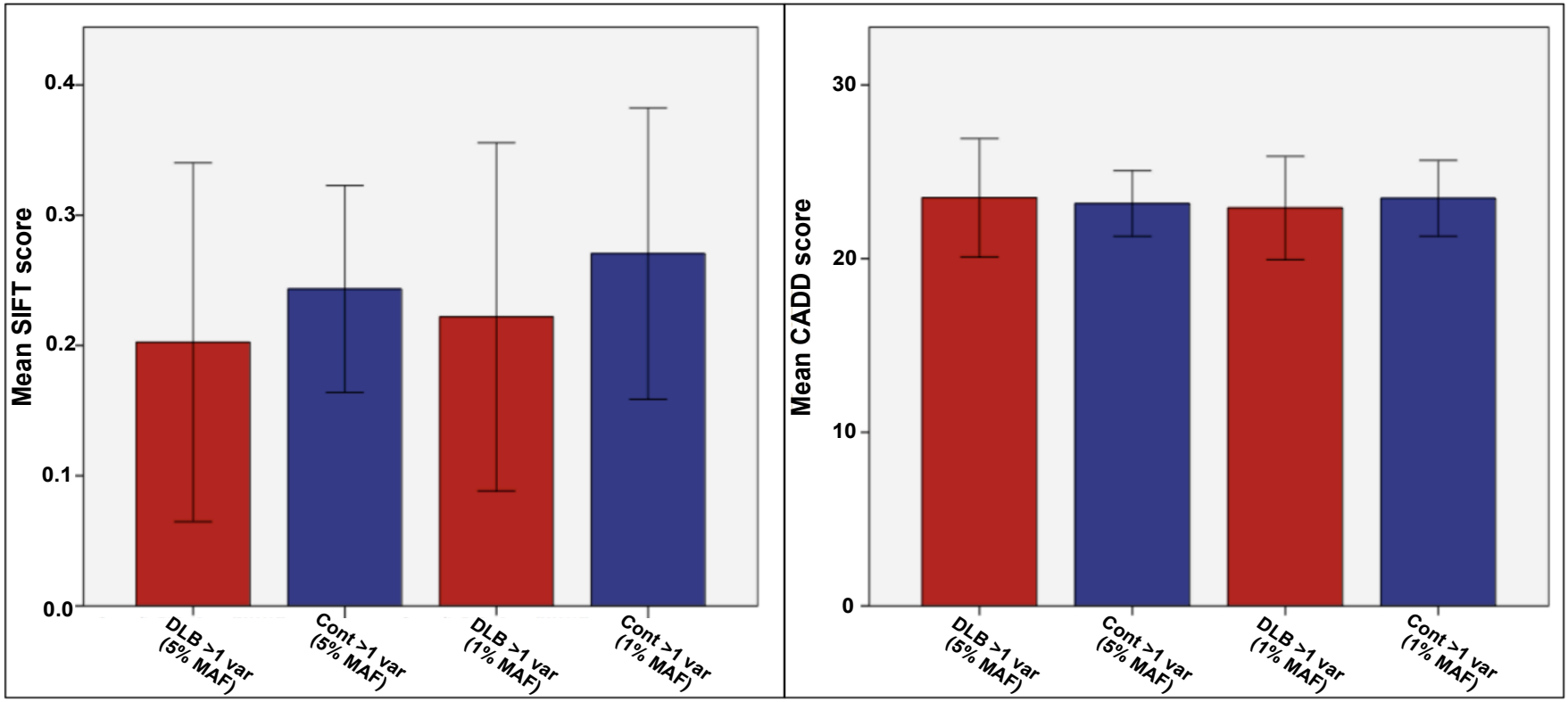
Appendix 48 The proportion of PolyPhen2 predicted damaging variants for FTD-ALS and controls.

The proportion of predicted benign, possibly damaging or damaging with PolyPhen2 for variants in FTD-ALS cases (n=211) and controls (n=362) who have >1 variant, and of which neither variant was deemed to be either a pathogenic, or likely pathogenic variant based on ACMG criteria, nor an established risk factor for disease. There were no significant differences between cases or controls in either criteria at the 5% MAF or 1% MAF threshold (P.0.05, Fisher's exact test).



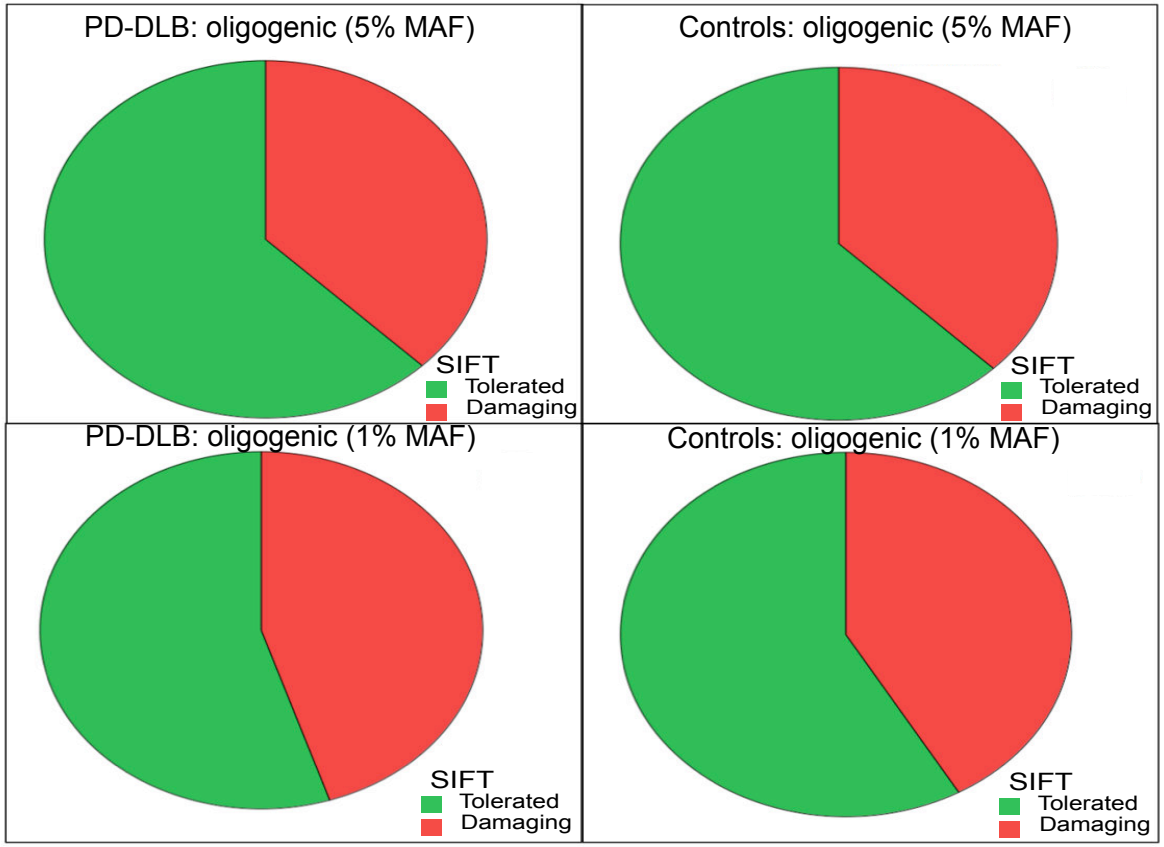
Appendix 49 Mean SIFT and CADD scores for PD-DLB cases and controls.

Mean SIFT and CADD *in-silico* pathogenicity scores for variants in PD-DLB cases (n=97) and controls (n=362) who have >1 variant, and of which neither variant was deemed to be either a pathogenic, or likely pathogenic variant based on ACMG criteria, nor an established risk factor for disease. There were no significant differences between cases or controls in either criteria at the 5% MAF or 1% MAF threshold (p>0.05, un-paired t-test). Error bars indicate standard deviation from mean.



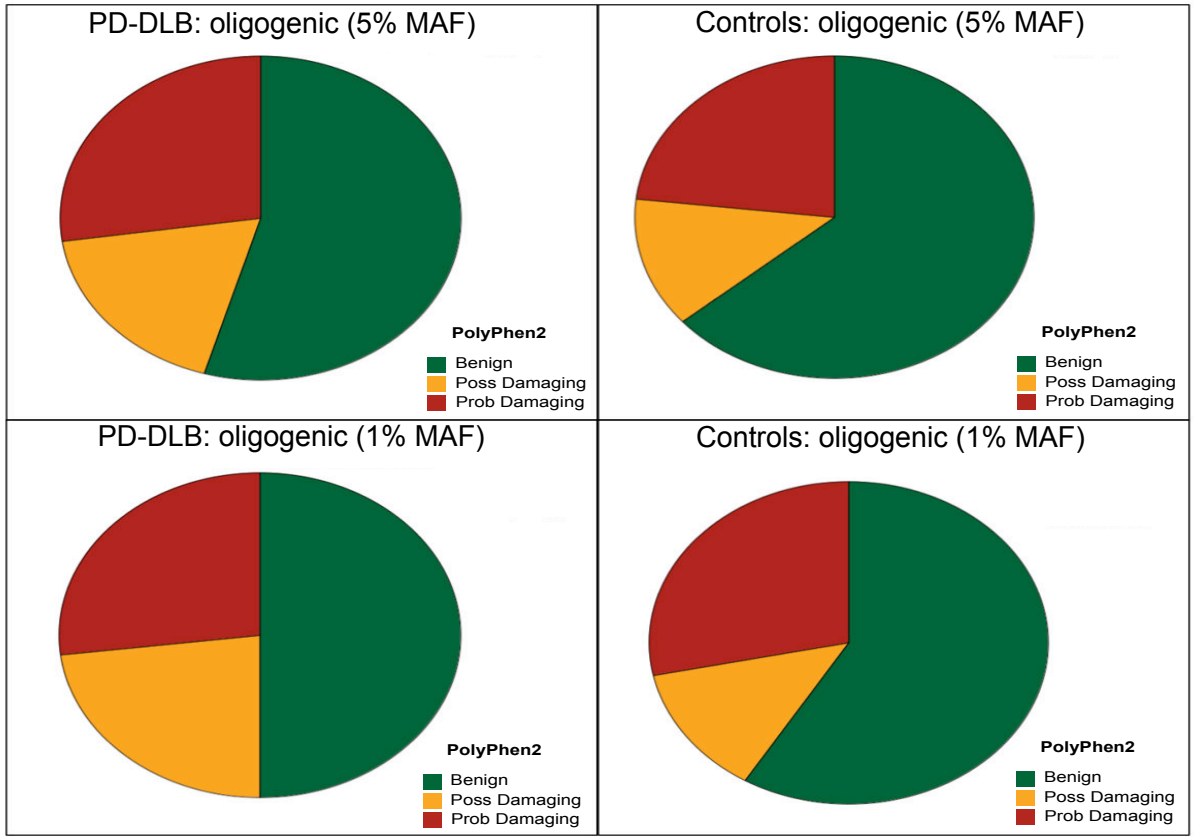
Appendix 50 The proportion of SIFT predicted damaging variants in PD-DLB cases and controls.

The proportion of predicted damaging and tolerant SIFT *in-silico* pathogenicity scores for variants in PD-DLB (n=97) and controls (n=362) who have >1 variant, and of which neither variant was deemed to be either a pathogenic, or likely pathogenic variant based on ACMG criteria, nor an established risk factor for disease. There were no significant differences between cases or controls in either criteria at the 5% MAF or 1% MAF threshold (P.0.05, Fisher's exact test).



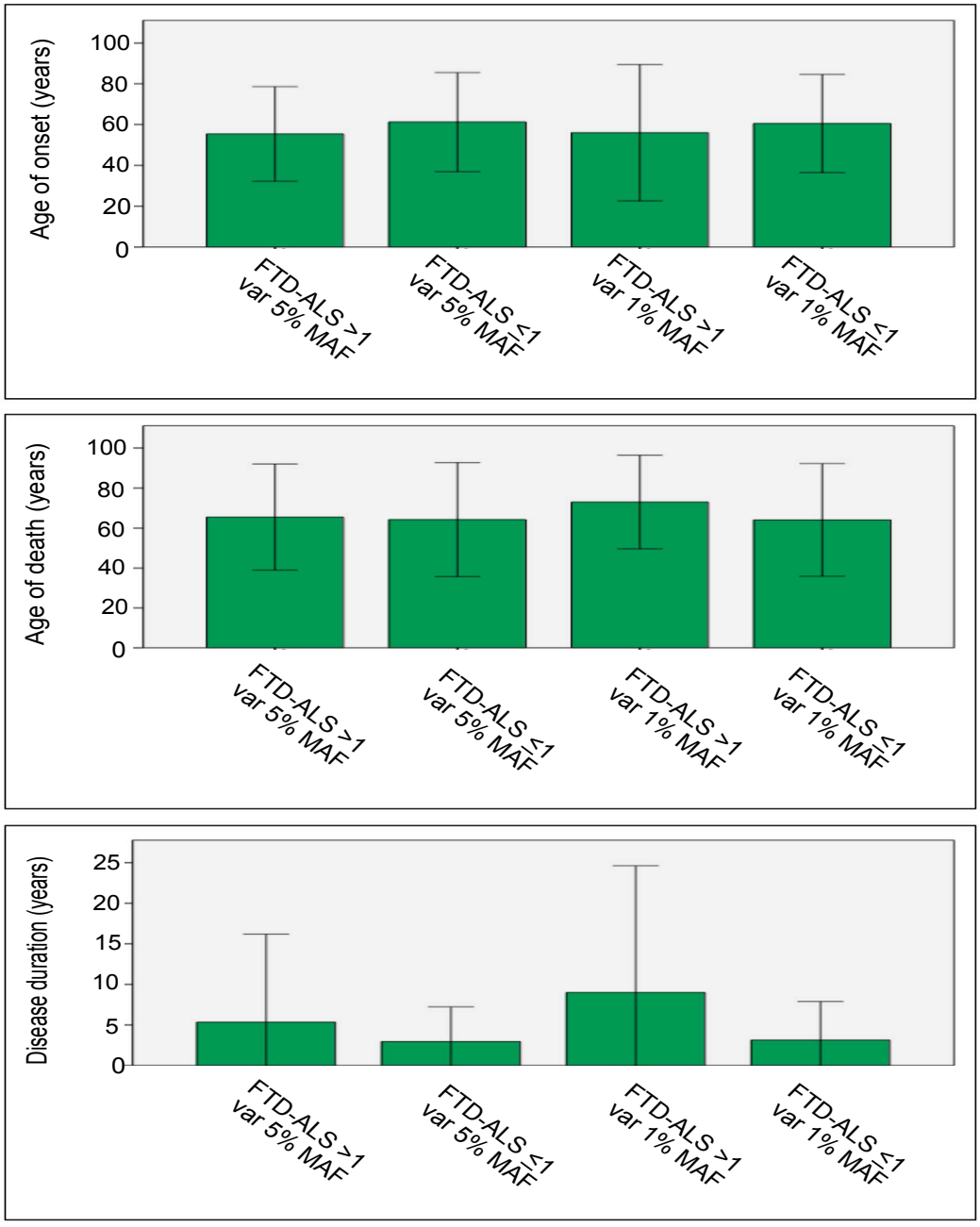
Appendix 51 The proportion of PolyPhen2 predicted damaging variants for PD-DLB and controls.

The proportion of predicted benign, possibly damaging or damaging with PolyPhen2 for variants in PD-DLB cases (n=97) and controls (n=362) who have >1 variant, and of which neither variant was deemed to be either a pathogenic, or likely pathogenic variant based on ACMG criteria, nor an established risk factor for disease. There were no significant differences between cases or controls in either criteria at the 5% MAF or 1% MAF threshold (P.0.05, Fisher's exact test).



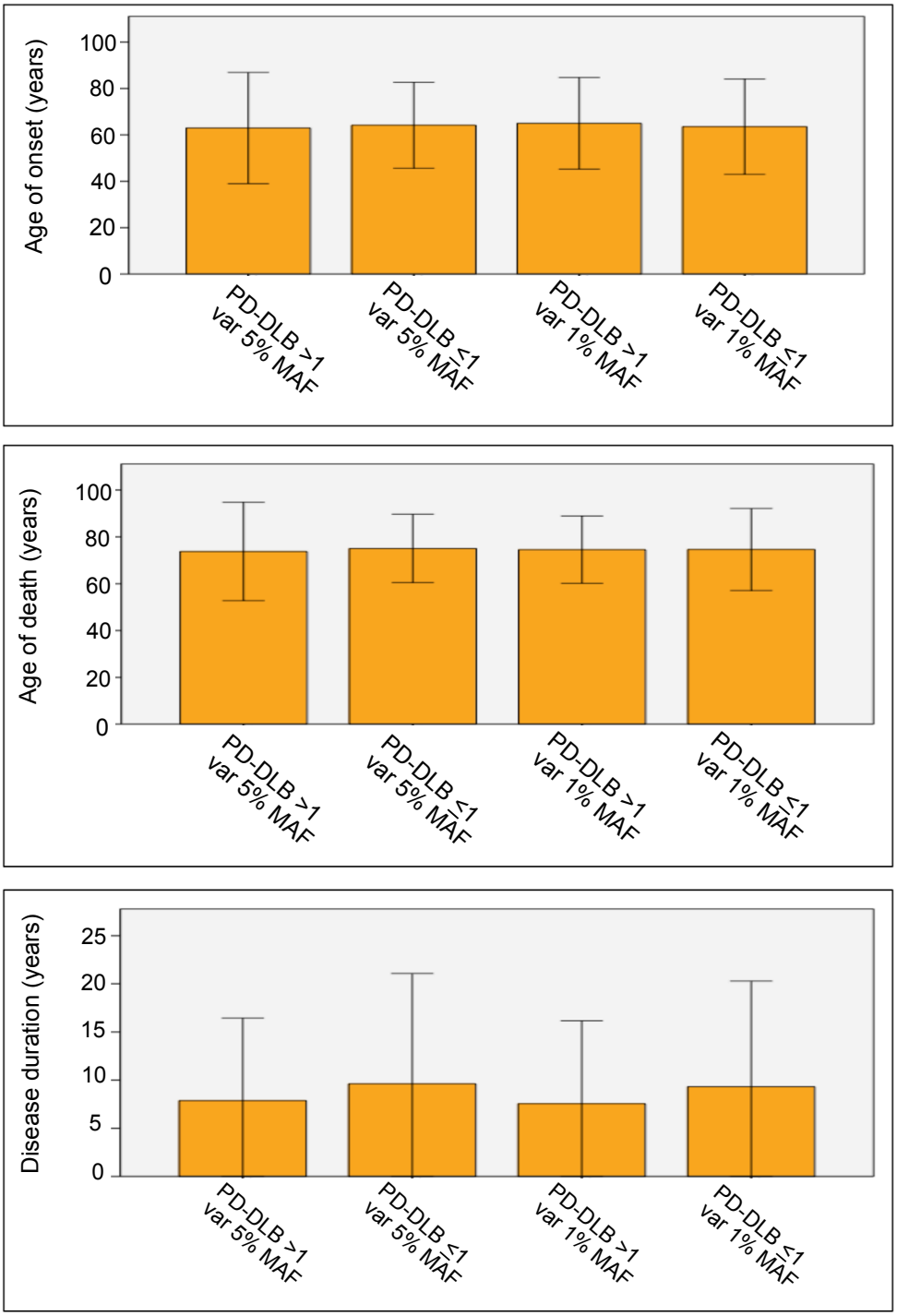
Appendix 52 Clinical data for oligogenic FTD-ALS cases.

Mean age of disease onset (top panel), death (middle panel) and mean disease duration (bottom panel) for all cases that have either ≥2 variants in the FTD-ALS gene panel (n=28) at the defined MAF threshold, compared to those with ≤1 variant. There were no differences between cohorts for any criteria. Error bars indicate standard deviation from mean.



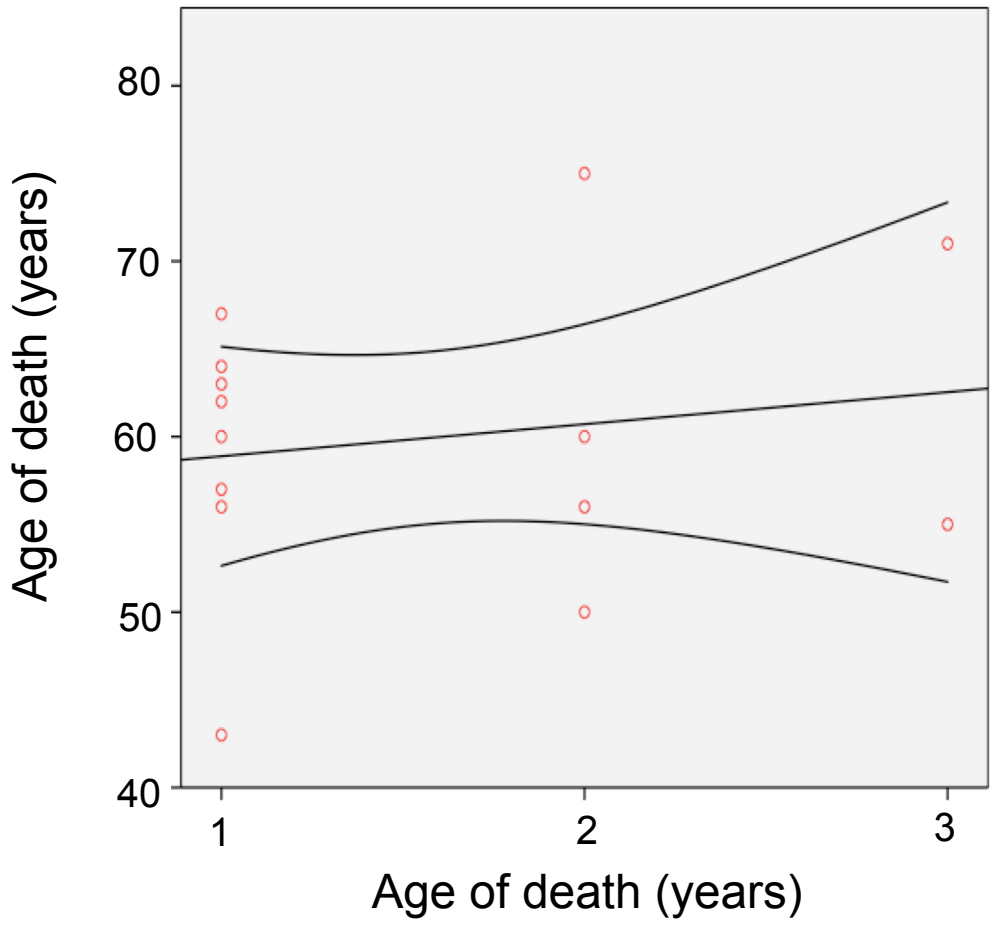
Appendix 53 Clinical data for oligogenic PD-DLB cases.

The mean age of disease onset (top panel), death (middle panel) and mean disease duration (bottom panel) for all cases that have either ≥2 variants in the PD-DLB gene panel (n=20) at the defined MAF threshold, compared to those with ≤1 variant. There were no differences between cohorts for any criteria. Error bars indicate standard deviation from mean.



Appendix 54 Age of death for C9orf72 repeat expansion carriers.

The mean age of death for all cases (n=14) that carried the C9orf72 mutation against the number of additional non-synonymous variants they possessed within the full FTD-ALS panel at 1% MAF. The line of best fit together with 95% CI is shown. There was no association between the age of death and the number of variants ($r^2 = 0.0064$).



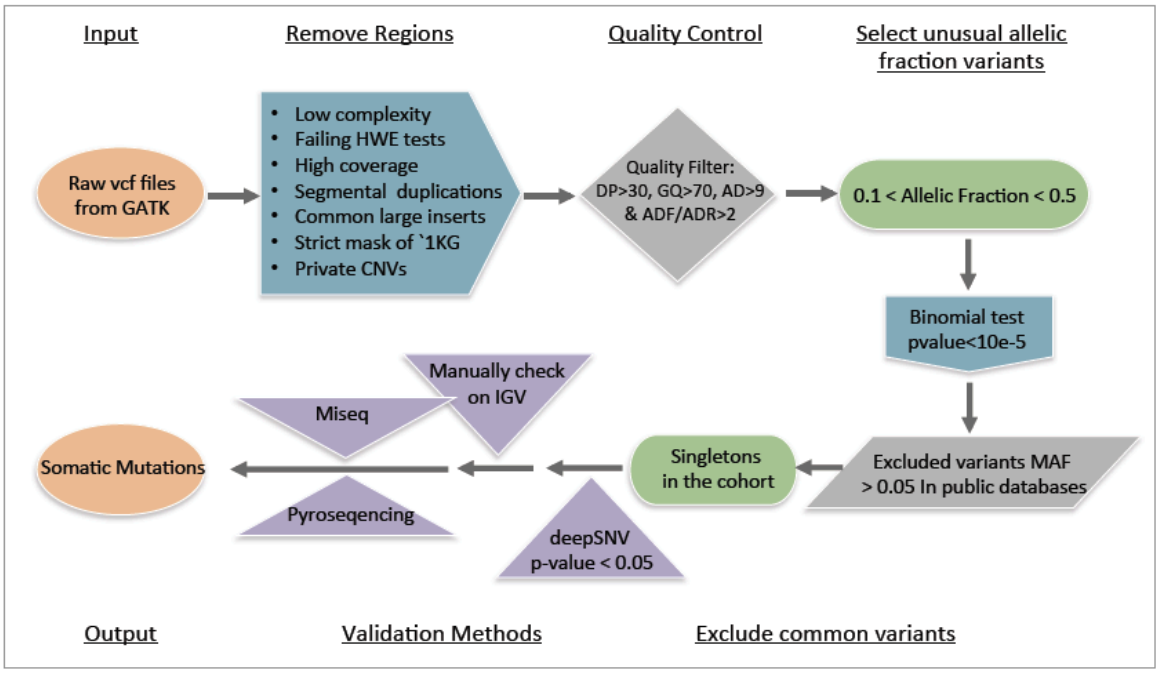
Appendix 55 A table of the cases utilized within the somatic sequencing study.

The total number of cases within each cohort together with the mean age of onset and death (years) are shown. Key: AD – Alzheimer's disease, CJD - Creutzfeldt Jakob Disease; FTD-ALS, Frontotemporal dementia – Amyotrophic lateral sclerosis; PD-DLB, Parkinson's disease – Dementia with Lewy Bodies

Disease Group	Total cohort	Male (n) (%)	Female (n) (%)	Age onset (years) (SD)	Age Death (years) (SD)	Disease duration (years)
AD	277	131 (47.3)	146 (52.7)	65.4 (10.2)	77.7 (11.7)	7.3 (4.7)
CJD	228	123 (53.9)	105 (46.1)	52.0 (20.1)	53.0 (19.6)	1.0 (1.0)
Control	362	232 (64.1)	130 (35.9)	N/A	63.3 (18.9)	N/A
FTD-ALS	244	143 (58.6)	101 (41.4)	59.4 (11.8)	64.6 (11.7)	6.5 (5.6)
Other	253	131 (51.8)	122 (48.2)	66.7 (20.2)	79.7 (14.3)	10.6 (15.8)
PD-DLB	97	64 (66.0)	33 (34.0)	63.7 (10.1)	74.9 (8.2)	9.2 (5.8)
Total	1461	824	637	59.3 (17.8)	68.2 (18.0)	5.6 (8.9)

Appendix 56 An overview of the pipeline to detect somatic variants within the dataset.

Abbreviations: VCF, Variant Call Format; GATK, Genome Analysis Tool Kit; HEW, Hardy-Weinberg; CNV, Copy Number Variant; DP, Depth; GQ, Genotype Quality; AD, Allele Depth; ADF/ADR, Depth of bases supporting variant on forward/reverse strand; MAF, Minor Allele Frequency.



Appendix 57 All PCR primers for putative somatic variants.

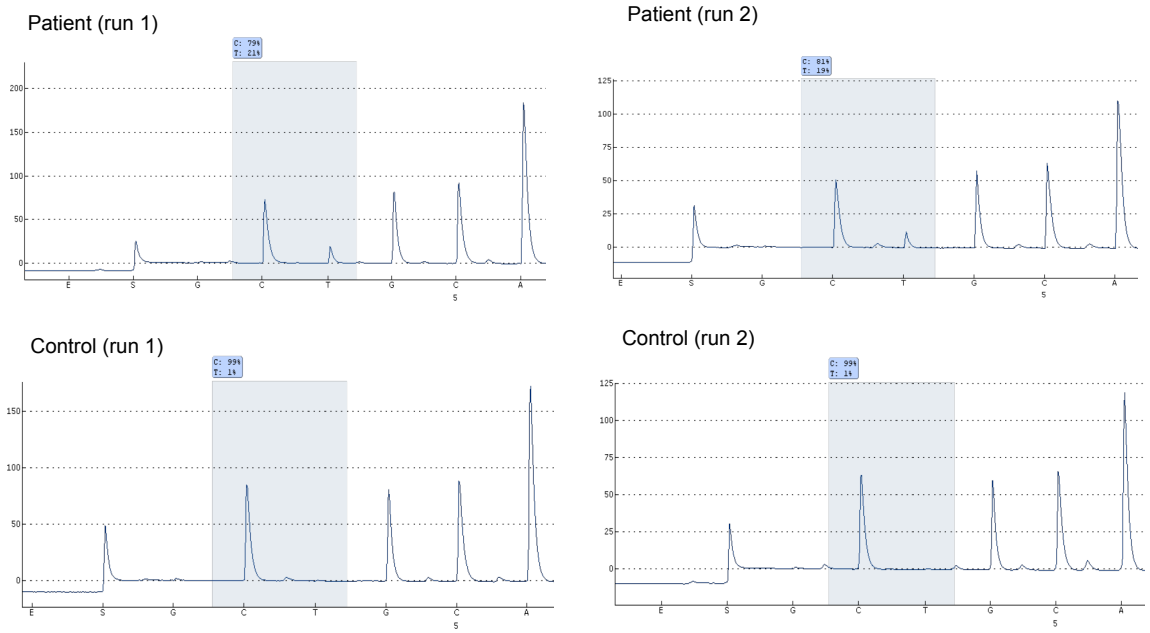
All forward and reverse primers spanning the position of putative somatic alleles (hg19) are shown.

Chr	Position	Ref	Var	Forward	Reverse
chr1	17570577	T	C	CTTTTGCAGGTTAACATGGTGGT	CAGGTGCAGATGCAAAGTCATTA
chr1	24125194	G	A	GCAAAGTGGATGACCGCCATA	CCATAATGCCTTCCGTGGTGA
chr1	43166558	C	T	CATGGATACTAAAGGCCTGGGG	TGAAGGGTGGCCAAATATGACT
chr1	46872017	C	T	TACCTCCCTCACCTCTCTGC	GCACCTTTGAAAGCCTGTGG
chr1	108742671	T	G	CACTTCTGTTACCGGGGTCG	CGATCTCCCGTGACTTCCTC
chr1	248224344	C	T	TCCCTTCAGGAAAGAGCACAC	GAGTCACCATTGCTGGGACA
chr10	96535189	G	C	GTGCAAGTGCCTGTTTCAGC	AGGGCCATTTCCACCATGTCC
chr11	1258327	G	A	TCTCCTACAAGTTCACCAGGCA	GGACATGGTCAGACCCTCCTTG
chr11	1718844	T	C	GAAATGGCACAAACAGGAACCAG	CCAAGAACCTGGAGTCATCACCT
chr11	56344581	G	T	TTGTTACCCCACAGGCACTTAG	TACGCTGGATGCTTCACACA
chr11	64323727	C	T	CACTCCTGGACACAAGATAAGG	GGCTGGCGAATGTCCAGAT
chr11	104905100	T	G	CCTTGCACTCAGCAAAAATAAATCC	CTTCAGGTCACTCCATGCACA
chr12	6138596	C	T	TCCCCAACAAGATGAAGCAAGA	TGTGCAGACATGTGAGGGATAG
chr12	9243951	A	G	CTTTGACCATTCCCCCGGAT	TGAGTGTTTTTAAGTCATGGCAGTA
chr12	42512876	T	G	ACTTACAAGGGCAGTGTTACAGA	TTTTTGAGCATGATGAAAGTACACA
chr12	104376635	C	T	TGTTTACGTGTGAGGTGGTGG	ACACTGCTATTCGTGGCTGAT
chr13	21742240	C	A	CCTTTTAACAGAGCTGCTTACC	CTAGTACACGAGCAAGAAGCC
chr14	23844983	G	A	ACCTGCTACCCTCATTTACAGTG	CTCCAGAAATGGGCAGAACTTTG
chr16	4833750	A	G	GGGGTAGACGTCGATGCAG	TCAACGAGCAATACGAGCAGT
chr16	88712548	G	A	GTGGTGCGGGTTAATGAGGA	AGAAGTACATGACCGCCGTG
chr17	13400048	G	C	TTGAGGTTGAAAGGCCGGTAG	ATCTCTGACATGCCTTTTTGCG
chr17	39502849	T	G	TAGAGTTGAGGTGTGACTGATGC	CTTGGAGCTTGTGAGTTCTTTGG
chr17	39521517	C	T	GAGTATCGCTGTGGTGGGAAA	GGCTTCAACTCATGTTTCCTAAG
chr17	45234387	A	G	AAGAGCTGCTGGTCCTCCTAA	TTGATTAGCAGAAGTGGACAGGAG
chr17	48070896	C	A	TGTACACGTGCTAACGCAGA	ATCAAGTCGCAGAGGGCTTG
chr17	76499013	G	A	AAGTCAGGGCCACCTGAAAG	AATCCTTGACCTGTCTGCCC
chr19	1440103	C	T	GCCTTAGTTCTCTGTCCGCC	GACACCAAGAGACCATCCACA
chr19	1440165	C	G	GCCTTAGTTCTCTGTCCGCC	GACACCAAGAGACCATCCACA
chr19	1440171	G	A	TTTAGAGGCCTTTGGCCCGA	CCCCGAACGTCCAACCTGA
chr19	1440183	C	T	TTTAGAGGCCTTTGGCCCGA	CCCCGAACGTCCAACCTGA
chr19	1818772	G	C	GCCTGAAGAGGACACGAGG	CCAAGACCAGCTTCTCGCTC
chr19	8999476	G	C	CTCTCTCTGAAGAGACTTGGGC	CCAAACCTAGCTCTTCACCTGT
chr19	9006365	C	A	CACATACAGACGTGTCCTAGCC	GCCCCATACAACCTACTGATGT
chr19	9361855	G	A	CAAAAGGTGTCCGAGCTACA	CAGGTGAGCTGTAACATAATCAAAT
chr19	14877162	T	C	TGTGATCAGATGTTTGTGGGTCT	TAAGAAGGGGAACCTCCACCATA
chr19	36275201	G	A	TGGAGTTTCTGCTCACCCATGT	GGGCTTCTTTCGAGGGACACT
chr19	50170347	G	A	CAGTTTGAGGTCAGAGGTGTTG	AGTGCAAAGAACGTCTTCCAGC
chr2	26702178	C	T	TTCTGGAAAAATCAGGGAACCCA	CCTGATTCAGAACGCAAGTGAT
chr2	85991195	C	T	GTCCTGAGGGCACAGAGTAAAA	GCATAGCCTCAGATAACCCACA
chr20	60718900	C	T	GCGACTCTCAAAAGCGCAC	GTAACGATGCCCCGCAAATG
chr20	60888258	G	A	ATCTACAGGACCAGTGGGGG	CTTCGTCTTCTACGTCGGGG
chr22	50752254	G	A	ATCTGACGGCCGCTTCTTC	TCAGAAGCCTCGTTGAGTGT
chr3	45837911	T	C	CTTCGGACAAATCACGCTCG	CCTGTCCTCTGCAATCTCGG
chr3	122629742	T	C	GTACACATTGGTCTGCTGCAAT	TACCTAGTCCTACGTCTGAGCG
chr5	115177753	A	G	ATGCGCACACTCCAGTTAGG	GGAAAGGATGGGAGTCGGC
chr5	140242885	C	G	AGCTATGACGCCTGGTTGTC	GCCGTAATGTTGACCTTGGG
chr6	5004177	G	A	TGATTTGGGGAAAAAGCGAGGT	AAATTGTGCTTTACCTTCCCGC
chr6	26458871	T	C	AGAATCGTCGAGAACCAGCG	AACTTGAGCCGTGCAATAGGA
chr6	27114458	T	A	GGAAGTATCCATTTTCGCGCC	AACCTCTGACGTCACCCTAAT
chr6	27114534	G	A	GGTCGAGCGCTTGTTGTAAT	TCTGACGTCACCCTAATAACCA
chr6	54066945	A	C	TGTGTTGACCAAATGCACCAG	ACCTGTTCAGAAAAGAGGTCG
chr6	56497767	A	C	ATTCGTTGTTTGTCCAACCAGT	GCATATACCACACGACCAGTTTG
chr6	136590614	G	A	GCATTGAAGTCCAACCTCCTG	GGCATGACATTGATCGCCG
chr6	136590646	T	C	GCATTGAAGTCCAACCTCCTG	GGCATGACATTGATCGCCG
chr6	136594277	G	A	GTATTGGCTGCAACTACCCAGT	TGGCAAGGTATTAGAGCATCCA
chr6	136600993	G	A	GCTCTGCCATTTACTAGAGCTTG	TAAGTGGGGCAGTCCGTAAA
chr7	1535876	C	T	CCCATAATCAAACACCCGGACT	ATCTCTGTTGACTGAACCTCCC
chr7	150815676	C	T	TGAAGCGGTGCACCTACTATG	CCAACCAATGTTTCCAAGCAAC

chr7	157178323	C	T	CCCAGAGTGTGAAGTAGAGGTT	TTAACTTCACAGCAGGCCGAC
chr8	144921555	T	C	CCCAAAGTCCGCACTTACCA	CAGAGTACGTGTCATCAGGCA
chr8	144940260	C	T	CCGATGGCATGATGGACTGA	GGACGACCGCGTCAAGC
chr9	95481489	C	T	CACGGATGATAGCGATCAGGT	GGATGGGGACTACTACGAGGT
chr9	125486851	T	A	AGGCTCCACACCCCTATGTA	TTTGAATGCCATGTGGAAAAGGGTT
chr9	125486893	G	A	CAGGCTCCACACCCCTATGT	TGACACTCCCATAGAACAGGACCA
chr9	125486925	C	T	AGGCTCCACACCCCTATGTA	ACACTCCCATAGAACAGGACC
chrX	135960166	G	A	ATACACCAAACCCGAGGTCC	ATGCAGTGGGTTCACCTTATGT

Appendix 58 Pyrogram traces from a confirmed somatic variant and control patient.

The pyrogram trace for both run one and run two for the same patient and control are shown. The mean variant allele frequency was 20% for the non-reference T allele.



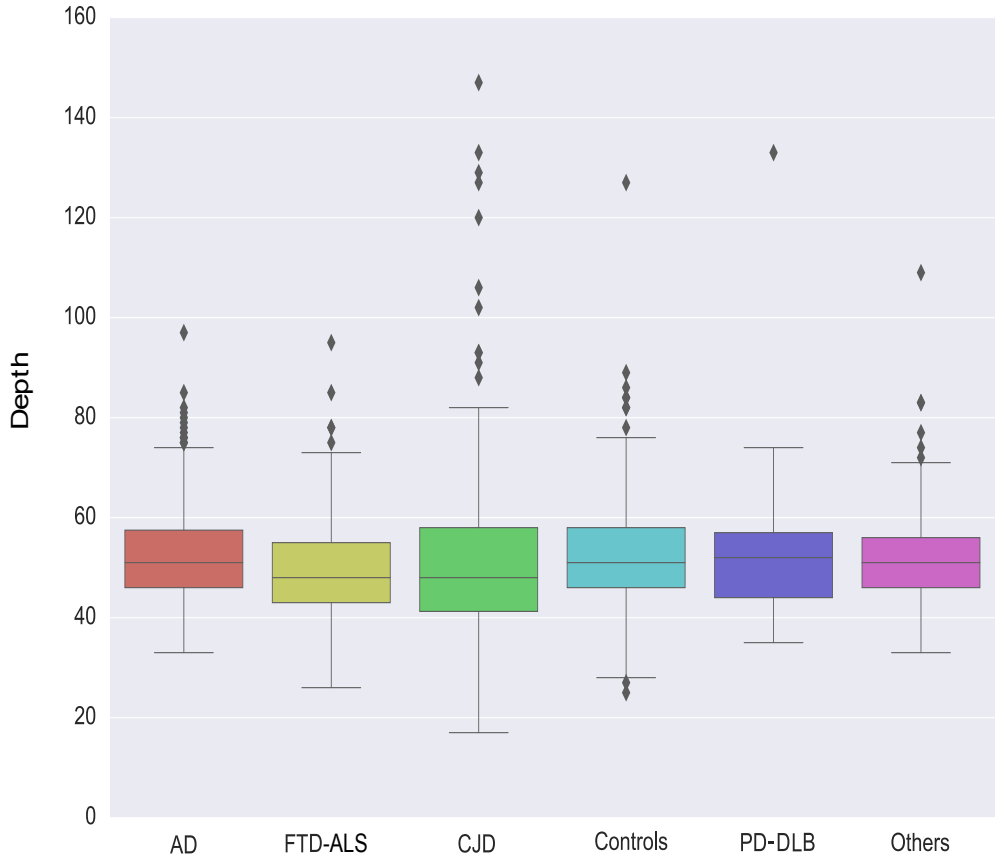
Appendix 59 Expected and observed variants occurring within different cohorts of genes as grouped by established brain expression data.

Binomial testing was performed between observed and expected variant proportions

Brain proteome data			
Brain proteome gene set	**Observed variants (total validated)**	**Expected mutation frequency**	**binomial test (p-value) ***
Elevated in Brain (n=1224)	1	1.1	1.0
Expressed in all (n=8588)	5	7.8	0.48
Mixed expression pattern (n=4404)	4	4	1.0
Not detected in brain (n=1318)	2	1.2	0.36
Not detected in any tissue (n=4157)	5	3.8	0.56

Appendix 60 Average sequencing depth for each group within the study.

Average sequencing depth for each group within the study. There were no significant differences (one-way ANOVA). Abbreviations: AD, Alzheimer's disease; CJD, Creutzfeldt Jakob Disease; FTD-ALS, Frontotemporal dementia – Amyotrophic lateral sclerosis; PD-DLB, Parkinson's disease – Dementia with Lewy Bodies.



Appendix 61 Genes and inheritance patterns causing their relevant neurodegenerative disease.

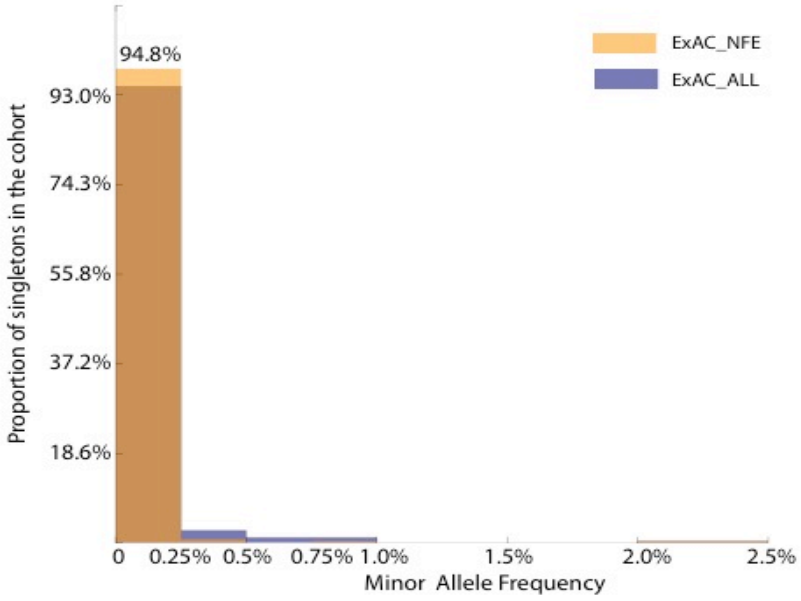
The 56 putative somatic variants detected in this study. Chromosome, base position (with reference to hg19 build), reference and observed alternate allele, ratio of alternate to reference allele, gene name, and whether the mutation is synonymous or non-synonymous are shown. Finally, the results of the validation experiment are shown. Abbreviations: Chr - Chromosome, Pos-Position, Ref - Reference allele, Alt -Alternative allele, Var-Variant Allele Frequencies (VAF)

No.	Chr	Pos	Ref	Alt	VAF	Gene name	Nonsynonymous/ synonymous SNV	Validation outcome
1	chr1	17570577	T	C	25.29%	PADI1	nonsynonymous SNV	Validated
2	chr1	24125194	G	A	21.67%	GALE	nonsynonymous SNV	Validated
3	chr1	43166558	C	T	17.57%	YBX1	nonsynonymous SNV	Technical failure
4	chr1	46872017	C	T	19.67%	FAAH	nonsynonymous SNV	Heterozygous
5	chr1	108742671	T	G	27.19%	SLC25A24	synonymous SNV	Heterozygous
6	chr1	228447313	C	G	28.26%	OBSCN	nonsynonymous SNV	Technical failure
7	chr1	248224344	C	T	15.15%	OR2L3	nonsynonymous SNV	Validated
8	chr10	96535189	G	C	27.72%	CYP2C19	nonsynonymous SNV	Heterozygous
9	chr11	1258327	G	A	25.41%	MUC5B	nonsynonymous SNV	Not present
10	chr11	1718844	T	C	14.86%	KRTAP5-6	synonymous SNV	Validated
11	chr11	56344581	G	T	13.04%	OR5M10	nonsynonymous SNV	Validated
12	chr11	104905100	T	G	16.90%	CASP1	nonsynonymous SNV	Validated
13	chr12	6138596	C	T	19.54%	VWF	nonsynonymous SNV	Validated
14	chr12	9243951	A	G	14.71%	A2M	nonsynonymous SNV	Not present
15	chr12	42512876	T	G	16.13%	GXYLT1	nonsynonymous SNV	Technical failure
16	chr12	104376635	C	T	18.52%	TDG	synonymous SNV	Not present
17	chr13	21742240	C	A	22.95%	SKA3	nonsynonymous SNV	Heterozygous
18	chr14	23844983	G	A	27.34%	IL25	nonsynonymous SNV	Heterozygous
19	chr16	4833750	A	G	22.77%	SETP12	nonsynonymous SNV	Validated
20	chr16	88712548	G	A	23.53%	CYBA	synonymous SNV	Validated
21	chr17	13400048	G	C	27.40%	HS3ST3A1	nonsynonymous SNV	Heterozygous
22	chr17	39502849	T	G	22.06%	KRT33A	nonsynonymous SNV	Validated
23	chr17	39521517	C	T	29.93%	KRT33B	synonymous SNV	Heterozygous
24	chr17	45234387	A	G	16.85%	CDC27	nonsynonymous SNV	Not present
25	chr17	48070896	C	A	30.25%	DLX3	nonsynonymous SNV	Technical failure
26	chr17	76499013	G	A	28.24%	DNAH17	synonymous SNV	Validated
27	chr19	8999476	G	C	20.21%	MUC16	nonsynonymous SNV	Heterozygous
28	chr19	9006365	C	A	17.24%	MUC16	nonsynonymous SNV	Not present
29	chr19	9361855	G	A	30.15%	OR7E24	nonsynonymous SNV	Validated
30	chr19	14877162	T	C	20.00%	EMR2	synonymous SNV	Technical failure
31	chr19	36275201	G	A	26.04%	ARHGAP33	nonsynonymous SNV	Validated
32	chr19	50170347	G	A	22.34%	BCL2L12	nonsynonymous SNV	Technical failure
33	chr2	26702178	C	T	28.11%	OTOF	nonsynonymous SNV	Heterozygous
34	chr2	85991195	C	T	28.30%	ATOH8	nonsynonymous SNV	Validated
35	chr20	60718900	C	T	19.70%	SS18L1	synonymous SNV	Technical failure
36	chr20	60888258	G	A	22.95%	LAMA5	synonymous SNV	Validated
37	chr21	44836731	C	T	27.42%	SIK1	nonsynonymous SNV	Technical failure
38	chr22	50752254	G	A	20.16%	DENND6B	nonsynonymous SNV	Validated
39	chr3	45837911	T	C	20.31%	SLC6A20	start lost	Validated
40	chr3	122629742	T	C	19.28%	SEMA5B	nonsynonymous SNV	Validated
41	chr5	115177753	A	G	18.18%	AP3S1	nonsynonymous SNV	Technical failure
42	chr5	140242885	C	G	27.56%	AX746964	nonsynonymous SNV	Heterozygous
43	chr6	5004177	G	A	22.92%	RPP40	synonymous SNV	Validated
44	chr6	26458871	T	C	22.45%	BTN2A1	nonsynonymous SNV	Technical failure
45	chr6	54066945	A	C	18.18%	MLIP	nonsynonymous SNV	Heterozygous
46	chr6	56497767	A	C	14.81%	DST	nonsynonymous SNV	Technical failure
47	chr6	136594277	G	A	20.83%	BCLAF1	nonsynonymous SNV	Technical failure

No.	Chr	Pos	Ref	Alt	VAF	Gene name	Nonsynonymous/ synonymous SNV	Validation outcome
48	chr7	1535876	C	T	14.74%	INTS1	nonsynonymous SNV	Validated
49	chr7	2617927	C	A	15.35%	IQCE	nonsynonymous SNV	Technical failure
50	chr7	150815676	C	T	22.32%	AGAP3	nonsynonymous SNV	Validated
51	chr7	157178323	C	T	26.15%	DNAJB6	nonsynonymous SNV	Technical failure
52	chr8	144921555	T	C	30.41%	NRBP2	nonsynonymous SNV	Validated
53	chr8	144940260	C	T	12.14%	EPPK1	nonsynonymous SNV	Technical failure
54	chr9	95481489	C	T	26.92%	BICD2	nonsynonymous SNV	Heterozygous
55	chrX	135960166	G	A	25.25%	RBMX	nonsynonymous SNV	Heterozygous
56	chrX	153416209	C	T	21.17%	OPN1LW	nonsynonymous SNV	Technical failure

Appendix 62 A histogram of the minor allele frequency of all putative somatic variants.

A histogram showing the proportion of putative somatic variants (before further filtering) against their minor allele frequency (MAF) in the population within the ExAC database are shown. The frequency for both non-Finnish Europeans (NFE) and all individuals (ALL) are shown.



Appendix 63 A table of all cases included within the CNV study.

A table of the clinicopathological diagnosis of all cases within the study. All individuals in each cohort had both clinical and pathological diagnosis of each disorder. The men age of onset (from clinical notes) and age of death is shown together with the number of cases for which this information was available. Cases included in the cohort statistical analysis are included above the grey row in the table.

Disease cohort	Total	Gender		Mean age of onset (SD) (number of cases)	Mean age of death (SD) (number of cases)
		Female	Male		
AD	242	124	118	63.67 (10.48) (n=75)	77.01 (12.06) (n=241)
CJD	213	92	121	51.58(19.85) (n=210)	52.50(19.43) (n=213)
Control	349	122	227		63.08 (18.97) (n=349)
PD-DLB	38	60	29	59.00 (9.98) (n=29)	71.74 (8.74) (n=38)
FTD-ALS	238	100	138	59.08 (11.34) (n=74)	64.42 (11.62) (n=234)
Control (High Braak)	36	24	12		87.97 (8.09) (n=36)
CBD	12	3	9	60.07 (11.40) (n=7)	69.60 (12.35) (n=10)
HD	7	3	4	59.80 (13.18) (n=5)	66.71 (10.25) (n=7)
MSA	9	1	8	58.40 (6.50) (n=5)	66.44 (7.60) (n=9)
Other neurological disorders (Supplementary table 2)	68	29	39	54.19 (25.64) (n=31)	70.91 (19.17) (n=66)
PSP	17	9	8	68.70 (12.13 (n=10)	77.12 (10.27) (n=17)
Vascular disease	39	23	16	77.89 (9.36) (n=9)	85.69 (6.36) (n=39)
Vascular disease / AD	5	3	2	80.33 (3.22 (n=3)	84.20 (7.73 (n=5)
Vascular disease / Control	18	8	10	81.20 (13.66) (n=5)	89.83 (7.25 (n=18)
Total	**1342**	**570**	**772**	**57.82 (17.56) (n=501)**	**67.34 (18.08) (n=1331)**

Appendix 64 A table of all cases that did not fulfil criteria for inclusion in a major disease cohort.

A table of the clinicopathological diagnoses of cases within the cohort that did not fulfill the diagnostic criteria for any of the major cohorts.

Number	Gender	Age of onset	Age of death	Clinicopathological diagnosis
1	Female	90	91	Unspecified dementia - Vascular disease / Dementia with
2	Male		67	Atypical tauopathy
3	Male		89	Atypical tauopathy
4	Female		64	Atypical tauopathy
5	Male	62	66	Atypical tauopathy
6	Female		65	Atypical tauopathy
7	Male	64	80	Atypical tauopathy
8	Female	63	72	Superficial siderosis
9	Female			Spino Muscular Atrophy
10	Female		58	Spinocerebellar Ataxia Type 7
11	Female		55	Spinocerebellar Ataxia (no molecular diagnosis)
12	Male		81	Spinocerebellar Ataxia Type 2
13	Male		103	Spinocerebellar Ataxia Type 14
14	Male		76	Spinocerebellar Ataxia Type I
15	Male		50	Spinocerebellar Ataxia
16	Female		95	High Braak stage pathology – clinical history unknown
17	Female		79	Frontotemporal dementia pathologically – no *ante-mortem*
18	Male	67	77	Pathological DLB – no clinical history
19	Female		96	Pathological DLB – no clinical history
20	Male		68	Atypical tauopathy – no *antemortem* history
21	Male	52	64	Uncategorized dementia
22	Female	62	82	Probable AD – no clinical history
23	Male		82	Possible pre-symptomatic Parkinson's disease
24	Male		87	Possible pre-symptomatic Parkinson's disease
25	Male			Mixed Parkinson's disease and Motor Neuron Disease
26	Male	64	65	Normal Pressure Hydrocephalus
27	Male	81	89	Neurofibrillary tangle only dementia
28	Male	88	96	Neurofibrillary tangle only dementia
29	Male	72	78	Neurofibrillary tangle only dementia
30	Female	14	16	Neuroaxonal Dystrophy
31	Male		73	Neorcortical Lewy Body Deposition
32	Male	44	58	Multiple Sclerosis
33	Female	15	52	Multiple Sclerosis
34	Male	44	75	Multiple Sclerosis
35	Male	79	77	Mixed pathology
36	Female	63	75	Alzheimer's Disease and Dementia with Lewy Bodies
37	Male	0	3	Mitochondrial disease
38	Male	13	49	MELAS
39	Male		85	Alzheimer's disease pathology – mild cognitive impairment
40	Male	10	54	Learning difficulties and epilepsy
41	Female		90	Lewy body disease
42	Female		90	Lewy body disease
43	Male	43	58	Kuf's disease
44	Female		52	Undiagnosed neurological disorder
45	Female	40	62	Huntington's disease phenocopy
46	Female		24	Status epilepticus
47	Male		34	Demyelinating disease (uncategorized)
48	Female		53	Demyelinating disease (uncategorized)
49	Male		69	Chronic Traumatic Encephalopathy
50	Male		40	Chorea-acanthocytosis
51	Female	12	58	Cerebello-olivary degeneration
52	Female		75	Cerebello-olivary degeneration
53	Female		84	Corticobasal degeneration and Alzheimer's disease
54	Male	40	62	CADASIL
55	Female	76	83	Probable Alzheimer's disease
56	Male	60	73	Atypical substantia tauopathy – possible CBD
57	Female	79	82	Atypical substantia nigral degeneration
58	Male		85	Atypical nigral degeneration

Number	Gender	Age of onset	Age of death	Clinicopathological diagnosis
59	Male		86	Atypical appearances
60	Male	55	65	Mild cognitive impairment
61	Female	85	85	Atypical neurodegenerative syndrome
62	Female		84	Amyloid angiopathy
63	Male		86	Argyrophilic Grain disease
64	Male		86	Argyrophilic Grain disease
65	Female	78	88	Argyrophilic Grain disease
66	Female		90	Argyrophilic Grain disease
67	Male	65	71	Adult onset gangliosidosis
68	Male		73	Alzheimer's disease and frontotemporal dementia

Appendix 65 Clinical demographics and characteristics of all CJD cases within the study.

The total number of individuals, together with the age of onset and death (in years) for which clinical data were available is shown.

	Total	Gender		Age of onset (SD)	Number of cases	Age of death (SD)	(n=number of cases)
		Female	Male				
Familial CJD	15	6	9	56.92 (7.21)	(n=13)	55.13 (9.36)	(n=15)
Iatrogenic CJD	21	6	15	31.52 (5.63)	(n=21)	32.62 (5.41)	(n=21)
Sporadic CJD	108	52	57	66.31 (9.51)	(n=107)	66.97 (9.19)	(n=108)
Sporadic CJD, panencephalopathic	8	5	4	54.25 (19.04	(n=8)	56.50 (18.37)	(n=8)
Variable Protease Sensitive Prionopathy	4	3	1	64.75 (10.72)	(n=4)	67.00 (9.83)	(n=4)
Variant CJD	57	20	37	28.79 (10.54)	(n=57)	30.14 (10.41)	(n=57)
Total	213	92	123	51.58 (19.85)	(n=210)	52.50 (19.43)	n=213

Appendix 66 The Prion Protein (PrP) isoforms of all cases within each subtype of CJD.

The number of individuals with each clinical diagnosis (first column) are shown within the relevant PrP isoform column.

	1A(+2)	1+2(A)	1A	1A (+LMWt)	1A (doublet) (+2)	1A/B	1B	2A	2A(+1)	2A (+LMWt)	2A (doublet)	2A/B	2B	LMWt	LMWt (+2A)	Total
Familial CJD	0	0	6	0	0	2	4	2	0	0	0	1	0	0	0	15
Iatrogenic CJD	4	3	3	1	0	0	0	8	2	0	0	0	0	0	0	21
Sporadic CJD	14	2	48	1	1	0	0	29	9	1	3	0	0	0	0	108
Sporadic CJD, panencephalopathic	0	0	6	0	0	0	0	1	1	0	0	0	0	0	0	8
Variable Protease Sensitive Prionopathy	0	0	0	0	0	0	0	0	0	0	0	0	0	2	2	4
Variant CJD	0	0	0	0	0	0	0	0	0	0	0	0	57	0	0	57
Total	18	5	64	2	1	2	4	41	12	1	3	1	57	2	2	215

Appendix 67 The PRNP codon 129 genotype of all CJD cases.

	MM	MM DEL34	MM 4XINS	MM 6XINS	MM 7XINS	MM D167G	MM E200K	MM E211Q	MV	MV A117A	MV E200K	VV	VV A117A	Total
Familial CJD	0	0	4	1	1	1	4	1	0	0	3	0	0	15
Iatrogenic CJD	4	0	0	0	0	0	0	0	10	0	0	6	0	20
Sporadic CJD	75	0	0	0	0	0	0	0	15	2	0	15	1	108
Sporadic CJD, panencephalopathic	5	1	0	0	0	0	0	0	2	0	0	0	0	8
Variable Protease Sensitive Prionopathy	0	0	0	0	0	0	0	0	1	0	0	3	0	4
Variant CJD	57	0	0	0	0	0	0	0	0	0	0	0	0	57
Total	141	1	4	1	1	1	4	1	28	2	3	24	1	213

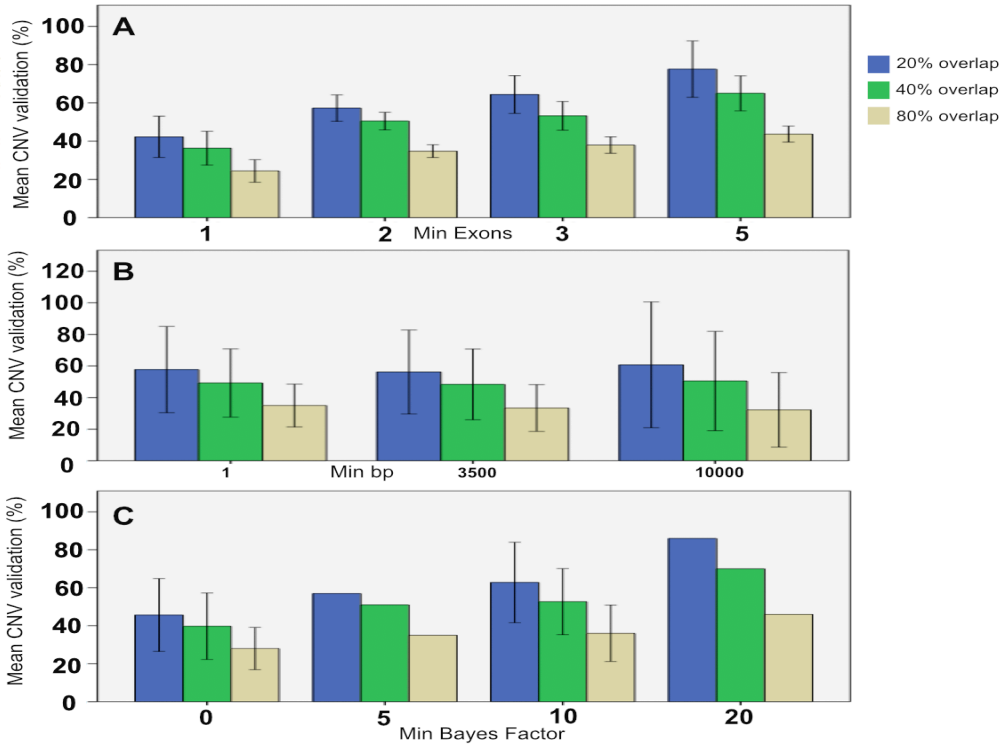
Appendix 68 The overlap of CNVs called by SNP genotyping and validated by exome sequencing when stratified by calling parameters.

A table showing the overlap of CNVs called by SNP genotyping by those called by exome sequencing. The proportion of cases overlapping at different thresholds of CNV calling confidence, length, and number of exons are shown.

Percentage overlap of CNVs	Minimum Bayes Factor	Minimum CNV length	Minimum exons	Mean % of CNVs that overlap	Number of calls	Number of individuals
20	0	1	1	34	1980	89
20	10	1	1	46	1039	89
20	0	1	2	54	1005	89
20	5	1	2	57	785	89
20	0	1	3	57	738	89
20	10	1	3	67	494	85
20	10	1	5	75	369	84
20	10	1	5	72	369	84
40	0	1	1	30	1980	89
40	10	1	1	40	1039	89
40	0	1	2	48	1005	89
40	5	1	2	51	785	89
40	0	1	3	50	738	89
40	10	1	3	50	494	85
40	10	1	5	64	369	84
40	10	1	5	61	369	84
80	0	1	1	23	1980	89
80	10	1	1	29	1039	89
80	0	1	2	33	1005	89
80	5	1	2	35	785	89
80	0	1	3	35	738	89
80	10	1	3	40	494	85
80	10	1	5	43	369	84
80	10	1	5	42	369	84
20	0	3500	1	41	1177	89
20	10	3500	2	61	570	88
20	10	3500	3	67	471	85
40	0	3500	1	36	1177	89
40	10	3500	2	52	570	88
40	10	3500	3	57	471	85
80	0	3500	1	25	1177	89
80	10	3500	2	36	570	88
80	10	3500	3	39	471	85
20	0	10000	1	42	811	89
20	10	10000	1	48	565	89
20	10	10000	3	67	384	85
20	20	10000	5	86	234	77
40	0	10000	1	35	811	89
40	10	10000	1	41	565	89
40	10	10000	3	56	384	85
40	20	10000	5	70	234	77
80	0	10000	1	24	811	89
80	10	10000	1	21	565	89
80	10	10000	3	38	384	85
80	20	10000	5	46	234	77

Appendix 69 The overlap between CNVs called by SNP genotyping and validated by Exome sequencing.

The overlap between CNVs called by SNP genotyping and validated by Exome sequencing stratified by minimum exons (A), length (B) and Bayes factor (C). The minimum overlap was varied in each case.



Appendix 70 A table of CNVs that were significantly enriched within the CJD cohort compared to controls.

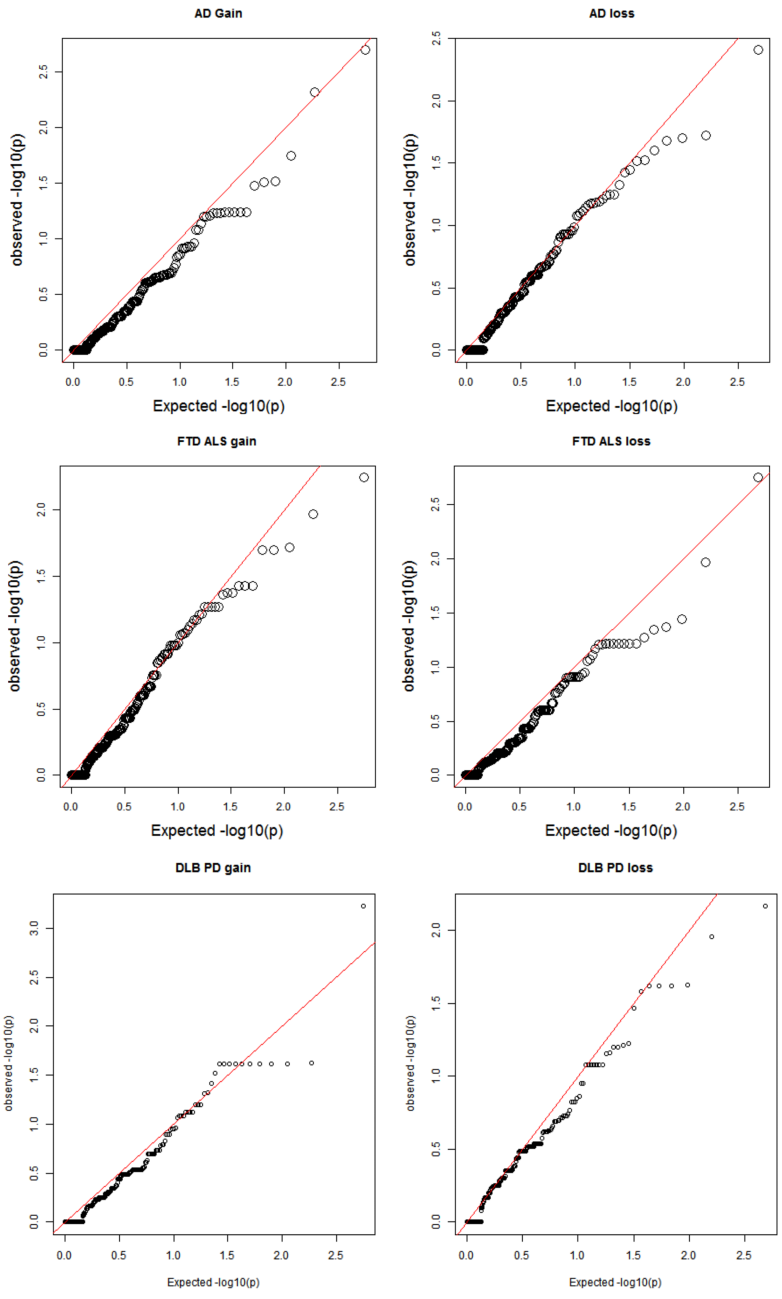
A table of all individuals who possessed a CNV in any of the three genes that were associated with the development of CJD. Case data is shown together with the start and end position of the CNVs called by both SNP genotyping and exome sequencing together with the Bayes factor for each platform.

Case data				CNVs called from SNP genotyping								CNVs called from Exome Sequencing							
Case ID	Number of CNVs	Number of validated CNVs	Form of CJD	CNV(gene)	Chromosome	Start position (hg19)	End position (hg19)	Length (bases)	Bayes factor	Copy number	Number of probes	Start position (hg19)	End position (hg19)	Number of exons	Length (bases)	Bayes Factor	No reads expected	No reads observed	Read ratio
1	1	1	sCJD	LAMA5	20	60885840	60899224	13385	136.758	3	89	60885722	60900602	35	14880	41.6	2418	332	1.38
2	1	1	sCJD	LAMA5	20	60890155	60897748	7594	94.4411	3	43	60891005	60895946	10	4941	8.11	320	451	1.41
3	1	1	sCJD	LAMA5	20	60891045	60894806	3762	66.261	3	23	60891005	60892834	5	1829	7.52	255	368	1.44
4	1	1	sCJD	LAMA5	20	60890375	60897748	7374	35.6858	3	40	60888706	60895946	15	7240	10.8	633	885	1.4
5	2	1	sCJD	LAMA5	20	60890375	60897465	7091	85.4015	3	37	60891728	60897224	10	5496	10.6	481	668	1.39
				EXD3	9	14024256	14024589	3330	16.4247	3	16								
6	3	3	sCJD	LAMA5	20	60885840	60890262	4423	52.7717	3	40	60886219	60895946	25	9727	30	1216	1754	1.44
				LAMA5	20	60890375	60899525	9151	219.967	4	51								
				TRPM2	21	45815307	45826588	11282	52.6353	3	26	45815298	45833957	9	18659	22	520	905	1.74
				EXD3	9	14024252	14025081	8294	129.969	3	39	140242524	140250820	10	8296	12	407	620	1.52
7	2	1	sCJD (pan)	LAMA5	20	60891045	60894806	3762	84.8535	3	23	60891939	60898912	13	6973	8.54	455	623	1.37
				EXD3	9	14024269	14025081	8127	50.1128	3	34								
8	1	1	sCJD	LAMA5	20	60890564	60899224	8661	85.5617	3	48	60886671	60893706	19	7035	19.8	1388	1873	1.35
9	3	1	sCJD	LAMA5	20	60890808	60895688	4881	74.646	3	27	60891005	60894851	8	3846	7.39	370	508	1.37
				TRPM2	21	45825103	45826588	1486	18.3264	3	8								
				EXD3	9	14023888	14024921	10327	17.942	3	34								
10	3	1	sCJD (pan)	LAMA5	20	60890564	60894806	4243	118.71	4	25	60886977	60900602	30	13625	43.2	2292	3211	1.4
				LAMA5	20	60895620	60899525	3906	71.9482	3	25								
				TRPM2	21	45819277	45826588	7312	23.4973	3	17								

Case data				CNVs called from SNP genotyping								CNVs called from Exome Sequencing							
Case ID	Number of CNVs	Number of validated CNVs	Form of CJD	CNV(gene)	Chromosome	Start position (hg19)	End position (hg19)	Length (bases)	Bayes factor	Copy number	Number of probes	Start position (hg19)	End position (hg19)	Number of exons	Length (bases)	Bayes Factor	No reads expected	No reads observed	Read ratio
				EXD3	9	14024252	14025761	15086	31.0872	3	40								
11	3	3	iCJD	LAMA5	20	60890564	60898905	8342	96.5046	3	42	60887230	60900602	29	13372	34.5	3072	406	1.32
				TRPM2	21	45815307	45826588	11282	24.4188	3	26	45811156	45855100	18	43944	28.1	1930	270	1.4
				EXD3	9	14021821	14026750	49292	75.3204	3	54	140243822	140250820	8	6998	14.7	793	113	1.44
12	1	1	fCJD	LAMA5	20	60890375	60892813	2439	33.3155	3	17	60882653	60942301	84	59648	39.4	8526	112	1.32
13	2	1	vCJD	LAMA5	20	60890155	60897748	7594	132.697	3	43	60886671	60897224	24	10553	32.9	1569	222	1.42
				EXD3	9	14024031	14024921	8902	15.2049	3	33								
14	1	0	sCJD	LAMA5	20	60890155	60899206	9052	90.0849	3	50								
15	1	0	sCJD	LAMA5	20	60890375	60892813	2439	15.5536	3	17								
16	3	2	sCJD	LAMA5	20	60891045	60894806	3762	59.346	3	23								
				TRPM2	21	45825103	45826588	1486	25.2829	3	8	45811156	45833957	10	22801	12.1	704	100	1.43
				EXD3	9	14024252	14026097	18446	50.7445	3	41	140246519	140250820	5	4301	9.34	313	494	1.58
17	1	0	sCJD	LAMA5	20	60891783	60894806	3024	32.3525	3	22								
18	1	0	sCJD	LAMA5	20	60891783	60897465	5683	19.1389	3	33								
19	1	0	sCJD	LAMA5	20	60891801	60892813	1013	11.7545	3	12								
20	1	0	sCJD	LAMA5	20	60892450	60893527	1078	17.5156	3	10								
21	1	0	sCJD	LAMA5	20	60892450	60894806	2357	14.9806	3	18								
22	1	0	sCJD	LAMA5	20	60892450	60895688	3239	41.3156	3	21								

Appendix 71 A Q-Q plot of each cohort against all other cases for CNV gains.

Q-Q plots of each cohort (as indicated) against all other cohorts. Copy number gains are shown on the left and losses on the right. All CNVs were filtered to include only those present in less than 1% of the population using DBVar .



Appendix 72 The number of sCJD individuals with LAMA5, EXD3 and TRPM2 copy number gains by PrP isoform.

PpP Isoform	LAMA5 CN gains	Non-LAMA5	Chi-squared test		EXD3 CN gains	Non-EXD3	Chi-squared test		TRPM2 CN gains	Non-TRPM2	Chi-squared test
1A	14	57	p=0.305		6	65	p=0.08		4	67	p=0.105
2A	5	40			0	45			0	45	
Total	**20**	**98**			**7**	**111**			**5**	**113**	

Appendix 73 The number of sCJD individuals with LAMA5, EXD3 and TRPM2 copy number gains by PRNP codon 129 genotype.

PRNP Genotype	LAMA5 CN gains	Non-LAMA5	Chi-squared test		EXD3 CN gains	Non-EXD3	Chi-squared test		TRPM2	Non-TRPM2	Chi-squared test
MM	15	66			6	75			4	77	
MV	1	18	**p=0.359**		0	19	**p=0.255**		0	19	**p=0.409**
VV	3	13			0	16			0	16	
Total	**19**	**97**			**7**	**111**			**5**	**113**	

Appendix 74 The mean age of onset and death of all sCJD cases by PRNP codon 129 genotype irrespective of CNV status.

PRNP genotype	Age onset (mean (SD))	Number of cases	ANOVA	Age death (mean (SD))	Number of cases	ANOVA
MM	65.80 (9.20)	80	**p=0.707**	66.48 (8.80)	81	**p=0.258**
MV	64.63 (16.53)	19		65.89 (15.78)	19	
VV	64.81 (10.25)	16		65.50 (10.07)	16	
Total	**65.44 (10.78)**	**115**		**66.22 (10.31)**	**116**	

Appendix 75 The age of onset and death for cases with and without the LAMA5, EXD3 or TRPM2 CNV by genotype.

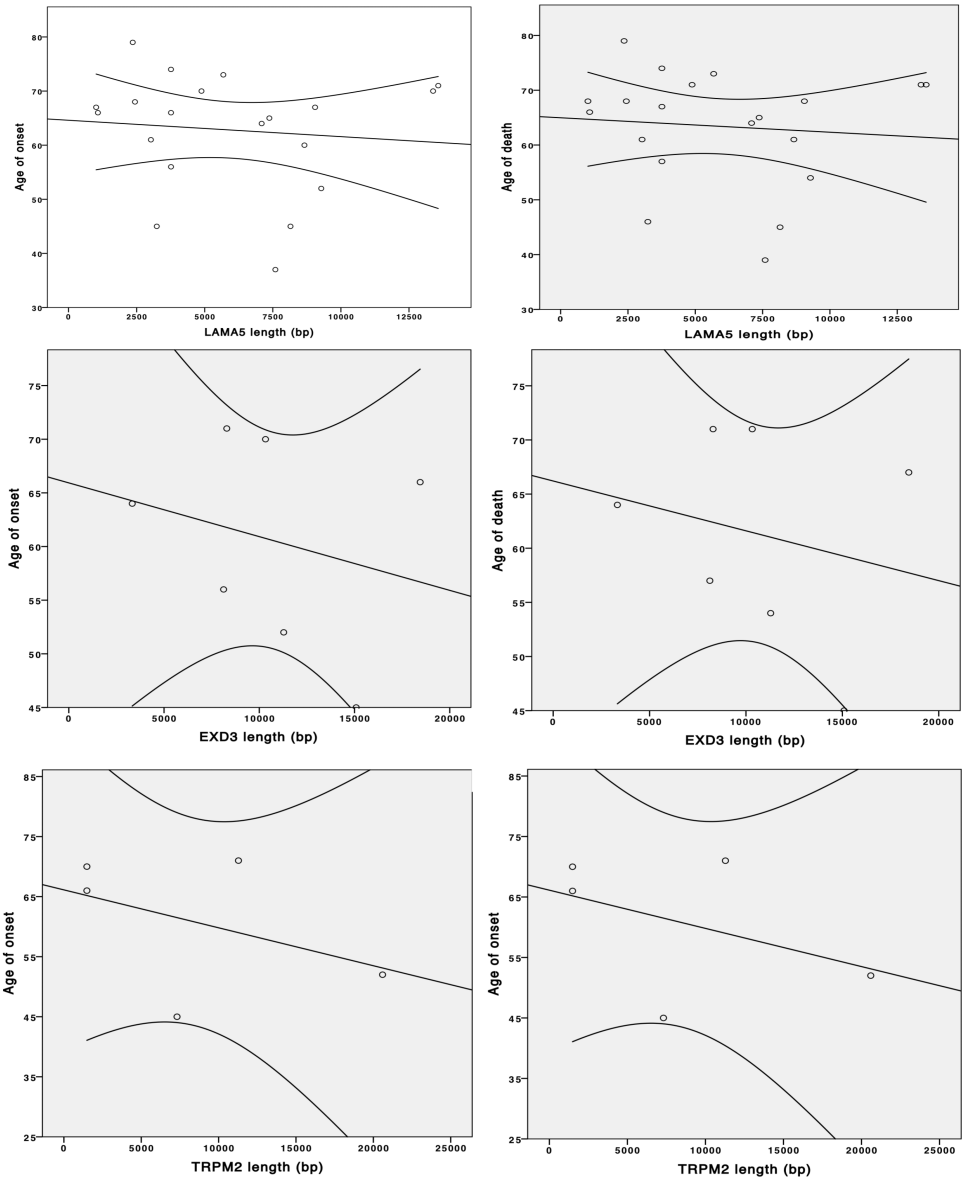
PRNP genotype	Clinical history	CNV	N	Mean	Std. Deviation	T-test
LAMA5						
MM	**Age of onset**	N	65	66.20	8.78	P=0.422
		Y	15	64.07	11.00	
	Age of death	N	66	66.92	8.344	P=0.345
		Y	15	64.53	10.66	
MV	**Age of onset**	N	18	64.83	16.978	P=0.829
		Y	1	61.00	.	
	Age of death	N	18	66.17	16.191	P=0.760
		Y	1	61.00	.	
VV	**Age of onset**	N	13	65.77	9.748	P=0.456
		Y	3	60.67	13.650	
	Age of death	N	13	66.38	9.544	P=0.482
		Y	3	61.67	13.650	
EXD3						
MM	**Age of onset**	N	74	66.11	9.145	P=0.296
		Y	6	62.00	9.90	
	Age of death	N	75	66.8	8.687	P=0.252
		Y	6	62.50	10.03	
MV	**Age of onset**	N	19	64.63	16.523	N/A
		Y	0	.	.	
	Age of death	N	19	65.89	15.779	N/A
		Y	0	.	.	
VV	**Age of onset**	N	16	64.81	10.252	N/A
		Y	0	.	.	
	Age of death	N	16	65.50	10.066	N/A
		Y	0	.	.	
TRPM2						
MM	**Age of onset**	N	76	65.95	9.098	P=0.536
		Y	4	63.00	12.19	
	Age of death	N	77	66.63	8.649	P=0.490
		Y	4	63.50	12.48	
MV	**Age of onset**	N	19	64.63	16.523	N/A
		Y	0[a]	.	.	
	Age of death	N	19	65.89	15.779	N/A
		Y	0	.	.	
VV	**Age of onset**	N	16	64.81	10.252	N/A
		Y	0	.	.	
	Age of death	N	16	65.50	10.066	N/A
		Y	0	.	.	

Appendix 76 The age of onset and death for cases with and without the *LAMA5, EXD3* or *TRPM2* CNV by PrP isoform.

PrP isoform	Clinical history	CNV	N	Mean	Std. Deviation	p-value
LAMA5						
1A	**Age of onset**	N	57	66.07	12.41	0.984
		Y	14	66.00	8.37	
	Age of death	N	57	66.63	11.92	0.936
		Y	14	66.36	8.28	
2A	**Age of onset**	N	39	65.62	7.91	0.211
		Y	5	56.00	14.35	
	Age of death	N	40	66.83	7.37	0.192
		Y	5	57.00	13.95	
EXD3						
1A	**Age of onset**	N	65	66.43	11.82	0.377
		Y	6	62.00	9.90	
	Age of death	N	65	66.95	11.35	0.357
		Y	6	62.50	10.04	
2A	**Age of onset**	N	44	64.52	9.16	N/A
		Y	0		.	
	Age of death	N	45	65.73	8.69	N/A
		Y	0		.	
TRPM2						
1A	**Age of onset**	N	67	66.24	11.71	0.593
		Y	4	63.00	12.19	
	Age of death	N	67	66.76	11.25	0.577
		Y	4	63.50	12.48	
2A	**Age of onset**	N	44	64.52	9.16	N/A
		Y	0		.	
	Age of death	N	45	65.73	8.69	N/A
		Y	0		.	

Appendix 77 The association between the age of disease onset and CNV length in CJD.

Linear regression analysis of disease age of onset and death by CNV length in LAMA5 (top), EXD3 (middle) and TRPM2 (bottom). 95% CI (curved lines) are shown for each linear regression. ($p>0.05$, R^2 0-0.2 in all cases.



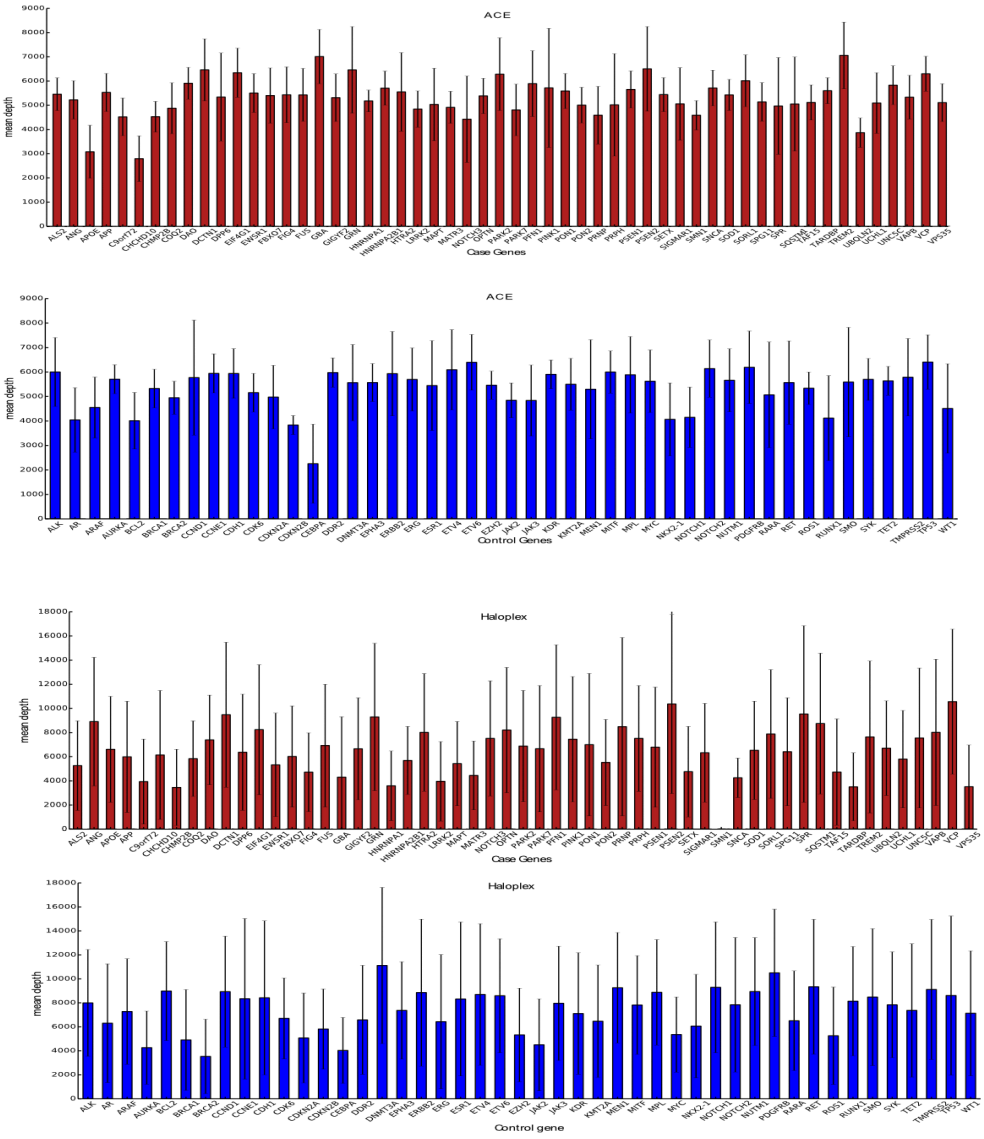
Appendix 78 Neurodegenerative disease and cancer associated genes sequenced.

The 102 Neurodegenerative disease and cancer genes sequenced. Both sets were included on the Accuracy and Content Enhanced (ACE) and Haloplex^{HS} platforms. Left = neurodegenerative disease genes. Right = cancer genes. Key – AD: Alzheimer's disease. DLB: Dementia with Lewy Bodies. FTD-ALS: Frontotemporal dementia – Amyotrophic lateral sclerosis. PD: Parkinson's disease. CJD – Creutzfeldt Jakob Disease.

Case Gene	Disorder		Control Gene	Associated disorder	Associated with haematological malignancy
ANG	FTD-ALS		*ALK*	Lymphoma	Y (Stein, Foss et al. 2000)
APOE	AD / DLB		*AR*	Prostate cancer / Androgen insensitivity	N
APP	AD		*CCND1*	B-cell leukaemia	Y (Aukema, Siebert et al. 2011)
DAO	FTD-ALS		*BCL2*	B-cell leukaemia	Y(Aukema, Siebert et al. 2011)
DCTN1	FTD-ALS		*BRCA1*	Breast cancer	N
EIF4G1	PD		*BRCA2*	Breast cancer	N
EWSR1	FTD-ALS		*CCNE1*	Breast cancer	N
FUS	FTD-ALS		*CDH1*	Stomach cancer	N
GBA	PD/DLB		*CDK6*	Leukaemia	Y (Scheicher, Hoelbl-Kovacic et al. 2015)
GRN	FTD-ALS		*CDKN2A*	Pancreatic cancer	N
HNRNPA1	FTD-ALS		*CDKN2B*	Myelodysplastic syndrome	Y (Kim, Kook et al. 2013)
HNRNPA2B1	FTD-ALS		*DNMT3A*	Acute Myeloid Leukaemia / Clonal haematopoesis	Y (Genovese, Kahler et al. 2014)
PON3	FTD-ALS		*ERBB2*	Lung cancer	N
SQSTM1	FTD-ALS		*ERG*	Acute myeloid leukaemia	Y (Martens 2011)
NOTCH3	CADASIL		*ESR1*	Breast cancer	N
PINK1	PD		*ETV4*	Sarcoma	N
UBQLN2	FTD-ALS		*ETV6*	Acute myeloid leukaemia	Y (Zhang, Churpek et al. 2015)
GIGYF2	PD		*EZH2*	Myelodysplastic syndrome	Y (Nikoloski, Langemeijer et al. 2010)
DPP6	FTD-ALS		*JAK2*	Acute myeloid leukaemia	Y (Baxter, Scott et al. 2005)
C9orf72	FTD-ALS		*JAK3*	Severe Combined Immunodeficiency	Y (Macchi, Villa et al. 1995)
MAPT	FTD-ALS		*KDR*	Haemangioma	N
PARK2	PD		*MEN1*	Multiple Endocrine Neoplasia	N
PFN1	FTD-ALS		*MITF*	Melanoma	N
PON1	FTD-ALS		*MPL*	Myeloproliferation	Y (Ballmaier, Germeshausen et al. 2001)
PON2	FTD-ALS		*MYC*	Burkitt Lymphoma	Y (Bhatia, Huppi et al. 1993)
PRNP	CJD		*DDR2*	No known cancer	N
PRPH	FTD-ALS		*PDGFRB*	Myofibromatosis / IBGC	N
PSEN1	AD		*RARA*	Acute myeloid leukaemia	Y (Borrow, Goddard et al. 1990)
PSEN2	AD		*RET*	Multiple Endocrine Neoplasia	N
SMN1	FTD-ALS		*ROS1*	Myocardial infarction	N
SNCA	PD		*SMO*	Basal cell carcinoma	N
SOD1	FTD-ALS		*AURKA*	Colon cancer	N
SORL1	AD		*SYK*	No known cancer	N
SPR	PD		*NKX2-1*	Medullary cancer	N
UCHL1	PD		*TMPRSS2*	Thyroid cancer	N
VCP	FTD-ALS		*TP53*	Clonal haematopoesis	Y (Wong, Ramsingh et al. 2015)
TAF15	FTD-ALS		*TET2*	Myelodysplastic syndrome	Y (Genovese, Kahler et al. 2014)
UNC5C	AD		*NUTM1*	Midline tumours	N
VAPB	FTD-ALS		*KMT2A*	Acute myeloid leukaemia	Y (Cimino, Moir et al. 1991)
MATR3	FTD-ALS		*CEBPA*	Acute myeloid leukaemia	Y (Smith, Cavenagh et al. 2004)
FIG4	FTD-ALS		*NOTCH2*	Alagille Syndrome	N
OPTN	FTD-ALS		*RUNX1*	Acute myeloid leukaemia	Y (Michaud, Wu et al. 2002)
SIGMAR1	FTD-ALS		*NOTCH1*	Adams-Oliver syndrome	N
PARK7	PD		*WT1*	Wilms tumour	N
SETX	FTD-ALS		*ARAF*	No known cancer	N
TARDBP	FTD-ALS		*EPHA3*	No known cancer	N
FBXO7	PD				
CHMP2B	FTD-ALS				
COQ2	FTD-ALS				
HTRA2	PD				
TREM2	PD/DLB				
VPS35	PD				
ALS2	FTD-ALS				
SPG11	PD				
CHCHD10	FTD-ALS				
LRRK2	PD				

Appendix 79 Additional data on sequencing depth and coverage.

Mean coverage (+SD) for each neurodegenerative and cancer gene on both the ACE (Accuracy and Content Enhanced) and Haloplex^{HS} platforms across all samples



Appendix 80 Genes Clinical and Neuropathological data for all cases and controls.

Both *ante mortem* and *post mortem* diagnoses are given, together with the Braak tau(Braak and Braak 1991) , Thal phase(Thal, Rub et al. 2002), CERAD score(Mirra, Heyman et al. 1991), Braak Lewy body stage(Braak, Del Tredici et al. 2003), and the McKeith stage(McKeith 2006) for each brain. Key – AD: Alzheimer's disease, PD: Parkinson's disease, DLB: Dementia with Lewy Bodies, MCI: Mild Cognitive Impairment.

Case	Clinical diagnosis	Neuro-path diagnosis	Age at death (years)	Braak stage (tau)	Thal Phase	CERAD	Braak Lewy Body Stage	McKeith Stage
1	Control	Control	65	1	0	negative	0	negative
2	Control	Control	88	2	3	negative	0	negative
3	Control	Control	64	1	0	negative	0	negative
4	Control	Control	88	2	1	negative	0	negative
5	Control	Control	78	0	1	negative	0	negative
6	Control	Control	89	3	2	negative	0	negative
7	Control	Control	97	2	2	negative	0	negative
8	Control	Control	85	1	2	negative	0	negative
9	Control	Control	73	0	0	negative	0	negative
10	Control	Control	80	3	2	negative	0	negative
11	Control	Control	81	1	0	negative	0	negative
12	Control	Control	81	2	0	negative	0	negative
13	Control	Control	80	1	1	negative	0	negative
14	Control	Control	93	3	1	negative	0	negative
15	AD	AD	68	6	5	frequent	0	negative
16	AD	AD	96	6	5	frequent	0	negative
17	AD	AD	91	6	5	frequent	0	negative
18	AD	AD	83	6	5	frequent	0	negative
19	AD	AD	78	6	5	frequent	0	negative
20	AD	AD	90	6	5	frequent	0	negative
21	AD	AD	89	6	5	frequent	0	negative
22	AD	AD	86	6	5	frequent	0	negative
23	AD	AD	92	6	5	frequent	0	negative
24	AD	AD	92	6	5	frequent	0	negative
25	AD	AD	81	6	5	frequent	0	negative
26	AD	AD	75	6	5	frequent	0	negative
27	AD	AD	84	6	5	frequent	0	negative
28	AD	AD	85	6	5	frequent	0	negative
29	AD	AD	93	6	5	frequent	0	negative
30	AD	AD	86	6	5	frequent	0	negative
31	AD	AD	84	6	5	frequent	0	negative
32	AD	AD	83	6	4	frequent	0	negative
33	AD	AD	77	6	5	frequent	0	negative
34	AD	AD	85	6	4	frequent	0	negative
35	PD	PD	90	2	0	negative	4	limbic
36	PD+MCI	DLB	80	3	0	negative	6	neocortical
37	DLB	DLB	92	1	0	negative	5	limbic
38	PDD	DLB	86	1	3	negative	5	neocortical
39	PD	PD	70	2	0	negative	4	limbic
40	DLB	DLB	76	2	3	negative	4	limbic
41	DLB	DLB	78	3	4	moderate	6	neocortical
42	DLB	DLB+AD	79	6	5	frequent	6	neocortical
43	PDD	DLB+AD	77	6	5	frequent	6	neocortical
44	DLB	DLB+AD	78	6	5	frequent	4	limbic
45	DLB	DLB+AD	78	6	5	frequent	4	neocortical
46	DLB	DLB+AD	67	6	5	frequent	6	neocortical
47	DLB	DLB	81	3	4	moderate	6	neocortical
48	DLB	DLB	81	3	3	moderate	6	neocortical
49	PDD	DLB	76	3	1	negative	5	neocortical
50	PDD	DLB	83	4	5	moderate	6	neocortical
51	DLB	DLB	73	3	1	negative	6	neocortical
52	DLB	DLB+AD	78	5	4	frequent	6	neocortical
53	DLB	DLB	91	3	4	moderate	6	neocortical
54	PD	PD	83	3	3	negative	4	limbic
Cont: Mean (SD)			81.6 (9.58)	1.6 (1.0)	1.1 (1.0)		0.00 (0.00)	
AD: Mean (SD)			84.9 (6.8)	**6.0 (0.0) ****	**4.9 (0.3)****		0.00 (0.00)	
LB: Mean (SD)			79.9 (6.5)	**3.6 (1.7) ****	**3.00 (1.9)***		**5.25 (0.9)****	

Appendix 81 Cohort sample sizes following quality control.

In total 173 brain samples and 6 paired blood samples from 54 individuals remained (Controls: n=14, Alzheimer's disease: n=20, Lewy body disease: n=20).

	Cerebellum	Entorhinal cortex	Frontal cortex	Medulla	Cingulate	Blood	Total within each disease cohort
Controls	14	14	14	7	5	2	56
Alzheimer disease	20	19	18	0	0	1	58
Lewy body disease	20	20	0	17	5	3	65
Total	54	53	32	24	10	6	179

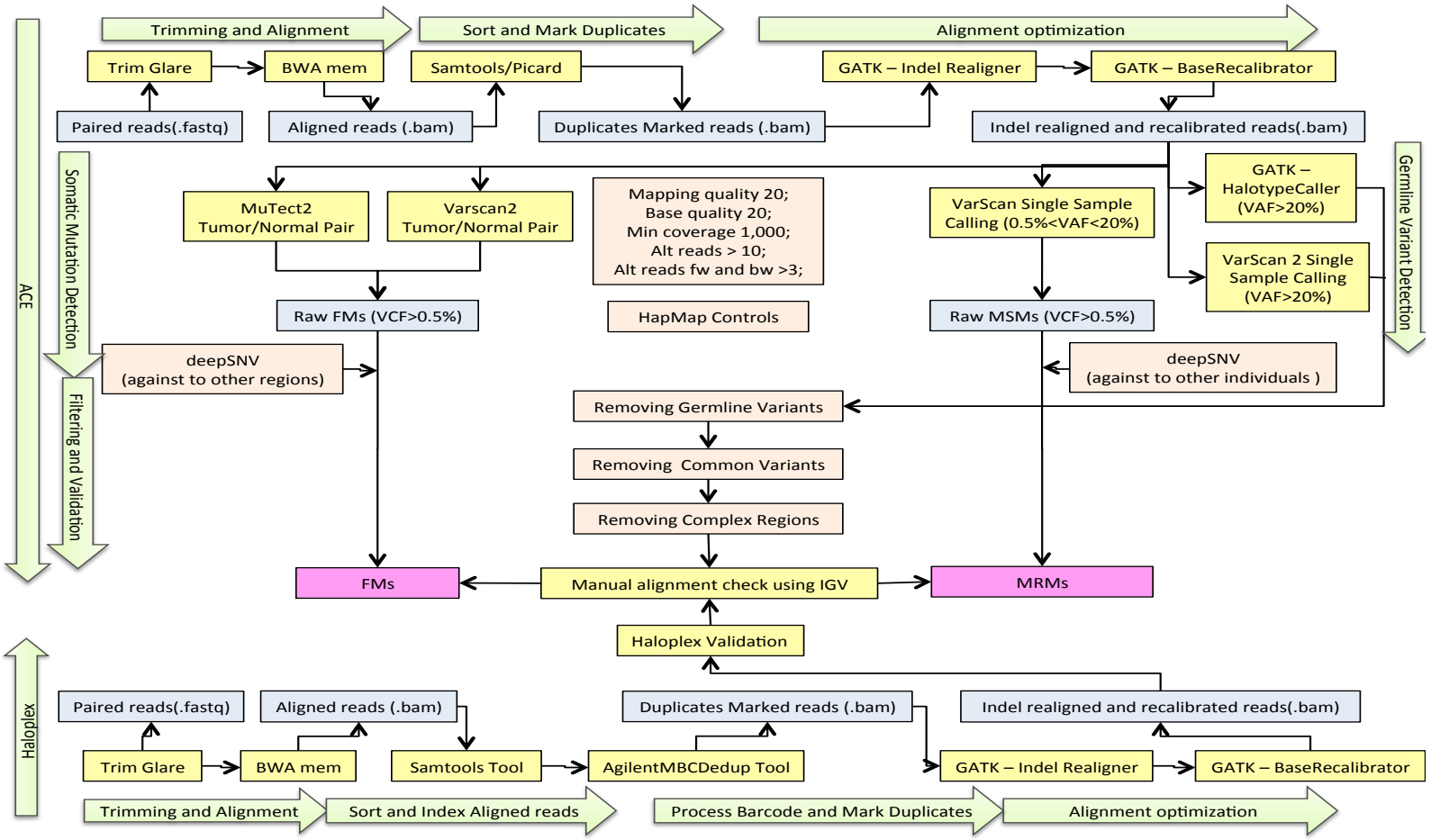
Appendix 82 Sensitivity and specificity of each calling algorithm.

Determined using HapMap CEPH cell lines NA12877 & NA12878

VAF	Caller	Sensitivity	Specificity
0.2%	Mutect	0.3158	0.9999
0.5%	Mutect	0.8070	1.0000
1.0%	Mutect	0.9737	1.0000
2.0%	Mutect	0.9825	1.0000
5.0%	Mutect	1.0000	1.0000
0.2%	Mutect and Varscan (concordant call)	0.2544	1.0000
0.5%	Mutect and Varscan (concordant call)	0.7807	1.0000
1.0%	Mutect and Varscan (concordant call)	0.9386	1.0000
2.0%	Mutect and Varscan (concordant call)	0.9474	1.0000
5.0%	Mutect and Varscan (concordant call)	1.0000	1.0000
0.2%	Mutect or Varscan	0.5614	0.9998
0.5%	Mutect or Varscan	0.9298	1.0000
1.0%	Mutect or Varscan	0.9825	1.0000
2.0%	Mutect or Varscan	1.0000	1.0000
5.0%	Mutect or Varscan	1.0000	1.0000
0.2%	Varscan	0.5000	0.9998
0.5%	Varscan	0.9035	1.0000
1.0%	Varscan	0.9474	1.0000
2.0%	Varscan	0.9649	1.0000
5.0%	Varscan	1.0000	1.0000
0.2%	deepSNV	0.1579	1.0000
0.5%	deepSNV	0.5965	1.0000
1.0%	deepSNV	0.7368	1.0000
2.0%	deepSNV	0.9298	1.0000
5.0%	deepSNV	1.0000	1.0000

Appendix 83 Bioinformatic workflow.

The work-flow for the ACE platform can be seen in the top portion of the figure moving down (identified on the left), and the Haloplex^{HS} pipeline from the bottom up. Green arrows indicate the process, yellow boxes - the bioinformatics programmes used, blue boxes - file outputs, pink boxes - methods. In summary, *ACE platform alignment and processing:* (1) Trimming of the raw paired-end reads (Trim Galore *http://www.bioinformatics.babraham.ac.uk/projects/trim_galore/*) with default settings. Three base pairs (bp) were removed from the 3' end of both paired reads after adapter/quality trimming has been performed. (2) Individual read groups were aligned to GRCh37 human genome by Burrows-Wheeler Aligner (BWA (v0.7.12)). (3) Aligned reads were indexed and sorted with Samtools (v1.3). (4) Duplicate reads were marked by Picard (v1.130) (*https://github.com/broadinstitute/picard*). (5) Genome Analysis Toolkit (v3.5) was used to recalibrate base quality scores and perform local realignment around known insertions and deletions. *Haloplex platform alignment and processing:* (1) As the HaloPlex^{HS} protocol uses a different vector adding an extra base to the adapter, this base was removed from the beginning of raw read 2. Both the read 1 adapter (GAGATCGGAAGAGCACACGTCTGAACTCCAGTCAC) and read 2 adapter (AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT) were trimmed using Trim Galore. (2) Three base pairs (bp) were removed from both 3' end of paired reads and 5' end of paired reads after adapter/quality trimming, and individual read groups were aligned to GRCh37 human genome by Burrows-Wheeler Aligner (BWA (v0.7.12)). (3) Aligned reads were indexed and sorted with Samtools(v1.3)(Li and Durbin 2009). (4) AgilentMBCDedup (*http://www.genomics.agilent.com/en/NGS-Data-Analysis-Software/AgilentMBCDedup-Tool/?cid=AG-PT-154&tabId=prod2510002*) was used to process the Molecular Barcode (MBC) information of HaloPlex^{HS}, and functions to tag read pairs in a bam/sam file with their MBC sequences read out of the index 2 FASTQ files, and mark MBC duplicates from that sam/bam file. (5) Genome Analysis Toolkit (GATK v3.5) was used to recalibrate base quality scores and perform local realignment around known insertions and deletions.(6) BAM files were analysed for coverage depth QC using bedtools (v 2.25) to calculate the depth per base and the histogram of coverage from the target regions using custom scripts. *Key:* BWA – Burrow Wheeler Aligner, GATK – Genome Analysis Tool Kit, FM – Focal Mutation, MRM – Multiple Region Mutation, VCF – Variant Calling Format, VAF – Variant Allele Frequency.


ACE
Somatic Mutation Detection
Filtering and Validation
Germline Variant Detection
Haloplex
Trimming and Alignment
Sort and Mark Duplicates
Alignment optimization
Trim Glare
BWA mem
Samtools/Picard
GATK – Indel Realigner
GATK – BaseRecalibrator
Paired reads(.fastq)
Aligned reads (.bam)
Duplicates Marked reads (.bam)
Indel realigned and recalibrated reads(.bam)
MuTect2 Tumor/Normal Pair
Varscan2 Tumor/Normal Pair
Mapping quality 20;
Base quality 20;
Min coverage 1,000;
Alt reads > 10;
Alt reads fw and bw >3;
VarScan Single Sample Calling (0.5%<VAF<20%)
GATK – HalotypeCaller (VAF>20%)
VarScan 2 Single Sample Calling (VAF>20%)
Raw FMs (VCF>0.5%)
HapMap Controls
Raw MSMs (VCF>0.5%)
deepSNV (against to other regions)
deepSNV (against to other individuals)
Removing Germline Variants
Removing Common Variants
Removing Complex Regions
FMs
Manual alignment check using IGV
MRMs
Haloplex Validation
Paired reads(.fastq)
Aligned reads (.bam)
Duplicates Marked reads (.bam)
Indel realigned and recalibrated reads(.bam)
Trim Glare
BWA mem
Samtools Tool
AgilentMBCDedup Tool
GATK – Indel Realigner
GATK – BaseRecalibrator
Trimming and Alignment
Sort and Index Aligned reads
Process Barcode and Mark Duplicates
Alignment optimization

Appendix 84 Somatic variants detected within the study.

Gene, and base position (in GrCh37 build) are described, together with the relative minor allele frequency within the ExAC and ESP6500 server. Four *in-silico* prediction tools for pathogenicity are also shown (SIFT (Kumar, Henikoff et al. 2009), LRT (Chun and Fay 2009), MutPred (Li, Krishnan et al. 2009) and CADD (Kircher, Witten et al. 2014)) are also shown. Right: relevant clinical data for the specified variant. Both case number relating to the specific case (Appendix 77), the mutation class (Single Region Mutations, SRM; or Multiple Region Mutations, MRM), and whether the variant was sequenced at not present (red) or sequenced and present (blue). Grey boxes signify that the region was not sequenced.

Variant data																Case data								
Case /Control Gene	Gene	Chr	Base Position	Ref allele	Alt allele	Mutation	ExAC MAF	ESP6500 MAF	SIFT_Score	SIFT_Pred	LRT_Score	LRT_Pred	MUT_Score	Mut_Pred	CADD_phred	Case number	Cohort	Mutation class	Cerebellum	Entorhinal Cortex	Frontal Cortex	Medulla	Cingulate gyrus	Blood
Case	*EIF4G1*	3	184052604	G	A	NS	.	.	0.1	T	0	D	1	D	21.8	28	AD	F						
Case	*LRRK2*	12	40709048	G	A	S	.	.	.	.	.	.	.	.	.	24	AD	F						
Case	*NOTCH3*	19	15276289	C	T	NS	0.00000830	.	0	D	0.007	N	1	D	28.2	45	LB	F						
Case	*SETX*	9	135206659	C	A	I	.	.	.	.	.	.	.	.	.	34	AD	F						
Case	*SORL1*	11	121421298	C	T	NS	0.00000827	.	0.03	D	0	D	1	D	22	9	Cont	F						
Case	*UCHL1*	4	41262768	C	T	S	0.00002471	0.00008	.	.	.	.	.	.	.	17	AD	F						
Case	*VPS35*	16	46705662	G	A	S	.	.	.	.	.	.	.	.	.	18	AD	F						
Case	*TAF15*	17	34171702	A	C	S	0.00001069	.	.	.	.	.	.	.	.	12	Cont	MRM						
Control	*ARAF*	X	47424757	T	A	UTR	.	.	.	.	.	.	.	.	.	51	LB	F						
Control	*BRCA1*	17	41234513	C	T	NS	0.00000824	.	0.05	D	0.015	N	1	D	18.54	41	LB	F						
Control	*BRCA2*	13	32950860	C	T	NS	0.00004970	0.00008	0.22	T	0.513	N	1	N	11.47	46	LB	F						
Control	*DNMT3A*	2	25464529	C	T	NS	0.00000839	.	0	D	0	D	1	D	36	49	LB	F						
Control	*ERBB2*	17	37882110	C	T	I	0.00010000	.	.	.	.	.	.	.	.	34	AD	F						
Control	*PDGFRB*	5	149497364	A	ATGGC	FSI	.	.	.	.	.	.	.	.	.	17	AD	F						
Control	*RET*	10	43597857	C	T	S	0.00020000	0.00050	.	.	.	.	.	.	.	7	Cont	F						
Control	*ROS1*	6	117724357	C	T	S	0.00007415	.	.	.	.	.	.	.	.	34	AD	F						
Control	*ROS1*	6	117678995	T	C	NS	.	.	0.71	T	0.36	N	1	D	6.403	35	LB	F						
Control	*SYK*	9	93607773	G	A	NS	0.00000824	.	0.18	T	0	D	1	D	28.5	53	LB	F						
Control	*TET2*	4	106158207	C	T	S	0.00003301	.	.	.	.	.	.	.	.	18	AD	F						
Control	*TET2*	4	106155806	AT	A	FSD	.	.	.	.	.	.	.	.	.	8	Cont	F						
Control	*TET2*	4	106190900	C	T	NS	.	.	0	D	.	.	1	D	28	8	Cont	F						
Control	*TP53*	17	7578281	G	T	NS	.	.	0	D	0	D	1	D	15.26	38	LB	F						
Control	*TP53*	17	7578410	T	A	NS	.	.	0	D	0.006	N	0.969	D	19.94	54	LB	F						
Control	*DNMT3A*	2	25461999	C	G	NS	.	.	.	.	0	D	1	D	33	29	AD	MRM						
Control	*DNMT3A*	2	25462012	G	A	NS	0.00001677	.	0.01	D	0	D	1	D	25.7	37	LB	MRM						
Control	*DNMT3A*	2	25467408	C	T	SSA	0.00003315	.	.	.	.	.	1	D	13.71	41	LB	MRM						

Variant data															Case data								
Case /Control Gene	Gene	Chr	Base Position	Ref allele	Alt allele	Mutation	ExAC MAF	ESP6500 MAF	SIFT_Score	SIFT_Pred	LRT_Score	LRT_Pred	MUT_Score	Mut_Pred	CADD_phred	Case number	Cohort	Mutation class	Cerebellum	Entorhinal Cortex	Frontal Cortex	Medulla	Cingulate gyrus	Blood
Control	*DNMT3A*	2	25462017	T	C	NS	0.00000837	.	0	D	0	D	1	D	23.8	44	LB	MRM						
Control	*DNMT3A*	2	25457242	C	T	NS	0.00050000	0.00070	0.03	D	0	D	1	D	23.1	51	LB	MRM						
Control	*DNMT3A*	2	25463583	G	A	NS	0.00001649	.	0	D	0	D	1	D	23.7	51	LB	MRM						
Control	*KDR*	4	55963895	C	T	NS	.	.	0.47	T	0	D	1	D	36	22	AD	MRM						
Control	*KMT2A*	11	118378293	T	G	NS	0.00000827	.	0.54	T	0.183	N	1	N	7.474	5	Cont	MRM						
Control	*TET2*	4	106158219	A	ATTTGA CCGCTC	FSI	.	.	.	.	.	.	.	.	.	18	AD	MRM						
Control	*TET2*	4	106164068	G	A	NS	.	.	0	D	.	.	1	D	22	20	AD	MRM						
Control	*TET2*	4	106190811	G	A	S	.	.	.	.	.	.	.	.	.	4	Cont	MRM						
Control	*TET2*	4	106182914	A	G	SSA	.	.	.	.	.	.	1	D	21.5	37	LB	MRM						
Control	*TET2*	4	106158372	A	AC	FSI	.	.	.	.	.	.	.	.	.	50	LB	MRM						
Control	*TET2*	4	106158509	G	A	NS	0.00003322	.	0.27	T	0.032	N	1	D	21.7	20	AD	MRM						
Control	*TP53*	17	7578526	C	T	NS	.	.	0	D	0	D	1	D	25.4	31	AD	MRM						
Control	*TP53*	17	7578503	C	T	NS	.	.	0	D	0.016	N	0.995	D	13.43	54	LB	MRM						

Appendix 85 Annotation of the 39 variants detected.

The mutation number, together with individual case number (Appendix 77), and brain region (or blood) are shown together with the amino-acid change are shown. In addition, the Variant Allele Frequency (VAF), together with the corrected p-vale for this variant compared against all other samples from other individuals and other samples in the same individual are shown. Finally, whether the same variant was previously detected by Jaiswal et al (Jaiswal, Fontanillas et al. 2014) , and the final variant classification (SRM=Single Regional Mutation, NP = Not Present, MRM = Multiple Region Variant) are shown.

Mutation number	Case / Control gene	Case number	Cohort	Region	Gene	Chr	Base Position	Ref allele	Alt allele	Amino-acid change	VAF (%)	deepSNV vs all other cases (corrected p-value)	deepSNV vs other regions (corrected p-value)	Previously identified as clonal haematopetic variant	Variant classification
1	Case	28	AD	Cerebellum	EIF4G1	3	184052604	G	A	p.A1375T	0.14%	p > 0.05	p > 0.05	N	NP
1	Case	28	AD	Entorhinal Cortex	EIF4G1	3	184052604	G	A	p.A1375T	0.09%	p > 0.05	p > 0.05	N	NP
1	Case	28	AD	Frontal Cortex	EIF4G1	3	184052604	G	A	p.A1375T	0.85%	**3.16E-14**	**2.90E-06**	N	**SRM**
2	Case	24	AD	Cerebellum	LRRK2	12	40709048	G	A	p.Q1591Q	0.12%	p > 0.05	p > 0.05	N	NP
2	Case	24	AD	Entorhinal Cortex	LRRK2	12	40709048	G	A	p.Q1591Q	0.62%	**7.76E-05**	**0.002614339**	N	**SRM**
2	Case	24	AD	Frontal Cortex	LRRK2	12	40709048	G	A	p.Q1591Q	0.24%	p > 0.05	p > 0.05	N	NP
3	Case	45	LB	Cerebellum	NOTCH3	19	15276289	C	T	p.R1902H	1.12%	**6.10E-12**	**4.67E-09**	N	**SRM**
3	Case	45	LB	Entorhinal Cortex	NOTCH3	19	15276289	C	T	p.R1902H	0.21%	p > 0.05	p > 0.05	N	NP
3	Case	45	LB	Medulla	NOTCH3	19	15276289	C	T	p.R1902H	0.14%	p > 0.05	p > 0.05	N	NP
3	Case	45	LB	Cingulate	NOTCH3	19	15276289	C	T	p.R1902H	0.18%	p > 0.05	p > 0.05	N	NP
4	Case	34	AD	Cerebellum	SETX	9	135206659	C	A	.	0.02%	p > 0.05	p > 0.05	N	NP
4	Case	34	AD	Entorhinal Cortex	SETX	9	135206659	C	A	.	0.60%	**0.003016751**	**3.81E-13**	N	**SRM**
4	Case	34	AD	Frontal Cortex	SETX	9	135206659	C	A	.	0.10%	p > 0.05	p > 0.05	N	NP
5	Case	9	Cont	Cerebellum	SORL1	11	121421298	C	T	p.R729W	0.07%	p > 0.05	p > 0.05	N	NP
5	Case	9	Cont	Entorhinal Cortex	SORL1	11	121421298	C	T	p.R729W	0.47%	**0.00339433**	**0.000229225**	N	**SRM**
5	Case	9	Cont	Frontal Cortex	SORL1	11	121421298	C	T	p.R729W	0.25%	p > 0.05	p > 0.05	N	NP
5	Case	9	Cont	Medulla	SORL1	11	121421298	C	T	p.R729W	0.05%	p > 0.05	p > 0.05	N	NP
6	Case	12	Cont	Cerebellum	TAF15	17	34171702	A	C	p.R464R	4.37%	**1.09E-27**	**1.54E-06**	N	**MRM**
6	Case	12	Cont	Entorhinal Cortex	TAF15	17	34171702	A	C	p.R464R	6.98%	**9.07E-72**	**0.000840152**	N	**MRM**
6	Case	12	Cont	Frontal Cortex	TAF15	17	34171702	A	C	p.R464R	7.35%	**1.29E-61**	**5.32E-05**	N	**MRM**
7	Case	17	AD	Cerebellum	UCHL1	4	41262768	C	T	p.I93I	1.56%	**2.09E-23**	**8.32E-22**	N	**SRM**
7	Case	17	AD	Entorhinal Cortex	UCHL1	4	41262768	C	T	p.I93I	0.09%	p > 0.05	p > 0.05	N	NP
7	Case	17	AD	Frontal Cortex	UCHL1	4	41262768	C	T	p.I93I	0.07%	p > 0.05	p > 0.05	N	NP

Mutation number	Case / Control gene	Case number	Cohort	Region	Gene	Chr	Base Position	Ref allele	Alt allele	Amino-acid change	VAF (%)	deepSNV vs all other cases (corrected p-value)	deepSNV vs other regions (corrected p-value)	Previously identified as clonal haematopoetic variant	Variant classification
8	Case	18	AD	Cerebellum	VPS35	16	46705662	G	A	p.R493R	0.03%	p > 0.05	p > 0.05	N	NP
8	Case	18	AD	Entorhinal Cortex	VPS35	16	46705662	G	A	p.R493R	0.49%	**4.64E-05**	**0.00024859**	N	**SRM**
8	Case	18	AD	Frontal Cortex	VPS35	16	46705662	G	A	p.R493R	0.28%	p > 0.05	p > 0.05	N	NP
9	Control	51	LB	Cerebellum	ARAF	X	47424757	T	A	.	0.05%	p > 0.05	p > 0.05	N	NP
9	Control	51	LB	Medulla	ARAF	X	47424757	T	A	.	0.10%	p > 0.05	p > 0.05	N	NP
9	Control	51	LB	Entorhinal Cortex	ARAF	X	47424757	T	A	.	0.32%	p > 0.05	p > 0.05	N	NP
9	Control	51	LB	Blood	ARAF	X	47424757	T	A	.	0.78%	**3.00E-09**	**3.50E-06**	N	**SRM**
10	Control	41	LB	Cerebellum	BRCA1	17	41234513	C	T	p.G193E	0.07%	p > 0.05	p > 0.05	N	NP
10	Control	41	LB	Entorhinal Cortex	BRCA1	17	41234513	C	T	p.G193E	0.63%	p > 0.05	**6.76E-05**	N	NP
10	Control	41	LB	Medulla	BRCA1	17	41234513	C	T	p.G193E	2.50%	**2.22E-18**	**3.46E-33**	N	**SRM**
11	Control	46	LB	Cerebellum	BRCA2	13	32950860	C	T	p.R2896C	0.15%	p > 0.05	p > 0.05	N	NP
11	Control	46	LB	Entorhinal Cortex	BRCA2	13	32950860	C	T	p.R2896C	0.61%	**6.88E-05**	**2.95E-05**	N	**SRM**
11	Control	46	LB	Cingulate	BRCA2	13	32950860	C	T	p.R2896C	0.07%	p > 0.05	p > 0.05	N	NP
12	Control	29	AD	Cerebellum	DNMT3A	2	25461999	C	G	p.R580T	0.05%	p > 0.05	p > 0.05	N	NP
12	Control	29	AD	Entorhinal Cortex	DNMT3A	2	25461999	C	G	p.R580T	0.55%	**4.12E-11**	**2.78E-08**	N	**MRM**
12	Control	29	AD	Frontal Cortex	DNMT3A	2	25461999	C	G	p.R580T	0.19%	**0.037680238**	p > 0.05	N	**MRM**
13	Control	37	LB	Cerebellum	DNMT3A	2	25462012	G	A	p.P576S	0.37%	**0.001019881**	p > 0.05	N	**MRM**
13	Control	37	LB	Entorhinal Cortex	DNMT3A	2	25462012	G	A	p.P576S	3.49%	**3.37E-81**	**3.90E-39**	N	**MRM**
13	Control	37	LB	Medulla	DNMT3A	2	25462012	G	A	p.P576S	7.23%	**9.94E-186**	**9.61E-111**	N	**MRM**
14	Control	41	LB	Entorhinal Cortex	DNMT3A	2	25467408	C	T	.	1.44%	**2.53E-20**	**1.26E-09**	N	**MRM**
14	Control	41	LB	Medulla	DNMT3A	2	25467408	C	T	.	4.91%	**9.09E-63**	**1.68E-44**	N	**MRM**
14	Control	41	LB	Cerebellum	DNMT3A	2	25467408	C	T	.	0.14%	p > 0.05	p > 0.05	N	NP
15	Control	44	LB	Cerebellum	DNMT3A	2	25462017	T	C	p.N574S	0.26%	p > 0.05	p > 0.05	N	NP
15	Control	44	LB	Entorhinal Cortex	DNMT3A	2	25462017	T	C	p.N574S	2.50%	**3.09E-55**	**1.25E-40**	N	**MRM**
15	Control	44	LB	Cingulate	DNMT3A	2	25462017	T	C	p.N574S	3.12%	**7.72E-63**	**2.01E-59**	N	**MRM**
16	Control	49	LB	Cerebellum	DNMT3A	2	25464529	C	T	p.A439T	0.08%	p > 0.05	p > 0.05	N	NP
16	Control	49	LB	Entorhinal Cortex	DNMT3A	2	25464529	C	T	p.A439T	0.17%	p > 0.05	p > 0.05	N	NP
16	Control	49	LB	Medulla	DNMT3A	2	25464529	C	T	p.A439T	0.50%	**0.000115263**	**0.001837809**	N	**SRM**
17	Control	51	LB	Cerebellum	DNMT3A	2	25457242	C	T	p.R659H	0.20%	p > 0.05	p > 0.05	N	NP
17	Control	51	LB	Entorhinal Cortex	DNMT3A	2	25457242	C	T	p.R659H	1.44%	**1.32E-14**	**1.62E-11**	N	**MRM**
17	Control	51	LB	Blood	DNMT3A	2	25457242	C	T	p.R659H	3.02%	**8.30E-39**	**1.85E-37**	N	**MRM**
17	Control	51	LB	Medulla	DNMT3A	2	25457242	C	T	p.R659H	1.37%	**1.07E-11**	**1.20E-10**	N	**MRM**

Mutation number	Case / Control gene	Case number	Cohort	Region	Gene	Chr	Base Position	Ref allele	Alt allele	Amino-acid change	VAF (%)	deepSNV vs all other cases (corrected p-value)	deepSNV vs other regions (corrected p-value)	Previously identified as clonal haematopoetic variant	Variant classification
18	Control	51	LB	Cerebellum	DNMT3A	2	25463583	G	A	p.P477L	0.24%	p > 0.05	p > 0.05	N	NP
18	Control	51	LB	Entorhinal Cortex	DNMT3A	2	25463583	G	A	p.P477L	0.98%	**2.37E-18**	**2.63E-07**	N	**MRM**
18	Control	51	LB	Blood	DNMT3A	2	25463583	G	A	p.P477L	10.58%	**8.80E-220**	**4.99E-246**	N	**MRM**
18	Control	51	LB	Medulla	DNMT3A	2	25463583	G	A	p.P477L	1.51%	**5.75E-28**	**2.16E-17**	N	**MRM**
19	Control	34	AD	Cerebellum	ERBB2	17	37882110	C	T	.	0.08%	p > 0.05	p > 0.05	N	NP
19	Control	34	AD	Entorhinal Cortex	ERBB2	17	37882110	C	T	.	0.71%	**0.000150048**	**4.29E-09**	N	**SRM**
19	Control	34	AD	Frontal Cortex	ERBB2	17	37882110	C	T	.	0.04%	p > 0.05	p > 0.05	N	NP
20	Control	22	AD	Cerebellum	KDR	4	55963895	C	T	p.E850K	0.52%	**6.04E-05**	p > 0.05	N	**MRM**
20	Control	22	AD	Entorhinal Cortex	KDR	4	55963895	C	T	p.E850K	0.81%	**2.23E-09**	p > 0.05	N	**MRM**
20	Control	22	AD	Frontal Cortex	KDR	4	55963895	C	T	p.E850K	1.50%	**4.32E-16**	**3.24E-05**	N	**MRM**
21	Control	5	Cont	Cerebellum	KMT2A	11	118378293	T	G	p.S3561A	0.04%	p > 0.05	p > 0.05	N	NP
21	Control	5	Cont	Entorhinal Cortex	KMT2A	11	118378293	T	G	p.S3561A	0.49%	**1.34E-05**	**0.000522044**	N	**MRM**
21	Control	5	Cont	Frontal Cortex	KMT2A	11	118378293	T	G	p.S3561A	0.34%	p > 0.05	**0.044828901**	N	NP
21	Control	5	Cont	Medulla	KMT2A	11	118378293	T	G	p.S3561A	1.26%	**5.41E-17**	**8.33E-16**	N	**MRM**
22	Control	17	AD	Cerebellum	PDGFRB	5	149497364	A	ATGGC	p.I985fs	0.64%	p > 0.05	p > 0.05	N	**SRM**
22	Control	17	AD	Entorhinal Cortex	PDGFRB	5	149497364	A	ATGGC	p.I985fs	0.06%	p > 0.05	p > 0.05	N	NP
22	Control	17	AD	Frontal Cortex	PDGFRB	5	149497364	A	ATGGC	p.I985fs	0.08%	p > 0.05	p > 0.05	N	NP
23	Control	7	Cont	Cerebellum	RET	10	43597857	C	T	p.G135G	0.10%	p > 0.05	p > 0.05	N	NP
23	Control	7	Cont	Entorhinal Cortex	RET	10	43597857	C	T	p.G135G	0.19%	p > 0.05	p > 0.05	N	NP
23	Control	7	Cont	Frontal Cortex	RET	10	43597857	C	T	p.G135G	0.96%	**3.32E-07**	**6.59E-10**	N	**SRM**
23	Control	7	Cont	Medulla	RET	10	43597857	C	T	p.G135G	0.07%	p > 0.05	p > 0.05	N	NP
24	Control	35	LB	Cerebellum	ROS1	6	117678995	T	C	p.I1276V	0.04%	p > 0.05	p > 0.05	N	NP
24	Control	35	LB	Entorhinal Cortex	ROS1	6	117678995	T	C	p.I1276V	0.09%	p > 0.05	p > 0.05	N	NP
24	Control	35	LB	Blood	ROS1	6	117678995	T	C	p.I1276V	0.56%	**9.90E-06**	**1.75E-05**	N	**SRM**
24	Control	35	LB	Medulla	ROS1	6	117678995	T	C	p.I1276V	0.08%	p > 0.05	p > 0.05	N	NP
25	Control	34	AD	Cerebellum	ROS1	6	117724357	C	T	p.A174A	0.94%	**1.42E-15**	**1.54E-13**	N	**SRM**
25	Control	34	AD	Entorhinal Cortex	ROS1	6	117724357	C	T	p.A174A	0.12%	p > 0.05	p > 0.05	N	NP
25	Control	34	AD	Frontal Cortex	ROS1	6	117724357	C	T	p.A174A	0.21%	p > 0.05	p > 0.05	N	NP
26	Control	53	LB	Cerebellum	SYK	9	93607773	G	A	p.A159T	0.50%	**0.000311436**	**0.00137584**	N	**SRM**
26	Control	53	LB	Entorhinal Cortex	SYK	9	93607773	G	A	p.A159T	0.07%	p > 0.05	p > 0.05	N	NP
26	Control	53	LB	Medulla	SYK	9	93607773	G	A	p.A159T	0.21%	p > 0.05	p > 0.05	N	NP
27	Control	4	Cont	Cerebellum	TET2	4	106190811	G	A	p.K1363K	0.24%	p > 0.05	p > 0.05	N	NP

Mutation number	Case / Control gene	Case number	Cohort	Region	Gene	Chr	Base Position	Ref allele	Alt allele	Amino-acid change	VAF (%)	deepSNV vs all other cases (corrected p-value)	deepSNV vs other regions (corrected p-value)	Previously identified as clonal haematopoetic variant	Variant classification
27	Control	4	Cont	Entorhinal Cortex	TET2	4	106190811	G	A	p.K1363K	3.63%	**2.83E-54**	**1.19E-38**	N	**MRM**
27	Control	4	Cont	Frontal Cortex	TET2	4	106190811	G	A	p.K1363K	2.87%	**1.04E-37**	**1.46E-28**	N	**MRM**
27	Control	4	Cont	Medulla	TET2	4	106190811	G	A	p.K1363K	8.38%	**4.70E-139**	**2.49E-108**	N	**MRM**
28	Control	8	Cont	Cerebellum	TET2	4	106155806	AT	A	p.D236fs	0.03%	p > 0.05	p > 0.05	N	NP
28	Control	8	Cont	Entorhinal Cortex	TET2	4	106155806	AT	A	p.D236fs	0.09%	p > 0.05	p > 0.05	N	NP
28	Control	8	Cont	Blood	TET2	4	106155806	AT	A	p.D236fs	0.87%	**NA**	**NA**	N	**SRM**
28	Control	8	Cont	Frontal Cortex	TET2	4	106155806	AT	A	p.D236fs	0.11%	p > 0.05	p > 0.05	N	NP
28	Control	8	Cont	Cingulate	TET2	4	106155806	AT	A	p.D236fs	0.13%	p > 0.05	p > 0.05	N	NP
29	Control	8	Cont	Cerebellum	TET2	4	106190900	C	T	p.T1393I	0%	p > 0.05	p > 0.05	N	NP
29	Control	8	Cont	Entorhinal Cortex	TET2	4	106190900	C	T	p.T1393I	0.08%	p > 0.05	p > 0.05	N	NP
29	Control	8	Cont	Blood	TET2	4	106190900	C	T	p.T1393I	0.50%	**0.000516169**	**0.001211125**	N	**SRM**
29	Control	8	Cont	Frontal Cortex	TET2	4	106190900	C	T	p.T1393I	0.10%	p > 0.05	p > 0.05	N	NP
29	Control	8	Cont	Cingulate	TET2	4	106190900	C	T	p.T1393I	0.10%	p > 0.05	p > 0.05	N	NP
30	Control	18	AD	Cerebellum	TET2	4	106158207	C	T	p.H1036H	0.84%	**1.51E-10**	**2.30E-08**	N	**SRM**
30	Control	18	AD	Entorhinal Cortex	TET2	4	106158207	C	T	p.H1036H	0.06%	p > 0.05	p > 0.05	N	NP
30	Control	18	AD	Frontal Cortex	TET2	4	106158207	C	T	p.H1036H	0.04%	p > 0.05	p > 0.05	N	NP
31	Control	18	AD	Cerebellum	TET2	4	106158219	A	ATTTGACCGCTC	p.L1040fs	0.17%	p > 0.05	p > 0.05	N	NP
31	Control	18	AD	Entorhinal Cortex	TET2	4	106158219	A	ATTTGACCGCTC	p.L1040fs	2.71%	**NA**	**NA**	N	**MRM**
31	Control	18	AD	Frontal Cortex	TET2	4	106158219	A	ATTTGACCGCTC	p.L1040fs	2.73%	**NA**	**NA**	N	**MRM**
32	Control	20	AD	Cerebellum	TET2	4	106158509	G	A	p.G1137D	0.15%	p > 0.05	p > 0.05	N	NP
32	Control	20	AD	Entorhinal Cortex	TET2	4	106158509	G	A	p.G1137D	1.49%	**1.55E-06**	**5.35E-13**	N	**MRM**
32	Control	20	AD	Blood	TET2	4	106158509	G	A	p.G1137D	17.85%	**P<1E-100**	**1.30E-304**	N	**MRM**
32	Control	20	AD	Frontal Cortex	TET2	4	106158509	G	A	p.G1137D	0.56%	p > 0.05	p > 0.05	N	NP
33	Control	20	AD	Cerebellum	TET2	4	106164068	G	A	p.V1163I	0.01%	p > 0.05	p > 0.05	N	NP
33	Control	20	AD	Frontal Cortex	TET2	4	106164068	G	A	p.V1163I	0.51%	p > 0.05	**1.93E-06**	N	NP
33	Control	20	AD	Entorhinal Cortex	TET2	4	106164068	G	A	p.V1163I	1.64%	**6.25E-11**	**2.95E-28**	N	**MRM**
33	Control	20	AD	Blood	TET2	4	106164068	G	A	p.V1163I	17.38%	**<1E-100**	**<1E-100**	N	**MRM**
34	Control	37	LB	Cerebellum	TET2	4	106182914	A	G	.	0.49%	p > 0.05	p > 0.05	N	NP
34	Control	37	LB	Entorhinal Cortex	TET2	4	106182914	A	G	.	4.97%	**2.54E-23**	**2.19E-33**	N	**MRM**
34	Control	37	LB	Medulla	TET2	4	106182914	A	G	.	8.30%	**9.82E-45**	**1.59E-60**	N	**MRM**

Mutation number	Case / Control gene	Case number	Cohort	Region	Gene	Chr	Base Position	Ref allele	Alt allele	Amino-acid change	VAF (%)	deepSNV vs all other cases (corrected p-value)	deepSNV vs other regions (corrected p-value)	Previously identified as clonal haematopetic variant	Variant classification
35	Control	50	LB	Cerebellum	TET2	4	106158372	A	AC	p.T1091fs	0.15%	NA	NA	N	NP
35	Control	50	LB	Entorhinal Cortex	TET2	4	106158372	A	AC	p.T1091fs	1.15%	**NA**	**NA**	N	**MRM**
35	Control	50	LB	Medulla	TET2	4	106158372	A	AC	p.T1091fs	2.69%	**NA**	**NA**	N	**MRM**
36	Control	31	AD	Cerebellum	TP53	17	7578526	C	T	p.C3Y	0.10%	p > 0.05	p > 0.05	N	NP
36	Control	31	AD	Entorhinal Cortex	TP53	17	7578526	C	T	p.C3Y	1.57%	**2.62E-33**	**3.50E-26**	N	**MRM**
36	Control	31	AD	Frontal Cortex	TP53	17	7578526	C	T	p.C3Y	0.68%	**6.79E-05**	**1.69E-05**	N	**MRM**
37	Control	38	LB	Cerebellum	TP53	17	7578281	G	T	p.P58T	0.06%	p > 0.05	p > 0.05	N	NP
37	Control	38	LB	Entorhinal Cortex	TP53	17	7578281	G	T	p.P58T	0.07%	p > 0.05	p > 0.05	N	NP
37	Control	38	LB	Blood	TP53	17	7578281	G	T	p.P58T	0.27%	p > 0.05	p > 0.05	N	NP
37	Control	38	LB	Medulla	TP53	17	7578281	G	T	p.P58T	0.78%	**4.58E-11**	**4.70E-10**	N	**SRM**
38	Control	54	LB	Cerebellum	TP53	17	7578410	T	A	p.R42W	0.05%	p > 0.05	p > 0.05	N	NP
38	Control	54	LB	Entorhinal Cortex	TP53	17	7578410	T	A	p.R42W	0.46%	**7.99E-07**	**4.82E-05**	N	**SRM**
38	Control	54	LB	Medulla	TP53	17	7578410	T	A	p.R42W	0.21%	p > 0.05	p > 0.05	N	NP
39	Control	54	LB	Cerebellum	TP53	17	7578503	C	T	p.V11M	0.08%	p > 0.05	p > 0.05	N	NP
39	Control	54	LB	Entorhinal Cortex	TP53	17	7578503	C	T	p.V11M	0.86%	**1.41E-14**	**1.07E-13**	N	**MRM**
39	Control	54	LB	Medulla	TP53	17	7578503	C	T	p.V11M	1.75%	**3.15E-33**	**1.75E-30**	N	**MRM**

Appendix 86 Mutational signatures of somatic mutations identified.

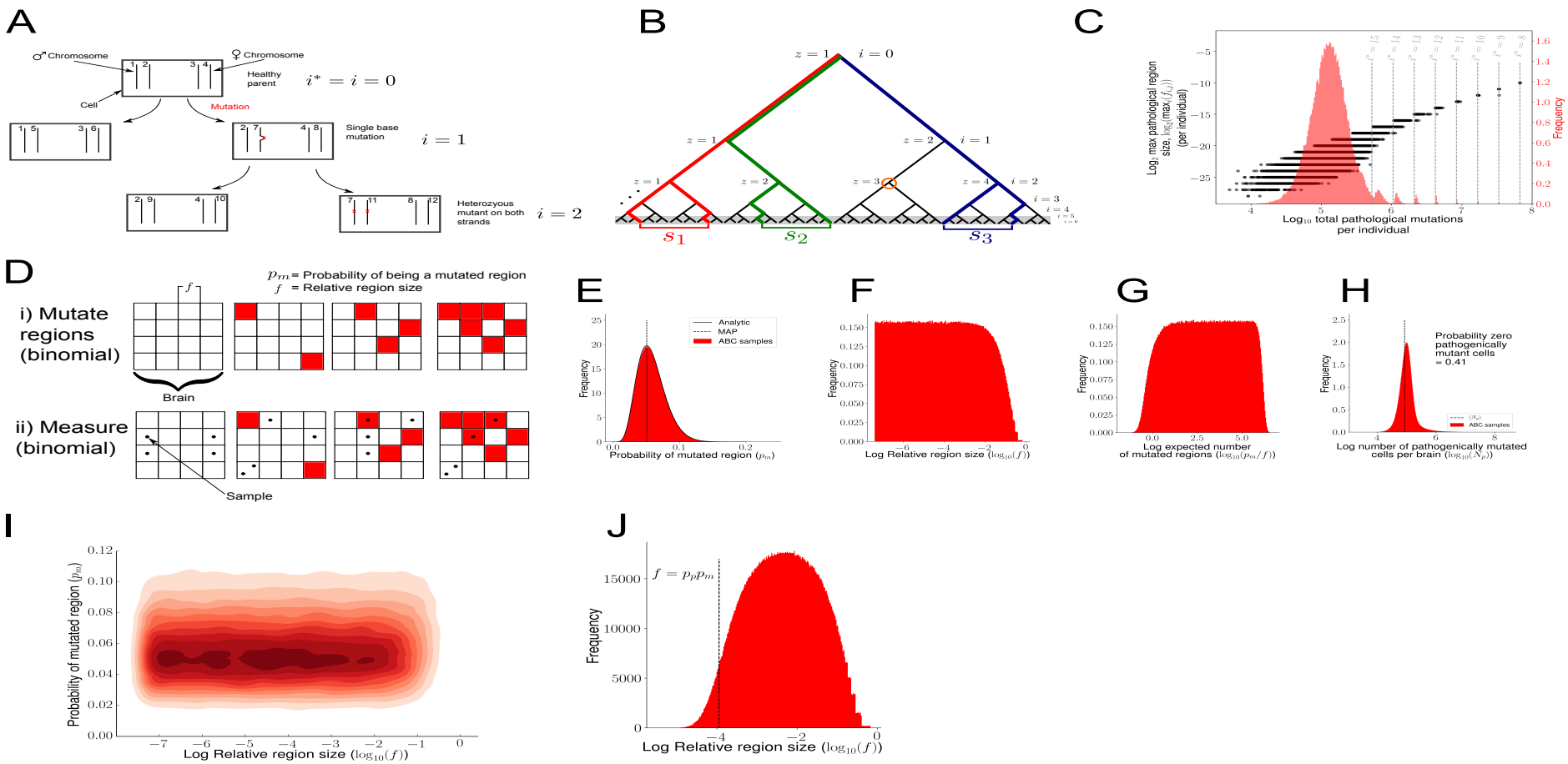
Comparison of the Pearson correlation product moment correlation coefficient values (r^2) between the 21 different mutational signatures observed in cancer as determined by Alexandrov et al(Alexandrov, Nik-Zainal et al. 2013), and the mutational signatures seen in both Single Regional Mutations (SRMs) and Multiple Regional Mutations (MRMs) in our study.

Mutational signature from Alexandrov et al	**Focal mutation (r^2 value)**	**MRM (r^2 value)**
Signature1A	0.617813	0.153022
Signature1B	0.5608662	0.2580154
Signature2	0.05990177	0.1291803
Signature3	-0.1655428	0.08372614
Signature4	-0.04539417	-0.07720455
Signature5	0.1114828	0.2058733
Signature6	0.6615663	0.1960743
Signature7	0.1259546	0.2942886
Signature8	0.1418824	-0.05672829
Signature9	-0.06170881	-0.1156328
Signature10	0.2350385	-0.04068971
Signature11	0.05486709	0.3558389
Signature12	-0.06881005	0.04614821
Signature13	0.01351007	0.03222178
Signature14	0.428101	0.2307443
Signature15	0.6327862	0.2275191
Signature16	-0.064039	0.02431248
Signature17	-0.01968848	-0.01370575
Signature18	0.006402309	-0.08938579
Signature19	0.1848091	0.2813491
Signature20	0.08050852	0.07071721
Signature21	0.01208709	0.08217952

Appendix 87 Mathematical models of somatic mutagenesis in neurodevelopment.

(a-c) Branching process model of neurodevelopment. (a) We idealise the cell as consisting of two chromosomes, one from each parent. We only consider the 132617 bp of the case genes. During division each strand is copied. To illustrate this, strands are labelled with integers and new strands are labelled with a new integer. Neurodevelopment is modelled to begin with a single founder cell, the generation number beginning at $i = 0$. Consider in this example a copying error which generates a SNV in strand 7 whilst copying strand 2 at generation $i^* = i = 0$. Strand 7 is subsequently copied in the following generation to create a heterozygous mutant. Thus, a mutation occurring during copying at generation $i^* = 0$ causes 1 in 4 daughters at generation $i = 2$ (and all subsequent generations) to harbour a heterozygous mutant. In general, the relative region size (f) may be expressed in terms of the mutant generation number as $f = 2^{-(i^*+2)}$. (b) Simulation strategy for approximate Bayesian computation of mutation rate. In this illustrative example, assume that the human brain consists of 2^6 cells and 3 samples (s_1, s_2 and s_3) of 11 cells are taken from the brain. We may assign integer labels to cells from left to right for a given generation number ($z \in [1,2^i]$ for integer z), giving each cell in the lineage a unique coordinate (z,i). Given the coordinates of the sampled cells, we only wish to simulate mutation events in the lineage of parents which give rise to sampled cells: these are the cells which lie on or between the coloured boundaries. For instance, at generation $i = 1$, we must simulate mutations for both cells $z = 1$ and $z = 2$. However, cell (3,2) (circled) has no influence over the sampled cells, and therefore does not need to be simulated. Cells in the final generation ($i = 6$ in this case) form the adult brain and are not replicated. Replication of cells in the penultimate generation cannot give rise to heterozygous mutants on both strands of DNA (see a), and therefore may also be neglected (grey shaded region) since mutations were checked for their presence on the reverse strand. Thus, in this example, we need only consider cells with $i \leq 4$. (c) Multimodality in the distribution of total pathological mutations per individual is driven by the largest pathological region per individual. Correlation between the largest pathological region over all mutations i, for individuals j, and the total number of pathological mutations per individual (grey point for each individual), overlaid with the distribution of total pathological mutations per individual. The number of cells corresponding to regions seeded at generation $i^* = 8,9,\ldots,13$ are shown in grey. At fixed i^*, e.g. $i^* = 11$, for individuals whose largest mutant region occurred at $i^* = 12$ (corresponding to $log_2(f) = -13$, see A), the width of the corresponding mode of the distribution of total pathological mutations is small relative to the number of mutant cells associated with e.g. mutations seeded at $i^* = 12$. Hence, individuals with a mutation seeded at $i^* = 12$ have sufficiently more pathologically mutated cells to induce a separate mode. (d-j) Constant region size model of neurodevelopment. (d) We use a simplified model of mutagenesis in the brain, where the brain consists of independent 'regions' of relative size f, where f is an unknown

constant. (i) Each region (small square) has an independent probability of being homogeneously mutated in any of the case genes (p_m, red square), with the number of mutated regions being binomially distributed. (ii) We model the measurement process as each sample (black circle) being an independent trial, with the number of samples being positive for a mutation as binomially distributed. (e) The marginal posterior distribution of p_m is constrained by the data to be p_m = 0.024–0.11 (95% B.C.I.). Displaying the analytic posterior distribution if p_m is treated independently of f (solid black line), and the maximum a posteriori estimate (MAP) $(p_m)_{MAP} = 6/119$ (dotted black line). This is supported by the lack of covariation observed between p_m and f (see (I)). (f) The marginal posterior distribution of f is not strongly constrained by the data, with $f = 5.7 \times 10^{-8}$–0.11 (95% B.C.I.), although support reduces markedly as $f \to 1$. (g) However, the mean number of mutated regions per individual ($\langle x \rangle = p_m/f$) is constrained strongly enough such that the lower bound of $\langle x \rangle \approx 1$ ($\langle x \rangle = 0.942 \pm 6.6 \times 10^{-3}$, 5th percentile of approximate posterior ± error. The assigned error was determined by splitting the posterior samples into 10 subsets of equal size, and taking the sample standard deviation of the 5th percentile of each subset). This means that individuals have, on average, at least ~ 1 mutant region, but potentially many more. (h) Using the probability that a region is pathologically mutated (= 298 / 132617 = the number of pathological SNVs across the case genes), the number of pathologically mutated neurons per individual may be simulated. 37% of simulated individuals had zero pathologically mutated cells (corresponding to when $f \gtrsim \tilde{p}p_m$ (see J), but those that did tended to have ~ 10^5 mutated neurons. $\langle N_p \rangle = 1.7 \times 10^5$ (black dotted, see Mathematical Supplement (Appendix 86) for derivation), which is in good agreement with Fig. 4E. (i) Joint distribution of p_m and $log_{10}(f)$ shows that p_m and f are well-approximated as being independent. (j) (e) shows that 41% of simulated individuals had zero pathologically mutated cells. For these individuals, we may plot the posterior distribution of relative region sizes, i.e. $P((log_{10}(f)|N_p = 0, \rho(S_m(D), S_d(D)) = 0))$. The support decays when $f \ll \tilde{p}p_m \approx \tilde{p} \cdot (p_m)_{MAP}$ =1.7 × 10^4 (black dotted line). This is the limit that the mean number of pathologically mutated regions per individual is much less than 1, i.e. $\langle \tilde{Y} \rangle \ll 1$, where $\tilde{Y}$ is defined in Eq.(28) of Mathematical Supplement (Appendix 88).


A
♂ Chromosome
♀ Chromosome
Healthy parent
$i^* = i = 0$
Cell
Mutation
Single base mutation
$i = 1$
Heterozyous mutant on both strands
$i = 2$
B
$z = 1$
$i = 0$
$z = 2$
$i = 1$
$z = 3$
$z = 4$
$i = 2$
$i = 3$
s_1
s_2
s_3
C
Log, max pathological region size, $\log_2(\max(J_{ij}))$ (per individual)
Frequency
$\log_{10}$ total pathological mutations per individual
D
p_m = Probability of being a mutated region
f = Relative region size
i) Mutate regions (binomial)
Brain
ii) Measure (binomial)
Sample
E
Analytic
MAP
ABC samples
Frequency
Probability of mutated region (p_m)
F
Frequency
Log Relative region size ($\log_{10}(f)$)
G
Frequency
Log expected number of mutated regions ($\log_{10}(p_m/f)$)
H
Probability zero pathogenically mutant cells = 0.41
ABC samples
Frequency
Log number of pathogenically mutated cells per brain ($\log_{10}(N_p)$)
I
Probability of mutated region (p_m)
Log Relative region size ($\log_{10}(f)$)
J
$f = p_p p_m$
Frequency
Log Relative region size ($\log_{10}(f)$)

Appendix 88 Mathematical supplement for somatic mutagenesis.

Mathematical Supplement

Contents

Supporting Methods

1. Approximate Bayesian computation for crude branching model of neurodevelopment

The genetic information we have allows us to make preliminary statements about the number of cells that might be affected by pathological mutations. The objective here is to place the empirical observations in a simplistic theoretical context, not to provide a definitive set of estimates. For more definitive estimates, more refined models, inference and data would be required. That noted, the inference in Fig. 4B, and outlined below, can be decoupled from Figs. 4C-F by simply using literature estimates for the mutation rates. Fig. 4B can then be interpreted as a check that our data is consistent with mutation rates that others have observed: providing support that the AD/PD/FTD case genes show a similar mutation rate to other genes. The conclusions we make do not then change qualitatively if we use literature values in Figs. 4C-F.

We consider neurodevelopment as a deterministic branching process where every cell has exactly two daughters per generation and the probability of death is zero. We assume that the brain consists purely of neurons, and that neurodevelopment begins with a single founder cell which doubles exactly 36 times, yielding $2^{36} \approx 68.7 \times 10^9$ neurons in the adult brain. We use an integer number of doublings for computational convenience, and chose 2^{36} as the closest number of neurons to experimental measurements of the adult brain, being approximately $86.1 \pm 8.1 \times 10^9$ neurons ($\pm$ SD) [1]. The number of neurons in the adult brain under our model is therefore compatible with this estimate, being within 2 standard deviations of Azevedo *et al.* [1].

Our simulation proceeds by sampling a mutation rate (λ_m, per base per cell division) from a prior distribution, which was motivated by the literature. A study by Roach *et al.* involved family-based whole genome sequencing in peripheral blood cells to deduce the human intergeneration mutation rate as $\sim 1.1 \times 10^{-8}$ per bp per generation per haploid genome [2]. In order to approximate the mutation rate per base per cell division, we estimate the number of cell divisions per generation to form the adult brain as 36 (given our model). Since both strands of DNA are replicated during cell

division, we estimate $\lambda_m \approx 1.1 \times 10^{-8}/(2 \times 36) \approx 10^{-10}$ per base per cell division. However, this order of magnitude estimate is subject to large uncertainty, since it is unclear the extent to which mutation rates observable in the blood between generations is representative of the mutation rate during neurodevelopment. Consequently, we chose to search ± 2 orders of magnitude around this estimate, using the following broad uninformative prior on λ_m:

$$P(\ln(\lambda_m)) = \text{unif}(\ln(10^{-12}), \ln(10^{-8})) \tag{1}$$

where $\text{unif}(a, b)$ is a uniform distribution between a and b. Note that the prior of λ_m is uniform on log-space, and therefore encodes our belief that λ_m is uncertain over 4 orders of magnitude

Given a sample of the mutation rate from the prior, we are then able to simulate neurodevelopment for an individual. Our strategy will be to simulate 40 such individuals (the number of individuals in the case cohort) and, loosely speaking, if a simulation is 'close enough' to the data we will accept the draw from the prior as a sample from the approximate posterior distribution for λ_m, which captures all of our uncertainty in the parameter given the data. This is the essence of approximate Bayesian computation (ABC).

The way in which extraction of tissue samples maps to the simplistic tree structure of our model of the brain is clearly non-trivial. We expect that cells sampled from a tissue sample are close in an evolutionary sense. We therefore make the marked simplification of assuming a 1D embedding of the leaves of the tree and select contiguous sets of 3415 leaves in this embedding (corresponding to the average number of cells per sample, see Extended Figure 3B). By using the experimental mean number of cells per sample in our simulation, we are assuming that all cells in the sample are neuronal. This is a simplifying assumption in the absence of knowledge of the identity (i.e. neuronal or non-neuronal) of the cells in the sample. If neuronal and non-neuronal cells are well-mixed in the brain, then we would expect that experimental samples are effectively a constant fraction smaller than has been observed. Whilst this may increase the variability of the inference presented in Fig. 4B, we would not expect it to affect the location of the distribution.

In Extended Figure 3B we show that it is only necessary to consider the set of parent cells which give rise to the cells in the sample. Once the set of parental cells have been determined, we seek to compute the locations of mutation events in the lineage. Given a draw of the mutation rate λ_m from the prior (Eq.(1)), we assume that each base has a probability λ_m of mutating during replication. The total number of bases which are copied during development (B) is $B = 4N_bN_p$ where N_b is the number of base pairs to be considered (132617 bp in the case genes) and N_p is the number of parent cells which gives rise to the sampled cells. The factor of 4 counts for the ploidy of the human genome (2 genomes) and the number of strands in DNA (2 strands) (see Extended Figure 3A). The total number of mutation events which occur in the lineage (M) is therefore modelled to be

$$P(M) = \text{Binom}(B, \lambda_m) \tag{2}$$

where $\text{Binom}(N, p)$ is the binomial distribution with N trials and probability of success p.

In order to find the locations of mutation events in the lineage, $R_1, R_2, \ldots, R_m$, we wish to sample from the joint distribution $P(R_1 = r_1, R_2 = r_2, \ldots, R_m = r_m, M = m)$ where upper case variables are random variables and lower case variables are realizations of the corresponding random variable. Using the product rule of probability, $P(R_1 = r_1, \ldots, R_m = r_m, M = m) = P(R_1 = r_1, \ldots, R_m = r_m | M = m)P(M = m)$. $P(M = m)$ is given in Eq.(2). Since we have assumed that each base is equally likely to mutate during division, the conditional distribution is assumed to be

$$P(R_1 = r_1, \ldots, R_m = r_m | M = m) = \Pi_{k=1}^{m} \text{unif}\{1, B\} \tag{3}$$

where $\text{unif}\{a, b\}$ denotes the discrete uniform distribution over $[a, b]$ (note an analogous result is provably true in continuous space for Poisson-distributed events [3]). Thus, to sample from the joint distribution, we first draw a Binomial random variable for the total number of mutation events,

and the location of each mutation event will be uniformly distributed across all of the bases in the lineage.

In doing this, we are assuming that mutations never affect the same cell at the same base more than once (although the same mutation may affect multiple different cells). This assumption is justified with the following calculation. Throughout the lineage history each base is assumed to be copied 36 times. If we assume a worst-case mutation rate of the upper-bound of our prior in Eq.(1), $\lambda_m = 1 \times 10^{-8}$, then the total number of times a single base is mutated is $X \sim \text{Binom}(36-2, \lambda_m)$ (the factor of (-2) is due to considerations discussed in Extended Figure 3A). It follows that $P(X \geq 2) = 6.3 \times 10^{-14}$. We thus calculate an upper bound on the total number of times bases are mutated more than once across the entire brain of an individual by assuming that every base in every neuron has an independent chance of being mutated, and therefore the total number of such events is $2^{36} \cdot 4N_b \cdot P(X \geq 2) \approx 2300$ multiple-mutated bases per individual. This is a strong overestimate since neurons share lineage history, so in effect the factor of 2^{36} is actually much lower. Given that the total number of mutation events in the brain is on average $4N_b \cdot \lambda_m \cdot \sum_{k=0}^{34} 2^k \approx 1.8 \times 10^8$ mutations per individual then, even for this overestimate, bases being mutated more than once per cell is an extremely rare event ($\mathcal{O}(10^{-5})$ of mutation events) and is thus neglected.

Finally, once the position of mutations in the lineage have been determined using Eq.(2) and Eq.(3), the overlap between the daughters of the parental mutant cell (resulting in a mutated region of relative size $f = 2^{-(i^*+2)}$, where i^*= mutant generation number, see Extended Figure 3A) and the sample can be determined to find the total number of mutations in sample s from patient p from mutational event j, in the cohort denoted as $X_{j,p,s}$. Note that $\sum_j X_{j,p,s}$ is not necessarily equivalent to the total number of mutated cells per sample, since a single cell may harbour multiple different mutations (e.g. consider the case of the founder cell of the brain containing a mutation, as well as one of its daughters containing a different mutation, then a subset of the final population of cells harbour two mutations per cell).

To determine whether the sample from the prior λ_m may be accepted as a sample from the approximate posterior distribution, we use summary statistics to compare the simulation with data ($\mathcal{D}$). Experimentally, the threshold for detection of a SNV per sample was $\sim 0.5\%$ VAF. Using an average sample size of 3415 cells, this corresponds to a detection threshold of ≥ 17 cells in order for a mutation to be detectable within the sample. Before defining the summary statistics, we define the following indicator functions

$$\Phi(x) = \begin{cases} x \text{ if } x \geq 17 \\ 0 \text{ otherwise} \end{cases} \tag{4}$$

$$I(x) = \begin{cases} 1 \text{ if } x > 0 \\ 0 \text{ otherwise} \end{cases} \tag{5}$$

where $\Phi(x)$ is based on the experimental detection threshold. The three summary statistics $\mathbf{S} = (\Gamma_1, \Gamma_2, \Gamma_3)$ of the model are defined as:

$$\Gamma_1 = \sum_{p=1}^{40} \sum_{s=1}^{s_p} \sum_{j=1}^{M} \Phi(X_{j,p,s}) \tag{6}$$

$$\Gamma_2 = \sum_{p=1}^{40} I\left(\sum_{s=1}^{s_p} \sum_{j=1}^{M} \Phi(X_{j,p,s}) \right) \tag{7}$$

$$\Gamma_3 = \frac{1}{\sum_p \sum_s \sum_j I(\Phi(X_{j,p,s}))} \sum_p \sum_s \sum_j \frac{\Phi(X_{j,p,s})}{3415} \tag{8}$$

where s_p is the number of samples taken for patient p. Γ_1 measures the total number of detectable mutations across all samples and all patients, Γ_2 measures the total number of patients with any

detectable mutations and Γ_3 is the average VAF for all detectable mutations across all samples and patients. Using a hat to denote data, these summary statistics were found experimentally to be:

$$\hat{\Gamma}_1 = 6 \tag{9}$$
$$\hat{\Gamma}_2 = 6 \tag{10}$$
$$\hat{\Gamma}_3 = 0.87 \times 10^{-2}. \tag{11}$$

Thus, the acceptance criteria we use are

$$|\Gamma_1 - \hat{\Gamma}_1| \leq \epsilon_{sa} \tag{12}$$
$$|\Gamma_2 - \hat{\Gamma}_2| \leq \epsilon_{pa} \tag{13}$$
$$|\Gamma_3 - \hat{\Gamma}_3| \leq \epsilon_{v} \tag{14}$$

where $\epsilon_{sa} = \epsilon_{pa} = 1$ and $\epsilon_v = 2\text{SEM}_v$ (SEM_v = standard error in the mean for the VAF across all positive experimental samples), and we demand that all three of the above criteria must be satisfied simultaneously. Given that these criteria are met, we accept λ_m as a draw from the approximate posterior distribution $P(\lambda_m|\rho(\mathbf{S},\mathcal{D}))$ where ρ denotes the acceptance criteria given above. It is possible that a more compact and/or optimal set of summary statistics could be obtained.

Using the draws from the approximate posterior distribution, we may then forward-simulate each individual using a single-draw from $P(\lambda_m|\rho(\mathbf{S},\mathcal{D}))$ and determine the distribution of mutant region sizes and the number of mutated cells per individual (see Fig. 4). The number of mutations that arise at generation i^* (X_{i^*}) is

$$X_{i^*} \sim \text{Binom}(4N_b \cdot 2^{i^*}, \lambda_m) \tag{15}$$

where the number of trials is the total number of bases across all cells at generation i^* (see Fig. 4C).

With this result, we may determine the probability of observing one or more regions seeded at generation i^*, given our posterior estimate for λ_m (see Fig. 4F). We define Q_{i^*} as the proportion of N individuals with one or more mutated regions from generation i^*. Relabelling $\lambda_m \equiv \lambda$, and indexing each individual by j, corresponding to a single draw from the approximate posterior $\lambda_j \sim P(\lambda_j|\rho(\mathbf{S},\mathcal{D}))$, we are able to simulate as many individuals (N) as we have available samples from the approximate posterior ($N \approx 10^5$). We may therefore write

$$Q_{i^*} = \frac{1}{N}\sum_{j=1}^{N} I(X_{i^*,j}), \tag{16}$$

which is the approximate posterior mean of the indicator function. We may then determine the standard deviation of Q_{i^*}. Observing that

$$\langle I(X_{i^*,j})\rangle = 0 \cdot P(X_{i^*,j}=0) + 1 \cdot P(X_{i^*,j}=1) + 1 \cdot P(X_{i^*,j}=2)\ldots \tag{17}$$
$$= 1 - P(X_{i^*,j}=0) \tag{18}$$

it follows that, since all $X_{i^*,j}$ are independent,

$$\langle Q_{i^*}\rangle = \frac{1}{N}\sum_{j=0}^{N}(1 - P(X_{i^*,j}=0)) \tag{19}$$

$$\left[\langle Q_{i^*}^2\rangle - \langle Q_{i^*}\rangle^2\right]^{1/2} = \left[\frac{1}{N^2}\sum_{j=0}^{N}\left\{(1 - P(X_{i^*,j}=0) - [1 - P(X_{i^*,j}=0)]^2\right\}\right]^{1/2} \tag{20}$$

where

$$P(X_{i^*,j} = 0) = (1 - \lambda_j)^{4N_b 2^{i^*}}. \tag{21}$$

Note that the standard deviation of Q_{i^*} scales as $1/N$ due to uncertainty arising from sampling noise in λ_j which diminishes as more samples from the approximate posterior distribution are obtained. Also note that the number of bases used in Eq.(21) is $N_b = 298$, as this is the number of pathological bases in the case genes. An empirical estimate of the standard error may be obtained by treating each individual as a Bernoulli trial with probability $p = P(X_{i^*} \geq 1)$, and SEM $\sqrt{p(1-p)/N}$. These are compatible with our uncertainty in the approximate posterior mean of the indicator function, shown in Eq.(20).

The visual representation in Fig. 4H was created by taking the mean of the approximate posterior samples for the mutation rate $\langle \lambda_m | \mathcal{D} \rangle$ as the mutation rate of the individual. The number of mutated regions for $22 \leq i^* \leq 27$ were drawn using Eq.(15), where $X_{i^*=22}$ was redrawn until $X_{i^*=22} \geq 1$. This is still a reasonable representation of a typical individual, since Fig. 4F shows that $\sim 81\%$ of all simulated individuals possess at least one region of this size. Each circular mutated region has an area proportional to the number of pathologically mutated cells in the corresponding region: $k\pi r^2 = 2^{36} \cdot 2^{-(i^*+2)}$ where k is chosen arbitrarily. The radius and angle of each region relative to the centre of the whole brain was chosen via a uniform random distribution. Although the relative sizes of each region are to scale, the area of the whole brain is not to scale with the mutated regions.

5×10^7 ABC iterations were performed, all code was written in `Python`, and is available from the authors upon request.

Supporting Information

S1. Constant region size model

S1.1. Summary

In this section, we consider an alternative model where the brain is modelled as consisting of independent regions of some characteristic size f, which are homogeneously mutated at the mean VAF experimentally observed in the case genes in the case patients. $f = 1$ suggests that if the brain is mutated at a particular base pair then this mutation is present across the entire brain. However, if $f < 1$ then there exists $1/f$ regions in the brain, each consisting of Nf cells, where N is the number of cells in the brain. We assume that each region has a probability p_m of being homogeneously mutated at a particular base pair in the case genes (Extended Figure 3D). This model does not make a commitment to a tree-like development instead assuming that mutations occupy characteristic volumes of tissue. This makes the model less realistic but it has the virtue of being a different perspective that allows us to probe the robustness of our observations.

The parameters f and p_m were considered as shared amongst all patients, for parsimony. We investigated the effect of varying f by orders of magnitude through approximate Bayesian computation (see Section S1.2). We again restricted our attention to patients with neurodegenerative disease to avoid sampling bias in our extrapolations. We found that, under this model, the relative size of independent regions was not strongly constrained by the data (Extended Figure 3F), finding only that large region sizes $\gtrsim 10\%$ are unlikely. (The tree-like structure considered in the maintext model puts a much stronger constraint on the distribution of sizes of patches of mutated cells).

The data did constrain the mean number of mutated regions per individual such that the lower bound was approximately 1 region per individual ($0.942 \pm 6.6 \times 10^{-3}$ mutated regions per individual, 5th percentile of posterior distribution, see Extended Figure 3G). Consequently, we found that 63% of simulated individuals carried non-zero levels of pathologically mutated cells in case genes, carrying between 0.33–5.0 $\times 10^5$ pathologically mutated cells per affected individual (95% Bayesian credible interval, see Extended Figure 3H). This number of cells with pathological mutations is in broad agreement with Fig. 4E. Thus even this remarkably crude model, that makes no commitments to a

developmental hierarchy, suggests that blocks of pathological mutations occur in most people. That we observe a subset of simulated individuals lacking mutated regions (and therefore pathologically mutated cells) is due to our broad priors in f and p_m allowing for independent brain regions to be large enough that it is unlikely for any pathological regions to arise in an individual (this limit is $f \gtrsim \tilde{p}p_m$ where $\tilde{p} = 298/132617$ is the probability of a region being pathologically mutated, given that it is mutated, see Section. S1.2 and Extended Figure 3J).

S1.2. Approximate Bayesian computation for fixed region size model

We consider the brain as consisting of 'regions', each having a relative mass (f) where $0 < f \leq 1$, implying that each brain contains $1/f$ such regions. For the subset of the data we considered (the case genes in the case cohort), all samples measuring positive for somatic mutations corresponded to unique genes in unique patients. Consequently, we used the parsimonious model that positive samples were homogeneously mutated in a particular case gene at the average VAF (v) determined experimentally ($v = 0.87\%$). We modelled each region as having an independent probability p_m of being homogeneously mutated at a particular base of the case genes.

Defining Y_i as the number of mutated regions in the brain of a single individual (i), then Y_i follows a binomial distribution

$$Y_i \sim \text{Binom}\left(\frac{1}{f}, p_m\right) \tag{22}$$

where $\text{Binom}(n, p)$ is a binomial trial with n attempts each with probability p of success. Note that $1/f$ is constrained to be an integer, and may be obtained by rounding $1/f$ for a given f. We then model the measurement process for each individual as consisting of n_i samples. Since there are Y_i patches, each of which occupies a fraction f of the brain, then fY_i gives the total fraction of the brain which is mutated, where $0 \leq fY_i \leq 1$. This results in M_i samples positive for mutation for individual i, with

$$M_i \sim \text{Binom}(n_i, fX_i). \tag{23}$$

Note that we have modelled all individuals as sharing the only two free parameters in the model: f and p_m.

To infer these parameters, we used an approximate Bayesian computation (ABC) rejection algorithm. We used the uninformative priors

$$\log(f) \sim \text{unif}\left(\log\left(\frac{3415}{86 \times 10^9}\right), 0\right) \tag{24}$$

$$p_m \sim \text{unif}(0, 1) \tag{25}$$

where we chose to use a prior for f which was uniform over log-space, since we were uncertain of its value over orders of magnitude. The lower limit of the prior for f was chosen such that f could not be smaller than the approximate relative size occupied by a single sample i.e. 3415 cells out of the total number of cells in the brain. We have idealized the brain as consisting purely of neurons, of which there are approximately 86×10^9 per brain [1].

We simulated each patient by drawing (f, p_m) from the prior, and determining M_i with Eq.(23). We then used the following summary statistics for the simulated model ($\mathbf{S}$)

$$\mathbf{S} = \left(\sum_{i=1}^{40} M_i, \sum_{i=1}^{40} I(M_i)\right) \tag{26}$$

where I is the indicator function defined in Eq.(5), and the upper summation limit corresponds to the number of case patients. The first element counts the total number of positive samples across

all patients, whereas the second element counts the total number of affected patients. The summary statistic for the data is $\hat{\mathbf{S}} = (6, 6)$. We used the distance metric

$$\rho(\mathbf{S}, \hat{\mathbf{S}}) = \sum |\mathbf{S} - \hat{\mathbf{S}}| \leq \epsilon \tag{27}$$

and required the tolerance $\epsilon = 0$ to accept $(f,\ p_m)$ as a draw from the approximate posterior $P(f, p_m|\rho(\mathbf{S}, \hat{\mathbf{S}}(\mathcal{D})) = 0) \approx P(f, p_m|\mathcal{D})$ for data $\mathcal{D}$. This therefore enforces the ABC rejection algorithm to only accept samples where every mutation occurs once in each affected patient (as observed in the data), and that the correct number of mutant samples and patients are observed overall.

Given posterior samples of (f, p_m), we may also infer the distribution of the number of pathologically mutated neurons per brain. Dropping the patient subscript i, and using Eq.(22), the number of pathologically mutated regions per patient $(\tilde{Y})$ is

$$\tilde{Y} \sim \text{Binom}(Y, \tilde{p}) \tag{28}$$

where $\tilde{p}$ is the probability of a mutated region being pathological, which we take as 298/132617, being the proportion of base pairs in the case genes being associated with pathology. From this, we approximate the number of pathologically mutated cells (N_p) as

$$N_p \approx \tilde{Y} \cdot (Nf) \cdot v \tag{29}$$

where N = number of neurons per brain and v = is the proportion of mutant cells in a positive sample, which we take as the average VAF $v = 0.87 \times 10^{-2}$. The term (Nf) is the number of neurons per region.

It is instructive to observe that the mean number of mutant cells, $\langle N_p \rangle$, for fixed f and p_m is $\langle N_p \rangle = Nfv\langle\tilde{Y}\rangle = Nfv\tilde{p}\langle Y\rangle$. Using Eq.(22), $\langle N_p \rangle = Nfv\tilde{p}p_m/f = Nv\tilde{p}p_m$ *which is independent of* f. Note that if $f \ll \tilde{p}p_m$ then $\langle\tilde{Y}\rangle \gg 1$, in which case an individual is likely to harbour at least one pathological mutation (see Extended Figure 3J). Substituting $N = 86 \times 10^9$, $v = 0.87 \times 10^{-2}$, and the maximum *a posteriori* estimate of $(p_m)_{MAP} = 6/119$, yields $\langle N_p \rangle = 1.70 \times 10^5$ pathologically mutated cells per brain, which approximately corresponds to the peak shown in Extended Figure 3H.

References

[1] **Azevedo FA, Carvalho LR, Grinberg LT, Farfel JM,** et al. 2009. Equal numbers of neuronal and nonneuronal cells make the human brain an isometrically scaled-up primate brain. *J Comp Neurol* **513**: 532–541.

[2] **Roach JC, Glusman G, Smit AF, Huff CD,** et al. 2010. Analysis of genetic inheritance in a family quartet by whole-genome sequencing. *Science* **328**: 636–639.

[3] **Ross SM,** et al. 1996. *Stochastic processes*, volume 2. John Wiley & Sons New York.

References

. "Exome Aggregation Consortium (ExAC)." Retrieved September, 2015, from *http://exac.broadinstitute.org*.

. "NHLBI GO Exome Sequencing Project (ESP)." Retrieved 12/12/2013, 2013, from URL: *http://evs.gs.washington.edu/EVS/*.

Abeliovich, A., Y. Schmitz, I. Farinas, D. Choi-Lundberg, W. H. Ho, P. E. Castillo, N. Shinsky, J. M. Verdugo, M. Armanini, A. Ryan, M. Hynes, H. Phillips, D. Sulzer and A. Rosenthal (2000). "Mice lacking alpha-synuclein display functional deficits in the nigrostriatal dopamine system." Neuron **25**(1): 239-252.

Adzhubei, I. A., S. Schmidt, L. Peshkin, V. E. Ramensky, A. Gerasimova, P. Bork, A. S. Kondrashov and S. R. Sunyaev (2010). "A method and server for predicting damaging missense mutations." Nat Methods **7**(4): 248-249.

Aggleton, J. P., M. M. Albasser, D. J. Aggleton, G. L. Poirier and J. M. Pearce (2010). "Lesions of the rat perirhinal cortex spare the acquisition of a complex configural visual discrimination yet impair object recognition." Behav Neurosci **124**(1): 55-68.

Ahmadi, A., M. Fredrikson, H. Jerregard, A. Akerback, P. A. Fall, A. Rannug, O. Axelson and P. Soderkvist (2000). "GSTM1 and mEPHX polymorphisms in Parkinson's disease and age of onset." Biochem Biophys Res Commun **269**(3): 676-680.

Ahmed, Z., Y. T. Asi, A. Sailer, A. J. Lees, H. Houlden, T. Revesz and J. L. Holton (2012). "The neuropathology, pathophysiology and genetics of multiple system atrophy." Neuropathol Appl Neurobiol **38**(1): 4-24.

Al-Chalabi, A., P. M. Andersen, B. Chioza, C. Shaw, P. C. Sham, W. Robberecht, G. Matthijs, W. Camu, S. L. Marklund, L. Forsgren, G. Rouleau, N. G. Laing, P. V. Hurse, T. Siddique, P. N. Leigh and J. F. Powell (1998). "Recessive amyotrophic lateral sclerosis families with the D90A SOD1 mutation share a common founder: evidence for a linked protective factor." Hum Mol Genet **7**(13): 2045-2050.

Al-Chalabi, A., F. Fang, M. F. Hanby, P. N. Leigh, C. E. Shaw, W. Ye and F. Rijsdijk (2010). "An estimate of amyotrophic lateral sclerosis heritability using twin data." J Neurol Neurosurg Psychiatry **81**(12): 1324-1326.

Al-Chalabi, A., L. H. van den Berg and J. Veldink (2017). "Gene discovery in amyotrophic lateral sclerosis: implications for clinical management." Nat Rev Neurol **13**(2): 96-104.

Alexandrov, L. B., S. Nik-Zainal, D. C. Wedge, S. A. Aparicio, S. Behjati, A. V. Biankin, G. R. Bignell, N. Bolli, A. Borg, A. L. Borresen-Dale, S. Boyault, B. Burkhardt, A. P. Butler, C. Caldas, H. R. Davies, C. Desmedt, R. Eils, J. E. Eyfjord, J. A. Foekens, M. Greaves, F. Hosoda, B. Hutter, T. Ilicic, S. Imbeaud, M. Imielinski, N. Jager, D. T. Jones, D. Jones, S. Knappskog, M. Kool, S. R. Lakhani, C. Lopez-Otin, S. Martin, N. C. Munshi, H. Nakamura, P. A. Northcott, M. Pajic, E. Papaemmanuil, A. Paradiso, J. V. Pearson, X. S. Puente, K. Raine, M. Ramakrishna, A. L. Richardson, J. Richter, P. Rosenstiel, M. Schlesner, T. N. Schumacher, P. N. Span, J. W. Teague, Y. Totoki, A. N. Tutt, R. Valdes-Mas, M. M. van Buuren, L. van 't Veer, A. Vincent-Salomon, N. Waddell, L. R. Yates, I. Australian Pancreatic Cancer Genome, I. B. C. Consortium, I. M.-S. Consortium, I. PedBrain, J. Zucman-Rossi, P. A. Futreal, U. McDermott, P. Lichter, M. Meyerson, S. M. Grimmond, R. Siebert, E. Campo, T. Shibata, S. M. Pfister, P. J. Campbell and M. R. Stratton (2013). "Signatures of mutational processes in human cancer." Nature **500**(7463): 415-421.

Alonso, A., G. Logroscino, S. S. Jick and M. A. Hernan (2009). "Incidence and lifetime risk of motor neuron disease in the United Kingdom: a population-based study." Eur J Neurol **16**(6): 745-751.

Alzheimer, A. (1907). "Über eine eigenartige Erkrankung der Hirnrinde." Allge Zeitsch Psychiatr und Psychisch-Gerich Med(64): 146-148.

Alzualde, A., F. Moreno, P. Martinez-Lage, I. Ferrer, A. Gorostidi, D. Otaegui, L. Blazquez, B. Atares, S. Cardoso, M. Martinez de Pancorbo, R. Juste, A. B. Rodriguez-Martinez, B. Indakoetxea and A. Lopez de Munain (2010). "Somatic mosaicism in a case of apparently sporadic Creutzfeldt-Jakob disease carrying a

de novo D178N mutation in the PRNP gene." Am J Med Genet B Neuropsychiatr Genet **153B**(7): 1283-1291.

Ameur, A., J. B. Stewart, C. Freyer, E. Hagstrom, M. Ingman, N. G. Larsson and U. Gyllensten (2011). "Ultra-deep sequencing of mouse mitochondrial DNA: mutational patterns and their origins." PLoS Genet **7**(3): e1002028.

An, J. J., Y. P. Lee, S. Y. Kim, S. H. Lee, D. W. Kim, M. J. Lee, M. S. Jeong, S. H. Jang, J. H. Kang, H. Y. Kwon, T. C. Kang, M. H. Won, S. W. Cho, O. S. Kwon, K. S. Lee, J. Park, W. S. Eum and S. Y. Choi (2008). "Transduction of familial amyotrophic lateral sclerosis-related mutant PEP-1-SOD proteins into neuronal cells." Mol Cells **25**(1): 55-63.

Andersen, P. M., L. Forsgren, M. Binzer, P. Nilsson, V. Ala-Hurula, M. L. Keranen, L. Bergmark, A. Saarinen, T. Haltia, I. Tarvainen, E. Kinnunen, B. Udd and S. L. Marklund (1996). "Autosomal recessive adult-onset amyotrophic lateral sclerosis associated with homozygosity for Asp90Ala CuZn-superoxide dismutase mutation. A clinical and genealogical study of 36 patients." Brain **119 (Pt 4)**: 1153-1172.

Andersen, P. M., P. Nilsson, V. Ala-Hurula, M. L. Keranen, I. Tarvainen, T. Haltia, L. Nilsson, M. Binzer, L. Forsgren and S. L. Marklund (1995). "Amyotrophic lateral sclerosis associated with homozygosity for an Asp90Ala mutation in CuZn-superoxide dismutase." Nat Genet **10**(1): 61-66.

Andrews, R. M., I. Kubacka, P. F. Chinnery, R. N. Lightowlers, D. M. Turnbull and N. Howell (1999). "Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA." Nat Genet **23**(2): 147.

Andrews, S. (2010). FastQC: A quality control tool for high throughput sequence data.

Andrews, S. J. and J. A. Rothnagel (2014). "Emerging evidence for functional peptides encoded by short open reading frames." Nat Rev Genet **15**(3): 193-204.

Antunes, M. and G. Biala (2012). "The novel object recognition memory: neurobiology, test procedure, and its modifications." Cogn Process **13**(2): 93-110.

Appel-Cresswell, S., C. Vilarino-Guell, M. Encarnacion, H. Sherman, I. Yu, B. Shah, D. Weir, C. Thompson, C. Szu-Tu, J. Trinh, J. O. Aasly, A. Rajput, A. H. Rajput, A. Jon Stoessl and M. J. Farrer (2013). "Alpha-synuclein p.H50Q, a novel pathogenic mutation for Parkinson's disease." Mov Disord **28**(6): 811-813.

Arendt, T., J. T. Stieler and M. Holzer (2016). "Tau and tauopathies." Brain Res Bull **126**(Pt 3): 238-292.

Ashburner, J. and K. J. Friston (2000). "Voxel-based morphometry--the methods." Neuroimage **11**(6 Pt 1): 805-821.

Association, A. s. (2015). "2015 Alzheimer's disease facts and figures." Alzheimers Dement **11**(3): 332-384.

Attems, J., K. Jellinger, D. R. Thal and W. Van Nostrand (2011). "Review: sporadic cerebral amyloid angiopathy." Neuropathol Appl Neurobiol **37**(1): 75-93.

Aukema, S. M., R. Siebert, E. Schuuring, G. W. van Imhoff, H. C. Kluin-Nelemans, E. J. Boerma and P. M. Kluin (2011). "Double-hit B-cell lymphomas." Blood **117**(8): 2319-2331.

Avila, J., J. J. Lucas, M. Perez and F. Hernandez (2004). "Role of tau protein in both physiological and pathological conditions." Physiol Rev **84**(2): 361-384.

Ayers, J., H. Lelie, A. Workman, M. Prudencio, H. Brown, S. Fromholt, J. Valentine, J. Whitelegge and D. Borchelt (2014). "Distinctive features of the D101N and D101G variants of superoxide dismutase 1; two mutations that produce rapidly progressing motor neuron disease." J Neurochem **128**(2): 305-314.

Bailey, J. A., Z. Gu, R. A. Clark, K. Reinert, R. V. Samonte, S. Schwartz, M. D. Adams, E. W. Myers, P. W. Li and E. E. Eichler (2002). "Recent segmental duplications in the human genome." Science **297**(5583): 1003-1007.

Bailey, J. A., G. Liu and E. E. Eichler (2003). "An Alu transposition model for the origin and expansion of human segmental duplications." Am J Hum Genet **73**(4): 823-834.

Bailey, J. A., A. M. Yavor, H. F. Massa, B. J. Trask and E. E. Eichler (2001). "Segmental duplications: organization and impact within the current human genome project assembly." Genome Res **11**(6): 1005-1017.

Baillie, J. K., M. W. Barnett, K. R. Upton, D. J. Gerhardt, T. A. Richmond, F. De Sapio, P. M. Brennan, P. Rizzu, S. Smith, M. Fell, R. T. Talbot, S. Gustincich, T. C. Freeman, J. S. Mattick, D. A. Hume, P. Heutink, P. Carninci, J. A. Jeddeloh and G. J. Faulkner (2011). "Somatic retrotransposition alters the genetic landscape of the human brain." Nature **479**(7374): 534-537.

Baines, H. L., J. B. Stewart, C. Stamp, A. Zupanic, T. B. Kirkwood, N. G. Larsson, D. M. Turnbull and L. C. Greaves (2014). "Similar patterns of clonally expanded somatic mtDNA mutations in the colon of heterozygous mtDNA mutator mice and ageing humans." Mech Ageing Dev **139**: 22-30.

Baizabal-Carvallo, J. F. and J. Jankovic (2016). "Parkinsonism, movement disorders and genetics in frontotemporal dementia." Nat Rev Neurol **12**(3): 175-185.

Baker, M., I. R. Mackenzie, S. M. Pickering-Brown, J. Gass, R. Rademakers, C. Lindholm, J. Snowden, J. Adamson, A. D. Sadovnick, S. Rollinson, A. Cannon, E. Dwosh, D. Neary, S. Melquist, A. Richardson, D. Dickson, Z. Berger, J. Eriksen, T. Robinson, C. Zehr, C. A. Dickey, R. Crook, E. McGowan, D. Mann, B. Boeve, H. Feldman and M. Hutton (2006). "Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17." Nature **442**(7105): 916-919.

Ballmaier, M., M. Germeshausen, H. Schulze, K. Cherkaoui, S. Lang, A. Gaudig, S. Krukemeier, M. Eilers, G. Strauss and K. Welte (2001). "c-mpl mutations are the cause of congenital amegakaryocytic thrombocytopenia." Blood **97**(1): 139-146.

Bamshad, M. J., S. B. Ng, A. W. Bigham, H. K. Tabor, M. J. Emond, D. A. Nickerson and J. Shendure (2011). "Exome sequencing as a tool for Mendelian disease gene discovery." Nat Rev Genet **12**(11): 745-755.

Bao, R., L. Huang, J. Andrade, W. Tan, W. A. Kibbe, H. Jiang and G. Feng (2014). "Review of current methods, applications, and data management for the bioinformatics analysis of whole exome sequencing." Cancer Inform **13**(Suppl 2): 67-82.

Baxter, E. J., L. M. Scott, P. J. Campbell, C. East, N. Fourouclas, S. Swanton, G. S. Vassiliou, A. J. Bench, E. M. Boyd, N. Curtin, M. A. Scott, W. N. Erber, A. R. Green and P. Cancer Genome (2005). "Acquired mutation of the tyrosine kinase JAK2 in human myeloproliferative disorders." Lancet **365**(9464): 1054-1061.

Beal, M. F. (2005). "Mitochondria take center stage in aging and neurodegeneration." Ann Neurol **58**(4): 495-505.

Beck, J., J. D. Rohrer, T. Campbell, A. Isaacs, K. E. Morrison, E. F. Goodall, E. K. Warrington, J. Stevens, T. Revesz, J. Holton, S. Al-Sarraj, A. King, R. Scahill, J. D. Warren, N. C. Fox, M. N. Rossor, J. Collinge and S. Mead (2008). "A distinct clinical, neuropsychological and radiological phenotype is associated with progranulin gene mutations in a large UK series." Brain **131**(Pt 3): 706-720.

Beck, J. A., M. Poulter, T. A. Campbell, G. Adamson, J. B. Uphill, R. Guerreiro, G. S. Jackson, J. C. Stevens, H. Manji, J. Collinge and S. Mead (2010). "PRNP allelic series from 19 years of prion protein gene sequencing at the MRC Prion Unit." Hum Mutat **31**(7): E1551-1563.

Beck, J. A., M. Poulter, T. A. Campbell, J. B. Uphill, G. Adamson, J. F. Geddes, T. Revesz, M. B. Davis, N. W. Wood, J. Collinge and S. J. Tabrizi (2004). "Somatic and germline mosaicism in sporadic early-onset Alzheimer's disease." Hum Mol Genet **13**(12): 1219-1224.

Bell, C. J., D. L. Dinwiddie, N. A. Miller, S. L. Hateley, E. E. Ganusova, J. Mudge, R. J. Langley, L. Zhang, C. C. Lee, F. D. Schilkey, V. Sheth, J. E. Woodward, H. E. Peckham, G. P. Schroth, R. W. Kim and S. F. Kingsmore (2011). "Carrier testing for severe childhood recessive diseases by next-generation sequencing." Sci Transl Med **3**(65): 65ra64.

Belzil, V. V., H. Daoud, J. St-Onge, A. Desjarlais, J. P. Bouchard, N. Dupre, L. Lacomblez, F. Salachas, P. F. Pradat, V. Meininger, W. Camu, P. A. Dion and G. A. Rouleau (2011). "Identification of novel FUS mutations in sporadic cases of amyotrophic lateral sclerosis." Amyotroph Lateral Scler **12**(2): 113-117.

Bender, A., K. J. Krishnan, C. M. Morris, G. A. Taylor, A. K. Reeve, R. H. Perry, E. Jaros, J. S. Hersheson, J. Betts, T. Klopstock, R. W. Taylor and D. M. Turnbull (2006). "High levels of mitochondrial DNA deletions in substantia nigra neurons in aging and Parkinson disease." Nat Genet **38**(5): 515-517.

Benitez, B. A., D. Alvarado, Y. Cai, K. Mayo, S. Chakraverty, J. Norton, J. C. Morris, M. S. Sands, A. Goate and C. Cruchaga (2011). "Exome-sequencing confirms DNAJC5 mutations as cause of adult neuronal ceroid-lipofuscinosis." PLoS One **6**(11): e26741.

Bennett, E. J., T. A. Shaler, B. Woodman, K. Y. Ryu, T. S. Zaitseva, C. H. Becker, G. P. Bates, H. Schulman and R. R. Kopito (2007). "Global changes to the ubiquitin system in Huntington's disease." Nature **448**(7154): 704-708.

Bernardi, L., C. Tomaino, M. Anfossi, M. Gallo, S. Geracitano, A. Costanzo, R. Colao, G. Puccio, F. Frangipane, S. A. Curcio, M. Mirabelli, N. Smirne, D. Iapaolo, R. G. Maletta and A. C. Bruni (2009). "Novel PSEN1 and PGRN mutations in early-onset familial frontotemporal dementia." Neurobiol Aging **30**(11): 1825-1833.

Bersano, A., M. Ranieri, A. Ciammola, C. Cinnante, S. Lanfranconi, M. T. Dotti, L. Candelise, C. Baschirotto, I. Ghione, E. Ballabio, N. Bresolin and M. T. Bassi (2012). "Considerations on a mutation in the NOTCH3 gene sparing a cysteine residue: a rare polymorphism rather than a CADASIL variant." Funct Neurol **27**(4): 247-252.

Bertoni, J. M., P. Brown, L. G. Goldfarb, R. Rubenstein and D. C. Gajdusek (1992). "Familial Creutzfeldt-Jakob disease (codon 200 mutation) with supranuclear palsy." JAMA **268**(17): 2413-2415.

Bhatia, K., K. Huppi, G. Spangler, D. Siwarski, R. Iyer and I. Magrath (1993). "Point mutations in the c-Myc transactivation domain are common in Burkitt's lymphoma and mouse plasmacytomas." Nat Genet **5**(1): 56-61.

Bird, T. D. (2008). "Genetic aspects of Alzheimer disease." Genet Med **10**(4): 231-239.

Bishop, M. T., C. Pennington, C. A. Heath, R. G. Will and R. S. Knight (2009). "PRNP variation in UK sporadic and variant Creutzfeldt Jakob disease highlights genetic risk factors and a novel non-synonymous polymorphism." BMC Med Genet **10**: 146.

Blauwendraat, C., C. Wilke, J. Simon-Sanchez, I. E. Jansen, A. Reifschneider, A. Capell, C. Haass, M. Castillo-Lizardo, S. Biskup, W. Maetzler, P. Rizzu, P. Heutink and M. Synofzik (2017). "The wide genetic landscape of clinical frontotemporal dementia: systematic combined sequencing of 121 consecutive subjects." Genet Med.

Bogenhagen, D. F. (2012). "Mitochondrial DNA nucleoid structure." Biochim Biophys Acta **1819**(9-10): 914-920.

Borchelt, D. R., T. Ratovitski, J. van Lare, M. K. Lee, V. Gonzales, N. A. Jenkins, N. G. Copeland, D. L. Price and S. S. Sisodia (1997). "Accelerated amyloid deposition in the brains of transgenic mice coexpressing mutant presenilin 1 and amyloid precursor proteins." Neuron **19**(4): 939-945.

Borrow, J., A. D. Goddard, D. Sheer and E. Solomon (1990). "Molecular analysis of acute promyelocytic leukemia breakpoint cluster region on chromosome 17." Science **249**(4976): 1577-1580.

Bossy-Wetzel, E., M. J. Barsoum, A. Godzik, R. Schwarzenbacher and S. A. Lipton (2003). "Mitochondrial fission in apoptosis, neurodegeneration and aging." Curr Opin Cell Biol **15**(6): 706-716.

Botstein, D. and N. Risch (2003). "Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease." Nat Genet **33 Suppl**: 228-237.

Bower, J. H., D. M. Maraganore, S. K. McDonnell and W. A. Rocca (1997). "Incidence of progressive supranuclear palsy and multiple system atrophy in Olmsted County, Minnesota, 1976 to 1990." Neurology **49**(5): 1284-1288.

Braak, E., D. Sandmann-Keil, U. Rub, W. P. Gai, R. A. de Vos, E. N. Steur, K. Arai and H. Braak (2001). "alpha-synuclein immunopositive Parkinson's disease-related inclusion bodies in lower brain stem nuclei." Acta Neuropathol **101**(3): 195-201.

Braak, H., I. Alafuzoff, T. Arzberger, H. Kretzschmar and K. Del Tredici (2006). "Staging of Alzheimer disease-associated neurofibrillary pathology using paraffin sections and immunocytochemistry." Acta Neuropathol **112**(4): 389-404.

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

Braak, H., K. Del Tredici, U. Rub, R. A. de Vos, E. N. Jansen Steur and E. Braak (2003). "Staging of brain pathology related to sporadic Parkinson's disease." Neurobiol Aging **24**(2): 197-211.

Bras, J., J. Simon-Sanchez, M. Federoff, A. Morgadinho, C. Januario, M. Ribeiro, L. Cunha, C. Oliveira and A. B. Singleton (2009). "Lack of replication of association between GIGYF2 variants and Parkinson disease." Hum Mol Genet **18**(2): 341-346.

Brooks, B. R. (1994). "El Escorial World Federation of Neurology criteria for the diagnosis of amyotrophic lateral sclerosis. Subcommittee on Motor Neuron Diseases/Amyotrophic Lateral Sclerosis of the World Federation of Neurology Research Group on Neuromuscular Diseases and the El Escorial "Clinical limits of amyotrophic lateral sclerosis" workshop contributors." J Neurol Sci **124 Suppl**: 96-107.

Brouwers, N., K. Sleegers, S. Engelborghs, S. Maurer-Stroh, I. Gijselinck, J. van der Zee, B. A. Pickut, M. Van den Broeck, M. Mattheijssens, K. Peeters, J. Schymkowitz, F. Rousseau, J. J. Martin, M. Cruts, P. P. De Deyn and C. Van

Broeckhoven (2008). "Genetic variability in progranulin contributes to risk for clinically diagnosed Alzheimer disease." Neurology **71**(9): 656-664.

Brown, P., M. Preece, J. P. Brandel, T. Sato, L. McShane, I. Zerr, A. Fletcher, R. G. Will, M. Pocchiari, N. R. Cashman, J. H. d'Aignaux, L. Cervenakova, J. Fradkin, L. B. Schonberger and S. J. Collins (2000). "Iatrogenic Creutzfeldt-Jakob disease at the millennium." Neurology **55**(8): 1075-1081.

Bruce, M. E., R. G. Will, J. W. Ironside, I. McConnell, D. Drummond, A. Suttie, L. McCardle, A. Chree, J. Hope, C. Birkett, S. Cousens, H. Fraser and C. J. Bostock (1997). "Transmissions to mice indicate that 'new variant' CJD is caused by the BSE agent." Nature **389**(6650): 498-501.

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

Budka, H., A. Aguzzi, P. Brown, J. M. Brucher, O. Bugiani, F. Gullotta, M. Haltia, J. J. Hauw, J. W. Ironside, K. Jellinger and et al. (1995). "Neuropathological diagnostic criteria for Creutzfeldt-Jakob disease (CJD) and other human spongiform encephalopathies (prion diseases)." Brain Pathol **5**(4): 459-466.

Bunton-Stasyshyn, R. K., R. A. Saccon, P. Fratta and E. M. Fisher (2015). "SOD1 Function and Its Implications for Amyotrophic Lateral Sclerosis Pathology: New and Renascent Themes." Neuroscientist **21**(5): 519-529.

Burrell, J. R., M. C. Kiernan, S. Vucic and J. R. Hodges (2011). "Motor neuron dysfunction in frontotemporal dementia." Brain **134**(Pt 9): 2582-2594.

Busque, L., J. P. Patel, M. E. Figueroa, A. Vasanthakumar, S. Provost, Z. Hamilou, L. Mollica, J. Li, A. Viale, A. Heguy, M. Hassimi, N. Socci, P. K. Bhatt, M. Gonen, C. E. Mason, A. Melnick, L. A. Godley, C. W. Brennan, O. Abdel-Wahab and R. L. Levine (2012). "Recurrent somatic TET2 mutations in normal elderly individuals with clonal hematopoiesis." Nat Genet **44**(11): 1179-1181.

Bystrom, R., P. M. Andersen, G. Grobner and M. Oliveberg (2010). "SOD1 mutations targeting surface hydrogen bonds promote amyotrophic lateral sclerosis without reducing apo-state stability." J Biol Chem **285**(25): 19544-19552.

Cabin, D. E., S. Gispert-Sanchez, D. Murphy, G. Auburger, R. R. Myers and R. L. Nussbaum (2005). "Exacerbated synucleinopathy in mice expressing A53T SNCA on a Snca null background." Neurobiol Aging **26**(1): 25-35.

Cadieux-Dion, M., E. Andermann, P. Lachance-Touchette, O. Ansorge, C. Meloche, A. Barnabe, R. I. Kuzniecky, F. Andermann, E. Faught, S. Leonberg, J. A. Damiano, S. F. Berkovic, G. A. Rouleau and P. Cossette (2013). "Recurrent mutations in DNAJC5 cause autosomal dominant Kufs disease." Clin Genet **83**(6): 571-575.

Cady, J., P. Allred, T. Bali, A. Pestronk, A. Goate, T. M. Miller, R. D. Mitra, J. Ravits, M. B. Harms and R. H. Baloh (2015). "Amyotrophic lateral sclerosis onset is influenced by the burden of rare variants in known amyotrophic lateral sclerosis genes." Ann Neurol **77**(1): 100-113.

Cai, X., G. D. Evrony, H. S. Lehmann, P. C. Elhosary, B. K. Mehta, A. Poduri and C. A. Walsh (2015). "Single-cell, genome-wide sequencing identifies clonal somatic copy-number variation in the human brain." Cell Rep **10**(4): 645.

Caldecott, K. W. (2008). "Single-strand break repair and genetic disease." Nat Rev Genet **9**(8): 619-631.

Callinan, P. A. and M. A. Batzer (2006). "Retrotransposable elements and human disease." Genome Dyn **1**: 104-115.

Canet-Aviles, R. M., M. A. Wilson, D. W. Miller, R. Ahmad, C. McLendon, S. Bandyopadhyay, M. J. Baptista, D. Ringe, G. A. Petsko and M. R. Cookson (2004). "The Parkinson's disease protein DJ-1 is neuroprotective due to cysteine-sulfinic acid-driven mitochondrial localization." Proc Natl Acad Sci U S A **101**(24): 9103-9108.

Canter, R. G., J. Penney and L. H. Tsai (2016). "The road to restoring neural circuits for the treatment of Alzheimer's disease." Nature **539**(7628): 187-196.

Cao, M., Y. Wu, G. Ashrafi, A. J. McCartney, H. Wheeler, E. A. Bushong, D. Boassa, M. H. Ellisman, T. A. Ryan and P. De Camilli (2017). "Parkinson Sac Domain Mutation in Synaptojanin 1 Impairs Clathrin Uncoating at Synapses and

Triggers Dystrophic Changes in Dopaminergic Axons." Neuron **93**(4): 882-896 e885.

Carecchio, M., C. Fenoglio, M. De Riz, I. Guidi, C. Comi, F. Cortini, E. Venturelli, I. Restelli, C. Cantoni, N. Bresolin, F. Monaco, E. Scarpini and D. Galimberti (2009). "Progranulin plasma levels as potential biomarker for the identification of GRN deletion carriers. A case with atypical onset as clinical amnestic Mild Cognitive Impairment converted to Alzheimer's disease." J Neurol Sci **287**(1-2): 291-293.

Caslake, R., J. N. Moore, J. C. Gordon, C. E. Harris and C. Counsell (2008). "Changes in diagnosis with follow-up in an incident cohort of patients with parkinsonism." J Neurol Neurosurg Psychiatry **79**(11): 1202-1207.

Celsi, F., P. Pizzo, M. Brini, S. Leo, C. Fotino, P. Pinton and R. Rizzuto (2009). "Mitochondria, calcium and cell death: a deadly triad in neurodegeneration." Biochim Biophys Acta **1787**(5): 335-344.

Chapman, J., A. Arlazoroff, L. G. Goldfarb, L. Cervenakova, M. Y. Neufeld, E. Werber, M. Herbert, P. Brown, D. C. Gajdusek and A. D. Korczyn (1996). "Fatal insomnia in a case of familial Creutzfeldt-Jakob disease with the codon 200(Lys) mutation." Neurology **46**(3): 758-761.

Chare, L., J. R. Hodges, C. E. Leyton, C. McGinley, R. H. Tan, J. J. Kril and G. M. Halliday (2014). "New criteria for frontotemporal dementia syndromes: clinical and pathological diagnostic implications." J Neurol Neurosurg Psychiatry **85**(8): 865-870.

Chartier-Harlin, M. C., J. C. Dachsel, C. Vilarino-Guell, S. J. Lincoln, F. Lepretre, M. M. Hulihan, J. Kachergus, A. J. Milnerwood, L. Tapia, M. S. Song, E. Le Rhun, E. Mutez, L. Larvor, A. Duflot, C. Vanbesien-Mailliot, A. Kreisler, O. A. Ross, K. Nishioka, A. I. Soto-Ortolaza, S. A. Cobb, H. L. Melrose, B. Behrouz, B. H. Keeling, J. A. Bacon, E. Hentati, L. Williams, A. Yanagiya, N. Sonenberg, P. J. Lockhart, A. C. Zubair, R. J. Uitti, J. O. Aasly, A. Krygowska-Wajs, G. Opala, Z. K. Wszolek, R. Frigerio, D. M. Maraganore, D. Gosal, T. Lynch, M. Hutchinson, A. R. Bentivoglio, E. M. Valente, W. C. Nichols, N. Pankratz, T. Foroud, R. A. Gibson, F. Hentati, D. W. Dickson, A. Destee and M. J. Farrer (2011). "Translation initiator

EIF4G1 mutations in familial Parkinson disease." Am J Hum Genet **89**(3): 398-406.

Chaudhuri, K. R., D. G. Healy, A. H. Schapira and E. National Institute for Clinical (2006). "Non-motor symptoms of Parkinson's disease: diagnosis and management." Lancet Neurol **5**(3): 235-245.

Chaussenot, A., I. Le Ber, S. Ait-El-Mkadem, A. Camuzat, A. de Septenville, S. Bannwarth, E. C. Genin, V. Serre, G. Auge, F. T. D. French research network on, A. L. S. Ftd, A. Brice, J. Pouget and V. Paquis-Flucklinger (2014). "Screening of CHCHD10 in a French cohort confirms the involvement of this gene in frontotemporal dementia with amyotrophic lateral sclerosis patients." Neurobiol Aging **35**(12): 2884 e2881-2884.

Chen, J., W. Li, T. Zhang, Y. J. Wang, X. J. Jiang and Z. Q. Xu (2014). "Glucocerebrosidase gene mutations associated with Parkinson's disease: a meta-analysis in a Chinese population." PLoS One **9**(12): e115747.

Chen, S., P. Sayana, X. Zhang and W. Le (2013). "Genetics of amyotrophic lateral sclerosis: an update." Mol Neurodegener **8**: 28.

Chen, W. T., Y. F. Hsieh, Y. J. Huang, C. C. Lin, Y. T. Lin, Y. C. Liu, C. C. Lien and I. H. Cheng (2014). "G206D Mutation of Presenilin-1 Reduces Pen2 Interaction, Increases Abeta42/Abeta40 Ratio and Elevates ER Ca Accumulation." Mol Neurobiol.

Chen-Plotkin, A. S., M. Martinez-Lage, P. M. Sleiman, W. Hu, R. Greene, E. M. Wood, S. Bing, M. Grossman, G. D. Schellenberg, K. J. Hatanpaa, M. F. Weiner, C. L. White, 3rd, W. S. Brooks, G. M. Halliday, J. J. Kril, M. Gearing, T. G. Beach, N. R. Graff-Radford, D. W. Dickson, R. Rademakers, B. F. Boeve, S. M. Pickering-Brown, J. Snowden, J. C. van Swieten, P. Heutink, H. Seelaar, J. R. Murrell, B. Ghetti, S. Spina, J. Grafman, J. A. Kaye, R. L. Woltjer, M. Mesulam, E. Bigio, A. Llado, B. L. Miller, A. Alzualde, F. Moreno, J. D. Rohrer, I. R. Mackenzie, H. H. Feldman, R. L. Hamilton, M. Cruts, S. Engelborghs, P. P. De Deyn, C. Van Broeckhoven, T. D. Bird, N. J. Cairns, A. Goate, M. P. Frosch, P. F. Riederer, N. Bogdanovic, V. M. Lee, J. Q. Trojanowski and V. M. Van Deerlin (2011). "Genetic

and clinical features of progranulin-associated frontotemporal lobar degeneration." Arch Neurol **68**(4): 488-497.

Chesselet, M. F. and F. Richter (2011). "Modelling of Parkinson's disease in mice." Lancet Neurol **10**(12): 1108-1118.

Chio, A., A. Calvo, L. Mazzini, R. Cantello, G. Mora, C. Moglia, L. Corrado, S. D'Alfonso, E. Majounie, A. Renton, F. Pisano, I. Ossola, M. Brunetti, B. J. Traynor, G. Restagno and Parals (2012). "Extensive genetics of ALS: a population-based study in Italy." Neurology **79**(19): 1983-1989.

Chio, A., B. J. Traynor, F. Lombardo, M. Fimognari, A. Calvo, P. Ghiglione, R. Mutani and G. Restagno (2008). "Prevalence of SOD1 mutations in the Italian ALS population." Neurology **70**(7): 533-537.

Chun, S. and J. C. Fay (2009). "Identification of deleterious mutations within three human genomes." Genome Res **19**(9): 1553-1561.

Cibulskis, K., M. S. Lawrence, S. L. Carter, A. Sivachenko, D. Jaffe, C. Sougnez, S. Gabriel, M. Meyerson, E. S. Lander and G. Getz (2013). "Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples." Nat Biotechnol **31**(3): 213-219.

Cimino, G., D. T. Moir, O. Canaani, K. Williams, W. M. Crist, S. Katzav, L. Cannizzaro, B. Lange, P. C. Nowell, C. M. Croce and et al. (1991). "Cloning of ALL-1, the locus involved in leukemias with the t(4;11)(q21;q23), t(9;11)(p22;q23), and t(11;19)(q23;p13) chromosome translocations." Cancer Res **51**(24): 6712-6714.

Clark, J., E. L. Clore, K. Zheng, A. Adame, E. Masliah and D. K. Simon (2010). "Oral N-acetyl-cysteine attenuates loss of dopaminergic terminals in alpha-synuclein overexpressing mice." PLoS One **5**(8): e12333.

Cleveland, D. W. and J. D. Rothstein (2001). "From Charcot to Lou Gehrig: deciphering selective motor neuron death in ALS." Nat Rev Neurosci **2**(11): 806-819.

Coe, B. P., K. Witherspoon, J. A. Rosenfeld, B. W. van Bon, A. T. Vulto-van Silfhout, P. Bosco, K. L. Friend, C. Baker, S. Buono, L. E. Vissers, J. H. Schuurs-Hoeijmakers, A. Hoischen, R. Pfundt, N. Krumm, G. L. Carvill, D. Li, D. Amaral, N. Brown, P. J. Lockhart, I. E. Scheffer, A. Alberti, M. Shaw, R. Pettinato, R. Tervo, N. de Leeuw, M. R. Reijnders, B. S. Torchia, H. Peeters, B. J. O'Roak, M. Fichera, J. Y. Hehir-Kwa, J. Shendure, H. C. Mefford, E. Haan, J. Gecz, B. B. de Vries, C. Romano and E. E. Eichler (2014). "Refining analyses of copy number variation identifies specific genes associated with developmental delay." Nat Genet **46**(10): 1063-1071.

Colella, S., C. Yau, J. M. Taylor, G. Mirza, H. Butler, P. Clouston, A. S. Bassett, A. Seller, C. C. Holmes and J. Ragoussis (2007). "QuantiSNP: an Objective Bayes Hidden-Markov Model to detect and accurately map copy number variation using SNP genotyping data." Nucleic Acids Res **35**(6): 2013-2025.

Collinge, J. and A. R. Clarke (2007). "A general model of prion strains and their pathogenicity." Science **318**(5852): 930-936.

Collinge, J., M. S. Palmer, T. Campbell, K. C. Sidle, D. Carroll and A. Harding (1993). "Inherited prion disease (PrP lysine 200) in Britain: two case reports." BMJ **306**(6873): 301-302.

Collins, S. J., P. Sanchez-Juan, C. L. Masters, G. M. Klug, C. van Duijn, A. Poleggi, M. Pocchiari, S. Almonti, N. Cuadrado-Corrales, J. de Pedro-Cuesta, H. Budka, E. Gelpi, M. Glatzel, M. Tolnay, E. Hewer, I. Zerr, U. Heinemann, H. A. Kretszchmar, G. H. Jansen, E. Olsen, E. Mitrova, A. Alperovitch, J. P. Brandel, J. Mackenzie, K. Murray and R. G. Will (2006). "Determinants of diagnostic investigation sensitivities across the clinical spectrum of sporadic Creutzfeldt-Jakob disease." Brain **129**(Pt 9): 2278-2287.

Comas-Herrera, A., R. Wittenberg, L. Pickard and M. Knapp (2007). "Cognitive impairment in older people: future demand for long-term care services and the associated costs." Int J Geriatr Psychiatry **22**(10): 1037-1045.

Conte, A., S. Lattante, M. Zollino, G. Marangi, M. Luigetti, A. Del Grande, S. Servidei, F. Trombetta and M. Sabatelli (2012). "P525L FUS mutation is

consistently associated with a severe form of juvenile amyotrophic lateral sclerosis." Neuromuscul Disord **22**(1): 73-75.

Cook, E. H., Jr. and S. W. Scherer (2008). "Copy-number variations associated with neuropsychiatric conditions." Nature **455**(7215): 919-923.

Cookson, M. R., J. Hardy and P. A. Lewis (2008). "Genetic neuropathology of Parkinson's disease." Int J Clin Exp Pathol **1**(3): 217-231.

Cookson, M. R., P. J. Lockhart, C. McLendon, C. O'Farrell, M. Schlossmacher and M. J. Farrer (2003). "RING finger 1 mutations in Parkin produce altered localization of the protein." Hum Mol Genet **12**(22): 2957-2965.

Cooper, N. J., C. J. Shtir, D. J. Smyth, H. Guo, A. D. Swafford, M. Zanda, M. E. Hurles, N. M. Walker, V. Plagnol, J. D. Cooper, J. M. Howson, O. S. Burren, S. Onengut-Gumuscu, S. S. Rich and J. A. Todd (2015). "Detection and correction of artefacts in estimation of rare copy number variants and analysis of rare deletions in type 1 diabetes." Hum Mol Genet **24**(6): 1774-1790.

Copois, V., F. Bibeau, C. Bascoul-Mollevi, N. Salvetat, P. Chalbos, C. Bareil, L. Candeil, C. Fraslon, E. Conseiller, V. Granci, P. Maziere, A. Kramar, M. Ychou, B. Pau, P. Martineau, F. Molina and M. Del Rio (2007). "Impact of RNA degradation on gene expression profiles: assessment of different methods to reliably determine RNA quality." J Biotechnol **127**(4): 549-559.

Cordaux, R. and M. A. Batzer (2009). "The impact of retrotransposons on human genome evolution." Nat Rev Genet **10**(10): 691-703.

Cornejo-Olivas, M. R., C. E. Yu, P. Mazzetti, I. F. Mata, M. Meza, S. Lindo-Samanamud, J. B. Leverenz and T. D. Bird (2014). "Clinical and molecular studies reveal a PSEN1 mutation (L153V) in a Peruvian family with early-onset Alzheimer's disease." Neurosci Lett **563**: 140-143.

Corradin, O. and P. C. Scacheri (2014). "Enhancer variants: evaluating functions in common disease." Genome Med **6**(10): 85.

Couthouis, J., A. R. Raphael, R. Daneshjou and A. D. Gitler (2014). "Targeted exon capture and sequencing in sporadic amyotrophic lateral sclerosis." PLoS Genet **10**(10): e1004704.

Cox, L. E., L. Ferraiuolo, E. F. Goodall, P. R. Heath, A. Higginbottom, H. Mortiboys, H. C. Hollinger, J. A. Hartley, A. Brockington, C. E. Burness, K. E. Morrison, S. B. Wharton, A. J. Grierson, P. G. Ince, J. Kirby and P. J. Shaw (2010). "Mutations in CHMP2B in lower motor neuron predominant amyotrophic lateral sclerosis (ALS)." PLoS One **5**(3): e9872.

Coxhead, J., M. Kurzawa-Akanbi, R. Hussain, A. Pyle, P. Chinnery and G. Hudson (2016). "Somatic mtDNA variation is an important component of Parkinson's disease." Neurobiol Aging **38**: 217 e211-217 e216.

Crabtree, B., N. Thiyagarajan, S. H. Prior, P. Wilson, S. Iyer, T. Ferns, R. Shapiro, K. Brew, V. Subramanian and K. R. Acharya (2007). "Characterization of human angiogenin variants implicated in amyotrophic lateral sclerosis." Biochemistry **46**(42): 11810-11818.

Cruchaga, C., G. Haller, S. Chakraverty, K. Mayo, F. L. Vallania, R. D. Mitra, K. Faber, J. Williamson, T. Bird, R. Diaz-Arrastia, T. M. Foroud, B. F. Boeve, N. R. Graff-Radford, P. St Jean, M. Lawson, M. G. Ehm, R. Mayeux, A. M. Goate and N.-L. N. F. S. Consortium (2012). "Rare variants in APP, PSEN1 and PSEN2 increase risk for AD in late-onset Alzheimer's disease families." PLoS One **7**(2): e31039.

Cruts, M., J. Theuns and C. Van Broeckhoven (2012). "Locus-specific mutation databases for neurodegenerative brain diseases." Hum Mutat **33**(9): 1340-1344.

Cruts, M. and C. Van Broeckhoven (1998). "Molecular genetics of Alzheimer's disease." Ann Med **30**(6): 560-565.

Cuyvers, E., K. Bettens, S. Philtjens, T. Van Langenhove, I. Gijselinck, J. van der Zee, S. Engelborghs, M. Vandenbulcke, J. Van Dongen, N. Geerts, G. Maes, M. Mattheijssens, K. Peeters, P. Cras, R. Vandenberghe, P. P. De Deyn, C. Van Broeckhoven, M. Cruts, K. Sleegers and B. consortium (2014). "Investigating the

role of rare heterozygous TREM2 variants in Alzheimer's disease and frontotemporal dementia." Neurobiol Aging **35**(3): 726 e711-729.

de Pedro-Cuesta, J., A. Rabano, P. Martinez-Martin, M. Ruiz-Tovar, E. Alcalde-Cabero, J. Almazan-Isla, F. Avellanal and M. Calero (2015). "Correction: Comparative Incidence of Conformational, Neurodegenerative Disorders." PLoS One **10**(10): e0140304.

De Strooper, B., P. Saftig, K. Craessaerts, H. Vanderstichele, G. Guhde, W. Annaert, K. Von Figura and F. Van Leuven (1998). "Deficiency of presenilin-1 inhibits the normal cleavage of amyloid precursor protein." Nature **391**(6665): 387-390.

Deng, H., M. W. Dodson, H. Huang and M. Guo (2008). "The Parkinson's disease genes pink1 and parkin promote mitochondrial fission and/or inhibit fusion in Drosophila." Proc Natl Acad Sci U S A **105**(38): 14503-14508.

Deng, H., K. Gao and J. Jankovic (2014). "The role of FUS gene variants in neurodegenerative diseases." Nat Rev Neurol **10**(6): 337-348.

Deoni, S. C. (2010). "Quantitative relaxometry of the brain." Top Magn Reson Imaging **21**(2): 101-113.

Desai, B. S., A. J. Monahan, P. M. Carvey and B. Hendey (2007). "Blood-brain barrier pathology in Alzheimer's and Parkinson's disease: implications for drug therapy." Cell Transplant **16**(3): 285-299.

Dichgans, M., H. Ludwig, J. Muller-Hocker, A. Messerschmidt and T. Gasser (2000). "Small in-frame deletions and missense mutations in CADASIL: 3D models predict misfolding of Notch3 EGF-like repeat domains." Eur J Hum Genet **8**(4): 280-285.

Dimas, A. S., S. Deutsch, B. E. Stranger, S. B. Montgomery, C. Borel, H. Attar-Cohen, C. Ingle, C. Beazley, M. Gutierrez Arcelus, M. Sekowska, M. Gagnebin, J. Nisbett, P. Deloukas, E. T. Dermitzakis and S. E. Antonarakis (2009). "Common regulatory variation impacts gene expression in a cell type-dependent manner." Science **325**(5945): 1246-1250.

Diomedi-Camassei, F., S. Di Giandomenico, F. M. Santorelli, G. Caridi, F. Piemonte, G. Montini, G. M. Ghiggeri, L. Murer, L. Barisoni, A. Pastore, A. O. Muda, M. L. Valente, E. Bertini and F. Emma (2007). "COQ2 nephropathy: a newly described inherited mitochondriopathy with primary renal involvement." J Am Soc Nephrol **18**(10): 2773-2780.

Diskin, S. J., M. Li, C. Hou, S. Yang, J. Glessner, H. Hakonarson, M. Bucan, J. M. Maris and K. Wang (2008). "Adjustment of genomic waves in signal intensities from whole-genome SNP genotyping platforms." Nucleic Acids Res **36**(19): e126.

Doetschman, T. (2009). "Influence of genetic background on genetically engineered mouse phenotypes." Methods Mol Biol **530**: 423-433.

Dubois, B., H. H. Feldman, C. Jacova, H. Hampel, J. L. Molinuevo, K. Blennow, S. T. DeKosky, S. Gauthier, D. Selkoe, R. Bateman, S. Cappa, S. Crutch, S. Engelborghs, G. B. Frisoni, N. C. Fox, D. Galasko, M. O. Habert, G. A. Jicha, A. Nordberg, F. Pasquier, G. Rabinovici, P. Robert, C. Rowe, S. Salloway, M. Sarazin, S. Epelbaum, L. C. de Souza, B. Vellas, P. J. Visser, L. Schneider, Y. Stern, P. Scheltens and J. L. Cummings (2014). "Advancing research diagnostic criteria for Alzheimer's disease: the IWG-2 criteria." Lancet Neurol **13**(6): 614-629.

Duffy, P., J. Wolf, G. Collins, A. G. DeVoe, B. Streeten and D. Cowen (1974). "Letter: Possible person-to-person transmission of Creutzfeldt-Jakob disease." N Engl J Med **290**(12): 692-693.

Duran, R., N. E. Mencacci, A. V. Angeli, M. Shoai, E. Deas, H. Houlden, A. Mehta, D. Hughes, T. M. Cox, P. Deegan, A. H. Schapira, A. J. Lees, P. Limousin, P. R. Jarman, K. P. Bhatia, N. W. Wood, J. Hardy and T. Foltynie (2013). "The glucocerobrosidase E326K variant predisposes to Parkinson's disease, but does not cause Gaucher's disease." Mov Disord **28**(2): 232-236.

Elson, J. L., D. C. Samuels, D. M. Turnbull and P. F. Chinnery (2001). "Random intracellular drift explains the clonal expansion of mitochondrial DNA mutations with age." Am J Hum Genet **68**(3): 802-806.

Ennaceur, A. (2010). "One-trial object recognition in rats and mice: methodological and theoretical issues." Behav Brain Res **215**(2): 244-254.

Exome_Variant_Server. (2016). "NHLBI GO Exome Sequencing Project (ESP)." Retrieved 1/2/16, 2016, from *http://evs.gs.washington.edu/EVS/*.

Falvo, M. J. and G. M. Earhart (2009). "Six-minute walk distance in persons with Parkinson disease: a hierarchical regression model." Arch Phys Med Rehabil **90**(6): 1004-1008.

Fan, H. C., L. I. Ho, C. S. Chi, S. J. Chen, G. S. Peng, T. M. Chan, S. Z. Lin and H. J. Harn (2014). "Polyglutamine (PolyQ) diseases: genetics to treatments." Cell Transplant **23**(4-5): 441-458.

Fares, M. B., B. Maco, A. Oueslati, E. Rockenstein, N. Ninkina, V. L. Buchman, E. Masliah and H. A. Lashuel (2016). "Induction of de novo alpha-synuclein fibrillization in a neuronal model for Parkinson's disease." Proc Natl Acad Sci U S A **113**(7): E912-921.

Farrer, M., J. Kachergus, L. Forno, S. Lincoln, D. S. Wang, M. Hulihan, D. Maraganore, K. Gwinn-Hardy, Z. Wszolek, D. Dickson and J. W. Langston (2004). "Comparison of kindreds with parkinsonism and alpha-synuclein genomic multiplications." Ann Neurol **55**(2): 174-179.

Fecto, F., J. Yan, S. P. Vemula, E. Liu, Y. Yang, W. Chen, J. G. Zheng, Y. Shi, N. Siddique, H. Arrat, S. Donkervoort, S. Ajroud-Driss, R. L. Sufit, S. L. Heller, H. X. Deng and T. Siddique (2011). "SQSTM1 mutations in familial and sporadic amyotrophic lateral sclerosis." Arch Neurol **68**(11): 1440-1446.

Fernandez-Santiago, R., S. Hoenig, P. Lichtner, A. D. Sperfeld, M. Sharma, D. Berg, O. Weichenrieder, T. Illig, K. Eger, T. Meyer, J. Anneser, C. Munch, S. Zierz, T. Gasser and A. Ludolph (2009). "Identification of novel Angiogenin (ANG) gene missense variants in German patients with amyotrophic lateral sclerosis." J Neurol **256**(8): 1337-1342.

Ferrari, R., D. Kapogiannis, E. D. Huey and P. Momeni (2011). "FTD and ALS: a tale of two diseases." Curr Alzheimer Res **8**(3): 273-294.

Ferri, A., M. Cozzolino, C. Crosio, M. Nencini, A. Casciati, E. B. Gralla, G. Rotilio, J. S. Valentine and M. T. Carri (2006). "Familial ALS-superoxide dismutases associate with mitochondria and shift their redox potentials." Proc Natl Acad Sci U S A **103**(37): 13860-13865.

Feuk, L., A. R. Carson and S. W. Scherer (2006). "Structural variation in the human genome." Nat Rev Genet **7**(2): 85-97.

Finch, N., M. Baker, R. Crook, K. Swanson, K. Kuntz, R. Surtees, G. Bisceglio, A. Rovelet-Lecrux, B. Boeve, R. C. Petersen, D. W. Dickson, S. G. Younkin, V. Deramecourt, J. Crook, N. R. Graff-Radford and R. Rademakers (2009). "Plasma progranulin levels predict progranulin mutation status in frontotemporal dementia patients and asymptomatic family members." Brain **132**(Pt 3): 583-591.

Finckh, U., T. Muller-Thomsen, U. Mann, C. Eggers, J. Marksteiner, W. Meins, G. Binetti, A. Alberici, C. Hock, R. M. Nitsch and A. Gal (2000). "High prevalence of pathogenic mutations in patients with early-onset dementia detected by sequence analyses of four different genes." Am J Hum Genet **66**(1): 110-117.

Forbes, N., S. Goodwin, K. Woodward, D. G. Morgan, L. Brady, M. B. Coulthart and M. A. Tarnopolsky (2014). "Evidence for synergistic effects of PRNP and ATP7B mutations in severe neuropsychiatric deterioration." BMC Med Genet **15**: 22.

Fratta, P., J. Charnock, T. Collins, A. Devoy, R. Howard, A. Malaspina, R. Orrell, K. Sidle, J. Clarke, M. Shoai, C. H. Lu, J. Hardy, V. Plagnol and E. M. Fisher (2014). "Profilin1 E117G is a moderate risk factor for amyotrophic lateral sclerosis." J Neurol Neurosurg Psychiatry **85**(5): 506-508.

Freichel, C., M. Neumann, T. Ballard, V. Muller, M. Woolley, L. Ozmen, E. Borroni, H. A. Kretzschmar, C. Haass, W. Spooren and P. J. Kahle (2007). "Age-dependent cognitive decline and amygdala pathology in alpha-synuclein transgenic mice." Neurobiol Aging **28**(9): 1421-1435.

Friedman-Levi, Y., Z. Meiner, T. Canello, K. Frid, G. G. Kovacs, H. Budka, D. Avrahami and R. Gabizon (2011). "Fatal prion disease in a mouse model of genetic E200K Creutzfeldt-Jakob disease." PLoS Pathog **7**(11): e1002350.

Frost, B., J. Ollesch, H. Wille and M. I. Diamond (2009). "Conformational diversity of wild-type Tau fibrils specified by templated conformation change." J Biol Chem **284**(6): 3546-3551.

Fujisawa, T., K. Homma, N. Yamaguchi, H. Kadowaki, N. Tsuburaya, I. Naguro, A. Matsuzawa, K. Takeda, Y. Takahashi, J. Goto, S. Tsuji, H. Nishitoh and H. Ichijo (2012). "A novel monoclonal antibody reveals a conformational alteration shared by amyotrophic lateral sclerosis-linked SOD1 mutants." Ann Neurol **72**(5): 739-749.

Fuke, S., M. Kametani, K. Yamada, T. Kasahara, M. Kubota-Sakashita, G. C. Kujoth, T. A. Prolla, S. Hitoshi and T. Kato (2014). "Heterozygous Polg mutation causes motor dysfunction due to mtDNA deletions." Ann Clin Transl Neurol **1**(11): 909-920.

Fuster, J. J., S. MacLauchlan, M. A. Zuriaga, M. N. Polackal, A. C. Ostriker, R. Chakraborty, C. L. Wu, S. Sano, S. Muralidharan, C. Rius, J. Vuong, S. Jacob, V. Muralidhar, A. A. Robertson, M. A. Cooper, V. Andres, K. K. Hirschi, K. A. Martin and K. Walsh (2017). "Clonal hematopoiesis associated with Tet2 deficiency accelerates atherosclerosis development in mice." Science.

Gao, X., K. C. Simon, J. Han, M. A. Schwarzschild and A. Ascherio (2009). "Genetic determinants of hair color and Parkinson's disease risk." Ann Neurol **65**(1): 76-82.

Garrison, E. M., G. (2012). "Haplotype-based variant detection from shortread

sequencing." arXiv: 1207.3907v1202.

Gass, J., A. Cannon, I. R. Mackenzie, B. Boeve, M. Baker, J. Adamson, R. Crook, S. Melquist, K. Kuntz, R. Petersen, K. Josephs, S. M. Pickering-Brown, N. Graff-Radford, R. Uitti, D. Dickson, Z. Wszolek, J. Gonzalez, T. G. Beach, E. Bigio, N. Johnson, S. Weintraub, M. Mesulam, C. L. White, 3rd, B. Woodruff, R. Caselli, G.

Y. Hsiung, H. Feldman, D. Knopman, M. Hutton and R. Rademakers (2006). "Mutations in progranulin are a major cause of ubiquitin-positive frontotemporal lobar degeneration." Hum Mol Genet **15**(20): 2988-3001.

Gasser, T. (2015). "Usefulness of Genetic Testing in PD and PD Trials: A Balanced Review." J Parkinsons Dis **5**(2): 209-215.

Gatchel, J. R. and H. Y. Zoghbi (2005). "Diseases of unstable repeat expansion: mechanisms and common principles." Nat Rev Genet **6**(10): 743-755.

Geisler, S., K. M. Holmstrom, A. Treis, D. Skujat, S. S. Weber, F. C. Fiesel, P. J. Kahle and W. Springer (2010). "The PINK1/Parkin-mediated mitophagy is compromised by PD-associated mutations." Autophagy **6**(7): 871-878.

Geller, L. N. and H. Potter (1999). "Chromosome missegregation and trisomy 21 mosaicism in Alzheimer's disease." Neurobiol Dis **6**(3): 167-179.

Gellera, C., C. Colombrita, N. Ticozzi, B. Castellotti, C. Bragato, A. Ratti, F. Taroni and V. Silani (2008). "Identification of new ANG gene mutations in a large cohort of Italian patients with amyotrophic lateral sclerosis." Neurogenetics **9**(1): 33-40.

Genomes Project, C., G. R. Abecasis, A. Auton, L. D. Brooks, M. A. DePristo, R. M. Durbin, R. E. Handsaker, H. M. Kang, G. T. Marth and G. A. McVean (2012). "An integrated map of genetic variation from 1,092 human genomes." Nature **491**(7422): 56-65.

Genovese, G., R. E. Handsaker, H. Li, E. E. Kenny and S. A. McCarroll (2013). "Mapping the human reference genome's missing sequence by three-way admixture in Latino genomes." Am J Hum Genet **93**(3): 411-421.

Genovese, G., A. K. Kahler, R. E. Handsaker, J. Lindberg, S. A. Rose, S. F. Bakhoum, K. Chambert, E. Mick, B. M. Neale, M. Fromer, S. M. Purcell, O. Svantesson, M. Landen, M. Hoglund, S. Lehmann, S. B. Gabriel, J. L. Moran, E. S. Lander, P. F. Sullivan, P. Sklar, H. Gronberg, C. M. Hultman and S. A. McCarroll (2014). "Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence." N Engl J Med **371**(26): 2477-2487.

George, S., S. S. Mok, M. Nurjono, S. Ayton, D. I. Finkelstein, C. L. Masters, Q. X. Li and J. G. Culvenor (2010). "alpha-Synuclein transgenic mice reveal compensatory increases in Parkinson's disease-associated proteins DJ-1 and parkin and have enhanced alpha-synuclein and PINK1 levels after rotenone treatment." J Mol Neurosci **42**(2): 243-254.

Gerstung, M., C. Beisel, M. Rechsteiner, P. Wild, P. Schraml, H. Moch and N. Beerenwinkel (2012). "Reliable detection of subclonal single-nucleotide variants in tumour cell populations." Nat Commun **3**: 811.

Gerstung, M., E. Papaemmanuil and P. J. Campbell (2014). "Subclonal variant calling with multiple samples and prior knowledge." Bioinformatics **30**(9): 1198-1204.

Geschwind, M. D. (2015). "Prion Diseases." Continuum (Minneap Minn) **21**(6 Neuroinfectious Disease): 1612-1638.

Ghani, M., A. E. Lang, L. Zinman, B. Nacmias, S. Sorbi, V. Bessi, A. Tedde, M. C. Tartaglia, E. I. Surace, C. Sato, D. Moreno, Z. Xi, R. Hung, M. A. Nalls, A. Singleton, P. St George-Hyslop and E. Rogaeva (2015). "Mutation analysis of patients with neurodegenerative disorders using NeuroX array." Neurobiol Aging **36**(1): 545 e549-514.

Ghanim, M., L. Guillot-Noel, F. Pasquier, L. Jornea, V. Deramecourt, B. Dubois, I. Le Ber, A. Brice, F. T. D. French Research Network on and Ftd/Mnd (2010). "CHMP2B mutations are rare in French families with frontotemporal lobar degeneration." J Neurol **257**(12): 2032-2036.

Giasson, B. I., J. E. Duda, S. M. Quinn, B. Zhang, J. Q. Trojanowski and V. M. Lee (2002). "Neuronal alpha-synucleinopathy with severe movement disorder in mice expressing A53T human alpha-synuclein." Neuron **34**(4): 521-533.

Giliberto, L., R. Borghi, A. Piccini, R. Mangerini, S. Sorbi, G. Cirmena, A. Garuti, B. Ghetti, F. Tagliavini, M. R. Mughal, M. P. Mattson, X. Zhu, X. Wang, M. Guglielmotto, E. Tamagno and M. Tabaton (2009). "Mutant presenilin 1 increases the expression and activity of BACE1." J Biol Chem **284**(14): 9027-9038.

Gitler, A. D. and H. Tsuiji (2016). "There has been an awakening: Emerging mechanisms of C9orf72 mutations in FTD/ALS." Brain Res **1647**: 19-29.

Gladman, M., C. Dharamshi and L. Zinman (2014). "Economic burden of amyotrophic lateral sclerosis: a Canadian study of out-of-pocket expenses." Amyotroph Lateral Scler Frontotemporal Degener **15**(5-6): 426-432.

Gleeson, J. G., S. Minnerath, R. I. Kuzniecky, W. B. Dobyns, I. D. Young, M. E. Ross and C. A. Walsh (2000). "Somatic and germline mosaic mutations in the doublecortin gene are associated with variable phenotypes." Am J Hum Genet **67**(3): 574-581.

Glessner, J. T., K. Wang, G. Cai, O. Korvatska, C. E. Kim, S. Wood, H. Zhang, A. Estes, C. W. Brune, J. P. Bradfield, M. Imielinski, E. C. Frackelton, J. Reichert, E. L. Crawford, J. Munson, P. M. Sleiman, R. Chiavacci, K. Annaiah, K. Thomas, C. Hou, W. Glaberson, J. Flory, F. Otieno, M. Garris, L. Soorya, L. Klei, J. Piven, K. J. Meyer, E. Anagnostou, T. Sakurai, R. M. Game, D. S. Rudd, D. Zurawiecki, C. J. McDougle, L. K. Davis, J. Miller, D. J. Posey, S. Michaels, A. Kolevzon, J. M. Silverman, R. Bernier, S. E. Levy, R. T. Schultz, G. Dawson, T. Owley, W. M. McMahon, T. H. Wassink, J. A. Sweeney, J. I. Nurnberger, H. Coon, J. S. Sutcliffe, N. J. Minshew, S. F. Grant, M. Bucan, E. H. Cook, J. D. Buxbaum, B. Devlin, G. D. Schellenberg and H. Hakonarson (2009). "Autism genome-wide copy number variation reveals ubiquitin and neuronal genes." Nature **459**(7246): 569-573.

Goedert, M. (2001). "Alpha-synuclein and neurodegenerative diseases." Nat Rev Neurosci **2**(7): 492-501.

Goedert, M. and M. G. Spillantini (2006). "A century of Alzheimer's disease." Science **314**(5800): 777-781.

Golbe, L. I., G. Di Iorio, V. Bonavita, D. C. Miller and R. C. Duvoisin (1990). "A large kindred with autosomal dominant Parkinson's disease." Ann Neurol **27**(3): 276-282.

Gomez-Isla, T., W. B. Growdon, M. J. McNamara, D. Nochlin, T. D. Bird, J. C. Arango, F. Lopera, K. S. Kosik, P. L. Lantos, N. J. Cairns and B. T. Hyman (1999). "The impact of different presenilin 1 andpresenilin 2 mutations on amyloid deposition, neurofibrillary changes and neuronal loss in the familial Alzheimer's disease brain: evidence for other phenotype-modifying factors." Brain **122 (Pt 9)**: 1709-1719.

Gomez-Isla, T., W. Wasco, W. P. Pettingell, S. Gurubhagavatula, S. D. Schmidt, P. D. Jondro, M. McNamara, L. A. Rodes, T. DiBlasi, W. B. Growdon, P. Seubert, D. Schenk, J. H. Growdon, B. T. Hyman and R. E. Tanzi (1997). "A novel presenilin-1 mutation: increased beta-amyloid and neurofibrillary changes." Ann Neurol **41**(6): 809-813.

Good, P. F., C. W. Olanow and D. P. Perl (1992). "Neuromelanin-containing neurons of the substantia nigra accumulate iron and aluminum in Parkinson's disease: a LAMMA study." Brain Res **593**(2): 343-346.

Goulart, B. K., M. N. de Lima, C. B. de Farias, G. K. Reolon, V. R. Almeida, J. Quevedo, F. Kapczinski, N. Schroder and R. Roesler (2010). "Ketamine impairs recognition memory consolidation and prevents learning-induced increase in hippocampal brain-derived neurotrophic factor levels." Neuroscience **167**(4): 969-973.

Gould, T. D. D., DT; Kovacsics, CE (2009). The Open Field Test. Mood and Anxiety Related Phenotypes in Mice. T. Gould. Totowa, NJ, Humana Press. **42**.

Grau-Rivera, O., E. Gelpi, C. Nos, C. Gaig, I. Ferrer, A. Saiz, A. Llado, J. L. Molinuevo, F. Graus, R. Sanchez-Valle and G. Neurological Tissue Bank Collaborative (2015). "Clinicopathological Correlations and Concomitant Pathologies in Rapidly Progressive Dementia: A Brain Bank Series." Neurodegener Dis **15**(6): 350-360.

Graveley, B. R. (2001). "Alternative splicing: increasing diversity in the proteomic world." Trends Genet **17**(2): 100-107.

Greaves, J., K. Lemonidis, O. A. Gorleku, C. Cruchaga, C. Grefen and L. H. Chamberlain (2012). "Palmitoylation-induced aggregation of cysteine-string protein mutants that cause neuronal ceroid lipofuscinosis." J Biol Chem **287**(44): 37330-37339.

Groen, E. J., K. Fumoto, A. M. Blokhuis, J. Engelen-Lee, Y. Zhou, D. M. van den Heuvel, M. Koppers, F. van Diggelen, J. van Heest, J. A. Demmers, J. Kirby, P. J. Shaw, E. Aronica, W. G. Spliet, J. H. Veldink, L. H. van den Berg and R. J. Pasterkamp (2013). "ALS-associated mutations in FUS disrupt the axonal distribution and function of SMN." Hum Mol Genet **22**(18): 3690-3704.

Gros-Louis, F., R. Lariviere, G. Gowing, S. Laurent, W. Camu, J. P. Bouchard, V. Meininger, G. A. Rouleau and J. P. Julien (2004). "A frameshift deletion in peripherin gene associated with amyotrophic lateral sclerosis." J Biol Chem **279**(44): 45951-45956.

Grounds, M. D., H. G. Radley, G. S. Lynch, K. Nagaraju and A. De Luca (2008). "Towards developing standard operating procedures for pre-clinical testing in the mdx mouse model of Duchenne muscular dystrophy." Neurobiol Dis **31**(1): 1-19.

Guella, I., A. Pistocchi, R. Asselta, V. Rimoldi, A. Ghilardi, F. Sironi, L. Trotta, P. Primignani, M. Zini, A. Zecchinelli, D. Coviello, G. Pezzoli, L. Del Giacco, S. Duga and S. Goldwurm (2011). "Mutational screening and zebrafish functional analysis of GIGYF2 as a Parkinson-disease gene." Neurobiol Aging **32**(11): 1994-2005.

Guerreiro, R., J. Bras and J. Hardy (2015). "SnapShot: Genetics of ALS and FTD." Cell **160**(4): 798 e791.

Guerreiro, R. and J. Hardy (2014). "Genetics of Alzheimer's disease." Neurotherapeutics **11**(4): 732-737.

Guerreiro, R., A. Wojtas, J. Bras, M. Carrasquillo, E. Rogaeva, E. Majounie, C. Cruchaga, C. Sassi, J. S. Kauwe, S. Younkin, L. Hazrati, J. Collinge, J. Pocock, T. Lashley, J. Williams, J. C. Lambert, P. Amouyel, A. Goate, R. Rademakers, K. Morgan, J. Powell, P. St George-Hyslop, A. Singleton, J. Hardy and G. Alzheimer

Genetic Analysis (2013). "TREM2 variants in Alzheimer's disease." N Engl J Med **368**(2): 117-127.

Guerreiro, R. J., M. Baquero, R. Blesa, M. Boada, J. M. Bras, M. J. Bullido, A. Calado, R. Crook, C. Ferreira, A. Frank, T. Gomez-Isla, I. Hernandez, A. Lleo, A. Machado, P. Martinez-Lage, J. Masdeu, L. Molina-Porcel, J. L. Molinuevo, P. Pastor, J. Perez-Tur, R. Relvas, C. R. Oliveira, M. H. Ribeiro, E. Rogaeva, A. Sa, L. Samaranch, R. Sanchez-Valle, I. Santana, L. Tarraga, F. Valdivieso, A. Singleton, J. Hardy and J. Clarimon (2010). "Genetic screening of Alzheimer's disease genes in Iberian and African samples yields novel mutations in presenilins and APP." Neurobiol Aging **31**(5): 725-731.

Guerreiro, R. J., N. Washecka, J. Hardy and A. Singleton (2010). "A thorough assessment of benign genetic variability in GRN and MAPT." Hum Mutat **31**(2): E1126-1140.

Guo, H. S., P. Zhu, L. Y. Yan, R. Li, B. Q. Hu, Y. Lian, J. Yan, X. L. Ren, S. L. Lin, J. S. Li, X. H. Jin, X. D. Shi, P. Liu, X. Y. Wang, W. Wang, Y. Wei, X. L. Li, F. Guo, X. L. Wu, X. Y. Fan, J. Yong, L. Wen, S. X. Xie, F. C. Tang and J. Qiao (2014). "The DNA methylation landscape of human early embryos." Nature **511**(7511): 606-+.

Hall, K., S. Yang, O. Sauchanka, M. G. Spillantini and O. Anichtchik (2015). "Behavioural deficits in transgenic mice expressing human truncated (1-120 amino acid) alpha-synuclein." Exp Neurol **264**: 8-13.

Hansson, O., H. Zetterberg, P. Buchhave, E. Londos, K. Blennow and L. Minthon (2006). "Association between CSF biomarkers and incipient Alzheimer's disease in patients with mild cognitive impairment: a follow-up study." Lancet Neurol **5**(3): 228-234.

Hastings, P. J., J. R. Lupski, S. M. Rosenberg and G. Ira (2009). "Mechanisms of change in gene copy number." Nat Rev Genet **10**(8): 551-564.

Hatters, D. M., C. A. Peters-Libeu and K. H. Weisgraber (2006). "Apolipoprotein E structure: insights into function." Trends Biochem Sci **31**(8): 445-454.

Hattori, N., T. Kitada, H. Matsumine, S. Asakawa, Y. Yamamura, H. Yoshino, T. Kobayashi, M. Yokochi, M. Wang, A. Yoritaka, T. Kondo, S. Kuzuhara, S. Nakamura, N. Shimizu and Y. Mizuno (1998). "Molecular genetic analysis of a novel Parkin gene in Japanese families with autosomal recessive juvenile parkinsonism: evidence for variable homozygous deletions in the Parkin gene in affected individuals." Ann Neurol **44**(6): 935-941.

Hauser, D. N., A. A. Dillman, J. Ding, Y. Li and M. R. Cookson (2014). "Post-translational decrease in respiratory chain proteins in the Polg mutator mouse brain." PLoS One **9**(4): e94646.

Head, S. R., H. K. Komori, S. A. LaMere, T. Whisenant, F. Van Nieuwerburgh, D. R. Salomon and P. Ordoukhanian (2014). "Library construction for next-generation sequencing: overviews and challenges." Biotechniques **56**(2): 61-64, 66, 68, passim.

Healy, D. G., M. Falchi, S. S. O'Sullivan, V. Bonifati, A. Durr, S. Bressman, A. Brice, J. Aasly, C. P. Zabetian, S. Goldwurm, J. J. Ferreira, E. Tolosa, D. M. Kay, C. Klein, D. R. Williams, C. Marras, A. E. Lang, Z. K. Wszolek, J. Berciano, A. H. Schapira, T. Lynch, K. P. Bhatia, T. Gasser, A. J. Lees, N. W. Wood and L. C. International (2008). "Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study." Lancet Neurol **7**(7): 583-590.

Hebert, L. E., P. A. Scherr, J. L. Bienias, D. A. Bennett and D. A. Evans (2003). "Alzheimer disease in the US population: prevalence estimates using the 2000 census." Arch Neurol **60**(8): 1119-1122.

Heilig, E. A., W. Xia, J. Shen and R. J. Kelleher, 3rd (2010). "A presenilin-1 mutation identified in familial Alzheimer disease with cotton wool plaques causes a nearly complete loss of gamma-secretase activity." J Biol Chem **285**(29): 22350-22359.

Henn, I. H., L. Bouman, J. S. Schlehe, A. Schlierf, J. E. Schramm, E. Wegener, K. Nakaso, C. Culmsee, B. Berninger, D. Krappmann, J. Tatzelt and K. F. Winklhofer

(2007). "Parkin mediates neuroprotection through activation of IkappaB kinase/nuclear factor-kappaB signaling." J Neurosci **27**(8): 1868-1878.

Hirtz, D., D. J. Thurman, K. Gwinn-Hardy, M. Mohamed, A. R. Chaudhuri and R. Zalutsky (2007). "How common are the "common" neurologic disorders?" Neurology **68**(5): 326-337.

Hoang, M. L., I. Kinde, C. Tomasetti, K. W. McMahon, T. A. Rosenquist, A. P. Grollman, K. W. Kinzler, B. Vogelstein and N. Papadopoulos (2016). "Genome-wide quantification of rare somatic mutations in normal human tissues using massively parallel sequencing." Proc Natl Acad Sci U S A **113**(35): 9846-9851.

Hocking, L. J., G. J. Lucas, A. Daroszewska, T. Cundy, G. C. Nicholson, J. Donath, J. P. Walsh, C. Finlayson, J. R. Cavey, B. Ciani, P. W. Sheppard, M. S. Searle, R. Layfield and S. H. Ralston (2004). "Novel UBA domain mutations of SQSTM1 in Paget's disease of bone: genotype phenotype correlation, functional analysis, and structural consequences." J Bone Miner Res **19**(7): 1122-1127.

Holman, R. C., E. D. Belay, K. Y. Christensen, R. A. Maddox, A. M. Minino, A. M. Folkema, D. L. Haberling, T. A. Hammett, K. D. Kochanek, J. J. Sejvar and L. B. Schonberger (2010). "Human prion diseases in the United States." PLoS One **5**(1): e8521.

Hoshi, K., H. Yoshino, J. Urata, Y. Nakamura, H. Yanagawa and T. Sato (2000). "Creutzfeldt-Jakob disease associated with cadaveric dura mater grafts in Japan." Neurology **55**(5): 718-721.

Houlden, H. and A. B. Singleton (2012). "The genetics and neuropathology of Parkinson's disease." Acta Neuropathol **124**(3): 325-338.

Hubley, R., R. D. Finn, J. Clements, S. R. Eddy, T. A. Jones, W. Bao, A. F. Smit and T. J. Wheeler (2016). "The Dfam database of repetitive DNA families." Nucleic Acids Res **44**(D1): D81-89.

Hurd, M. D., P. Martorell and K. M. Langa (2013). "Monetary costs of dementia in the United States." N Engl J Med **369**(5): 489-490.

Hutton, M., C. L. Lendon, P. Rizzu, M. Baker, S. Froelich, H. Houlden, S. Pickering-Brown, S. Chakraverty, A. Isaacs, A. Grover, J. Hackett, J. Adamson, S. Lincoln, D. Dickson, P. Davies, R. C. Petersen, M. Stevens, E. de Graaff, E. Wauters, J. van Baren, M. Hillebrand, M. Joosse, J. M. Kwon, P. Nowotny, L. K. Che, J. Norton, J. C. Morris, L. A. Reed, J. Trojanowski, H. Basun, L. Lannfelt, M. Neystat, S. Fahn, F. Dark, T. Tannenberg, P. R. Dodd, N. Hayward, J. B. Kwok, P. R. Schofield, A. Andreadis, J. Snowden, D. Craufurd, D. Neary, F. Owen, B. A. Oostra, J. Hardy, A. Goate, J. van Swieten, D. Mann, T. Lynch and P. Heutink (1998). "Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17." Nature **393**(6686): 702-705.

Hyman, B. T. and J. Q. Trojanowski (1997). "Consensus recommendations for the postmortem diagnosis of Alzheimer disease from the National Institute on Aging and the Reagan Institute Working Group on diagnostic criteria for the neuropathological assessment of Alzheimer disease." J Neuropathol Exp Neurol **56**(10): 1095-1097.

Ingelsson, M., K. Ramasamy, C. Russ, S. H. Freeman, J. Orne, S. Raju, T. Matsui, J. H. Growdon, M. P. Frosch, B. Ghetti, R. H. Brown, M. C. Irizarry and B. T. Hyman (2007). "Increase in the relative expression of tau with four microtubule binding repeat regions in frontotemporal lobar degeneration and progressive supranuclear palsy brains." Acta Neuropathol **114**(5): 471-479.

Inoue, I., T. Kitamoto, K. Doh-ura, H. Shii, I. Goto and J. Tateishi (1994). "Japanese family with Creutzfeldt-Jakob disease with codon 200 point mutation of the prion protein gene." Neurology **44**(2): 299-301.

Itoh, K., S. Weis, P. Mehraein and J. Muller-Hocker (1996). "Cytochrome c oxidase defects of the human substantia nigra in normal aging." Neurobiol Aging **17**(6): 843-848.

Jack, C. R., Jr., D. S. Knopman, W. J. Jagust, R. C. Petersen, M. W. Weiner, P. S. Aisen, L. M. Shaw, P. Vemuri, H. J. Wiste, S. D. Weigand, T. G. Lesnick, V. S. Pankratz, M. C. Donohue and J. Q. Trojanowski (2013). "Tracking pathophysiological processes in Alzheimer's disease: an updated hypothetical model of dynamic biomarkers." Lancet Neurol **12**(2): 207-216.

Jaenisch, R. and B. Mintz (1974). "Simian virus 40 DNA sequences in DNA of healthy adult mice derived from preimplantation blastocysts injected with viral DNA." Proc Natl Acad Sci U S A **71**(4): 1250-1254.

Jaiswal, S., P. Fontanillas, J. Flannick, A. Manning, P. V. Grauman, B. G. Mar, R. C. Lindsley, C. H. Mermel, N. Burtt, A. Chavez, J. M. Higgins, V. Moltchanov, F. C. Kuo, M. J. Kluk, B. Henderson, L. Kinnunen, H. A. Koistinen, C. Ladenvall, G. Getz, A. Correa, B. F. Banahan, S. Gabriel, S. Kathiresan, H. M. Stringham, M. I. McCarthy, M. Boehnke, J. Tuomilehto, C. Haiman, L. Groop, G. Atzmon, J. G. Wilson, D. Neuberg, D. Altshuler and B. L. Ebert (2014). "Age-related clonal hematopoiesis associated with adverse outcomes." N Engl J Med **371**(26): 2488-2498.

Jaiswal, S., P. Natarajan, A. J. Silver, C. J. Gibson, A. G. Bick, E. Shvartz, M. McConkey, N. Gupta, S. Gabriel, D. Ardissino, U. Baber, R. Mehran, V. Fuster, J. Danesh, P. Frossard, D. Saleheen, O. Melander, G. K. Sukhova, D. Neuberg, P. Libby, S. Kathiresan and B. L. Ebert (2017). "Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease." N Engl J Med **377**(2): 111-121.

Jamuar, S. S., A. T. Lam, M. Kircher, A. M. D'Gama, J. Wang, B. J. Barry, X. Zhang, R. S. Hill, J. N. Partlow, A. Rozzo, S. Servattalab, B. K. Mehta, M. Topcu, D. Amrom, E. Andermann, B. Dan, E. Parrini, R. Guerrini, I. E. Scheffer, S. F. Berkovic, R. J. Leventer, Y. Shen, B. L. Wu, A. J. Barkovich, M. Sahin, B. S. Chang, M. Bamshad, D. A. Nickerson, J. Shendure, A. Poduri, T. W. Yu and C. A. Walsh (2014). "Somatic mutations in cerebral cortical malformations." N Engl J Med **371**(8): 733-743.

Jankovic, J. (2008). "Parkinson's disease: clinical features and diagnosis." J Neurol Neurosurg Psychiatry **79**(4): 368-376.

Jankowsky, J. L., D. J. Fadale, J. Anderson, G. M. Xu, V. Gonzales, N. A. Jenkins, N. G. Copeland, M. K. Lee, L. H. Younkin, S. L. Wagner, S. G. Younkin and D. R. Borchelt (2004). "Mutant presenilins specifically elevate the levels of the 42 residue beta-amyloid peptide in vivo: evidence for augmentation of a 42-specific gamma secretase." Hum Mol Genet **13**(2): 159-170.

Janssen, J. C., J. A. Beck, T. A. Campbell, A. Dickinson, N. C. Fox, R. J. Harvey, H. Houlden, M. N. Rossor and J. Collinge (2003). "Early onset familial Alzheimer's disease: Mutation frequency in 31 families." Neurology **60**(2): 235-239.

Janus, C. and H. Welzl (2010). "Mouse models of neurodegenerative diseases: criteria and general methodology." Methods Mol Biol **602**: 323-345.

Jellinger, K. (2013). "Neuropathology of Dementia Disorders." J Alzheimers Dis & Parkinsonism **4**(1).

Jin, S. C., B. A. Benitez, C. M. Karch, B. Cooper, T. Skorupa, D. Carrell, J. B. Norton, S. Hsu, O. Harari, Y. Cai, S. Bertelsen, A. M. Goate and C. Cruchaga (2014). "Coding variants in TREM2 increase risk for Alzheimer's disease." Hum Mol Genet **23**(21): 5838-5846.

Johnson, J. O., J. Mandrioli, M. Benatar, Y. Abramzon, V. M. Van Deerlin, J. Q. Trojanowski, J. R. Gibbs, M. Brunetti, S. Gronka, J. Wuu, J. Ding, L. McCluskey, M. Martinez-Lage, D. Falcone, D. G. Hernandez, S. Arepalli, S. Chong, J. C. Schymick, J. Rothstein, F. Landi, Y. D. Wang, A. Calvo, G. Mora, M. Sabatelli, M. R. Monsurro, S. Battistini, F. Salvi, R. Spataro, P. Sola, G. Borghero, I. Consortium, G. Galassi, S. W. Scholz, J. P. Taylor, G. Restagno, A. Chio and B. J. Traynor (2010). "Exome sequencing reveals VCP mutations as a cause of familial ALS." Neuron **68**(5): 857-864.

Jonsson, P. A., K. S. Graffmo, P. M. Andersen, T. Brannstrom, M. Lindberg, M. Oliveberg and S. L. Marklund (2006). "Disulphide-reduced superoxide dismutase-1 in CNS of transgenic amyotrophic lateral sclerosis models." Brain **129**(Pt 2): 451-464.

Jonsson, P. A., K. S. Graffmo, T. Brannstrom, P. Nilsson, P. M. Andersen and S. L. Marklund (2006). "Motor neuron disease in mice expressing the wild type-like D90A mutant superoxide dismutase-1." J Neuropathol Exp Neurol **65**(12): 1126-1136.

Ju, Y. S., I. Martincorena, M. Gerstung, M. Petljak, L. B. Alexandrov, R. Rahbari, D. C. Wedge, H. R. Davies, M. Ramakrishna, A. Fullam, S. Martin, C. Alder, N.

Patel, S. Gamble, S. O'Meara, D. D. Giri, T. Sauer, S. E. Pinder, C. A. Purdie, A. Borg, H. Stunnenberg, M. van de Vijver, B. K. Tan, C. Caldas, A. Tutt, N. T. Ueno, L. J. van 't Veer, J. W. Martens, C. Sotiriou, S. Knappskog, P. N. Span, S. R. Lakhani, J. E. Eyfjord, A. L. Borresen-Dale, A. Richardson, A. M. Thompson, A. Viari, M. E. Hurles, S. Nik-Zainal, P. J. Campbell and M. R. Stratton (2017). "Somatic mutations reveal asymmetric cellular dynamics in the early human embryo." Nature **543**(7647): 714-718.

Jubault, T., S. M. Brambati, C. Degroot, B. Kullmann, A. P. Strafella, A. L. Lafontaine, S. Chouinard and O. Monchi (2009). "Regional brain stem atrophy in idiopathic Parkinson's disease detected by anatomical MRI." PLoS One **4**(12): e8247.

Kabashi, E., V. Bercier, A. Lissouba, M. Liao, E. Brustein, G. A. Rouleau and P. Drapeau (2011). "FUS and TARDBP but not SOD1 interact in genetic models of amyotrophic lateral sclerosis." PLoS Genet **7**(8): e1002214.

Kabashi, E., H. El Oussini, V. Bercier, F. Gros-Louis, P. N. Valdmanis, J. McDearmid, I. A. Mejier, P. A. Dion, N. Dupre, D. Hollinger, J. Sinniger, S. Dirrig-Grosch, W. Camu, V. Meininger, J. P. Loeffler, F. Rene, P. Drapeau, G. A. Rouleau and L. Dupuis (2013). "Investigating the contribution of VAPB/ALS8 loss of function in amyotrophic lateral sclerosis." Hum Mol Genet **22**(12): 2350-2360.

Kamboh, M. I., C. E. Aston, J. Perez-Tur, E. Kokmen, R. E. Ferrell, J. Hardy and S. T. DeKosky (1999). "A novel mutation in the apolipoprotein E gene (APOE*4 Pittsburgh) is associated with the risk of late-onset Alzheimer's disease." Neurosci Lett **263**(2-3): 129-132.

Karumbayaram, S., T. K. Kelly, A. A. Paucar, A. J. Roe, J. A. Umbach, A. Charles, S. A. Goldman, H. I. Kornblum and M. Wiedau-Pazos (2009). "Human embryonic stem cell-derived motor neurons expressing SOD1 mutants exhibit typical signs of motor neuron degeneration linked to ALS." Dis Model Mech **2**(3-4): 189-195.

Keeney, P. M., J. Xie, R. A. Capaldi and J. P. Bennett, Jr. (2006). "Parkinson's disease brain mitochondrial complex I has oxidatively damaged subunits and is functionally impaired and misassembled." J Neurosci **26**(19): 5256-5264.

Keller, M. F., L. Ferrucci, A. B. Singleton, P. J. Tienari, H. Laaksovirta, G. Restagno, A. Chio, B. J. Traynor and M. A. Nalls (2014). "Genome-wide analysis of the heritability of amyotrophic lateral sclerosis." JAMA Neurol **71**(9): 1123-1134.

Keller, M. F., M. Saad, J. Bras, F. Bettella, N. Nicolaou, J. Simon-Sanchez, F. Mittag, F. Buchel, M. Sharma, J. R. Gibbs, C. Schulte, V. Moskvina, A. Durr, P. Holmans, L. L. Kilarski, R. Guerreiro, D. G. Hernandez, A. Brice, P. Ylikotila, H. Stefansson, K. Majamaa, H. R. Morris, N. Williams, T. Gasser, P. Heutink, N. W. Wood, J. Hardy, M. Martinez, A. B. Singleton, M. A. Nalls, C. International Parkinson's Disease Genomics and C. Wellcome Trust Case Control (2012). "Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease." Hum Mol Genet **21**(22): 4996-5009.

Keogh, M. J. and P. F. Chinnery (2015). "Mitochondrial DNA mutations in neurodegeneration." Biochim Biophys Acta **1847**(11): 1401-1411.

Keogh, M. J., M. Kurzawa-Akanbi, H. Griffin, K. Douroudis, K. L. Ayers, R. I. Hussein, G. Hudson, A. Pyle, H. J. Cordell, J. Attems, I. G. McKeith, J. T. O'Brien, D. J. Burn, C. M. Morris, A. J. Thomas and P. F. Chinnery (2016). "Exome sequencing in dementia with Lewy bodies." Transl Psychiatry **6**: e728.

Keogh, M. J., A. Pyle, D. Daud, H. Griffin, K. Douroudis, G. Eglon, J. Miller, R. Horvath and P. F. Chinnery (2015). "Clinical heterogeneity of primary familial brain calcification due to a novel mutation in PDGFB." Neurology **84**(17): 1818-1820.

Keogh, M. J., H. Steele, K. Douroudis, A. Pyle, J. Duff, R. Hussain, T. Smertenko, H. Griffin, M. Santibanez-Koref, R. Horvath and P. F. Chinnery (2015). "Frequency of rare recessive mutations in unexplained late onset cerebellar ataxia." J Neurol **262**(8): 1822-1827.

Khoris, J., B. Moulard, V. Briolotti, M. Hayer, A. Durieux, P. Clavelou, A. Malafosse, G. A. Rouleau and W. Camu (2000). "Coexistence of dominant and recessive familial amyotrophic lateral sclerosis with the D90A Cu,Zn superoxide dismutase mutation within the same country." Eur J Neurol **7**(2): 207-211.

Khrapko, K. (2011). "The timing of mitochondrial DNA mutations in aging." Nat Genet **43**(8): 726-727.

Kim, H. J., H. Y. Kim, C. S. Ki and S. H. Kim (2010). "Presenilin 1 gene mutation (M139I) in a patient with an early-onset Alzheimer's disease: clinical characteristics and genetic identification." Neurol Sci **31**(6): 781-783.

Kim, H. J., N. C. Kim, Y. D. Wang, E. A. Scarborough, J. Moore, Z. Diaz, K. S. MacLea, B. Freibaum, S. Li, A. Molliex, A. P. Kanagaraj, R. Carter, K. B. Boylan, A. M. Wojtas, R. Rademakers, J. L. Pinkus, S. A. Greenberg, J. Q. Trojanowski, B. J. Traynor, B. N. Smith, S. Topp, A. S. Gkazi, J. Miller, C. E. Shaw, M. Kottlors, J. Kirschner, A. Pestronk, Y. R. Li, A. F. Ford, A. D. Gitler, M. Benatar, O. D. King, V. E. Kimonis, E. D. Ross, C. C. Weihl, J. Shorter and J. P. Taylor (2013). "Mutations in prion-like domains in hnRNPA2B1 and hnRNPA1 cause multisystem proteinopathy and ALS." Nature **495**(7442): 467-473.

Kim, M., H. Kook, H. J. Park, H. S. Ahn, K. C. Lee, K. S. Lee, S. K. Park, J. Y. Lim, H. K. Kim, D. K. Han, D. S. Lee and M. D. S. W. P. Korean Pediatric (2013). "Quantitative comparison of CDKN2B methylation in pediatric and adult myelodysplastic syndromes." Acta Haematol **130**(2): 115-121.

King, A., S. Al-Sarraj, C. Troakes, B. N. Smith, S. Maekawa, M. Iovino, M. G. Spillantini and C. E. Shaw (2013). "Mixed tau, TDP-43 and p62 pathology in FTLD associated with a C9ORF72 repeat expansion and p.Ala239Thr MAPT (tau) variant." Acta Neuropathol **125**(2): 303-310.

Kirby, J., J. R. Highley, L. Cox, E. F. Goodall, C. Hewitt, J. A. Hartley, H. C. Hollinger, M. Fox, P. G. Ince, C. J. McDermott and P. J. Shaw (2013). "Lack of unique neuropathology in amyotrophic lateral sclerosis associated with p.K54E angiogenin (ANG) mutation." Neuropathol Appl Neurobiol **39**(5): 562-571.

Kircher, M., D. M. Witten, P. Jain, B. J. O'Roak, G. M. Cooper and J. Shendure (2014). "A general framework for estimating the relative pathogenicity of human genetic variants." Nat Genet **46**(3): 310-315.

Klein, C., A. Djarmati, K. Hedrich, N. Schafer, C. Scaglione, R. Marchese, N. Kock, B. Schule, A. Hiller, T. Lohnau, S. Winkler, K. Wiegers, R. Hering, P. Bauer, O. Riess, G. Abbruzzese, P. Martinelli and P. P. Pramstaller (2005). "PINK1, Parkin, and DJ-1 mutations in Italian patients with early-onset parkinsonism." Eur J Hum Genet **13**(9): 1086-1093.

Klein, C. and A. Westenberger (2012). "Genetics of Parkinson's disease." Cold Spring Harb Perspect Med **2**(1): a008888.

Klement, I. A., P. J. Skinner, M. D. Kaytor, H. Yi, S. M. Hersch, H. B. Clark, H. Y. Zoghbi and H. T. Orr (1998). "Ataxin-1 nuclear localization and aggregation: role in polyglutamine-induced disease in SCA1 transgenic mice." Cell **95**(1): 41-53.

Kleopa, K. A., M. Sherman, B. Neal, G. J. Romano and T. Heiman-Patterson (1999). "Bipap improves survival and rate of pulmonary function decline in patients with ALS." J Neurol Sci **164**(1): 82-88.

Kloosterman, W. P., L. C. Francioli, F. Hormozdiari, T. Marschall, J. Y. Hehir-Kwa, A. Abdellaoui, E. W. Lameijer, M. H. Moed, V. Koval, I. Renkens, M. J. van Roosmalen, P. Arp, L. C. Karssen, B. P. Coe, R. E. Handsaker, E. D. Suchiman, E. Cuppen, D. T. Thung, M. McVey, M. C. Wendl, C. Genome of Netherlands, A. Uitterlinden, C. M. van Duijn, M. A. Swertz, C. Wijmenga, G. B. van Ommen, P. E. Slagboom, D. I. Boomsma, A. Schonhuth, E. E. Eichler, P. I. de Bakker, K. Ye and V. Guryev (2015). "Characteristics of de novo structural changes in the human genome." Genome Res **25**(6): 792-801.

Kobayashi, Y. M., E. P. Rader, R. W. Crawford and K. P. Campbell (2012). "Endpoint measures in the mdx mouse relevant for muscular dystrophy pre-clinical studies." Neuromuscul Disord **22**(1): 34-42.

Koboldt, D. C., K. Chen, T. Wylie, D. E. Larson, M. D. McLellan, E. R. Mardis, G. M. Weinstock, R. K. Wilson and L. Ding (2009). "VarScan: variant detection in massively parallel sequencing of individual and pooled samples." Bioinformatics **25**(17): 2283-2285.

Koboldt, D. C., Q. Zhang, D. E. Larson, D. Shen, M. D. McLellan, L. Lin, C. A. Miller, E. R. Mardis, L. Ding and R. K. Wilson (2012). "VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing." Genome Res **22**(3): 568-576.

Kovacs, G. G. (2015). "Invited review: Neuropathology of tauopathies: principles and practice." Neuropathol Appl Neurobiol **41**(1): 3-23.

Kraytsberg, Y., E. Kudryavtseva, A. C. McKee, C. Geula, N. W. Kowall and K. Khrapko (2006). "Mitochondrial DNA deletions are abundant and cause functional impairment in aged human substantia nigra neurons." Nat Genet **38**(5): 518-520.

Kujoth, G. C., A. Hiona, T. D. Pugh, S. Someya, K. Panzer, S. E. Wohlgemuth, T. Hofer, A. Y. Seo, R. Sullivan, W. A. Jobling, J. D. Morrow, H. Van Remmen, J. M. Sedivy, T. Yamasoba, M. Tanokura, R. Weindruch, C. Leeuwenburgh and T. A. Prolla (2005). "Mitochondrial DNA mutations, oxidative stress, and apoptosis in mammalian aging." Science **309**(5733): 481-484.

Kukull, W. A., G. D. Schellenberg, J. D. Bowen, W. C. McCormick, C. E. Yu, L. Teri, J. D. Thompson, E. S. O'Meara and E. B. Larson (1996). "Apolipoprotein E in Alzheimer's disease risk and case detection: a case-control study." J Clin Epidemiol **49**(10): 1143-1148.

Kumar, P., S. Henikoff and P. C. Ng (2009). "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." Nat Protoc **4**(7): 1073-1081.

Kurihara, N., Y. Hiruma, H. Zhou, M. A. Subler, D. W. Dempster, F. R. Singer, S. V. Reddy, H. E. Gruber, J. J. Windle and G. D. Roodman (2007). "Mutation of the sequestosome 1 (p62) gene increases osteoclastogenesis but does not induce Paget disease." J Clin Invest **117**(1): 133-142.

Kwok, C. T., A. Morris and J. S. de Belleroche (2013). "Sequestosome-1 (SQSTM1) sequence variants in ALS cases in the UK: prevalence and coexistence of SQSTM1 mutations in ALS kindred with PDB." Eur J Hum Genet.

Lambert, J. C., C. A. Ibrahim-Verbaas, D. Harold, A. C. Naj, R. Sims, C. Bellenguez, A. L. DeStafano, J. C. Bis, G. W. Beecham, B. Grenier-Boley, G. Russo, T. A. Thorton-Wells, N. Jones, A. V. Smith, V. Chouraki, C. Thomas, M. A. Ikram, D. Zelenika, B. N. Vardarajan, Y. Kamatani, C. F. Lin, A. Gerrish, H. Schmidt, B. Kunkle, M. L. Dunstan, A. Ruiz, M. T. Bihoreau, S. H. Choi, C. Reitz, F. Pasquier, C. Cruchaga, D. Craig, N. Amin, C. Berr, O. L. Lopez, P. L. De Jager, V. Deramecourt, J. A. Johnston, D. Evans, S. Lovestone, L. Letenneur, F. J. Moron, D. C. Rubinsztein, G. Eiriksdottir, K. Sleegers, A. M. Goate, N. Fievet, M. W. Huentelman, M. Gill, K. Brown, M. I. Kamboh, L. Keller, P. Barberger-Gateau, B. McGuiness, E. B. Larson, R. Green, A. J. Myers, C. Dufouil, S. Todd, D. Wallon, S. Love, E. Rogaeva, J. Gallacher, P. St George-Hyslop, J. Clarimon, A. Lleo, A. Bayer, D. W. Tsuang, L. Yu, M. Tsolaki, P. Bossu, G. Spalletta, P. Proitsi, J. Collinge, S. Sorbi, F. Sanchez-Garcia, N. C. Fox, J. Hardy, M. C. Deniz Naranjo, P. Bosco, R. Clarke, C. Brayne, D. Galimberti, M. Mancuso, F. Matthews, I. European Alzheimer's Disease, Genetic, D. Environmental Risk in Alzheimer's, C. Alzheimer's Disease Genetic, H. Cohorts for, E. Aging Research in Genomic, S. Moebus, P. Mecocci, M. Del Zompo, W. Maier, H. Hampel, A. Pilotto, M. Bullido, F. Panza, P. Caffarra, B. Nacmias, J. R. Gilbert, M. Mayhaus, L. Lannefelt, H. Hakonarson, S. Pichler, M. M. Carrasquillo, M. Ingelsson, D. Beekly, V. Alvarez, F. Zou, O. Valladares, S. G. Younkin, E. Coto, K. L. Hamilton-Nelson, W. Gu, C. Razquin, P. Pastor, I. Mateo, M. J. Owen, K. M. Faber, P. V. Jonsson, O. Combarros, M. C. O'Donovan, L. B. Cantwell, H. Soininen, D. Blacker, S. Mead, T. H. Mosley, Jr., D. A. Bennett, T. B. Harris, L. Fratiglioni, C. Holmes, R. F. de Bruijn, P. Passmore, T. J. Montine, K. Bettens, J. I. Rotter, A. Brice, K. Morgan, T. M. Foroud, W. A. Kukull, D. Hannequin, J. F. Powell, M. A. Nalls, K. Ritchie, K. L. Lunetta, J. S. Kauwe, E. Boerwinkle, M. Riemenschneider, M. Boada, M. Hiltuenen, E. R. Martin, R. Schmidt, D. Rujescu, L. S. Wang, J. F. Dartigues, R. Mayeux, C. Tzourio, A. Hofman, M. M. Nothen, C. Graff, B. M. Psaty, L. Jones, J. L. Haines, P. A. Holmans, M. Lathrop, M. A. Pericak-Vance, L. J. Launer, L. A. Farrer, C. M. van Duijn, C. Van Broeckhoven, V. Moskvina, S. Seshadri, J. Williams, G. D. Schellenberg and P. Amouyel (2013). "Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease." Nat Genet **45**(12): 1452-1458.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, N. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J. Grimwood, D. R. Cox, M. V. Olson, R. Kaul, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp,

D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowski, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi and Y. J. Chen (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Lander, E. S., L. M. Linton, B. Birren, C. Nusbaum, M. C. Zody, J. Baldwin, K. Devon, K. Dewar, M. Doyle, W. FitzHugh, R. Funke, D. Gage, K. Harris, A. Heaford, J. Howland, L. Kann, J. Lehoczky, R. LeVine, P. McEwan, K. McKernan, J. Meldrim, J. P. Mesirov, C. Miranda, W. Morris, J. Naylor, C. Raymond, M. Rosetti, R. Santos, A. Sheridan, C. Sougnez, Y. Stange-Thomann, N. Stojanovic, A. Subramanian, D. Wyman, J. Rogers, J. Sulston, R. Ainscough, S. Beck, D. Bentley, J. Burton, C. Clee, N. Carter, A. Coulson, R. Deadman, P. Deloukas, A. Dunham, I. Dunham, R. Durbin, L. French, D. Grafham, S. Gregory, T. Hubbard, S. Humphray, A. Hunt, M. Jones, C. Lloyd, A. McMurray, L. Matthews, S. Mercer, S. Milne, J. C. Mullikin, A. Mungall, R. Plumb, M. Ross, R. Shownkeen, S. Sims, R. H. Waterston, R. K. Wilson, L. W. Hillier, J. D. McPherson, M. A. Marra, E. R. Mardis, L. A. Fulton, A. T. Chinwalla, K. H. Pepin, W. R. Gish, S. L. Chissoe, M. C. Wendl, K. D. Delehaunty, T. L. Miner, A. Delehaunty, J. B. Kramer, L. L. Cook, R. S. Fulton, D. L. Johnson, P. J. Minx, S. W. Clifton, T. Hawkins, E. Branscomb, P. Predki, P. Richardson, S. Wenning, T. Slezak, N. Doggett, J. F. Cheng, A. Olsen, S. Lucas, C. Elkin, E. Uberbacher, M. Frazier, R. A. Gibbs, D. M. Muzny, S. E. Scherer, J. B. Bouck, E. J. Sodergren, K. C. Worley, C. M. Rives, J. H. Gorrell, M. L. Metzker, S. L. Naylor, R. S. Kucherlapati, D. L. Nelson, G. M. Weinstock, Y. Sakaki, A. Fujiyama, M. Hattori, T. Yada, A. Toyoda, T. Itoh, C. Kawagoe, H. Watanabe, Y. Totoki, T. Taylor, J. Weissenbach, R. Heilig, W. Saurin, F. Artiguenave, P. Brottier, T. Bruls, E. Pelletier, C. Robert, P. Wincker, D. R. Smith, L. Doucette-Stamm, M. Rubenfield, K. Weinstock, H. M. Lee, J. Dubois, A. Rosenthal, M. Platzer, G. Nyakatura, S. Taudien, A. Rump, H. Yang, J. Yu, J. Wang, G. Huang, J. Gu, L. Hood, L. Rowen, A. Madan, S. Qin, R. W. Davis, N. A. Federspiel, A. P. Abola, M. J. Proctor, R. M. Myers, J. Schmutz, M. Dickson, J.

Grimwood, D. R. Cox, M. V. Olson, R. Kaul, C. Raymond, N. Shimizu, K. Kawasaki, S. Minoshima, G. A. Evans, M. Athanasiou, R. Schultz, B. A. Roe, F. Chen, H. Pan, J. Ramser, H. Lehrach, R. Reinhardt, W. R. McCombie, M. de la Bastide, N. Dedhia, H. Blocker, K. Hornischer, G. Nordsiek, R. Agarwala, L. Aravind, J. A. Bailey, A. Bateman, S. Batzoglou, E. Birney, P. Bork, D. G. Brown, C. B. Burge, L. Cerutti, H. C. Chen, D. Church, M. Clamp, R. R. Copley, T. Doerks, S. R. Eddy, E. E. Eichler, T. S. Furey, J. Galagan, J. G. Gilbert, C. Harmon, Y. Hayashizaki, D. Haussler, H. Hermjakob, K. Hokamp, W. Jang, L. S. Johnson, T. A. Jones, S. Kasif, A. Kaspryzk, S. Kennedy, W. J. Kent, P. Kitts, E. V. Koonin, I. Korf, D. Kulp, D. Lancet, T. M. Lowe, A. McLysaght, T. Mikkelsen, J. V. Moran, N. Mulder, V. J. Pollara, C. P. Ponting, G. Schuler, J. Schultz, G. Slater, A. F. Smit, E. Stupka, J. Szustakowki, D. Thierry-Mieg, J. Thierry-Mieg, L. Wagner, J. Wallis, R. Wheeler, A. Williams, Y. I. Wolf, K. H. Wolfe, S. P. Yang, R. F. Yeh, F. Collins, M. S. Guyer, J. Peterson, A. Felsenfeld, K. A. Wetterstrand, A. Patrinos, M. J. Morgan, P. de Jong, J. J. Catanese, K. Osoegawa, H. Shizuya, S. Choi, Y. J. Chen, J. Szustakowki and C. International Human Genome Sequencing (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.

Landers, J. E., A. L. Leclerc, L. Shi, A. Virkud, T. Cho, M. M. Maxwell, A. F. Henry, M. Polak, J. D. Glass, T. J. Kwiatkowski, A. Al-Chalabi, C. E. Shaw, P. N. Leigh, I. Rodriguez-Leyza, D. McKenna-Yasek, P. C. Sapp and R. H. Brown, Jr. (2008). "New VAPB deletion variant and exclusion of VAPB mutations in familial ALS." Neurology **70**(14): 1179-1185.

Langmead, B., C. Trapnell, M. Pop and S. L. Salzberg (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." Genome Biol **10**(3): R25.

Lappalainen, I., J. Lopez, L. Skipper, T. Hefferon, J. D. Spalding, J. Garner, C. Chen, M. Maguire, M. Corbett, G. Zhou, J. Paschall, V. Ananiev, P. Flicek and D. M. Church (2013). "DbVar and DGVa: public archives for genomic structural variation." Nucleic Acids Res **41**(Database issue): D936-941.

Larner, A. J., P. S. Ray and M. Doran (2007). "The R269H mutation in presenilin-1 presenting as late-onset autosomal dominant Alzheimer's disease." J Neurol Sci **252**(2): 173-176.

Lattante, S., I. Le Ber, A. Camuzat, A. Brice and E. Kabashi (2013). "Mutations in the PFN1 gene are not a common cause in patients with amyotrophic lateral sclerosis and frontotemporal lobar degeneration in France." Neurobiol Aging **34**(6): 1709 e1701-1702.

Lautier, C., S. Goldwurm, A. Durr, B. Giovannone, W. G. Tsiaras, G. Pezzoli, A. Brice and R. J. Smith (2008). "Mutations in the GIGYF2 (TNRC15) gene at the PARK11 locus in familial Parkinson disease." Am J Hum Genet **82**(4): 822-833.

Le Ber, I., A. Camuzat, R. Guerreiro, K. Bouya-Ahmed, J. Bras, G. Nicolas, A. Gabelle, M. Didic, A. De Septenville, S. Millecamps, T. Lenglet, M. Latouche, E. Kabashi, D. Campion, D. Hannequin, J. Hardy, A. Brice, C. French and F. F.-A. Genetic Research Network on (2013). "SQSTM1 Mutations in French Patients With Frontotemporal Dementia or Frontotemporal Dementia With Amyotrophic Lateral Sclerosis." JAMA Neurol **70**(11): 1403-1410.

Le Ber, I., J. van der Zee, D. Hannequin, I. Gijselinck, D. Campion, M. Puel, A. Laquerriere, T. De Pooter, A. Camuzat, M. Van den Broeck, B. Dubois, F. Sellal, L. Lacomblez, M. Vercelletto, C. Thomas-Anterion, B. F. Michel, V. Golfier, M. Didic, F. Salachas, C. Duyckaerts, M. Cruts, P. Verpillat, C. Van Broeckhoven, A. Brice and F. F.-M. French Research Network on (2007). "Progranulin null mutations in both sporadic and familial frontotemporal dementia." Hum Mutat **28**(9): 846-855.

Leblond, C. S., A. Webber, Z. Gan-Or, F. Moore, A. Dagher, P. A. Dion and G. A. Rouleau (2016). "De novo FUS P525L mutation in Juvenile amyotrophic lateral sclerosis with dysphonia and diplopia." Neurol Genet **2**(2): e63.

Lee, B. R. and T. Kamitani (2011). "Improved immunodetection of endogenous alpha-synuclein." PLoS One **6**(8): e23939.

Lee, J. H., M. Huynh, J. L. Silhavy, S. Kim, T. Dixon-Salazar, A. Heiberg, E. Scott, V. Bafna, K. J. Hill, A. Collazo, V. Funari, C. Russ, S. B. Gabriel, G. W. Mathern and J. G. Gleeson (2012). "De novo somatic mutations in components of the PI3K-AKT3-mTOR pathway cause hemimegalencephaly." Nat Genet **44**(8): 941-945.

Lee, V. M., M. Goedert and J. Q. Trojanowski (2001). "Neurodegenerative tauopathies." Annu Rev Neurosci **24**: 1121-1159.

Leger, M., A. Quiedeville, V. Bouet, B. Haelewyn, M. Boulouard, P. Schumann-Bard and T. Freret (2013). "Object recognition test in mice." Nat Protoc **8**(12): 2531-2537.

Lek, M., K. J. Karczewski, E. V. Minikel, K. E. Samocha, E. Banks, T. Fennell, A. H. O'Donnell-Luria, J. S. Ware, A. J. Hill, B. B. Cummings, T. Tukiainen, D. P. Birnbaum, J. A. Kosmicki, L. E. Duncan, K. Estrada, F. Zhao, J. Zou, E. Pierce-Hoffman, J. Berghout, D. N. Cooper, N. Deflaux, M. DePristo, R. Do, J. Flannick, M. Fromer, L. Gauthier, J. Goldstein, N. Gupta, D. Howrigan, A. Kiezun, M. I. Kurki, A. L. Moonshine, P. Natarajan, L. Orozco, G. M. Peloso, R. Poplin, M. A. Rivas, V. Ruano-Rubio, S. A. Rose, D. M. Ruderfer, K. Shakir, P. D. Stenson, C. Stevens, B. P. Thomas, G. Tiao, M. T. Tusie-Luna, B. Weisburd, H. H. Won, D. Yu, D. M. Altshuler, D. Ardissino, M. Boehnke, J. Danesh, S. Donnelly, R. Elosua, J. C. Florez, S. B. Gabriel, G. Getz, S. J. Glatt, C. M. Hultman, S. Kathiresan, M. Laakso, S. McCarroll, M. I. McCarthy, D. McGovern, R. McPherson, B. M. Neale, A. Palotie, S. M. Purcell, D. Saleheen, J. M. Scharf, P. Sklar, P. F. Sullivan, J. Tuomilehto, M. T. Tsuang, H. C. Watkins, J. G. Wilson, M. J. Daly, D. G. MacArthur and C. Exome Aggregation (2016). "Analysis of protein-coding genetic variation in 60,706 humans." Nature **536**(7616): 285-291.

Lesage, S., C. Condroyer, N. Hecham, M. Anheim, S. Belarbi, E. Lohman, F. Viallet, P. Pollak, M. Abada, A. Durr, M. Tazir, A. Brice and G. French Parkinson Disease Genetic (2011). "Mutations in the glucocerebrosidase gene confer a risk for Parkinson disease in North Africa." Neurology **76**(3): 301-303.

Lesage, S., C. Condroyer, S. Klebe, E. Lohmann, F. Durif, P. Damier, F. Tison, M. Anheim, A. Honore, F. Viallet, A. M. Bonnet, A. M. Ouvrard-Hernandez, M.

Vidailhet, A. Durr, A. Brice and G. French Parkinson's Disease Genetics Study (2012). "EIF4G1 in familial Parkinson's disease: pathogenic mutations or rare benign variants?" Neurobiol Aging **33**(9): 2233 e2231-2233 e2235.

Lesage, S., A. Durr, M. Tazir, E. Lohmann, A. L. Leutenegger, S. Janin, P. Pollak, A. Brice and G. French Parkinson's Disease Genetics Study (2006). "LRRK2 G2019S as a cause of Parkinson's disease in North African Arabs." N Engl J Med **354**(4): 422-423.

Li, B., V. G. Krishnan, M. E. Mort, F. Xin, K. K. Kamati, D. N. Cooper, S. D. Mooney and P. Radivojac (2009). "Automated inference of molecular mechanisms of disease from amino acid substitutions." Bioinformatics **25**(21): 2744-2750.

Li, B. and S. M. Leal (2008). "Methods for detecting associations with rare variants for common diseases: application to analysis of sequence data." Am J Hum Genet **83**(3): 311-321.

Li, H. (2014). "Toward better understanding of artifacts in variant calling from high-coverage samples." Bioinformatics **30**(20): 2843-2851.

Li, H. and R. Durbin (2009). "Fast and accurate short read alignment with Burrows-Wheeler transform." Bioinformatics **25**(14): 1754-1760.

Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan, N. Homer, G. Marth, G. Abecasis, R. Durbin and S. Genome Project Data Processing (2009). "The Sequence Alignment/Map format and SAMtools." Bioinformatics **25**(16): 2078-2079.

Li, H. F. and Z. Y. Wu (2016). "Genotype-phenotype correlations of amyotrophic lateral sclerosis." Transl Neurodegener **5**: 3.

Lieber, M. R. (2008). "The mechanism of human nonhomologous DNA end joining." J Biol Chem **283**(1): 1-5.

Lill, C. M. and L. Bertram (2011). "Towards unveiling the genetics of neurodegenerative diseases." Semin Neurol **31**(5): 531-541.

Lim, J. S., W. I. Kim, H. C. Kang, S. H. Kim, A. H. Park, E. K. Park, Y. W. Cho, S. Kim, H. M. Kim, J. A. Kim, J. Kim, H. Rhee, S. G. Kang, H. D. Kim, D. Kim, D. S. Kim and J. H. Lee (2015). "Brain somatic mutations in MTOR cause focal cortical dysplasia type II leading to intractable epilepsy." Nat Med **21**(4): 395-400.

Lin, M. K. and M. J. Farrer (2014). "Genetics and genomics of Parkinson's disease." Genome Med **6**(6): 48.

Lin, M. T., I. Cantuti-Castelvetri, K. Zheng, K. E. Jackson, Y. B. Tan, T. Arzberger, A. J. Lees, R. A. Betensky, M. F. Beal and D. K. Simon (2012). "Somatic mitochondrial DNA mutations in early Parkinson and incidental Lewy body disease." Ann Neurol **71**(6): 850-854.

Lindhurst, M. J., J. C. Sapp, J. K. Teer, J. J. Johnston, E. M. Finn, K. Peters, J. Turner, J. L. Cannons, D. Bick, L. Blakemore, C. Blumhorst, K. Brockmann, P. Calder, N. Cherman, M. A. Deardorff, D. B. Everman, G. Golas, R. M. Greenstein, B. M. Kato, K. M. Keppler-Noreuil, S. A. Kuznetsov, R. T. Miyamoto, K. Newman, D. Ng, K. O'Brien, S. Rothenberg, D. J. Schwartzentruber, V. Singhal, R. Tirabosco, J. Upton, S. Wientroub, E. H. Zackai, K. Hoag, T. Whitewood-Neal, P. G. Robey, P. L. Schwartzberg, T. N. Darling, L. L. Tosi, J. C. Mullikin and L. G. Biesecker (2011). "A mosaic activating mutation in AKT1 associated with the Proteus syndrome." N Engl J Med **365**(7): 611-619.

Litvan, I., D. A. Grimes, A. E. Lang, J. Jankovic, A. McKee, M. Verny, K. Jellinger, K. R. Chaudhuri and R. K. Pearce (1999). "Clinical features differentiating patients with postmortem confirmed progressive supranuclear palsy and corticobasal degeneration." J Neurol **246 Suppl 2**: II1-5.

Liu, R., X. Gao, Y. Lu and H. Chen (2011). "Meta-analysis of the relationship between Parkinson disease and melanoma." Neurology **76**(23): 2002-2009.

Liu, Z. J., H. F. Li, G. H. Tan, Q. Q. Tao, W. Ni, X. W. Cheng, Z. Q. Xiong and Z. Y. Wu (2014). "Identify mutation in amyotrophic lateral sclerosis cases using HaloPlex target enrichment system." Neurobiol Aging **35**(12): 2881 e2811-2885.

Lleo, A., R. Blesa, R. Queralt, M. Ezquerra, J. L. Molinuevo, J. Pena-Casanova, A. Rojo and R. Oliva (2002). "Frequency of mutations in the presenilin and amyloid precursor protein genes in early-onset Alzheimer disease in Spain." Arch Neurol **59**(11): 1759-1763.

Lodato, M. A., M. B. Woodworth, S. Lee, G. D. Evrony, B. K. Mehta, A. Karger, S. Lee, T. W. Chittenden, A. M. D'Gama, X. Cai, L. J. Luquette, E. Lee, P. J. Park and C. A. Walsh (2015). "Somatic mutation in single human neurons tracks developmental and transcriptional history." Science **350**(6256): 94-98.

Lohmueller, K. E., C. L. Pearce, M. Pike, E. S. Lander and J. N. Hirschhorn (2003). "Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease." Nat Genet **33**(2): 177-182.

Lopez, L. C., C. M. Quinzii, E. Area, A. Naini, S. Rahman, M. Schuelke, L. Salviati, S. Dimauro and M. Hirano (2010). "Treatment of CoQ(10) deficient fibroblasts with ubiquinone, CoQ analogs, and vitamin C: time- and compound-dependent effects." PLoS One **5**(7): e11897.

Lubbe, S. J., V. Escott-Price, J. R. Gibbs, M. A. Nalls, J. Bras, T. R. Price, A. Nicolas, I. E. Jansen, K. Y. Mok, A. M. Pittman, J. E. Tomkins, P. A. Lewis, A. J. Noyce, S. Lesage, M. Sharma, E. R. Schiff, A. P. Levine, A. Brice, T. Gasser, J. Hardy, P. Heutink, N. W. Wood, A. B. Singleton, N. M. Williams, H. R. Morris and C. for International Parkinson's Disease Genomics (2016). "Additional Rare Variant Analysis in Parkinson's Disease Cases with and Without Known Pathogenic Mutations: Evidence for Oligogenic Inheritance." Hum Mol Genet.

Luk, K. C., D. J. Covell, V. M. Kehm, B. Zhang, I. Y. Song, M. D. Byrne, R. M. Pitkin, S. C. Decker, J. Q. Trojanowski and V. M. Lee (2016). "Molecular and Biological Compatibility with Host Alpha-Synuclein Influences Fibril Pathogenicity." Cell Rep **16**(12): 3373-3387.

Mabrouk, O. S., I. J. Dripps, S. Ramani, C. Chang, J. L. Han, K. C. Rice and E. M. Jutkiewicz (2014). "Automated touch screen device for recording complex rodent behaviors." J Neurosci Methods **233**: 129-136.

MacArthur, D. G., T. A. Manolio, D. P. Dimmock, H. L. Rehm, J. Shendure, G. R. Abecasis, D. R. Adams, R. B. Altman, S. E. Antonarakis, E. A. Ashley, J. C. Barrett, L. G. Biesecker, D. F. Conrad, G. M. Cooper, N. J. Cox, M. J. Daly, M. B. Gerstein, D. B. Goldstein, J. N. Hirschhorn, S. M. Leal, L. A. Pennacchio, J. A. Stamatoyannopoulos, S. R. Sunyaev, D. Valle, B. F. Voight, W. Winckler and C. Gunter (2014). "Guidelines for investigating causality of sequence variants in human disease." Nature **508**(7497): 469-476.

Macchi, P., A. Villa, S. Giliani, M. G. Sacco, A. Frattini, F. Porta, A. G. Ugazio, J. A. Johnston, F. Candotti, J. J. O'Shea and et al. (1995). "Mutations of Jak-3 gene in patients with autosomal severe combined immune deficiency (SCID)." Nature **377**(6544): 65-68.

MacDonald, J. R., R. Ziman, R. K. Yuen, L. Feuk and S. W. Scherer (2014). "The Database of Genomic Variants: a curated collection of structural variation in the human genome." Nucleic Acids Res **42**(Database issue): D986-992.

Mackenzie, I. R., M. Neumann, A. Baborie, D. M. Sampathu, D. Du Plessis, E. Jaros, R. H. Perry, J. Q. Trojanowski, D. M. Mann and V. M. Lee (2011). "A harmonized classification system for FTLD-TDP pathology." Acta Neuropathol **122**(1): 111-113.

Mackenzie, I. R., M. Neumann, E. H. Bigio, N. J. Cairns, I. Alafuzoff, J. Kril, G. G. Kovacs, B. Ghetti, G. Halliday, I. E. Holm, P. G. Ince, W. Kamphorst, T. Revesz, A. J. Rozemuller, S. Kumar-Singh, H. Akiyama, A. Baborie, S. Spina, D. W. Dickson, J. Q. Trojanowski and D. M. Mann (2010). "Nomenclature and nosology for neuropathologic subtypes of frontotemporal lobar degeneration: an update." Acta Neuropathol **119**(1): 1-4.

Madsen, B. E. and S. R. Browning (2009). "A groupwise association test for rare mutations using a weighted sum statistic." PLoS Genet **5**(2): e1000384.

Majounie, E., A. E. Renton, K. Mok, E. G. Dopper, A. Waite, S. Rollinson, A. Chio, G. Restagno, N. Nicolaou, J. Simon-Sanchez, J. C. van Swieten, Y. Abramzon, J. O. Johnson, M. Sendtner, R. Pamphlett, R. W. Orrell, S. Mead, K. C. Sidle, H. Houlden, J. D. Rohrer, K. E. Morrison, H. Pall, K. Talbot, O. Ansorge, A. L. S. F. T.

D. C. Chromosome, F. F. A. French research network on, I. Consortium, D. G. Hernandez, S. Arepalli, M. Sabatelli, G. Mora, M. Corbo, F. Giannini, A. Calvo, E. Englund, G. Borghero, G. L. Floris, A. M. Remes, H. Laaksovirta, L. McCluskey, J. Q. Trojanowski, V. M. Van Deerlin, G. D. Schellenberg, M. A. Nalls, V. E. Drory, C. S. Lu, T. H. Yeh, H. Ishiura, Y. Takahashi, S. Tsuji, I. Le Ber, A. Brice, C. Drepper, N. Williams, J. Kirby, P. Shaw, J. Hardy, P. J. Tienari, P. Heutink, H. R. Morris, S. Pickering-Brown and B. J. Traynor (2012). "Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study." Lancet Neurol **11**(4): 323-330.

Majounie, E., B. J. Traynor, A. Chio, G. Restagno, J. Mandrioli, M. Benatar, J. P. Taylor and A. B. Singleton (2012). "Mutational analysis of the VCP gene in Parkinson's disease." Neurobiol Aging **33**(1): 209 e201-202.

Mandellos, D., G. Limbitaki, A. Papadimitriou and D. Anastasopoulos (2005). "Cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy (CADASIL) in a Greek family." Neurol Sci **26**(4): 278-281.

Manichaikul, A., J. C. Mychaleckyj, S. S. Rich, K. Daly, M. Sale and W. M. Chen (2010). "Robust relationship inference in genome-wide association studies." Bioinformatics **26**(22): 2867-2873.

Marcus, C., E. Mena and R. M. Subramaniam (2014). "Brain PET in the diagnosis of Alzheimer's disease." Clin Nucl Med **39**(10): e413-422; quiz e423-416.

Maroteaux, L., J. T. Campanelli and R. H. Scheller (1988). "Synuclein: a neuron-specific protein localized to the nucleus and presynaptic nerve terminal." J Neurosci **8**(8): 2804-2815.

Martens, J. H. (2011). "Acute myeloid leukemia: a central role for the ETS factor ERG." Int J Biochem Cell Biol **43**(10): 1413-1416.

Maruyama, H., H. Morino, H. Ito, Y. Izumi, H. Kato, Y. Watanabe, Y. Kinoshita, M. Kamada, H. Nodera, H. Suzuki, O. Komure, S. Matsuura, K. Kobatake, N. Morimoto, K. Abe, N. Suzuki, M. Aoki, A. Kawata, T. Hirai, T. Kato, K. Ogasawara, A. Hirano, T. Takumi, H. Kusaka, K. Hagiwara, R. Kaji and H. Kawakami (2010).

"Mutations of optineurin in amyotrophic lateral sclerosis." Nature **465**(7295): 223-226.

Masellis, M., P. Momeni, W. Meschino, R. Heffner, Jr., J. Elder, C. Sato, Y. Liang, P. St George-Hyslop, J. Hardy, J. Bilbao, S. Black and E. Rogaeva (2006). "Novel splicing mutation in the progranulin gene causing familial corticobasal syndrome." Brain **129**(Pt 11): 3115-3123.

Masliah, E., W. Dumaop, D. Galasko and P. Desplats (2013). "Distinctive patterns of DNA methylation associated with Parkinson disease: identification of concordant epigenetic changes in brain and peripheral blood leukocytes." Epigenetics **8**(10): 1030-1038.

Masliah, E., E. Rockenstein, I. Veinbergs, M. Mallory, M. Hashimoto, A. Takeda, Y. Sagara, A. Sisk and L. Mucke (2000). "Dopaminergic loss and inclusion body formation in alpha-synuclein mice: implications for neurodegenerative disorders." Science **287**(5456): 1265-1269.

McKeith, I. G. (2006). "Consensus guidelines for the clinical and pathologic diagnosis of dementia with Lewy bodies (DLB): report of the Consortium on DLB International Workshop." J Alzheimers Dis **9**(3 Suppl): 417-423.

McKeith, I. G., B. F. Boeve, D. W. Dickson, G. Halliday, J. P. Taylor, D. Weintraub, D. Aarsland, J. Galvin, J. Attems, C. G. Ballard, A. Bayston, T. G. Beach, F. Blanc, N. Bohnen, L. Bonanni, J. Bras, P. Brundin, D. Burn, A. Chen-Plotkin, J. E. Duda, O. El-Agnaf, H. Feldman, T. J. Ferman, D. Ffytche, H. Fujishiro, D. Galasko, J. G. Goldman, S. N. Gomperts, N. R. Graff-Radford, L. S. Honig, A. Iranzo, K. Kantarci, D. Kaufer, W. Kukull, V. M. Y. Lee, J. B. Leverenz, S. Lewis, C. Lippa, A. Lunde, M. Masellis, E. Masliah, P. McLean, B. Mollenhauer, T. J. Montine, E. Moreno, E. Mori, M. Murray, J. T. O'Brien, S. Orimo, R. B. Postuma, S. Ramaswamy, O. A. Ross, D. P. Salmon, A. Singleton, A. Taylor, A. Thomas, P. Tiraboschi, J. B. Toledo, J. Q. Trojanowski, D. Tsuang, Z. Walker, M. Yamada and K. Kosaka (2017). "Diagnosis and management of dementia with Lewy bodies: Fourth consensus report of the DLB Consortium." Neurology **89**(1): 88-100.

McKeith, I. G., D. W. Dickson, J. Lowe, M. Emre, J. T. O'Brien, H. Feldman, J. Cummings, J. E. Duda, C. Lippa, E. K. Perry, D. Aarsland, H. Arai, C. G. Ballard, B. Boeve, D. J. Burn, D. Costa, T. Del Ser, B. Dubois, D. Galasko, S. Gauthier, C. G. Goetz, E. Gomez-Tortosa, G. Halliday, L. A. Hansen, J. Hardy, T. Iwatsubo, R. N. Kalaria, D. Kaufer, R. A. Kenny, A. Korczyn, K. Kosaka, V. M. Lee, A. Lees, I. Litvan, E. Londos, O. L. Lopez, S. Minoshima, Y. Mizuno, J. A. Molina, E. B. Mukaetova-Ladinska, F. Pasquier, R. H. Perry, J. B. Schulz, J. Q. Trojanowski, M. Yamada and D. L. B. Consortium on (2005). "Diagnosis and management of dementia with Lewy bodies: third report of the DLB Consortium." Neurology **65**(12): 1863-1872.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, K. Garimella, D. Altshuler, S. Gabriel, M. Daly and M. A. DePristo (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data." Genome Res **20**(9): 1297-1303.

McKerrell, T., N. Park, T. Moreno, C. S. Grove, H. Ponstingl, J. Stephens, G. Understanding Society Scientific, C. Crawley, J. Craig, M. A. Scott, C. Hodkinson, J. Baxter, R. Rad, D. R. Forsyth, M. A. Quail, E. Zeggini, W. Ouwehand, I. Varela and G. S. Vassiliou (2015). "Leukemia-associated somatic mutations drive distinct patterns of age-related clonal hemopoiesis." Cell Rep **10**(8): 1239-1245.

McKhann, G. M., D. S. Knopman, H. Chertkow, B. T. Hyman, C. R. Jack, Jr., C. H. Kawas, W. E. Klunk, W. J. Koroshetz, J. J. Manly, R. Mayeux, R. C. Mohs, J. C. Morris, M. N. Rossor, P. Scheltens, M. C. Carrillo, B. Thies, S. Weintraub and C. H. Phelps (2011). "The diagnosis of dementia due to Alzheimer's disease: recommendations from the National Institute on Aging-Alzheimer's Association workgroups on diagnostic guidelines for Alzheimer's disease." Alzheimers Dement **7**(3): 263-269.

McLaughlin, R. L., A. Vajda and O. Hardiman (2015). "Heritability of Amyotrophic Lateral Sclerosis: Insights From Disparate Numbers." JAMA Neurol **72**(8): 857-858.

McPherson, J. D., M. Marra, L. Hillier, R. H. Waterston, A. Chinwalla, J. Wallis, M. Sekhon, K. Wylie, E. R. Mardis, R. K. Wilson, R. Fulton, T. A. Kucaba, C. Wagner-

McPherson, W. B. Barbazuk, S. G. Gregory, S. J. Humphray, L. French, R. S. Evans, G. Bethel, A. Whittaker, J. L. Holden, O. T. McCann, A. Dunham, C. Soderlund, C. E. Scott, D. R. Bentley, G. Schuler, H. C. Chen, W. Jang, E. D. Green, J. R. Idol, V. V. Maduro, K. T. Montgomery, E. Lee, A. Miller, S. Emerling, Kucherlapati, R. Gibbs, S. Scherer, J. H. Gorrell, E. Sodergren, K. Clerc-Blankenburg, P. Tabor, S. Naylor, D. Garcia, P. J. de Jong, J. J. Catanese, N. Nowak, K. Osoegawa, S. Qin, L. Rowen, A. Madan, M. Dors, L. Hood, B. Trask, C. Friedman, H. Massa, V. G. Cheung, I. R. Kirsch, T. Reid, R. Yonescu, J. Weissenbach, T. Bruls, R. Heilig, E. Branscomb, A. Olsen, N. Doggett, J. F. Cheng, T. Hawkins, R. M. Myers, J. Shang, L. Ramirez, J. Schmutz, O. Velasquez, K. Dixon, N. E. Stone, D. R. Cox, D. Haussler, W. J. Kent, T. Furey, S. Rogic, S. Kennedy, S. Jones, A. Rosenthal, G. Wen, M. Schilhabel, G. Gloeckner, G. Nyakatura, R. Siebert, B. Schlegelberger, J. Korenberg, X. N. Chen, A. Fujiyama, M. Hattori, A. Toyoda, T. Yada, H. S. Park, Y. Sakaki, N. Shimizu, S. Asakawa, K. Kawasaki, T. Sasaki, A. Shintani, A. Shimizu, K. Shibuya, J. Kudoh, S. Minoshima, J. Ramser, P. Seranski, C. Hoff, A. Poustka, R. Reinhardt and H. Lehrach (2001). "A physical map of the human genome." Nature **409**(6822): 934-941.

Mead, S. (2006). "Prion disease genetics." Eur J Hum Genet **14**(3): 273-281.

Meeus, B., J. Theuns and C. Van Broeckhoven (2012). "The genetics of dementia with Lewy bodies: what are we missing?" Arch Neurol **69**(9): 1113-1118.

Mehta, S. G., M. Khare, R. Ramani, G. D. Watts, M. Simon, K. E. Osann, S. Donkervoort, E. Dec, A. Nalbandian, J. Platt, M. Pasquali, A. Wang, T. Mozaffar, C. D. Smith and V. E. Kimonis (2013). "Genotype-phenotype studies of VCP-associated inclusion body myopathy with Paget disease of bone and/or frontotemporal dementia." Clin Genet **83**(5): 422-431.

Mellick, G. D., G. A. Siebert, M. Funayama, D. D. Buchanan, Y. Li, Y. Imamichi, H. Yoshino, P. A. Silburn and N. Hattori (2009). "Screening PARK genes for mutations in early-onset Parkinson's disease patients from Queensland, Australia." Parkinsonism Relat Disord **15**(2): 105-109.

Merner, N. D., S. L. Girard, H. Catoire, C. V. Bourassa, V. V. Belzil, J. B. Riviere, P. Hince, A. Levert, A. Dionne-Laporte, D. Spiegelman, A. Noreau, S. Diab, A. Szuto, H. Fournier, J. Raelson, M. Belouchi, M. Panisset, P. Cossette, N. Dupre, G. Bernard, S. Chouinard, P. A. Dion and G. A. Rouleau (2012). "Exome sequencing identifies FUS mutations as a cause of essential tremor." Am J Hum Genet **91**(2): 313-319.

Michaud, J., F. Wu, M. Osato, G. M. Cottles, M. Yanagida, N. Asou, K. Shigesada, Y. Ito, K. F. Benson, W. H. Raskind, C. Rossier, S. E. Antonarakis, S. Israels, A. McNicol, H. Weiss, M. Horwitz and H. S. Scott (2002). "In vitro analyses of known and novel RUNX1/AML1 mutations in dominant familial platelet disorder with predisposition to acute myelogenous leukemia: implications for mechanisms of pathogenesis." Blood **99**(4): 1364-1372.

Milholland, B., X. Dong, L. Zhang, X. Hao, Y. Suh and J. Vijg (2017). "Differences between germline and somatic mutation rates in humans and mice." Nat Commun **8**: 15183.

Milholland, B., X. Dong, L. Zhang, X. X. Hao, Y. Suh and J. Vijg (2017). "Differences between germline and somatic mutation rates in humans and mice." Nature Communications **8**.

Minoche, A. E., J. C. Dohm and H. Himmelbauer (2011). "Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and genome analyzer systems." Genome Biol **12**(11): R112.

Mirra, S. S., A. Heyman, D. McKeel, S. M. Sumi, B. J. Crain, L. M. Brownlee, F. S. Vogel, J. P. Hughes, G. van Belle and L. Berg (1991). "The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease." Neurology **41**(4): 479-486.

Mitchell, J., P. Paul, H. J. Chen, A. Morris, M. Payling, M. Falchi, J. Habgood, S. Panoutsou, S. Winkler, V. Tisato, A. Hajitou, B. Smith, C. Vance, C. Shaw, N. D. Mazarakis and J. de Belleroche (2010). "Familial amyotrophic lateral sclerosis is

associated with a mutation in D-amino acid oxidase." Proc Natl Acad Sci U S A **107**(16): 7556-7561.

Monahan, A. J., M. Warren and P. M. Carvey (2008). "Neuroinflammation and peripheral immune infiltration in Parkinson's disease: an autoimmune hypothesis." Cell Transplant **17**(4): 363-372.

Montine, T. J., C. H. Phelps, T. G. Beach, E. H. Bigio, N. J. Cairns, D. W. Dickson, C. Duyckaerts, M. P. Frosch, E. Masliah, S. S. Mirra, P. T. Nelson, J. A. Schneider, D. R. Thal, J. Q. Trojanowski, H. V. Vinters, B. T. Hyman, A. National Institute on and A. Alzheimer's (2012). "National Institute on Aging-Alzheimer's Association guidelines for the neuropathologic assessment of Alzheimer's disease: a practical approach." Acta Neuropathol **123**(1): 1-11.

Morishima-Kawashima, M., M. Hasegawa, K. Takio, M. Suzuki, H. Yoshida, A. Watanabe, K. Titani and Y. Ihara (1995). "Hyperphosphorylation of tau in PHF." Neurobiol Aging **16**(3): 365-371; discussion 371-380.

Morris, R. G. (1990). "Toward a representational hypothesis of the role of hippocampal synaptic plasticity in spatial and other forms of learning." Cold Spring Harb Symp Quant Biol **55**: 161-173.

Mules, E. H., S. Hayflick, C. S. Miller, L. W. Reynolds and G. H. Thomas (1992). "Six novel deleterious and three neutral mutations in the gene encoding the alpha-subunit of hexosaminidase A in non-Jewish individuals." Am J Hum Genet **50**(4): 834-841.

Muller, S. K., A. Bender, C. Laub, T. Hogen, F. Schlaudraff, B. Liss, T. Klopstock and M. Elstner (2013). "Lewy body pathology is associated with mitochondrial DNA damage in Parkinson's disease." Neurobiol Aging **34**(9): 2231-2233.

Munch, C., R. Sedlmeier, T. Meyer, V. Homberg, A. D. Sperfeld, A. Kurt, J. Prudlo, G. Peraus, C. O. Hanemann, G. Stumm and A. C. Ludolph (2004). "Point mutations of the p150 subunit of dynactin (DCTN1) gene in ALS." Neurology **63**(4): 724-726.

Murayama, O., T. Tomita, N. Nihonmatsu, M. Murayama, X. Sun, T. Honda, T. Iwatsubo and A. Takashima (1999). "Enhancement of amyloid beta 42 secretion by 28 different presenilin 1 mutations of familial Alzheimer's disease." Neurosci Lett **265**(1): 61-63.

Muresan, V., C. Villegas and Z. Ladescu Muresan (2014). "Functional interaction between amyloid-beta precursor protein and peripherin neurofilaments: a shared pathway leading to Alzheimer's disease and amyotrophic lateral sclerosis?" Neurodegener Dis **13**(2-3): 122-125.

Murphy, D. D., S. M. Rueter, J. Q. Trojanowski and V. M. Lee (2000). "Synucleins are developmentally expressed, and alpha-synuclein regulates the size of the presynaptic vesicular pool in primary hippocampal neurons." J Neurosci **20**(9): 3214-3220.

Nalls, M. A., N. Pankratz, C. M. Lill, C. B. Do, D. G. Hernandez, M. Saad, A. L. DeStefano, E. Kara, J. Bras, M. Sharma, C. Schulte, M. F. Keller, S. Arepalli, C. Letson, C. Edsall, H. Stefansson, X. Liu, H. Pliner, J. H. Lee, R. Cheng, C. International Parkinson's Disease Genomics, G. I. Parkinson's Study Group Parkinson's Research: The Organized, andMe, GenePd, C. NeuroGenetics Research, G. Hussman Institute of Human, I. Ashkenazi Jewish Dataset, H. Cohorts for, E. Aging Research in Genetic, C. North American Brain Expression, C. United Kingdom Brain Expression, C. Greek Parkinson's Disease, G. Alzheimer Genetic Analysis, M. A. Ikram, J. P. Ioannidis, G. M. Hadjigeorgiou, J. C. Bis, M. Martinez, J. S. Perlmutter, A. Goate, K. Marder, B. Fiske, M. Sutherland, G. Xiromerisiou, R. H. Myers, L. N. Clark, K. Stefansson, J. A. Hardy, P. Heutink, H. Chen, N. W. Wood, H. Houlden, H. Payami, A. Brice, W. K. Scott, T. Gasser, L. Bertram, N. Eriksson, T. Foroud and A. B. Singleton (2014). "Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease." Nat Genet **46**(9): 989-993.

Nath, U., Y. Ben-Shlomo, R. G. Thomson, A. J. Lees and D. J. Burn (2003). "Clinical features and natural history of progressive supranuclear palsy: a clinical cohort study." Neurology **60**(6): 910-916.

Navarro, A., A. Boveris, M. J. Bandez, M. J. Sanchez-Pino, C. Gomez, G. Muntane and I. Ferrer (2009). "Human brain cortex: mitochondrial oxidative damage and adaptive response in Parkinson disease and in dementia with Lewy bodies." Free Radic Biol Med **46**(12): 1574-1580.

Nelson, O., H. Tu, T. Lei, M. Bentahir, B. de Strooper and I. Bezprozvanny (2007). "Familial Alzheimer disease-linked mutations specifically disrupt Ca2+ leak function of presenilin 1." J Clin Invest **117**(5): 1230-1239.

Neumann, M., D. M. Sampathu, L. K. Kwong, A. C. Truax, M. C. Micsenyi, T. T. Chou, J. Bruce, T. Schuck, M. Grossman, C. M. Clark, L. F. McCluskey, B. L. Miller, E. Masliah, I. R. Mackenzie, H. Feldman, W. Feiden, H. A. Kretzschmar, J. Q. Trojanowski and V. M. Lee (2006). "Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis." Science **314**(5796): 130-133.

Nikoloski, G., S. M. Langemeijer, R. P. Kuiper, R. Knops, M. Massop, E. R. Tonnissen, A. van der Heijden, T. N. Scheele, P. Vandenberghe, T. de Witte, B. A. van der Reijden and J. H. Jansen (2010). "Somatic mutations of the histone methyltransferase gene EZH2 in myelodysplastic syndromes." Nat Genet **42**(8): 665-667.

Nodera, H., Y. Izumi and R. Kaji (2007). "[New diagnostic criteria of ALS (Awaji criteria)]." Brain Nerve **59**(10): 1023-1029.

Nussbaum, R. L. and C. E. Ellis (2003). "Alzheimer's disease and Parkinson's disease." N Engl J Med **348**(14): 1356-1364.

Nuytemans, K., J. Theuns, M. Cruts and C. Van Broeckhoven (2010). "Genetic etiology of Parkinson disease associated with mutations in the SNCA, PARK2, PINK1, PARK7, and LRRK2 genes: a mutation update." Hum Mutat **31**(7): 763-780.

Nygaard, H. B., C. F. Lippa, D. Mehdi and J. M. Baehring (2014). "A Novel Presenilin 1 Mutation in Early-Onset Alzheimer's Disease With Prominent Frontal Features." Am J Alzheimers Dis Other Demen **29**(5): 433-435.

O'Huallachain, M., K. J. Karczewski, S. M. Weissman, A. E. Urban and M. P. Snyder (2012). "Extensive genetic variation in somatic human tissues." Proc Natl Acad Sci U S A **109**(44): 18018-18023.

Obermann, M. and M. Lyon (2015). "Financial cost of amyotrophic lateral sclerosis: a case study." Amyotroph Lateral Scler Frontotemporal Degener **16**(1-2): 54-57.

Orrell, R. W., S. L. Marklund and J. S. deBelleroche (1997). "Familial ALS is associated with mutations in all exons of SOD1: a novel mutation in exon 3 (Gly72Ser)." J Neurol Sci **153**(1): 46-49.

Ostrow, S. L., R. Barshir, J. DeGregori, E. Yeger-Lotem and R. Hershberg (2014). "Cancer evolution is associated with pervasive positive selection on globally expressed genes." PLoS Genet **10**(3): e1004239.

Padhi, A. K., H. Kumar, S. V. Vasaikar, B. Jayaram and J. Gomes (2012). "Mechanisms of loss of functions of human angiogenin variants implicated in amyotrophic lateral sclerosis." PLoS One **7**(2): e32479.

Padhi, A. K., S. V. Vasaikar, B. Jayaram and J. Gomes (2013). "Fast prediction of deleterious angiogenin mutations causing amyotrophic lateral sclerosis." FEBS Lett **587**(12): 1762-1766.

Paisan-Ruiz, C. (2009). "LRRK2 gene variation and its contribution to Parkinson disease." Hum Mutat **30**(8): 1153-1160.

Paisan-Ruiz, C., S. Jain, E. W. Evans, W. P. Gilks, J. Simon, M. van der Brug, A. Lopez de Munain, S. Aparicio, A. M. Gil, N. Khan, J. Johnson, J. R. Martinez, D. Nicholl, I. M. Carrera, A. S. Pena, R. de Silva, A. Lees, J. F. Marti-Masso, J. Perez-Tur, N. W. Wood and A. B. Singleton (2004). "Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease." Neuron **44**(4): 595-600.

Paisan-Ruiz, C., A. Li, S. A. Schneider, J. L. Holton, R. Johnson, D. Kidd, J. Chataway, K. P. Bhatia, A. J. Lees, J. Hardy, T. Revesz and H. Houlden (2012). "Widespread Lewy body and tau accumulation in childhood and adult onset

dystonia-parkinsonism cases with PLA2G6 mutations." Neurobiol Aging **33**(4): 814-823.

Pallier, P. N., C. J. Drew and A. J. Morton (2009). "The detection and measurement of locomotor deficits in a transgenic mouse model of Huntington's disease are task- and protocol-dependent: influence of non-motor factors on locomotor function." Brain Res Bull **78**(6): 347-355.

Palmer, M. S., J. A. Beck, T. A. Campbell, C. B. Humphries, P. K. Roques, N. C. Fox, R. Harvey, M. N. Rossor and J. Collinge (1999). "Pathogenic presenilin 1 mutations (P436S & I143F) in early-onset Alzheimer's disease in the UK. Mutations in brief no. 223. Online." Hum Mutat **13**(3): 256.

Palmer, M. S., A. J. Dryden, J. T. Hughes and J. Collinge (1991). "Homozygous prion protein genotype predisposes to sporadic Creutzfeldt-Jakob disease." Nature **352**(6333): 340-342.

Pangalos, C., D. Avramopoulos, J. L. Blouin, O. Raoul, M. C. deBlois, M. Prieur, A. A. Schinzel, M. Gika, D. Abazis and S. E. Antonarakis (1994). "Understanding the mechanism(s) of mosaic trisomy 21 by using DNA polymorphism analysis." Am J Hum Genet **54**(3): 473-481.

Pansuriya, T. C., R. van Eijk, P. d'Adamo, M. A. van Ruler, M. L. Kuijjer, J. Oosting, A. M. Cleton-Jansen, J. G. van Oosterwijk, S. L. Verbeke, D. Meijer, T. van Wezel, K. H. Nord, L. Sangiorgi, B. Toker, B. Liegl-Atzwanger, M. San-Julian, R. Sciot, N. Limaye, L. G. Kindblom, S. Daugaard, C. Godfraind, L. M. Boon, M. Vikkula, K. C. Kurek, K. Szuhai, P. J. French and J. V. Bovee (2011). "Somatic mosaic IDH1 and IDH2 mutations are associated with enchondroma and spindle cell hemangioma in Ollier disease and Maffucci syndrome." Nat Genet **43**(12): 1256-1261.

Papp, M. I., J. E. Kahn and P. L. Lantos (1989). "Glial cytoplasmic inclusions in the CNS of patients with multiple system atrophy (striatonigral degeneration, olivopontocerebellar atrophy and Shy-Drager syndrome)." J Neurol Sci **94**(1-3): 79-100.

Papp, M. I. and P. L. Lantos (1994). "The distribution of oligodendroglial inclusions in multiple system atrophy and its relevance to clinical symptomatology." Brain **117 (Pt 2)**: 235-243.

Paques, F. and J. E. Haber (1999). "Multiple pathways of recombination induced by double-strand breaks in Saccharomyces cerevisiae." Microbiol Mol Biol Rev **63**(2): 349-404.

Parchi, P., R. Castellani, S. Capellari, B. Ghetti, K. Young, S. G. Chen, M. Farlow, D. W. Dickson, A. A. Sima, J. Q. Trojanowski, R. B. Petersen and P. Gambetti (1996). "Molecular basis of phenotypic variability in sporadic Creutzfeldt-Jakob disease." Ann Neurol **39**(6): 767-778.

Paw, B. H., S. M. Moskowitz, N. Uhrhammer, N. Wright, M. M. Kaback and E. F. Neufeld (1990). "Juvenile GM2 gangliosidosis caused by substitution of histidine for arginine at position 499 or 504 of the alpha-subunit of beta-hexosaminidase." J Biol Chem **265**(16): 9452-9457.

Payne, B. A., I. J. Wilson, P. Yu-Wai-Man, J. Coxhead, D. Deehan, R. Horvath, R. W. Taylor, D. C. Samuels, M. Santibanez-Koref and P. F. Chinnery (2013). "Universal heteroplasmy of human mitochondrial DNA." Hum Mol Genet **22**(2): 384-390.

Pearce, J. M. (2003). "Pick's disease." J Neurol Neurosurg Psychiatry **74**(2): 169.

Peoc'h, K., E. Levavasseur, E. Delmont, A. De Simone, I. Laffont-Proust, N. Privat, Y. Chebaro, C. Chapuis, P. Bedoucha, J. P. Brandel, A. Laquerriere, J. L. Kemeny, J. J. Hauw, M. Borg, H. Rezaei, P. Derreumaux, J. L. Laplanche and S. Haik (2012). "Substitutions at residue 211 in the prion protein drive a switch between CJD and GSS syndrome, a new mechanism governing inherited neurodegenerative disorders." Hum Mol Genet **21**(26): 5417-5428.

Peoc'h, K., P. Manivet, P. Beaudry, F. Attane, G. Besson, D. Hannequin, N. Delasnerie-Laupretre and J. L. Laplanche (2000). "Identification of three novel mutations (E196K, V203I, E211Q) in the prion protein gene (PRNP) in inherited prion diseases with Creutzfeldt-Jakob disease phenotype." Hum Mutat **15**(5): 482.

Perez, F. A. and R. D. Palmiter (2005). "Parkin-deficient mice are not a robust model of parkinsonism." Proc Natl Acad Sci U S A **102**(6): 2174-2179.

Perier, C., A. Bender, E. Garcia-Arumi, M. J. Melia, J. Bove, C. Laub, T. Klopstock, M. Elstner, R. B. Mounsey, P. Teismann, T. Prolla, A. L. Andreu and M. Vila (2013). "Accumulation of mitochondrial DNA deletions within dopaminergic neurons triggers neuroprotective mechanisms." Brain **136**(Pt 8): 2369-2378.

Pfeifer, G. P. (2006). "Mutagenesis at methylated CpG sequences." Curr Top Microbiol Immunol **301**: 259-281.

Piccini, A., G. Zanusso, R. Borghi, C. Noviello, S. Monaco, R. Russo, G. Damonte, A. Armirotti, M. Gelati, R. Giordano, P. Zambenedetti, C. Russo, B. Ghetti and M. Tabaton (2007). "Association of a presenilin 1 S170F mutation with a novel Alzheimer disease molecular phenotype." Arch Neurol **64**(5): 738-745.

Pick, A. (1906). "Über einen weiteren Symptomenkomplex im Rahman der

Dementia senilis, bedingt durch umschriebene starkere Hirnatrophie

(gemischte Apraxie)." Monat Psych Neurol (19): 97-108.

Pihlstrom, L., A. Rengmark, K. A. Bjornara and M. Toft (2014). "Effective variant detection by targeted deep sequencing of DNA pools: an example from Parkinson's disease." Ann Hum Genet **78**(3): 243-252.

Pinz, K. G., S. Shibutani and D. F. Bogenhagen (1995). "Action of mitochondrial DNA polymerase gamma at sites of base loss or oxidative damage." J Biol Chem **270**(16): 9202-9206.

Plagnol, V., J. Curtis, M. Epstein, K. Y. Mok, E. Stebbings, S. Grigoriadou, N. W. Wood, S. Hambleton, S. O. Burns, A. J. Thrasher, D. Kumararatne, R. Doffinger and S. Nejentsev (2012). "A robust model for read count data in exome sequencing experiments and implications for copy number variant calling." Bioinformatics **28**(21): 2747-2754.

Poduri, A., G. D. Evrony, X. Cai, P. C. Elhosary, R. Beroukhim, M. K. Lehtinen, L. B. Hills, E. L. Heinzen, A. Hill, R. S. Hill, B. J. Barry, B. F. Bourgeois, J. J. Riviello,

A. J. Barkovich, P. M. Black, K. L. Ligon and C. A. Walsh (2012). "Somatic activation of AKT3 causes hemispheric developmental brain malformations." Neuron **74**(1): 41-48.

Poduri, A., G. D. Evrony, X. Cai and C. A. Walsh (2013). "Somatic mutation, genomic variation, and neurological disease." Science **341**(6141): 1237758.

Poehler, A. M., W. Xiang, P. Spitzer, V. E. May, H. Meixner, E. Rockenstein, O. Chutna, T. F. Outeiro, J. Winkler, E. Masliah and J. Klucken (2014). "Autophagy modulates SNCA/alpha-synuclein release, thereby generating a hostile microenvironment." Autophagy **10**(12): 2171-2192.

Polymeropoulos, M. H., C. Lavedan, E. Leroy, S. E. Ide, A. Dehejia, A. Dutra, B. Pike, H. Root, J. Rubenstein, R. Boyer, E. S. Stenroos, S. Chandrasekharappa, A. Athanassiadou, T. Papapetropoulos, W. G. Johnson, A. M. Lazzarini, R. C. Duvoisin, G. Di Iorio, L. I. Golbe and R. L. Nussbaum (1997). "Mutation in the alpha-synuclein gene identified in families with Parkinson's disease." Science **276**(5321): 2045-2047.

Postuma, R. B., D. Berg, M. Stern, W. Poewe, C. W. Olanow, W. Oertel, J. Obeso, K. Marek, I. Litvan, A. E. Lang, G. Halliday, C. G. Goetz, T. Gasser, B. Dubois, P. Chan, B. R. Bloem, C. H. Adler and G. Deuschl (2015). "MDS clinical diagnostic criteria for Parkinson's disease." Mov Disord **30**(12): 1591-1601.

Pringle, C. E., A. J. Hudson, D. G. Munoz, J. A. Kiernan, W. F. Brown and G. C. Ebers (1992). "Primary lateral sclerosis. Clinical features, neuropathology and diagnostic criteria." Brain **115 (Pt 2)**: 495-520.

Prinz, M. and J. Priller (2017). "The role of peripheral immune cells in the CNS in steady state and disease." Nat Neurosci **20**(2): 136-144.

Proukakis, C., C. G. Dudzik, T. Brier, D. S. MacKay, J. M. Cooper, G. L. Millhauser, H. Houlden and A. H. Schapira (2013). "A novel alpha-synuclein missense mutation in Parkinson disease." Neurology **80**(11): 1062-1064.

Prudencio, M., P. J. Hart, D. R. Borchelt and P. M. Andersen (2009). "Variation in aggregation propensities among ALS-associated variants of SOD1: correlation to human disease." Hum Mol Genet **18**(17): 3217-3226.

Prusiner, S. B. (2001). "Shattuck lecture--neurodegenerative diseases and prions." N Engl J Med **344**(20): 1516-1526.

Prut, L. and C. Belzung (2003). "The open field as a paradigm to measure the effects of drugs on anxiety-like behaviors: a review." Eur J Pharmacol **463**(1-3): 3-33.

Pulst, S. M. (1999). "Genetic linkage analysis." Arch Neurol **56**(6): 667-672.

Puoti, G., A. Bizzi, G. Forloni, J. G. Safar, F. Tagliavini and P. Gambetti (2012). "Sporadic human prion diseases: molecular insights and diagnosis." Lancet Neurol **11**(7): 618-628.

Purcell, S., B. Neale, K. Todd-Brown, L. Thomas, M. A. Ferreira, D. Bender, J. Maller, P. Sklar, P. I. de Bakker, M. J. Daly and P. C. Sham (2007). "PLINK: a tool set for whole-genome association and population-based linkage analyses." Am J Hum Genet **81**(3): 559-575.

Qin, H., J. A. Buckley, X. Li, Y. Liu, T. H. Fox, 3rd, G. P. Meares, H. Yu, Z. Yan, A. S. Harms, Y. Li, D. G. Standaert and E. N. Benveniste (2016). "Inhibition of the JAK/STAT Pathway Protects Against alpha-Synuclein-Induced Neuroinflammation and Dopaminergic Neurodegeneration." J Neurosci **36**(18): 5144-5159.

Raben, N., K. Nagaraju, E. Lee and P. Plotz (2000). "Modulation of disease severity in mice with targeted disruption of the acid alpha-glucosidase gene." Neuromuscul Disord **10**(4-5): 283-291.

Rabinovici, G. D., P. N. Wang, J. Levin, L. Cook, M. Pravdin, J. Davis, S. J. DeArmond, N. M. Barbaro, J. Martindale, B. L. Miller and M. D. Geschwind (2006). "First symptom in sporadic Creutzfeldt-Jakob disease." Neurology **66**(2): 286-287.

Rademakers, R., H. Stewart, M. Dejesus-Hernandez, C. Krieger, N. Graff-Radford, M. Fabros, H. Briemberg, N. Cashman, A. Eisen and I. R. Mackenzie

(2010). "Fus gene mutations in familial and sporadic amyotrophic lateral sclerosis." Muscle Nerve **42**(2): 170-176.

Rahbari, R., A. Wuster, S. J. Lindsay, R. J. Hardwick, L. B. Alexandrov, S. Al Turki, A. Dominiczak, A. Morris, D. Porteous, B. Smith, M. R. Stratton, U. K. Consortium and M. E. Hurles (2016). "Timing, rates and spectra of human germline mutation." Nat Genet **48**(2): 126-133.

Rajakulendran, S., R. D. Pitceathly, J. W. Taanman, H. Costello, M. G. Sweeney, C. E. Woodward, Z. Jaunmuktane, J. L. Holton, T. S. Jacques, B. N. Harding, C. Fratter, M. G. Hanna and S. Rahman (2016). "A Clinical, Neuropathological and Genetic Study of Homozygous A467T POLG-Related Mitochondrial Disease." PLoS One **11**(1): e0145500.

Ramakrishnan, V. (2002). "Ribosome structure and the mechanism of translation." Cell **108**(4): 557-572.

Ramanan, V. K. and A. J. Saykin (2013). "Pathways to neurodegeneration: mechanistic insights from GWAS in Alzheimer's disease, Parkinson's disease, and related disorders." Am J Neurodegener Dis **2**(3): 145-175.

Rana, H. Q., M. Balwani, L. Bier and R. N. Alcalay (2013). "Age-specific Parkinson disease risk in GBA mutation carriers: information for genetic counseling." Genet Med **15**(2): 146-149.

Rands, C. M., S. Meader, C. P. Ponting and G. Lunter (2014). "8.2% of the Human genome is constrained: variation in rates of turnover across functional element classes in the human lineage." PLoS Genet **10**(7): e1004525.

Rascovsky, K., J. R. Hodges, D. Knopman, M. F. Mendez, J. H. Kramer, J. Neuhaus, J. C. van Swieten, H. Seelaar, E. G. Dopper, C. U. Onyike, A. E. Hillis, K. A. Josephs, B. F. Boeve, A. Kertesz, W. W. Seeley, K. P. Rankin, J. K. Johnson, M. L. Gorno-Tempini, H. Rosen, C. E. Prioleau-Latham, A. Lee, C. M. Kipps, P. Lillo, O. Piguet, J. D. Rohrer, M. N. Rossor, J. D. Warren, N. C. Fox, D. Galasko, D. P. Salmon, S. E. Black, M. Mesulam, S. Weintraub, B. C. Dickerson, J. Diehl-Schmid, F. Pasquier, V. Deramecourt, F. Lebert, Y. Pijnenburg, T. W.

Chow, F. Manes, J. Grafman, S. F. Cappa, M. Freedman, M. Grossman and B. L. Miller (2011). "Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia." Brain **134**(Pt 9): 2456-2477.

Rasmussen, K. D., G. Jia, J. V. Johansen, M. T. Pedersen, N. Rapin, F. O. Bagger, B. T. Porse, O. A. Bernard, J. Christensen and K. Helin (2015). "Loss of TET2 in hematopoietic cells leads to DNA hypermethylation of active enhancers and induction of leukemogenesis." Genes Dev **29**(9): 910-922.

Raux, G., R. Gantier, C. Martin, Y. Pothin, A. Brice, T. Frebourg and D. Campion (2000). "A novel presenilin 1 missense mutation (L153V) segregating with early-onset autosomal dominant Alzheimer's disease." Hum Mutat **16**(1): 95.

Rayaprolu, S., B. Mullen, M. Baker, T. Lynch, E. Finger, W. W. Seeley, K. J. Hatanpaa, C. Lomen-Hoerth, A. Kertesz, E. H. Bigio, C. Lippa, K. A. Josephs, D. S. Knopman, C. L. White, 3rd, R. Caselli, I. R. Mackenzie, B. L. Miller, M. Boczarska-Jedynak, G. Opala, A. Krygowska-Wajs, M. Barcikowska, S. G. Younkin, R. C. Petersen, N. Ertekin-Taner, R. J. Uitti, J. F. Meschia, K. B. Boylan, B. F. Boeve, N. R. Graff-Radford, Z. K. Wszolek, D. W. Dickson, R. Rademakers and O. A. Ross (2013). "TREM2 in neurodegeneration: evidence for association of the p.R47H variant with frontotemporal dementia and Parkinson's disease." Mol Neurodegener **8**: 19.

Reeve, A. K., T. K. Park, E. Jaros, G. R. Campbell, N. Z. Lax, P. D. Hepplewhite, K. J. Krishnan, J. L. Elson, C. M. Morris, I. G. McKeith and D. M. Turnbull (2012). "Relationship between mitochondria and alpha-synuclein: a study of single substantia nigra neurons." Arch Neurol **69**(3): 385-393.

Reger, M. L., D. A. Hovda and C. C. Giza (2009). "Ontogeny of Rat Recognition Memory measured by the novel object recognition task." Dev Psychobiol **51**(8): 672-678.

Renton, A. E., A. Chio and B. J. Traynor (2014). "State of play in amyotrophic lateral sclerosis genetics." Nat Neurosci **17**(1): 17-23.

Renton, A. E., E. Majounie, A. Waite, J. Simon-Sanchez, S. Rollinson, J. R. Gibbs, J. C. Schymick, H. Laaksovirta, J. C. van Swieten, L. Myllykangas, H. Kalimo, A. Paetau, Y. Abramzon, A. M. Remes, A. Kaganovich, S. W. Scholz, J. Duckworth, J. Ding, D. W. Harmer, D. G. Hernandez, J. O. Johnson, K. Mok, M. Ryten, D. Trabzuni, R. J. Guerreiro, R. W. Orrell, J. Neal, A. Murray, J. Pearson, I. E. Jansen, D. Sondervan, H. Seelaar, D. Blake, K. Young, N. Halliwell, J. B. Callister, G. Toulson, A. Richardson, A. Gerhard, J. Snowden, D. Mann, D. Neary, M. A. Nalls, T. Peuralinna, L. Jansson, V. M. Isoviita, A. L. Kaivorinne, M. Holtta-Vuori, E. Ikonen, R. Sulkava, M. Benatar, J. Wuu, A. Chio, G. Restagno, G. Borghero, M. Sabatelli, I. Consortium, D. Heckerman, E. Rogaeva, L. Zinman, J. D. Rothstein, M. Sendtner, C. Drepper, E. E. Eichler, C. Alkan, Z. Abdullaev, S. D. Pack, A. Dutra, E. Pak, J. Hardy, A. Singleton, N. M. Williams, P. Heutink, S. Pickering-Brown, H. R. Morris, P. J. Tienari and B. J. Traynor (2011). "A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD." Neuron **72**(2): 257-268.

Reu, P., A. Khosravi, S. Bernard, J. E. Mold, M. Salehpour, K. Alkass, S. Perl, J. Tisdale, G. Possnert, H. Druid and J. Frisen (2017). "The Lifespan and Turnover of Microglia in the Human Brain." Cell Rep **20**(4): 779-784.

Reya, T., S. J. Morrison, M. F. Clarke and I. L. Weissman (2001). "Stem cells, cancer, and cancer stem cells." Nature **414**(6859): 105-111.

Richards, S., N. Aziz, S. Bale, D. Bick, S. Das, J. Gastier-Foster, W. W. Grody, M. Hegde, E. Lyon, E. Spector, K. Voelkerding, H. L. Rehm and A. L. Q. A. Committee (2015). "Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology." Genet Med **17**(5): 405-424.

Rieger, R., F. Edenhofer, C. I. Lasmezas and S. Weiss (1997). "The human 37-kDa laminin receptor precursor interacts with the prion protein in eukaryotic cells." Nat Med **3**(12): 1383-1388.

Ringholz, G. M., S. H. Appel, M. Bradshaw, N. A. Cooke, D. M. Mosnik and P. E. Schulz (2005). "Prevalence and patterns of cognitive impairment in sporadic ALS." Neurology **65**(4): 586-590.

Robinson, J. T., H. Thorvaldsdottir, W. Winckler, M. Guttman, E. S. Lander, G. Getz and J. P. Mesirov (2011). "Integrative genomics viewer." Nat Biotechnol **29**(1): 24-26.

Robinson, M., B. Y. Lee and F. T. Hane (2017). "Recent Progress in Alzheimer's Disease Research, Part 2: Genetics and Epidemiology." J Alzheimers Dis **57**(2): 317-330.

Rockenstein, E., M. Mallory, M. Hashimoto, D. Song, C. W. Shults, I. Lang and E. Masliah (2002). "Differential neuropathological alterations in transgenic mice expressing alpha-synuclein from the platelet-derived growth factor and Thy-1 promoters." J Neurosci Res **68**(5): 568-578.

Rohrer, J. D., G. R. Ridgway, M. Modat, S. Ourselin, S. Mead, N. C. Fox, M. N. Rossor and J. D. Warren (2010). "Distinct profiles of brain atrophy in frontotemporal lobar degeneration caused by progranulin and tau mutations." Neuroimage **53**(3): 1070-1076.

Rosen, D. R., T. Siddique, D. Patterson, D. A. Figlewicz, P. Sapp, A. Hentati, D. Donaldson, J. Goto, J. P. O'Regan, H. X. Deng and et al. (1993). "Mutations in Cu/Zn superoxide dismutase gene are associated with familial amyotrophic lateral sclerosis." Nature **362**(6415): 59-62.

Ross, C. A. (2002). "Polyglutamine pathogenesis: emergence of unifying mechanisms for Huntington's disease and related disorders." Neuron **35**(5): 819-822.

Ross, C. A. and M. A. Poirier (2004). "Protein aggregation and neurodegenerative disease." Nat Med **10 Suppl**: S10-17.

Ross, J. M., J. B. Stewart, E. Hagstrom, S. Brene, A. Mourier, G. Coppotelli, C. Freyer, M. Lagouge, B. J. Hoffer, L. Olson and N. G. Larsson (2013). "Germline

mitochondrial DNA mutations aggravate ageing and can impair brain development." Nature **501**(7467): 412-415.

Ross, O. A., A. I. Soto-Ortolaza, M. G. Heckman, J. O. Aasly, N. Abahuni, G. Annesi, J. A. Bacon, S. Bardien, M. Bozi, A. Brice, L. Brighina, C. Van Broeckhoven, J. Carr, M. C. Chartier-Harlin, E. Dardiotis, D. W. Dickson, N. N. Diehl, A. Elbaz, C. Ferrarese, A. Ferraris, B. Fiske, J. M. Gibson, R. Gibson, G. M. Hadjigeorgiou, N. Hattori, J. P. Ioannidis, B. Jasinska-Myga, B. S. Jeon, Y. J. Kim, C. Klein, R. Kruger, E. Kyratzi, S. Lesage, C. H. Lin, T. Lynch, D. M. Maraganore, G. D. Mellick, E. Mutez, C. Nilsson, G. Opala, S. S. Park, A. Puschmann, A. Quattrone, M. Sharma, P. A. Silburn, Y. H. Sohn, L. Stefanis, V. Tadic, J. Theuns, H. Tomiyama, R. J. Uitti, E. M. Valente, S. van de Loo, D. K. Vassilatis, C. Vilarino-Guell, L. R. White, K. Wirdefeldt, Z. K. Wszolek, R. M. Wu, M. J. Farrer and C. Genetic Epidemiology Of Parkinson's Disease (2011). "Association of LRRK2 exonic variants with susceptibility to Parkinson's disease: a case-control study." Lancet Neurol **10**(10): 898-908.

Ross, O. A., A. I. Soto-Ortolaza, M. G. Heckman, C. Verbeeck, D. J. Serie, S. Rayaprolu, S. S. Rich, M. A. Nalls, A. Singleton, R. Guerreiro, E. Kinsella, Z. K. Wszolek, T. G. Brott, R. D. Brown, Jr., B. B. Worrall and J. F. Meschia (2013). "NOTCH3 variants and risk of ischemic stroke." PLoS One **8**(9): e75035.

Rovelet-Lecrux, A., D. Hannequin, G. Raux, N. Le Meur, A. Laquerriere, A. Vital, C. Dumanchin, S. Feuillette, A. Brice, M. Vercelletto, F. Dubas, T. Frebourg and D. Campion (2006). "APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy." Nat Genet **38**(1): 24-26.

Ruderfer, D. M., T. Hamamsy, M. Lek, K. J. Karczewski, D. Kavanagh, K. E. Samocha, C. Exome Aggregation, M. J. Daly, D. G. MacArthur, M. Fromer and S. M. Purcell (2016). "Patterns of genic intolerance of rare copy number variation in 59,898 human exomes." Nat Genet **48**(10): 1107-1111.

Sachidanandam, R., D. Weissman, S. C. Schmidt, J. M. Kakol, L. D. Stein, G. Marth, S. Sherry, J. C. Mullikin, B. J. Mortimore, D. L. Willey, S. E. Hunt, C. G. Cole, P. C. Coggill, C. M. Rice, Z. Ning, J. Rogers, D. R. Bentley, P. Y. Kwok, E.

R. Mardis, R. T. Yeh, B. Schultz, L. Cook, R. Davenport, M. Dante, L. Fulton, L. Hillier, R. H. Waterston, J. D. McPherson, B. Gilman, S. Schaffner, W. J. Van Etten, D. Reich, J. Higgins, M. J. Daly, B. Blumenstiel, J. Baldwin, N. Stange-Thomann, M. C. Zody, L. Linton, E. S. Lander and D. Altshuler (2001). "A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms." Nature **409**(6822): 928-933.

Samocha, K. E., E. B. Robinson, S. J. Sanders, C. Stevens, A. Sabo, L. M. McGrath, J. A. Kosmicki, K. Rehnstrom, S. Mallick, A. Kirby, D. P. Wall, D. G. MacArthur, S. B. Gabriel, M. DePristo, S. M. Purcell, A. Palotie, E. Boerwinkle, J. D. Buxbaum, E. H. Cook, Jr., R. A. Gibbs, G. D. Schellenberg, J. S. Sutcliffe, B. Devlin, K. Roeder, B. M. Neale and M. J. Daly (2014). "A framework for the interpretation of de novo mutation in human disease." Nat Genet **46**(9): 944-950.

Sankar, T. S., B. D. Wastuwidyaningtyas, Y. Dong, S. A. Lewis and J. D. Wang (2016). "The nature of mutations induced by replication-transcription collisions." Nature **535**(7610): 178-181.

Sassi, C., R. Guerreiro, R. Gibbs, J. Ding, M. K. Lupton, C. Troakes, S. Al-Sarraj, M. Niblock, J. M. Gallo, J. Adnan, R. Killick, K. S. Brown, C. Medway, J. Lord, J. Turton, J. Bras, U. K. C. Alzheimer's Research, K. Morgan, J. F. Powell, A. Singleton and J. Hardy (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**(12): 2881 e2881-2886.

Saudou, F., S. Finkbeiner, D. Devys and M. E. Greenberg (1998). "Huntingtin acts in the nucleus to induce apoptosis but death does not correlate with the formation of intranuclear inclusions." Cell **95**(1): 55-66.

Schapira, A. H. (1994). "Evidence for mitochondrial dysfunction in Parkinson's disease--a critical appraisal." Mov Disord **9**(2): 125-138.

Scheicher, R., A. Hoelbl-Kovacic, F. Bellutti, A. S. Tigan, M. Prchal-Murphy, G. Heller, C. Schneckenleithner, M. Salazar-Roa, S. Zochbauer-Muller, J. Zuber, M. Malumbres, K. Kollmann and V. Sexl (2015). "CDK6 as a key regulator of hematopoietic and leukemic stem cell activation." Blood **125**(1): 90-101.

Schenck, J. F. (1995). "Imaging of brain iron by magnetic resonance: T2 relaxation at different field strengths." J Neurol Sci **134 Suppl**: 10-18.

Schoenwolf, G., S. Bleyl, P. Brauer and P. Francis-West (2015). Larsen's Human Embryology, Churchill Livingstone.

Schols, L., P. Bauer, T. Schmidt, T. Schulte and O. Riess (2004). "Autosomal dominant cerebellar ataxias: clinical features, genetics, and pathogenesis." Lancet Neurol **3**(5): 291-304.

Scholz, J., Y. Niibori, P. W. Frankland and J. P. Lerch (2015). "Rotarod training in mice is associated with changes in brain structure observable with multimodal MRI." Neuroimage **107**: 182-189.

Schulte, E. C., B. Mollenhauer, A. Zimprich, B. Bereznai, P. Lichtner, D. Haubenberger, W. Pirker, T. Brucke, M. J. Molnar, A. Peters, C. Gieger, C. Trenkwalder and J. Winkelmann (2012). "Variants in eukaryotic translation initiation factor 4G1 in sporadic Parkinson's disease." Neurogenetics **13**(3): 281-285.

Schulz, J. B., T. Klockgether, D. Petersen, M. Jauch, W. Muller-Schauenburg, S. Spieker, K. Voigt and J. Dichgans (1994). "Multiple system atrophy: natural history, MRI morphology, and dopamine receptor imaging with 123IBZM-SPECT." J Neurol Neurosurg Psychiatry **57**(9): 1047-1056.

Schulz-Schaeffer, W. J. (2010). "The synaptic pathology of alpha-synuclein aggregation in dementia with Lewy bodies, Parkinson's disease and Parkinson's disease dementia." Acta Neuropathol **120**(2): 131-143.

Schwab, C., T. Arai, M. Hasegawa, S. Yu and P. L. McGeer (2008). "Colocalization of transactivation-responsive DNA-binding protein 43 and huntingtin in inclusions of Huntington disease." J Neuropathol Exp Neurol **67**(12): 1159-1165.

Sciacco, M., E. Bonilla, E. A. Schon, S. DiMauro and C. T. Moraes (1994). "Distribution of wild-type and common deletion forms of mtDNA in normal and

respiration-deficient muscle fibers from patients with mitochondrial myopathy." Hum Mol Genet **3**(1): 13-19.

Scott, S., J. E. Kranz, J. Cole, J. M. Lincecum, K. Thompson, N. Kelly, A. Bostrom, J. Theodoss, B. M. Al-Nakhala, F. G. Vieira, J. Ramasubbu and J. A. Heywood (2008). "Design, power, and interpretation of studies in the standard murine model of ALS." Amyotroph Lateral Scler **9**(1): 4-15.

Sebat, J., B. Lakshmi, J. Troge, J. Alexander, J. Young, P. Lundin, S. Maner, H. Massa, M. Walker, M. Chi, N. Navin, R. Lucito, J. Healy, J. Hicks, K. Ye, A. Reiner, T. C. Gilliam, B. Trask, N. Patterson, A. Zetterberg and M. Wigler (2004). "Large-scale copy number polymorphism in the human genome." Science **305**(5683): 525-528.

Sedelis, M., R. K. Schwarting and J. P. Huston (2001). "Behavioral phenotyping of the MPTP mouse model of Parkinson's disease." Behav Brain Res **125**(1-2): 109-125.

Seelen, M., A. E. Visser, D. J. Overste, H. J. Kim, A. Palud, T. H. Wong, J. C. van Swieten, P. Scheltens, N. C. Voermans, F. Baas, J. M. de Jong, A. J. van der Kooi, M. de Visser, J. H. Veldink, J. P. Taylor, M. A. Van Es and L. H. van den Berg (2014). "No mutations in hnRNPA1 and hnRNPA2B1 in Dutch patients with amyotrophic lateral sclerosis, frontotemporal dementia, and inclusion body myopathy." Neurobiol Aging **35**(8): 1956 e1959-1956 e1911.

Seto-Salvia, N., J. Pagonabarraga, H. Houlden, B. Pascual-Sedano, O. Dols-Icardo, A. Tucci, C. Paisan-Ruiz, A. Campolongo, S. Anton-Aguirre, I. Martin, L. Munoz, E. Bufill, L. Vilageliu, D. Grinberg, M. Cozar, R. Blesa, A. Lleo, J. Hardy, J. Kulisevsky and J. Clarimon (2012). "Glucocerebrosidase mutations confer a greater risk of dementia during Parkinson's disease course." Mov Disord **27**(3): 393-399.

Sherrington, R., S. Froelich, S. Sorbi, D. Campion, H. Chi, E. A. Rogaeva, G. Levesque, E. I. Rogaev, C. Lin, Y. Liang, M. Ikeda, L. Mar, A. Brice, Y. Agid, M. E. Percy, F. Clerget-Darpoux, S. Piacentini, G. Marcon, B. Nacmias, L. Amaducci, T. Frebourg, L. Lannfelt, J. M. Rommens and P. H. St George-Hyslop (1996).

"Alzheimer's disease associated with mutations in presenilin 2 is rare and variably penetrant." Hum Mol Genet **5**(7): 985-988.

Sherrington, R., E. I. Rogaev, Y. Liang, E. A. Rogaeva, G. Levesque, M. Ikeda, H. Chi, C. Lin, G. Li, K. Holman, T. Tsuda, L. Mar, J. F. Foncin, A. C. Bruni, M. P. Montesi, S. Sorbi, I. Rainero, L. Pinessi, L. Nee, I. Chumakov, D. Pollen, A. Brookes, P. Sanseau, R. J. Polinsky, W. Wasco, H. A. Da Silva, J. L. Haines, M. A. Perkicak-Vance, R. E. Tanzi, A. D. Roses, P. E. Fraser, J. M. Rommens and P. H. St George-Hyslop (1995). "Cloning of a gene bearing missense mutations in early-onset familial Alzheimer's disease." Nature **375**(6534): 754-760.

Shimura, H., N. Hattori, S. Kubo, Y. Mizuno, S. Asakawa, S. Minoshima, N. Shimizu, K. Iwai, T. Chiba, K. Tanaka and T. Suzuki (2000). "Familial Parkinson disease gene product, parkin, is a ubiquitin-protein ligase." Nat Genet **25**(3): 302-305.

Shiotsuki, H., K. Yoshimi, Y. Shimo, M. Funayama, Y. Takamatsu, K. Ikeda, R. Takahashi, S. Kitazawa and N. Hattori (2010). "A rotarod test for evaluation of motor skill learning." J Neurosci Methods **189**(2): 180-185.

Shirley, M. D., H. Tang, C. J. Gallione, J. D. Baugher, L. P. Frelin, B. Cohen, P. E. North, D. A. Marchuk, A. M. Comi and J. Pevsner (2013). "Sturge-Weber syndrome and port-wine stains caused by somatic mutation in GNAQ." N Engl J Med **368**(21): 1971-1979.

Shoffner, J. M., R. L. Watts, J. L. Juncos, A. Torroni and D. C. Wallace (1991). "Mitochondrial oxidative phosphorylation defects in Parkinson's disease." Ann Neurol **30**(3): 332-339.

Sicca, F., A. Kelemen, P. Genton, S. Das, D. Mei, F. Moro, W. B. Dobyns and R. Guerrini (2003). "Mosaic mutations of the LIS1 gene cause subcortical band heterotopia." Neurology **61**(8): 1042-1046.

Silva, A. (1997). "Mutant mice and neuroscience: recommendations concerning genetic background. Banbury Conference on genetic background in mice." Neuron **19**(4): 755-759.

Sim, N. L., P. Kumar, J. Hu, S. Henikoff, G. Schneider and P. C. Ng (2012). "SIFT web server: predicting effects of amino acid substitutions on proteins." Nucleic Acids Res **40**(Web Server issue): W452-457.

Simon, D. K., M. T. Lin, L. Zheng, G. J. Liu, C. H. Ahn, L. M. Kim, W. M. Mauck, F. Twu, M. F. Beal and D. R. Johns (2004). "Somatic mitochondrial DNA mutations in cortex and substantia nigra in aging and Parkinson's disease." Neurobiol Aging **25**(1): 71-81.

Simon-Sanchez, J., E. G. Dopper, P. E. Cohn-Hokke, R. K. Hukema, N. Nicolaou, H. Seelaar, J. R. de Graaf, I. de Koning, N. M. van Schoor, D. J. Deeg, M. Smits, J. Raaphorst, L. H. van den Berg, H. J. Schelhaas, C. E. De Die-Smulders, D. Majoor-Krakauer, A. J. Rozemuller, R. Willemsen, Y. A. Pijnenburg, P. Heutink and J. C. van Swieten (2012). "The clinical and pathological phenotype of C9ORF72 hexanucleotide repeat expansions." Brain **135**(Pt 3): 723-735.

Simsek, D., A. Furda, Y. Gao, J. Artus, E. Brunet, A. K. Hadjantonakis, B. Van Houten, S. Shuman, P. J. McKinnon and M. Jasin (2011). "Crucial role for DNA ligase III in mitochondria but not in Xrcc1-dependent repair." Nature **471**(7337): 245-248.

Singleton, A. and K. Gwinn-Hardy (2004). "Parkinson's disease and dementia with Lewy bodies: a difference in dose?" Lancet **364**(9440): 1105-1107.

Singleton, A. and J. Hardy (2016). "The Evolution of Genetics: Alzheimer's and Parkinson's Diseases." Neuron **90**(6): 1154-1163.

Singleton, A. B., M. Farrer, J. Johnson, A. Singleton, S. Hague, J. Kachergus, M. Hulihan, T. Peuralinna, A. Dutra, R. Nussbaum, S. Lincoln, A. Crawley, M. Hanson, D. Maraganore, C. Adler, M. R. Cookson, M. Muenter, M. Baptista, D. Miller, J. Blancato, J. Hardy and K. Gwinn-Hardy (2003). "alpha-Synuclein locus triplication causes Parkinson's disease." Science **302**(5646): 841.

Smith, M. L., J. D. Cavenagh, T. A. Lister and J. Fitzgibbon (2004). "Mutation of CEBPA in familial acute myeloid leukemia." N Engl J Med **351**(23): 2403-2407.

Snider, B. J., J. Norton, M. A. Coats, S. Chakraverty, C. E. Hou, R. Jervis, C. L. Lendon, A. M. Goate, D. W. McKeel, Jr. and J. C. Morris (2005). "Novel presenilin 1 mutation (S170F) causing Alzheimer disease with Lewy bodies in the third decade of life." Arch Neurol **62**(12): 1821-1830.

Sodeyama, N., T. Iwata, K. Ishikawa, H. Mizusawa, M. Yamada, Y. Itoh, E. Otomo, M. Matsushita and Y. Komatsuzaki (2001). "Very early onset Alzheimer's disease with spastic paraparesis associated with a novel presenilin 1 mutation (Phe237Ile)." J Neurol Neurosurg Psychiatry **71**(4): 556-557.

Sofic, E., P. Riederer, W. Killian and A. Rett (1987). "Reduced concentrations of ascorbic acid and glutathione in a single case of Rett syndrome: a postmortem brain study." Brain Dev **9**(5): 529-531.

Spillantini, M. G., J. R. Murrell, M. Goedert, M. R. Farlow, A. Klug and B. Ghetti (1998). "Mutation in the tau gene in familial multiple system tauopathy with presenile dementia." Proc Natl Acad Sci U S A **95**(13): 7737-7741.

Spillantini, M. G., M. L. Schmidt, V. M. Lee, J. Q. Trojanowski, R. Jakes and M. Goedert (1997). "Alpha-synuclein in Lewy bodies." Nature **388**(6645): 839-840.

Spudich, S., J. A. Mastrianni, M. Wrensch, R. Gabizon, Z. Meiner, I. Kahana, H. Rosenmann, E. Kahana and S. B. Prusiner (1995). "Complete penetrance of Creutzfeldt-Jakob disease in Libyan Jews carrying the E200K mutation in the prion protein gene." Mol Med **1**(6): 607-613.

Squire, L. R. (1992). "Memory and the hippocampus: a synthesis from findings with rats, monkeys, and humans." Psychol Rev **99**(2): 195-231.

Sreedharan, J., I. P. Blair, V. B. Tripathi, X. Hu, C. Vance, B. Rogelj, S. Ackerley, J. C. Durnall, K. L. Williams, E. Buratti, F. Baralle, J. de Belleroche, J. D. Mitchell, P. N. Leigh, A. Al-Chalabi, C. C. Miller, G. Nicholson and C. E. Shaw (2008). "TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis." Science **319**(5870): 1668-1672.

Sriram, S. R., X. Li, H. S. Ko, K. K. Chung, E. Wong, K. L. Lim, V. L. Dawson and T. M. Dawson (2005). "Familial-associated mutations differentially disrupt the

solubility, localization, binding and ubiquitination properties of parkin." Hum Mol Genet **14**(17): 2571-2586.

Statistics, O. f. N. (2017). "Overview of the UK population: July 2017."

Stefanis, L. (2012). "alpha-Synuclein in Parkinson's disease." Cold Spring Harb Perspect Med **2**(2): a009399.

Stein, H., H. D. Foss, H. Durkop, T. Marafioti, G. Delsol, K. Pulford, S. Pileri and B. Falini (2000). "CD30(+) anaplastic large cell lymphoma: a review of its histopathologic, genetic, and clinical features." Blood **96**(12): 3681-3695.

Stenson, P. D., M. Mort, E. V. Ball, K. Shaw, A. Phillips and D. N. Cooper (2014). "The Human Gene Mutation Database: building a comprehensive mutation repository for clinical and molecular genetics, diagnostic testing and personalized genomic medicine." Hum Genet **133**(1): 1-9.

Stevens, T., G. Livingston, G. Kitchen, M. Manela, Z. Walker and C. Katona (2002). "Islington study of dementia subtypes in the community." Br J Psychiatry **180**: 270-276.

Stewart, J. B., C. Freyer, J. L. Elson, A. Wredenberg, Z. Cansu, A. Trifunovic and N. G. Larsson (2008). "Strong purifying selection in transmission of mammalian mitochondrial DNA." PLoS Biol **6**(1): e10.

Stockmann, M., M. Meyer-Ohlendorf, K. Achberger, S. Putz, M. Demestre, H. Yin, C. Hendrich, L. Linta, J. Heinrich, C. Brunner, C. Proepper, G. F. Kuh, B. Baumann, T. Langer, B. Schwalenstocker, K. E. Braunstein, C. von Arnim, S. Schneuwly, T. Meyer, P. C. Wong, T. M. Boeckers, A. C. Ludolph and S. Liebau (2013). "The dynactin p150 subunit: cell biology studies of sequence changes found in ALS/MND and Parkinsonian syndromes." J Neural Transm **120**(5): 785-798.

Taipa, R., J. Pinho and M. Melo-Pires (2012). "Clinico-pathological correlations of the most common neurodegenerative dementias." Front Neurol **3**: 68.

Tanaka, A., L. T. Hoang, Y. Nishi, S. Maniwa, M. Oka and T. Yamano (2003). "Different attenuated phenotypes of GM2 gangliosidosis variant B in Japanese patients with HEXA mutations at codon 499, and five novel mutations responsible for infantile acute form." J Hum Genet **48**(11): 571-574.

Tapia, L., A. Milnerwood, A. Guo, F. Mills, E. Yoshida, C. Vasuta, I. R. Mackenzie, L. Raymond, M. Cynader, W. Jia and S. X. Bamji (2011). "Progranulin deficiency decreases gross neural connectivity but enhances transmission at individual synapses." J Neurosci **31**(31): 11126-11132.

Tatem, K. S., J. L. Quinn, A. Phadke, Q. Yu, H. Gordish-Dressman and K. Nagaraju (2014). "Behavioral and locomotor measurements using an open field activity monitoring system for skeletal muscle diseases." J Vis Exp(91): 51785.

Taylor, J. P., J. Hardy and K. H. Fischbeck (2002). "Toxic proteins in neurodegenerative disease." Science **296**(5575): 1991-1995.

Taylor, R. W., M. J. Barron, G. M. Borthwick, A. Gospel, P. F. Chinnery, D. C. Samuels, G. A. Taylor, S. M. Plusa, S. J. Needham, L. C. Greaves, T. B. Kirkwood and D. M. Turnbull (2003). "Mitochondrial DNA mutations in human colonic crypt stem cells." J Clin Invest **112**(9): 1351-1360.

Temperley, R., R. Richter, S. Dennerlein, R. N. Lightowlers and Z. M. Chrzanowska-Lightowlers (2010). "Hungry codons promote frameshifting in human mitochondrial ribosomes." Science **327**(5963): 301.

Thal, D. R., U. Rub, M. Orantes and H. Braak (2002). "Phases of A beta-deposition in the human brain and its relevance for the development of AD." Neurology **58**(12): 1791-1800.

Theuns, J., J. Del-Favero, B. Dermaut, C. M. van Duijn, H. Backhovens, M. V. Van den Broeck, S. Serneels, E. Corsmit, C. V. Van Broeckhoven and M. Cruts (2000). "Genetic variability in the regulatory region of presenilin 1 associated with risk for Alzheimer's disease and variable expression." Hum Mol Genet **9**(3): 325-331.

Thiyagarajan, N., R. Ferguson, V. Subramanian and K. R. Acharya (2012). "Structural and molecular insights into the mechanism of action of human angiogenin-ALS variants in neurons." Nat Commun **3**: 1121.

Thomas, B. and M. F. Beal (2007). "Parkinson's disease." Hum Mol Genet **16 Spec No. 2**: R183-194.

Thorvaldsdottir, H., J. T. Robinson and J. P. Mesirov (2013). "Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration." Brief Bioinform **14**(2): 178-192.

Ticozzi, N., A. L. LeClerc, P. J. Keagle, J. D. Glass, A. M. Wills, M. van Blitterswijk, D. A. Bosco, I. Rodriguez-Leyva, C. Gellera, A. Ratti, F. Taroni, D. McKenna-Yasek, P. C. Sapp, V. Silani, C. E. Furlong, R. H. Brown, Jr. and J. E. Landers (2010). "Paraoxonase gene mutations in amyotrophic lateral sclerosis." Ann Neurol **68**(1): 102-107.

Tomasetti, C., B. Vogelstein and G. Parmigiani (2013). "Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation." Proc Natl Acad Sci U S A **110**(6): 1999-2004.

Trifunovic, A., A. Wredenberg, M. Falkenberg, J. N. Spelbrink, A. T. Rovio, C. E. Bruder, Y. M. Bohlooly, S. Gidlof, A. Oldfors, R. Wibom, J. Tornell, H. T. Jacobs and N. G. Larsson (2004). "Premature ageing in mice expressing defective mitochondrial DNA polymerase." Nature **429**(6990): 417-423.

Trounce, I., E. Byrne and S. Marzuki (1989). "Decline in skeletal muscle mitochondrial respiratory chain function: possible factor in ageing." Lancet **1**(8639): 637-639.

Tsika, E., L. Glauser, R. Moser, A. Fiser, G. Daniel, U. M. Sheerin, A. Lees, J. C. Troncoso, P. A. Lewis, R. Bandopadhyay, B. L. Schneider and D. J. Moore (2014). "Parkinson's disease-linked mutations in VPS35 induce dopaminergic neurodegeneration." Hum Mol Genet **23**(17): 4621-4638.

Tsuboi, Y., H. Uchikado and D. W. Dickson (2007). "Neuropathology of Parkinson's disease dementia and dementia with Lewy bodies with reference to striatal pathology." Parkinsonism Relat Disord **13 Suppl 3**: S221-224.

Tsuji, S. (2010). "Genetics of neurodegenerative diseases: insights from high-throughput resequencing." Hum Mol Genet **19**(R1): R65-70.

Turner, M. R., A. Al-Chalabi, A. Chio, O. Hardiman, M. C. Kiernan, J. D. Rohrer, J. Rowe, W. Seeley and K. Talbot (2017). "Genetic screening in sporadic ALS and FTD." J Neurol Neurosurg Psychiatry.

Turner, T. N., Q. Yi, N. Krumm, J. Huddleston, K. Hoekzema, F. S. HA, A. L. Doebley, R. A. Bernier, D. A. Nickerson and E. E. Eichler (2017). "denovo-db: a compendium of human de novo variants." Nucleic Acids Res **45**(D1): D804-D811.

Twelves, D., K. S. Perkins and C. Counsell (2003). "Systematic review of incidence studies of Parkinson's disease." Mov Disord **18**(1): 19-31.

Tzoulis, C., G. T. Tran, T. Schwarzlmuller, K. Specht, K. Haugarvoll, N. Balafkan, P. K. Lilleng, H. Miletic, M. Biermann and L. A. Bindoff (2013). "Severe nigrostriatal degeneration without clinical parkinsonism in patients with polymerase gamma mutations." Brain **136**(Pt 8): 2393-2404.

Uddin, M., K. Tammimies, G. Pellecchia, B. Alipanahi, P. Hu, Z. Wang, D. Pinto, L. Lau, T. Nalpathamkalam, C. R. Marshall, B. J. Blencowe, B. J. Frey, D. Merico, R. K. Yuen and S. W. Scherer (2014). "Brain-expressed exons under purifying selection are enriched for de novo mutations in autism spectrum disorder." Nat Genet **46**(7): 742-747.

Uhlen, M., L. Fagerberg, B. M. Hallstrom, C. Lindskog, P. Oksvold, A. Mardinoglu, A. Sivertsson, C. Kampf, E. Sjostedt, A. Asplund, I. Olsson, K. Edlund, E. Lundberg, S. Navani, C. A. Szigyarto, J. Odeberg, D. Djureinovic, J. O. Takanen, S. Hober, T. Alm, P. H. Edqvist, H. Berling, H. Tegel, J. Mulder, J. Rockberg, P. Nilsson, J. M. Schwenk, M. Hamsten, K. von Feilitzen, M. Forsberg, L. Persson, F. Johansson, M. Zwahlen, G. von Heijne, J. Nielsen and F. Ponten (2015).

"Proteomics. Tissue-based map of the human proteome." Science **347**(6220): 1260419.

Ulrich, J. D. and D. M. Holtzman (2016). "TREM2 Function in Alzheimer's Disease and Neurodegeneration." ACS Chem Neurosci **7**(4): 420-427.

Unit, E. C. (2017). "Creutzfeldt-Jakob disease in the UK." from *http://www.cjd.ed.ac.uk/sites/default/files/figs2.pdf*.

Vadlamudi, L., L. M. Dibbens, K. M. Lawrence, X. Iona, J. M. McMahon, W. Murrell, A. Mackay-Sim, I. E. Scheffer and S. F. Berkovic (2010). "Timing of de novo mutagenesis--a twin study of sodium-channel mutations." N Engl J Med **363**(14): 1335-1340.

Valente, E. M., P. M. Abou-Sleiman, V. Caputo, M. M. Muqit, K. Harvey, S. Gispert, Z. Ali, D. Del Turco, A. R. Bentivoglio, D. G. Healy, A. Albanese, R. Nussbaum, R. Gonzalez-Maldonado, T. Deller, S. Salvi, P. Cortelli, W. P. Gilks, D. S. Latchman, R. J. Harvey, B. Dallapiccola, G. Auburger and N. W. Wood (2004). "Hereditary early-onset Parkinson's disease caused by mutations in PINK1." Science **304**(5674): 1158-1160.

van Blitterswijk, M., M. A. van Es, E. A. Hennekam, D. Dooijes, W. van Rheenen, J. Medic, P. R. Bourque, H. J. Schelhaas, A. J. van der Kooi, M. de Visser, P. I. de Bakker, J. H. Veldink and L. H. van den Berg (2012). "Evidence for an oligogenic basis of amyotrophic lateral sclerosis." Hum Mol Genet **21**(17): 3776-3784.

van Blitterswijk, M., L. Vlam, M. A. van Es, W. L. van der Pol, E. A. Hennekam, D. Dooijes, H. J. Schelhaas, A. J. van der Kooi, M. de Visser, J. H. Veldink and L. H. van den Berg (2012). "Genetic overlap between apparently sporadic motor neuron diseases." PLoS One **7**(11): e48983.

van de Warrenburg, B. P., M. Lammens, C. B. Lucking, P. Denefle, P. Wesseling, J. Booij, P. Praamstra, N. Quinn, A. Brice and M. W. Horstink (2001). "Clinical and pathologic abnormalities in a family with parkinsonism and parkin gene mutations." Neurology **56**(4): 555-557.

van der Giezen, M. and J. Tovar (2005). "Degenerate mitochondria." EMBO Rep **6**(6): 525-530.

van der Putten, H., K. H. Wiederhold, A. Probst, S. Barbieri, C. Mistl, S. Danner, S. Kauffmann, K. Hofele, W. P. Spooren, M. A. Ruegg, S. Lin, P. Caroni, B. Sommer, M. Tolnay and G. Bilbe (2000). "Neuropathology in mice expressing human alpha-synuclein." J Neurosci **20**(16): 6021-6029.

van der Zee, J., T. Van Langenhove, G. G. Kovacs, L. Dillen, W. Deschamps, S. Engelborghs, R. Matej, M. Vandenbulcke, A. Sieben, B. Dermaut, K. Smets, P. Van Damme, C. Merlin, A. Laureys, M. Van Den Broeck, M. Mattheijssens, K. Peeters, L. Benussi, G. Binetti, R. Ghidoni, B. Borroni, A. Padovani, S. Archetti, P. Pastor, C. Razquin, S. Ortega-Cubero, I. Hernandez, M. Boada, A. Ruiz, A. de Mendonca, G. Miltenberger-Miltenyi, F. S. do Couto, S. Sorbi, B. Nacmias, S. Bagnoli, C. Graff, H. H. Chiang, H. Thonberg, R. Perneczky, J. Diehl-Schmid, P. Alexopoulos, G. B. Frisoni, C. Bonvicini, M. Synofzik, W. Maetzler, J. M. vom Hagen, L. Schols, T. B. Haack, T. M. Strom, H. Prokisch, O. Dols-Icardo, J. Clarimon, A. Lleo, I. Santana, M. R. Almeida, B. Santiago, M. T. Heneka, F. Jessen, A. Ramirez, R. Sanchez-Valle, A. Llado, E. Gelpi, S. Sarafov, I. Tournev, A. Jordanova, E. Parobkova, G. M. Fabrizi, S. Testi, E. Salmon, T. Strobel, P. Santens, W. Robberecht, P. De Jonghe, J. J. Martin, P. Cras, R. Vandenberghe, P. P. De Deyn, M. Cruts, K. Sleegers and C. Van Broeckhoven (2014). "Rare mutations in SQSTM1 modify susceptibility to frontotemporal lobar degeneration." Acta Neuropathol **128**(3): 397-410.

van Es, M. A., F. P. Diekstra, J. H. Veldink, F. Baas, P. R. Bourque, H. J. Schelhaas, E. Strengman, E. A. Hennekam, D. Lindhout, R. A. Ophoff and L. H. van den Berg (2009). "A case of ALS-FTD in a large FALS pedigree with a K17I ANG mutation." Neurology **72**(3): 287-288.

Vance, C., B. Rogelj, T. Hortobagyi, K. J. De Vos, A. L. Nishimura, J. Sreedharan, X. Hu, B. Smith, D. Ruddy, P. Wright, J. Ganesalingam, K. L. Williams, V. Tripathi, S. Al-Saraj, A. Al-Chalabi, P. N. Leigh, I. P. Blair, G. Nicholson, J. de Belleroche, J. M. Gallo, C. C. Miller and C. E. Shaw (2009). "Mutations in FUS, an RNA

processing protein, cause familial amyotrophic lateral sclerosis type 6." Science **323**(5918): 1208-1211.

Vandamme, T. F. (2014). "Use of rodents as models of human diseases." J Pharm Bioallied Sci **6**(1): 2-9.

Vann Jones, S., O'Brien, JT (2014). "The prevalence and incidence of dementia with Lewy bodies: a systematic review of population and clinical studies." Psychological Medicine **44**(4): 673-683.

Venter, J. C., M. D. Adams, E. W. Myers, P. W. Li, R. J. Mural, G. G. Sutton, H. O. Smith, M. Yandell, C. A. Evans, R. A. Holt, J. D. Gocayne, P. Amanatides, R. M. Ballew, D. H. Huson, J. R. Wortman, Q. Zhang, C. D. Kodira, X. H. Zheng, L. Chen, M. Skupski, G. Subramanian, P. D. Thomas, J. Zhang, G. L. Gabor Miklos, C. Nelson, S. Broder, A. G. Clark, J. Nadeau, V. A. McKusick, N. Zinder, A. J. Levine, R. J. Roberts, M. Simon, C. Slayman, M. Hunkapiller, R. Bolanos, A. Delcher, I. Dew, D. Fasulo, M. Flanigan, L. Florea, A. Halpern, S. Hannenhalli, S. Kravitz, S. Levy, C. Mobarry, K. Reinert, K. Remington, J. Abu-Threideh, E. Beasley, K. Biddick, V. Bonazzi, R. Brandon, M. Cargill, I. Chandramouliswaran, R. Charlab, K. Chaturvedi, Z. Deng, V. Di Francesco, P. Dunn, K. Eilbeck, C. Evangelista, A. E. Gabrielian, W. Gan, W. Ge, F. Gong, Z. Gu, P. Guan, T. J. Heiman, M. E. Higgins, R. R. Ji, Z. Ke, K. A. Ketchum, Z. Lai, Y. Lei, Z. Li, J. Li, Y. Liang, X. Lin, F. Lu, G. V. Merkulov, N. Milshina, H. M. Moore, A. K. Naik, V. A. Narayan, B. Neelam, D. Nusskern, D. B. Rusch, S. Salzberg, W. Shao, B. Shue, J. Sun, Z. Wang, A. Wang, X. Wang, J. Wang, M. Wei, R. Wides, C. Xiao, C. Yan, A. Yao, J. Ye, M. Zhan, W. Zhang, H. Zhang, Q. Zhao, L. Zheng, F. Zhong, W. Zhong, S. Zhu, S. Zhao, D. Gilbert, S. Baumhueter, G. Spier, C. Carter, A. Cravchik, T. Woodage, F. Ali, H. An, A. Awe, D. Baldwin, H. Baden, M. Barnstead, I. Barrow, K. Beeson, D. Busam, A. Carver, A. Center, M. L. Cheng, L. Curry, S. Danaher, L. Davenport, R. Desilets, S. Dietz, K. Dodson, L. Doup, S. Ferriera, N. Garg, A. Gluecksmann, B. Hart, J. Haynes, C. Haynes, C. Heiner, S. Hladun, D. Hostin, J. Houck, T. Howland, C. Ibegwam, J. Johnson, F. Kalush, L. Kline, S. Koduru, A. Love, F. Mann, D. May, S. McCawley, T. McIntosh, I. McMullen, M. Moy, L. Moy, B. Murphy, K. Nelson, C. Pfannkoch, E. Pratts, V. Puri, H. Qureshi, M. Reardon, R. Rodriguez, Y. H. Rogers, D. Romblad, B. Ruhfel, R. Scott, C.

Sitter, M. Smallwood, E. Stewart, R. Strong, E. Suh, R. Thomas, N. N. Tint, S. Tse, C. Vech, G. Wang, J. Wetter, S. Williams, M. Williams, S. Windsor, E. Winn-Deen, K. Wolfe, J. Zaveri, K. Zaveri, J. F. Abril, R. Guigo, M. J. Campbell, K. V. Sjolander, B. Karlak, A. Kejariwal, H. Mi, B. Lazareva, T. Hatton, A. Narechania, K. Diemer, A. Muruganujan, N. Guo, S. Sato, V. Bafna, S. Istrail, R. Lippert, R. Schwartz, B. Walenz, S. Yooseph, D. Allen, A. Basu, J. Baxendale, L. Blick, M. Caminha, J. Carnes-Stine, P. Caulk, Y. H. Chiang, M. Coyne, C. Dahlke, A. Mays, M. Dombroski, M. Donnelly, D. Ely, S. Esparham, C. Fosler, H. Gire, S. Glanowski, K. Glasser, A. Glodek, M. Gorokhov, K. Graham, B. Gropman, M. Harris, J. Heil, S. Henderson, J. Hoover, D. Jennings, C. Jordan, J. Jordan, J. Kasha, L. Kagan, C. Kraft, A. Levitsky, M. Lewis, X. Liu, J. Lopez, D. Ma, W. Majoros, J. McDaniel, S. Murphy, M. Newman, T. Nguyen, N. Nguyen, M. Nodell, S. Pan, J. Peck, M. Peterson, W. Rowe, R. Sanders, J. Scott, M. Simpson, T. Smith, A. Sprague, T. Stockwell, R. Turner, E. Venter, M. Wang, M. Wen, D. Wu, M. Wu, A. Xia, A. Zandieh and X. Zhu (2001). "The sequence of the human genome." Science **291**(5507): 1304-1351.

Wahlsten, D., S. F. Cooper and J. C. Crabbe (2005). "Different rankings of inbred mouse strains on the Morris maze and a refined 4-arm water escape task." Behav Brain Res **165**(1): 36-51.

Wahlsten, D., P. Metten, T. J. Phillips, S. L. Boehm, 2nd, S. Burkhart-Kasch, J. Dorow, S. Doerksen, C. Downing, J. Fogarty, K. Rodd-Henricks, R. Hen, C. S. McKinnon, C. M. Merrill, C. Nolte, M. Schalomon, J. P. Schlumbohm, J. R. Sibert, C. D. Wenger, B. C. Dudek and J. C. Crabbe (2003). "Different data from different labs: lessons from studies of gene-environment interaction." J Neurobiol **54**(1): 283-311.

Walf, A. A. and C. A. Frye (2007). "The use of the elevated plus maze as an assay of anxiety-related behavior in rodents." Nat Protoc **2**(2): 322-328.

Walker, E. S., M. Martinez, A. L. Brunkan and A. Goate (2005). "Presenilin 2 familial Alzheimer's disease mutations result in partial loss of function and dramatic changes in Abeta 42/40 ratios." J Neurochem **92**(2): 294-301.

Walsh, R. C., RA (1976). "The open-field test: A critical Review." Psychological Bulletin **83**: 482-504.

Wang, K., M. Li and H. Hakonarson (2010). "ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data." Nucleic Acids Res **38**(16): e164.

Warburton, E. C. and M. W. Brown (2015). "Neural circuitry for rat recognition memory." Behav Brain Res **285**: 131-139.

Wei, W., M. J. Keogh, I. Wilson, J. Coxhead, S. Ryan, S. Rollinson, H. Griffin, M. Kurzawa-Akanbi, M. Santibanez-Koref, K. Talbot, M. R. Turner, C. A. McKenzie, C. Troakes, J. Attems, C. Smith, S. Al Sarraj, C. M. Morris, O. Ansorge, S. Pickering-Brown, J. W. Ironside and P. F. Chinnery (2017). "Mitochondrial DNA point mutations and relative copy number in 1363 disease and control human brains." Acta Neuropathol Commun **5**(1): 13.

Weisgraber, K. H., T. L. Innerarity and R. W. Mahley (1982). "Abnormal lipoprotein receptor-binding activity of the human E apoprotein due to cysteine-arginine interchange at a single site." J Biol Chem **257**(5): 2518-2521.

Wellcome Trust Case Control, C. (2007). "Genome-wide association study of 14,000 cases of seven common diseases and 3,000 shared controls." Nature **447**(7145): 661-678.

Wexler, N. S., L. Collett, A. R. Wexler, M. D. Rawlins, S. J. Tabrizi, I. Douglas, L. Smeeth and S. J. Evans (2016). "Incidence of adult Huntington's disease in the UK: a UK-based primary care study and a systematic review." BMJ Open **6**(2): e009070.

Wiesner, R. J., J. C. Ruegg and I. Morano (1992). "Counting target molecules by exponential polymerase chain reaction: copy number of mitochondrial DNA in rat tissues." Biochem Biophys Res Commun **183**(2): 553-559.

Will, R. G. (2003). "Acquired prion disease: iatrogenic CJD, variant CJD, kuru." Br Med Bull **66**: 255-265.

Wingo, T. S., J. J. Lah, A. I. Levey and D. J. Cutler (2012). "Autosomal recessive causes likely in early-onset Alzheimer disease." Arch Neurol **69**(1): 59-64.

Wolfe, M. S., W. Xia, B. L. Ostaszewski, T. S. Diehl, W. T. Kimberly and D. J. Selkoe (1999). "Two transmembrane aspartates in presenilin-1 required for presenilin endoproteolysis and gamma-secretase activity." Nature **398**(6727): 513-517.

Wong, T. N., G. Ramsingh, A. L. Young, C. A. Miller, W. Touma, J. S. Welch, T. L. Lamprecht, D. Shen, J. Hundal, R. S. Fulton, S. Heath, J. D. Baty, J. M. Klco, L. Ding, E. R. Mardis, P. Westervelt, J. F. DiPersio, M. J. Walter, T. A. Graubert, T. J. Ley, T. E. Druley, D. C. Link and R. K. Wilson (2015). "Role of TP53 mutations in the origin and evolution of therapy-related acute myeloid leukaemia." Nature **518**(7540): 552-555.

Wu, C., F. Ivars, P. Anderson, R. Hallmann, D. Vestweber, P. Nilsson, H. Robenek, K. Tryggvason, J. Song, E. Korpos, K. Loser, S. Beissert, E. Georges-Labouesse and L. M. Sorokin (2009). "Endothelial basement membrane laminin alpha5 selectively inhibits T lymphocyte extravasation into the brain." Nat Med **15**(5): 519-527.

Wu, C. H., C. Fallini, N. Ticozzi, P. J. Keagle, P. C. Sapp, K. Piotrowska, P. Lowe, M. Koppers, D. McKenna-Yasek, D. M. Baron, J. E. Kost, P. Gonzalez-Perez, A. D. Fox, J. Adams, F. Taroni, C. Tiloca, A. L. Leclerc, S. C. Chafe, D. Mangroo, M. J. Moore, J. A. Zitzewitz, Z. S. Xu, L. H. van den Berg, J. D. Glass, G. Siciliano, E. T. Cirulli, D. B. Goldstein, F. Salachas, V. Meininger, W. Rossoll, A. Ratti, C. Gellera, D. A. Bosco, G. J. Bassell, V. Silani, V. E. Drory, R. H. Brown, Jr. and J. E. Landers (2012). "Mutations in the profilin 1 gene cause familial amyotrophic lateral sclerosis." Nature **488**(7412): 499-503.

Wu, D., W. Yu, H. Kishikawa, R. D. Folkerth, A. J. Iafrate, Y. Shen, W. Xin, K. Sims and G. F. Hu (2007). "Angiogenin loss-of-function mutations in amyotrophic lateral sclerosis." Ann Neurol **62**(6): 609-617.

Wu, M. C., S. Lee, T. Cai, Y. Li, M. Boehnke and X. Lin (2011). "Rare-variant association testing for sequencing data with the sequence kernel association test." Am J Hum Genet **89**(1): 82-93.

Wu, Y. Y., I. H. Cheng, C. C. Lee, M. J. Chiu, M. J. Lee, T. F. Chen and J. L. Hsu (2011). "Clinical phenotype of G206D mutation in the presenilin 1 gene in pathologically confirmed familial Alzheimer's disease." J Alzheimers Dis **25**(1): 145-150.

Xue, Y., Y. Chen, Q. Ayub, N. Huang, E. V. Ball, M. Mort, A. D. Phillips, K. Shaw, P. D. Stenson, D. N. Cooper, C. Tyler-Smith and C. Genomes Project (2012). "Deleterious- and disease-allele prevalence in healthy individuals: insights from current predictions, mutation databases, and population-scale resequencing." Am J Hum Genet **91**(6): 1022-1032.

Yadav, V. K., J. DeGregori and S. De (2016). "The landscape of somatic mutations in protein coding genes in apparently benign human tissues carries signatures of relaxed purifying selection." Nucleic Acids Res **44**(5): 2075-2084.

Yamada, M., S. Tsuji and H. Takahashi (2000). "Pathology of CAG repeat diseases." Neuropathology **20**(4): 319-325.

Yan, J., H. X. Deng, N. Siddique, F. Fecto, W. Chen, Y. Yang, E. Liu, S. Donkervoort, J. G. Zheng, Y. Shi, K. B. Ahmeti, B. Brooks, W. K. Engel and T. Siddique (2010). "Frameshift and novel mutations in FUS in familial amyotrophic lateral sclerosis and ALS/dementia." Neurology **75**(9): 807-814.

Yang, L., R. Rau and M. A. Goodell (2015). "DNMT3A in haematological malignancies." Nat Rev Cancer **15**(3): 152-165.

Ye, J., G. Coulouris, I. Zaretskaya, I. Cutcutache, S. Rozen and T. L. Madden (2012). "Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction." BMC Bioinformatics **13**: 134.

Yokochi, M. (1997). "Familial juvenile parkinsonism." Eur Neurol **38 Suppl 1**: 29-33.

Yu, C. E., T. D. Bird, L. M. Bekris, T. J. Montine, J. B. Leverenz, E. Steinbart, N. M. Galloway, H. Feldman, R. Woltjer, C. A. Miller, E. M. Wood, M. Grossman, L. McCluskey, C. M. Clark, M. Neumann, A. Danek, D. R. Galasko, S. E. Arnold, A. Chen-Plotkin, A. Karydas, B. L. Miller, J. Q. Trojanowski, V. M. Lee, G. D. Schellenberg and V. M. Van Deerlin (2010). "The spectrum of mutations in progranulin: a collaborative study screening 545 cases of neurodegeneration." Arch Neurol **67**(2): 161-170.

Yuen, R. K., D. Merico, H. Cao, G. Pellecchia, B. Alipanahi, B. Thiruvahindrapuram, X. Tong, Y. Sun, D. Cao, T. Zhang, X. Wu, X. Jin, Z. Zhou, X. Liu, T. Nalpathamkalam, S. Walker, J. L. Howe, Z. Wang, J. R. MacDonald, A. Chan, L. D'Abate, E. Deneault, M. T. Siu, K. Tammimies, M. Uddin, M. Zarrei, M. Wang, Y. Li, J. Wang, J. Wang, H. Yang, M. Bookman, J. Bingham, S. S. Gross, D. Loy, M. Pletcher, C. R. Marshall, E. Anagnostou, L. Zwaigenbaum, R. Weksberg, B. A. Fernandez, W. Roberts, P. Szatmari, D. Glazer, B. J. Frey, R. H. Ring, X. Xu and S. W. Scherer (2016). "Genome-wide characteristics of de novo mutations in autism." NPJ Genom Med **1**: 160271-1602710.

Zeidler, M., G. Stewart, S. N. Cousens, K. Estibeiro and R. G. Will (1997). "Codon 129 genotype and new variant CJD." Lancet **350**(9078): 668.

Zhang, F., W. Gu, M. E. Hurles and J. R. Lupski (2009). "Copy number variation in human health, disease, and evolution." Annu Rev Genomics Hum Genet **10**: 451-481.

Zhang, F. and J. R. Lupski (2015). "Non-coding genetic variants in human disease." Hum Mol Genet **24**(R1): R102-110.

Zhang, M. Y., J. E. Churpek, S. B. Keel, T. Walsh, M. K. Lee, K. R. Loeb, S. Gulsuner, C. C. Pritchard, M. Sanchez-Bonilla, J. J. Delrow, R. S. Basom, M. Forouhar, B. Gyurkocza, B. S. Schwartz, B. Neistadt, R. Marquez, C. J. Mariani, S. A. Coats, I. Hofmann, R. C. Lindsley, D. A. Williams, J. L. Abkowitz, M. S. Horwitz, M. C. King, L. A. Godley and A. Shimamura (2015). "Germline ETV6 mutations in familial thrombocytopenia and hematologic malignancy." Nat Genet **47**(2): 180-185.

Zhang, W., S. W. Han, D. W. McKeel, A. Goate and J. Y. Wu (1998). "Interaction of presenilins with the filamin family of actin-binding proteins." J Neurosci **18**(3): 914-922.

Zhang, X., R. Joehanes, B. H. Chen, T. Huan, S. Ying, P. J. Munson, A. D. Johnson, D. Levy and C. J. O'Donnell (2015). "Identification of common genetic variants controlling transcript isoform variation in human whole blood." Nat Genet **47**(4): 345-352.

Zimprich, A., S. Biskup, P. Leitner, P. Lichtner, M. Farrer, S. Lincoln, J. Kachergus, M. Hulihan, R. J. Uitti, D. B. Calne, A. J. Stoessl, R. F. Pfeiffer, N. Patenge, I. C. Carbajal, P. Vieregge, F. Asmus, B. Muller-Myhsok, D. W. Dickson, T. Meitinger, T. M. Strom, Z. K. Wszolek and T. Gasser (2004). "Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology." Neuron **44**(4): 601-607.