GENETIC ANALYSIS OF FRONTOTEMPORAL DEMENTIA AND PROGRESSIVE SUPRANUCLEAR PALSY

Thesis submitted in fulfilment of the degree of Doctor of Philosophy

Department of Molecular Neuroscience and Reta Lila Weston Institute

Institute of Neurology

University College London

University of London

December 2013

RAFFAELE FERRARI

Declaration

I, Raffaele Ferrari, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that it has been indicated.

Abstract

Genome-wide association study (GWAS) is an effective method for mapping genetic variants underlying common and complex diseases.

This thesis describes the investigation of the disorders, frontotemporal dementia (FTD) and progressive supranuclear palsy (PSP). FTD affects the frontal/temporal lobes and presents behavioural changes (bvFTD), cognitive decline or language dysfunction (primary progressive aphasia [PPA]), whilst PSP affects predominantly the brain stem resulting in loss of balance, falls, and parkinsonian signatures. Beside recent advancements in the understanding of the clinical and neuropathological characteristics as well as the genetics associated with these diseases, their underlying pathogenic mechanisms are still unclear.

The main aims of the work in this thesis were to perform a complete GWAS on clinical FTD and a follow up targeted study of new loci identified in the recent PSP-GWAS in a cohort of ~100 pathologically confirmed PSP cases.

By completing these projects the main expected outcomes were to: **1** – Identify novel risk loci associated with FTD, and; **2** – Identify SNPs and haplotypes associated with increased risk of developing PSP and genetic variants possibly affecting expression and/or splicing of transcription elements relevant to PSP.

Dedication

In memory of my grandparents Caterina Damonte, Raffaele Ferrari (Sr), Helene Roehrig and Heinrich Luksza, and of my aunt and uncle Maria Ferrari and Mario Calcagno.

Moreover, this work is dedicated to my parents Ursula Luksza and Giambattista Ferrari, and to my uncle Michael Luksza.

Acknowledgements

When I joined the laboratory of Neurogenetics of Dr Parastoo Momeni at the Texas Tech University Health Sciences Center (TTUHSC), Lubbock, Texas, I discovered the desire to pursue a PhD in neurogenetics and Dr Momeni suggested taking part in the part-time PhD program at the University College London (UCL), with Drs John Hardy and Rohan de Silva primary and subsidiary supervisors, respectively. I will never be able to thank Dr Momeni enough for her mentorship, for her continuous support, for granting me freedom and for creating an ideal environment for my professional and personal growth during the years of this PhD. I am obviously also extremely grateful to Drs Hardy and de Silva for involving me in such interesting research projects, and for their mentorship and guidance throughout this experience. I would like to thank all the individuals that made these studies possible by sharing their own specimen and all the collaborators who participated in the project on FTD. In addition, I would like to thank Dr Andrew Singleton and the group at NIH for genotyping and, especially, Dr Michael Nalls for providing the necessary computational environment and expertise to perform the association analyses for the FTD project; Dr Jonathan Rohrer for advice during the course of the FTD study and Dr Stuart Pickering-Brown for financial support of the replication phase of the FTD study; Ms Kerra Pearce at the UCL Genomics core facility for help in genotyping of the FTD cases; Dr Mina Ryten for her valuable help in the PSP project; Ms June Smalley and Daniela Warr for their kind help during these years of PhD; Dr Lee Stanyer, Ms Jenny McGowan and Mr Mark Gaskin for their help in managing the laboratory matters at

UCL. I would also like to thank Ms June Howard from the Office of Sponsored Programs at TTUHSC for her valuable help in processing the MTAs for the FTD-GWAS.

Finally, I would like to thank all colleagues at TTUHSC and UCL for their support and help during this PhD, and for their valuable friendship.

Publications during this work

Raffaele Ferrari, Dena G Hernandez, Michael A Nalls, Jonathan D Rohrer, Adaikalavan Ramasamy, John BJ Kwok, Carol Dobson-Stone, William S Brooks, Peter R Schofield, Glenda M Halliday, John R Hodges, Olivier Piguet, Lauren Bartley, Elizabeth Thompson, Eric Haan, Isabel Hernández, Agustín Ruiz, Mercè Boada, Barbara Borroni, Alessandro Padovani, Carlos Cruchaga, Nigel J Cairns, Luisa Benussi, Giuliano Binetti, Roberta Ghidoni, Gianluigi Forloni, Daniela Galimberti, Chiara Fenoglio, Maria Serpente, Elio Scarpini, Jordi Clarimón, Alberto Lleó, Rafael Blesa, Maria Landqvist Waldö, Karin Nilsson, Christer Nilsson, Ian RA Mackenzie, Ging-Yuek R Hsiung, David MA Mann, Jordan Grafman, Christopher M Morris, Johannes Attems, Timothy D Griffiths, Ian G McKeith, Alan J Thomas, Pietro Pietrini, Edward D Huey, Eric M Wassermann, Atik Baborie, Evelyn Jaros, Michael C Tierney, Pau Pastor, Cristina Razquin, Sara Ortega-Cubero, Elena Alonso, Robert Perneczky, Janine Diehl-Schmid, Panagiotis Alexopoulos, Alexander Kurz, Innocenzo Rainero, Elisa Rubino, Lorenzo Pinessi, Ekaterina Rogaeva, Peter St George-Hyslop, Giacomina Rossi, Fabrizio Tagliavini, Giorgio Giaccone, James B Rowe, Johannes CM Schlachetzki, James Uphill, John Collinge, Simon Mead, Adrian Danek, Vivianna M Van Deerlin, Murray Grossman, John Q Trojanowski, Julie van der Zee, William Deschamps, Tim Van Langenhove, Marc Cruts, Christine Van Broeckhoven, The Belgian Neurology Consortium and the European Early-Onset Dementia consortium, Stefano F Cappa, Isabelle Le Ber, Didier Hannequin, Véronique Golfier, Martine Vercelletto, Alexis Brice, The French research network on FTLD/FTLD-ALS, Benedetta Nacmias, Sandro Sorbi, Silvia Bagnoli, Irene Piaceri, Jørgen E Nielsen, Lena E Hjermind, Matthias Riemenschneider, Manuel Mayhaus, Bernd Ibach, Gilles Gasparoni, Sabrina Pichler, Wei Gu, Martin N Rossor, Nick C Fox, Jason D Warren, Maria Grazia Spillantini, Huw R Morris, Patrizia Rizzu, Peter Heutink, Julie S Snowden, Sara Rollinson, Anna Richardson, Alexander Gerhard, Amalia C Bruni, Raffaele Maletta, Francesca Frangipane, Chiara Cupidi, Livia Bernardi, Maria Anfossi, Maura Gallo, Maria Elena Conidi, Nicoletta Smirne, Rosa Rademakers, Matt Baker, Dennis W Dickson, Neill R Graff-Radford, Ronald C Petersen, David Knopman, Keith A Josephs, Bradley F Boeve, Joseph E Parisi, William W Seeley, Bruce L Miller, Anna M Karydas, Howard Rosen, John C van Swieten, Elise GP Dopper, Harro Seelaar, Yolande AL Pijnenburg, Philip Scheltens, Giancarlo Logroscino, Rosa Capozzo, Valeria Novelli, Annibale A Puca, Massimo Franceschi, Alfredo Postiglione, Graziella Milan, Paolo Sorrentino, Mark Kristiansen, Huei-Hsin Chiang, Caroline Graff, Florence Pasquier, Adeline Rollin, Vincent Deramecourt, Florence Lebert, Dimitrios Kapogiannis, UK Brain Expression Consortium, North American Brain Expression Consortium, Luigi Ferrucci, Stuart Pickering-Brown, John Hardy, Parastoo Momeni and Andrew B Singleton. ***Genome-wide association study reveals lysosomal and immune system involvement in frontotemporal dementia***. Lancet Neurology (accepted for publication on May 6th, 2014).

Raffaele Ferrari, Mina Ryten, Roberto Simone, Daniah Trabzuni, Naiya Nicolaou, Geshanthi Hondhamuni, Adaikalavan Ramasamy, UK Brain Expression Consortium, Michael E. Weale, Andrew J Lees, Parastoo Momeni, John Hardy and Rohan de Silva. ***Assessment of common variability and expression quantitative trait loci for GWAS associations for progressive supranuclear palsy***. Neurobiol Aging, 2014 Jun;35(6):1514.e1-1514.e12.

Raffaele Ferrari, Mia Kero, Kin Mok, Anders Paetau, Tero Hiekkalinna, PenttiTienari, Olli Tynninen, Hannu Kalimo, John Hardy, Parastoo Momeni, Auli Verkkoniemi-Ahola, Liisa Myllykangas. ***FTD kindred associated with C9orf72 repeat expansion and dysplastic gangliocytoma***. Neurobiol Aging, 2014 Feb;35(2):444.e11-4.

Raffaele Ferrari, Saad Dawoodi, Merrill Raju, Avinash Thumma, Linda S Hynan, Shirin Hejazi Maasumi, Joan S Reisch, Sid O'Bryant, Marjorie Jenkins, Robert Barber and Parastoo Momeni. ***Androgen receptor gene and gender specific Alzheimer's disease***. Neurobiol Aging, 2013 Aug;34(8):2077.e19-20.

Raffaele Ferrari, Avinash Thumma and Parastoo Momeni (2013). ***Molecular genetics of Frontotemporal Dementia***. In: eLS. John Wiley & Sons, Ldt: Chichester. DOI: 10.1002/9780470015902.a0024457.

Corradi L, Porro I, Schenone A, Momeni P, **Ferrari R**, Nobili F, Ferrara M, Arnulfo G, Fato MM. ***A repository based on a dynamically extensible data model supporting multidisciplinary research in neuroscience***. BMC Med Inform Decis Mak. 2012 Oct 8;12:115. doi: 10.1186/1472-6947-12-115.

Raffaele Ferrari, Kin Mok, Jorge H Moreno, Stephanie Cosentino, Jill Goldman, Pierto Pietrini, Richard Mayeux, Michael C Tierney, Dimitrios Kapogiannis, Gregory A Jicha, Jill R Murrell, Bernardino Ghetti, Eric M Wassermann, Jordan Grafman, John Hardy, Edward D Huey and Parastoo Momeni. ***Screening for C9ORF72 repeat expansion in FTLD***. Neurobiol Aging, 2012 Aug;33(8):1850.e1-11.

Raffaele Ferrari, Jorge H Moreno, Abu T Minhajuddin, Sid E O'Bryant, Joan S Reisch, Robert C Barber and Parastoo Momeni. ***Implication of common and disease specific variants in CLU, CR1 and PICALM***. Neurobiol Aging, 2012 Mar 6. [Epub ahead of print].

Huey ED, **Ferrari R**, Moreno JH, Jensen C, Morris CM, Potocnik F, Kalaria RN, Tierney M, Wassermann EM, Hardy J, Grafman J, Momeni P. ***FUS and TDP43 genetic variability in FTD and CBS***. Neurobiol Aging, 2012 May;33(5):1016.e9-17.

Ferrari R, Hardy J, Momeni P. ***Frontotemporal dementia: from mendelian genetics towards genome-wide association studies***. J Mol Neurosci. 2011 Nov;45(3):500-15.

Hardy J, Guerreiro R, Wray S, **Ferrari R** and Momeni P. ***The genetics of Alzheimer's disease and other tauopathies***. J Alzheimers Dis. 2011;vol 23 (suppl 1):S33-9.

Raffaele Ferrari, Dimitrios Kapogiannis, Edward D Huey and Parastoo Momeni. ***FTD and ALS: a tale of two diseases***. Curr Alzheimer Res. 2011 May;8(3):273-94.

Baborie A, Griffiths TD, Jaros E, McKeith IG, Burn DJ, Richardson A, **Ferrari R**, Moreno J, Momeni P, Duplessis D, Pal P, Rollinson S, Pickering-Brown S, Thompson JC, Neary D, Snowden JS, Perry R, Mann DM. ***Pathological correlates of frontotemporal lobar degeneration in the elderly***. Acta Neuropathol. 2011 Mar;121(3):365-71.

Raffaele Ferrari, Dimitrios Kapogiannis, Edward D Huey, Jordan Grafman, John Hardy and Parastoo Momeni. ***Novel missense mutation in Charged Multivescicular Body Protein 2B in a patient with Frontotemporal Dementia***. Alzheimer Dis Assoc Disord. 2010 Jun 29. [Epub ahead of print].

Parastoo Momeni, **Raffaele Ferrari**. ***Genetics and blood biomarkers of Alzheimer's disease***. The Open Nuclear Medicine Journal, 2010, 2, 17-29.

Jonathan D Rohrer, Rita Guerreiro, Jana Vandrovcova, James Uphill, David Reiman, Jonathan Beck, Adrian M Isaacs, Astrid Authier, **Raffaele Ferrari**, Nick C Fox, Ian R A Mackenzie, John Collinge, Jason D Warren, Rohan de Silva, Janice Holton, Tamas Revesz, John Hardy, Simon Mead, Martin N Rossor. ***The heritability and genetics of frontotemporal lobar degeneration***. Neurology. 2009 Nov ;73(18):1451-6.

Table of contents

Abbreviations

Activating transcription factor 4	(ATF4)
Alzheimer's disease	(AD)
Alzheimer's Research UK	(ARUK)
Amplified fragment length polymorphism	(AFLP)
Amyloid beta (A4) precursor protein	(APP)
Amyotrophic lateral sclerosis	(ALS)
Apolipoprotein E	(ApoE)
Attention deficit disorder	(ADD)
Attention deficit hyperactivity disorder	(ADHD)
Atypical FTLD-U	(aFTLD-U)
Atypical PSP	(aPSP)
Base pairs	(bp)
Behavioural variant frontotemporal dementia	(bvFTD)
Butyrophilin-like 2 (MHC class II associated) gene	(BTNL2)
Cathepsin C	(CTSC)
Central nervous system	(CNS)
Centrosomal protein 57	(CEP57)
Cerebellar cortex	(CRBL)
Cerebrovascular accident	(CVA)
Charged multivesicular body protein 2B	(CHMP2B)
Chromosome 9 open reading frame 72	(C9orf72)

Clusterin	(CLU)
Complement component (3b/4b) receptor 1	(CR1)
Computed tomography	(CT)
Copy-neutral LOH	(cnLOH)
Copy number polymorphisms	(CNPs)
Copy number variations	(CNVs)
Corticobasal syndrome	(CBS)
Corticobasal degeneration	(CBD)
Corticotropin releasing hormone receptor 1	(CRHR1)
Cycle threshold	(Ct)
Degree of freedom	(df)
Dementia lacking distinctive histopathology	(DLDH)
Dementia with Lewy bodies	(DLB)
Deoxyribonucleic acid	(DNA)
Di-peptide repeat	(DPR)
Dystrophic neurites	(DNs)
Encyclopaedia of DNA elements	(ENCODE)
Endoplasmic reticulum	(ER)
Endosomal Sorting Complex required for Transport III	(ESCRT-III)
ER associated degradation	(ERAD)
Eukaryotic translation-initiation factor 2	(EIF-2)
Eukaryotic translation initiation factor 2-alpha kinase 3 gene	(EIF2AK3)
Expression quantitative trait loci	(eQTL)
Familial amyotrophic lateral sclerosis	(fALS)

Family with sequence similarity 76 member B	(FAM76B)
Fifth edition of the diagnostic and statistical manual of mental disorders	(DSM-5)
File transfer protocol	(FTP)
Frontal cortex	(FCTX)
Frontotemporal dementia	(FTD)
Frontotemporal dementia with motor neuron disease	(FTD-MND)
FTD associated with TDP-43	(FTD-TDP)
Frontotemporal lobar degeneration	(FTLD)
Frontotemporal lobar degeneration with tau pathology	(FTLD-tau)
Frontotemporal lobar degeneration with ubiquitin pathology	(FTLD-U)
FTLD associated with ubiquitin proteosome system (UPS)	(FTLD-UPS)
FTD linked to chromosome 3	(FTD-3)
Fused in sarcoma	(FUS)
Genome-wide association study	(GWAS)
Genomic inflation factor	(GIF)
Glycogen synthase kinases	(GSKs)
Granulin	(GRN)
Hardy-Weinberg Equilibrium	(HWE)
Heterogeneous nuclear ribonucleoprotein	(hnRNP)
High-income countries	(HIC)
Hippocampus	(HIPP)
Histone 3 acetylated at lysine 27	(H3K27Ac)
Human genome project	(HGP)
Human leukocyte antigen	(HLA)

Huntington's disease	(HD)
Identification	(ID)
Identity by descent	(IBD)
Inch of Mercury	(in.Hg)
Inclusion body myopathy	(IBM)
Inclusion body myopathy associated with Paget's disease of the bone and frontotemporal dementia	(IBMPFD)
Insertion/deletion	(indel)
Institute of Child Health	(ICH)
Institutional Review Board	(IRB)
Interferon γ	(INF-γ)
Interferon kappa	(IFNK)
International Conference on Alzheimer's Disease	(ICAD)
Intraflagellar transport 74	(IFT74)
Intraluminal vesicles	(ILVs)
Intramembrane protease 5	(IMP5)
1000 base pairs	(Kbp)
Kinase activator MOB 3B	(MOB3B or MOBKL2B)
Laboratory of Neurogenetics	(LNG)
Late onset Alzheimer's disease	(LOAD)
Linkage disequilibrium	(LD)
Logarithm of the odds	(LOD)
Logopenic progressive aphasia	(LPA)
Loss of heterozygosity	(LOH)

Low/middle income countries	(LMIC)
Luria Agar Base	(LGA)
Luria Broth	(LB)
Magnetic resonance imaging	(MRI)
Major histocompatibility	(MHC)
Major histocompatibility complex, class I-related	(MR1)
Major histocompatibility complex class II, DR alpha	(HLA-DRA)
Major histocompatibility complex class II, DR beta 5 genes	(HLA-DRB5)
Material Transfer Agreement	(MTA)
1000000 base pairs	(Mbp)
Medical Research Council	(MRC)
Medulla	(MEDU)
Messanger RNA	(mRNA)
Methylation quantitative trait loci	(mQTL)
Microtubules	(MTs)
Microtubule associated protein tau	(MAPT)
Microtubule interacting transport	(MIT)
Microtubule interacting transport (MIT)-interacting region	(MIR)
Minimental state examination	(MMSE)
Minor allele frequency	(MAF)
MIT-interacting motif	(MIM)
Motor neuron disease	(MND)
Multiple sclerosis	(MS)
Multiple system atrophy	(MSA)

Multiplex ligation-dependent probe amplification	(MLPA)
Multivesicular body	(MVB)
Myelin-associated oligodendrocyte basic protein	(MOBP)
Myotubularin-related 2	(MTMR2)
National Center for Biotechnology information	(NCBI)
National Institute on Ageing	(NIA)
National Institutes of Health	(NIH)
National Institute of Neurological Disorders and Stroke	(NINDS)
Natural killer	(NK)
NEL-like protein 2	(NELL2)
Nerve growth factor	(NGF)
N-ethylmaleimide-sensitive factor	(NSF)
Neurocognitive disorders	(NCDs)
Neurofibrillary tangles	(NFTs)
Neuronal cytoplasmic inclusions	(NCIs)
Neuronal intermediate filament inclusion disease	(NIFID)
Neuronal intranuclear inclusions	(NIIs)
Next generation sequencing	(NGS)
Non-sense mediated decay	(NMD)
North American Brain Expression Consortium	(NABEC)
Not otherwise specified	(NOS)
Obsessive compulsive disorder	(OCD)
Occipital cortex	(OCTX)
Odds ratio	(OR)

Online Mendelian Inheritance in Men	(OMIM)
Paget's disease of the bone	(PDB)
Parkinson's disease	(PD)
Phosphatase and tensin homolog gene	(PTEN)
Phosphatidylinositol binding clathrin assembly protein	(PICALM)
Pick's disease	(PiD)
Post-traumatic stress disorder	(PTSD)
Presenilin 1	(PSEN1)
Presenilin 2	(PSEN2)
Principal component analysis	(PCA)
Polymerase chain reaction	(PCR)
Polymorphism Phenotyping version 2	(PolyPhen-2)
Positron emission tomography	(PET)
Primary progressive aphasia	(PPA)
Principal components analysis	(PCA)
Principal investigator	(PI)
Progressive nonfluent aphasia	(PNFA)
Progranulin	(GRN)
Progressive supranuclear palsy	(PSP)
PSP-pure akinesia with gaze freezing	(PAGF)
PSP with parkinsonism	(PSP-P)
PSP with progressive nonfluent aphasia	(PSP-PNFA)
Putamen	(PUTM)
Quality control	(QC)

Quantile-Quantile plots	(QQ plots)
Quantitative-PCRs	(Q-PCRs)
RAB38, member RAS oncogene family	(RAB38)
Repeat-primed polymerase chain reaction	(RP-PCR)
Reverse transcriptase-PCR	(RT-PCR)
Ribonucleic acid	(RNA)
Saitohin	(STH)
Semantic dementia	(SD)
Sequence Detection System	(SDS)
Simple nucleotide variations	(SNVs)
Single nucleotide polymorphisms	(SNPs)
Single-photon emission computed tomography	(SPECT)
Soluble N-ethylmaleimide-sensitive factor-attachment protein receptor	(SNARE)
Solute carrier family 39 (zinc transporter), member 7 gene	(SLC39A7)
Syntaxin 6	(STX6)
SNP database	(dbSNP)
Sporadic amyotrophic lateral sclerosis	(sALS)
Structural variations	(SVs)
Substantia nigra	(SNIG)
Sun Health Research Institute	(SHRI)
Superoxide dismutase-1 gene	(SOD-1)
TAR DNA binding protein	(TARDBP or TDP-43)
Temporal cortex	(TCTX)
Texas Tech University Health Sciences Center	(TTUHSC)

Thalamus	(THAL)
Toll-like receptor	(TLR)
Translocase of outer mitochondrial membrane 40 homolog	(TOMM40)
Trans-Golgi network	(TGN)
Transient ischemic attack	(TIA)
Transmembrane protein 106B	(TMEM106B)
Traumatic brain injury	(TBI)
Tris/Borate/EDTA	(TBE)
Tumour necrosis factor α	(TNF-α)
Ubiquilin 2	(UBQLN2)
Ubiquitin pathology	(FTLD-U)
UK Brain Expression Consortium	(UKBEC)
Ultra violet	(UV)
Unfolded protein response	(UPR)
Uniparental disomy	(UPD)
University College London	(UCL)
Untranslated region	(UTR)
Vacuolar protein sorting 4 homolog A protein	(VPS4)
Vascular endothelial growth factor	(VEGF)
Valosin-containing protein	(VCP)
Wellcome Trust Case-Control Consortium	(WTCCC)
Western Blot	(WB)
White matter	(WHMT)
Whole exome sequencing	(WES)

Whole genome sequencing	(WGS)
Wild-type	(WT)
Wolcott-Rallison syndrome	(WRS)

List of tables

List of figures

CHAPTER 1 – Introduction

1.1 Overview

1.1.1 Ageing societies

Overall improved quality of life has led to unprecedented population growth over the past six decades in our societies accounting considerably for an increase of diseases of the older age.

In high-income countries (HIC) as well as low/middle income countries (LMIC) ageing is the major global risk factor contributing to dementia [1], whose incidence and prevalence is expected to dramatically increase in the next decades [2].

In the year 2010, 4.7% of the people over 60 years of age (~36 million people) were affected by dementia, worldwide [1].

An analysis of regional distribution of prevalence of dementia showed that 6.2% individuals were affected in Europe, 6.9% in North America, 8.5% in Latin America, ~4% in Asia, ~6% in Australasia and ~3% in Africa [1, 2] and it was estimated that the incidence of dementia would almost double every 20 years reaching ~115 million in 2050 [1] (**Table 1-1**).

Table 1-1. Prevalence and estimated incidence of dementia worldwide

World region	Prevalence of dementia (%) (in individuals >60 years of age)	Estimated incidence		
		2010	2030	2050
Africa	~3%	~36M	~66M	~115M
Asia	~4%			
Australasia	~6%			
Europe	~6%			
North America	~7%			
Central/South America	~8.5%			

Table 1-1. Data are adapted from [1, 2]. The prevalence of dementia is expressed in percentages and reflects the estimates of elderly over 60 years of age affected by dementia in the year 2010. The estimated incidence of dementia is expressed in millions (M).

The costs associated with dementia represent a financial burden that impacts families of affected individuals, the health/medical, and social services. It was estimated that in 2010 the overall costs of dementia worldwide reached $604 billion [3]. The distribution of these costs seemed to differ across HIC and LMIC: HIC (including Western Europe, North America and East/Pacific Asia) accounted for ~70% of total costs despite a lower prevalence of dementia compared to LMIC (~42% vs. 58%) [3]. It was suggested that such disparity was the result of differences in the average wages (~$30,000 in HIC vs. $903-12,000 in LMIC) and disparity in services or resources for dementia care [3]. In fact, one third to half of the affected individuals in HIC relies on formal care (residential or nursing homes), whilst in LMIC this service is essentially not offered and patients undergo informal care, i.e. caregivers are (unpaid) family members [3]. Also, different healthcare systems across different countries contribute to differences in direct medical/social costs as well as availability of medical/social assistance [3].

Another important socio-economic aspect is that the financial burden of dementia is currently higher than that of other chronic conditions. For instance, a study performed in the UK revealed that yearly societal costs for dementia (£23 billion) exceeded those for cancer (£12 billion) or heart disease (£8 billion) [3].

These statistics urge for a collective and global shift in the strategies to face dementia and advocate for intervention at the government level including the development of programs and policies targeting issues such as:

1 – Early diagnosis and intervention;

2 – Families/caregivers support and training, and;

3 – Availability of cost effective services (e.g. medical and social care) [1].

In summary, these statistics, together with the fact that there is still an almost complete lack in both the understanding of the mechanisms that lead to disease and the availability of disease-modifying or curative therapies, call for more investment in infrastructure to support research and advocacy for the care of those affected by dementia [3, 4].

1.1.2 Dementias

Dementia is among the main causes of impairment of broad cognitive functions in individuals of older age leading to a wide range of deficits that impact memory, cognition, behaviour as well as motor skills and, eventually, the whole body as a system. These syndromes, alone or collectively, contribute to a progressive deterioration in the performance of normal daily activities and leave the affected individual in need of assistance [1].

Recently, the fifth edition of the diagnostic and statistical manual of mental disorders (DSM-5) was published providing the most updated guidelines for defining and diagnosing the broad spectrum of mental disorders [5].

In the DSM-5, Alzheimer's and non-Alzheimer's types of dementias fall within the extended category of neurocognitive disorders (NCDs) [5]. The main feature driving the diagnosis is impaired cognition that is defined as an acquired rather than inborn condition [5]. In the DSM-5, NCDs have been divided into two groups: Major and mild.

Although the differences between the major and the mild form are rather subtle, a distinction may be ascribed to the presence of a "notable" (major NCDs) vs. "modest" (mild NCDs) cognitive decline compared to previous performances of the same individual and/or "interference" (major NCDs) vs. "no major interference" (mild NCDs) with normal everyday activities [5]. To diagnose the type of NCD as well as the "major" vs. the "mild" form, the specific abilities which are tested and assessed concern complex

attention, executive function, learning and memory, language and perceptual motor/social cognition [5].

The two distinct groups, Alzheimer's type and non-Alzheimer's type of dementia, are distinguished by their clinical presentation. The former is identified exclusively by Alzheimer's disease (AD) because of its relatively unique early clinical features that include short-term memory loss, visuo-spatial deficits as well as cognitive decline and language impairment. The latter encompasses a long list of conditions characterised by differential symptomatology that in the vast majority of cases do not overlap with those specific to AD. Non Alzheimer's type of dementia include, among others, vascular dementia (second in incidence and prevalence only to AD), dementia with Lewy bodies (DLB), Creutzfeld-Jakob (prion) disease, frontotemporal dementia (FTD) and other forms of neurodegenerative disorders that present a major motor component that may overlap with dementia symptoms such as the cases of Parkinson's disease (PD), progressive supranuclear palsy (PSP), corticobasal syndrome/degeneration (CBS/D) and amyotrophic lateral sclerosis (ALS). For each and every neurological disorder, the clinical picture is a reflection of the topographical areas of the brain that undergo degeneration. Subsequently, the neuropathology is still the gold standard for the definite diagnosis.

In this thesis the neurodegenerative disorders under study are frontotemporal dementia (FTD) and progressive supranuclear palsy (PSP). Their main clinical, pathological and known genetic features are summarised in the upcoming sections.

1.2 Frontotemporal dementia (FTD)

1.2.1 Relevance and statistics of the disease

In 1892 the neuropsychiatrist Arnold Pick (1851-1924) reported the case of an individual with signs of progressive cognitive impairment, bouts of aggressiveness and aphasia whose brain pathology assessment revealed atrophy in the left temporal lobe [6, 7]. This is the first patient with FTD-like features in medical history. In the early 1900s Pick presented a further case with stereotyped behaviour [8] and semantic dementia-like symptoms [9]. These types of neurodegenerative disorders were named Pick's disease (PiD). The study of the brain pathology revealed inclusions (Pick bodies) in neurons of the frontal lobes [10] that became the histopathological hallmark of PiD. Much later the microtubule associated protein tau (MAPT) and ubiquitin were shown to be the main component of these inclusions [11] [12]. PiD is currently considered a subgroup of FTD [13], as Pick bodies are found in 10-30% of sporadic FTD cases [14] and based on the historical case reports Arnold Pick is considered the father of FTD.

Worldwide, FTD is the second most common form of young-onset dementia in the population under 65 years and contributes to ~10-20% of all dementias [15] and appears with average onset age of mid to late 50s [16]. FTD occurs in approximately 3-15 out of 100,000 individuals that are in their mid to late 50s or early 60s [16]. The disease has insidious onset, it can present as familial (30-50% of cases) or sporadic (50-70% of cases) form [17, 18] and its incidence is almost equal among men and women [19].

1.2.2 Diagnostic criteria and clinical features

FTD is characterised by cognitive decline and behavioural dysfunction (bvFTD), which result in changes in personal and social conduct, and by language dysfunctions broadly called primary progressive aphasia (PPA). PPA is subdivided in semantic dementia (SD) (or semantic variant PPA), progressive nonfluent aphasia (PNFA) (or nonfluent/agrammatic variant PPA) and logopenic progressive aphasia (LPA) (or logopenic variant PPA) [20, 21]. Within this extended clinical picture, memory and visuo-spatial abnormalities remain initially intact.

1.2.2.1 Behavioral variant FTD (bvFTD)

Patients in this subtype may show disinhibition, emotional blunting, loss of insight, tendency to binge eating, loss of volition and inertia [21] and, in addition, deficits in executive functions, such as planning, reasoning and problem solving [21].

1.2.2.2 Semantic dementia (or semantic variant PPA)

This syndrome is a disorder of conceptual knowledge characterised by a severe naming and word comprehension impairment, whilst speech output remains fluent [20]. SD is classified as right or left SD temporal variant based on the temporal lobe affected by

atrophy [22] and signs of behavioural changes can be seen although distinctive from those observed in pure bvFTD [23].

1.2.2.3 Progressive non fluent aphasia (or nonfluent/agrammatic variant PPA)

This syndrome is a disorder of expressive language characterised by effortful non fluent speech, phonological and grammatical errors, and by difficulties in word retrieval as well as in reading and writing [20]. Behavioural changes are less common in PNFA but may emerge later while disease progresses [23].

1.2.2.4 Logopenic progressive aphasia (or logopenic variant PPA)

The main features of this syndrome are represented by word retrieval (although milder than in SD) and repetition deficits [20]. Speech is slow and interrupted because of word-finding problems, however the grammar is intact but accompanied by occasional phonological errors [24]. Caution has been suggested as this syndrome may be a subtle and atypical early presentation of AD [25].

The Neary criteria [21] have been the most common diagnostic criteria for FTD since 1998 (**Table 1-2**).

Table 1-2. Diagnostic criteria for FTD

Neary criteria	Reference
bvFTD	
Insidious onset and gradual progression	[21]
Early decline in social interpersonal conduct	
Early impairment in regulation of personal conduct	
Early emotional blunting	
Early loss of insight	
SD	
Insidious onset and gradual progression	[21]
Progressive, fluent, empty spontaneous speech	
Loss of word meaning (impaired naming and comprehension)	
Semantic paraphasias	
Impaired recognition of familiar faces (prospagnosia)	
Impaired recognition of objects	
Preserved perceptual matching and drawing reproduction	
Preserved single-word repetition	
Preserved ability to read aloud and write	
PNFA	
Insidious onset and gradual progression	[21]
Nonfluent spontaneous speech with at least one of the following: Agrammatism, phonemic paraphasias, anomia	
More details to be found in [21]	

Table 1-2. Summary of main diagnostic criteria for FTD. **Abbreviations**: BvFTD=behavioural variant FTD; SD=semantic dementia; PNFA=progressive nonfluent aphasia. Table taken, or adapted, from [26].

More recently, two international consortia developed revised guidelines specifically for the diagnosis of the bvFTD subtype [27] and the language variants [20]. Guidelines for bvFTD are structured in five categories, each, summarizing symptoms that must be present to meet criteria for bvFTD (**Table 1-3**).

Table 1-3. Revised diagnostic criteria for the bvFTD

bvFTD	Reference
Category I (Neurodegenerative disease)	[27]
Gradual deterioration of behaviour and/or cognition	
Category II (Possible bvFTD)	[27]
Disinhibition	
Apathy or inertia	
Loss of sympathy or empathy	
Stereotyped or ritualistic behavior	
Executive dysfunction	
Category III (Probable bvFTD)	[27]
Ascertained "Possible bvFTD" (category II)	
Significant functional decline	
Presence of frontal and/or anterior temporal atrophy (MRI) or hypometabolism (PET or SPECT)	
Category IV (bvFTD with specific FTLD pathology)	[27]
Ascertained "Possible bvFTD" (category II) or "Probable bvFTD" (category III)	
Presence of pathological hallmarks (e.g. FTLD-tau or FTLD-TDP)	
Presence of known pathogenic mutations	
Category V (Exclusion criteria for bvFTD)	[27]
Dysfunctions typical of other neurological disorders	
Behavioural changes resembling psychiatric disorders	
Biomarkers reflecting other neurological disorders	
More details to be found in [27]	

Table 1-3. Summary of the revised diagnostic criteria for the bvFTD subtype. **Abbreviations**: BvFTD=behavioural variant FTD; SD=semantic dementia; PNFA=progressive nonfluent aphasia, MRI-Magnetic resonance imaging; PET=positron emission tomography; SPECT=single-photon emission computed tomography; FTLD=frontotemporal lobar degeneration; FTLD-TDP=frontotemporal lobar degeneration with ubiquitin/TDP-43 pathology; FTLD-tau=frontotemporal lobar degeneration with tau pathology. Table taken, or adapted, from [26].

Guidelines for the diagnosis of the main PPA syndromes (SD [or semantic variant PPA], PNFA [or nonfluent/agrammatic variant PPA]) are summarised in **Table 1-4**.

Table 1-4. Revised diagnostic criteria for PPA

Language variant	Reference
Category I (SD or semantic variant PPA)	
Impaired naming Impaired single word comprehension Impaired knowledge of low familiar objects Impairment in learning or writing Spared repetition and speech production	[20]
Category II (SD or semantic variant PPA)	
Ascertained clinical diagnosis (category I) Presence of anterior temporal lobe atrophy (MRI) or hypometabolism (PET or SPECT)	[20]
Category III (SD or semantic variant PPA)	
Ascertained clinical diagnosis (category I) Presence of pathological hallmarks (FTLD-TDP most common) or pathogenic mutation (*MAPT* or *GRN* mutations)	[20]
Category I (PNFA or nofluent/agrammatic variant PPA)	
Agrammatism or apraxia of speech Impaired comprehension Spared single-word comprehension Spared object recognition	[20]
Category II (PNFA or nofluent/agrammatic variant PPA)	
Ascertained clinical diagnosis (category I) Presence of left posterior fronto-insular atrophy (MRI) or hypometabolism (PET or SPECT)	[20]
Category III (PNFA or nofluent/agrammatic variant PPA)	
Ascertained clinical diagnosis (category I) Presence of pathological hallmarks (FTLD-TDP or FTLD-tau) or pathogenic mutation (*GRN* mutations)	[20]
More details to be found in [20]	

Table 1-4. Summary of the revised diagnostic criteria for the PPA subtypes. **Abbreviations**: SD=semantic dementia; PNFA=progressive nonfluent aphasia; MRI-Magnetic resonance imaging; PET=positron emission tomography; SPECT=single-photon emission computed tomography; FTLD=frontotemporal lobar degeneration; FTLD-TDP=frontotemporal lobar degeneration with ubiquitin/TDP-43 pathology; FTLD-tau=frontotemporal lobar degeneration with tau pathology; MAPT=microtubule-associated protein tau; GRN=progranulin. Table taken, or adapted, from [26].

Both the revised diagnostic criteria for bvFTD [27] and for the language variants [20] represent an improvement of the Neary criteria [21] in that they increase the sensitivity of

the clinical diagnosis of each syndrome, they provide robust tools to diagnose the disease at insidious early stages, and they substantially contribute to uniform and harmonize the diagnosis of FTD across research centres worldwide.

1.2.3 Neuropathology

The brains of FTD patients show shrinkage in the frontal and temporal areas and present spongiform morphology due to neuronal loss that progressively increases during the course of the disease. Based on the topography of the lesions, frontotemporal dementia is also called FTLD (frontotemporal lobar degeneration).

In bvFTD atrophy affects frontal lobes bilaterally, i.e. the medial frontal lobes and the anterior temporal lobes [22]. In PPA, the SD subtype shows asymmetric atrophy in the middle, inferior and medial anterior temporal lobe [20, 22], whilst the PNFA subtype presents mainly with left posterior frontal and insular regions atrophy [20]. The molecular pathology in FTLD is characterised by abnormal accumulation of protein aggregates [13] that cause inclusion bodies (i.e. neuronal cytoplasmic inclusions [NCI] and/or neuronal intranuclear inclusions [NII]) in neurons and/or glial cells, and by dystrophic neurites (DN) in the affected areas [28]. It has been suggested that protein misfolding, cleavage, phosphorylation and changes in protein solubility are key factors leading to protein aggregates [13]. FTLD pathology is differentially classified, based on the protein inclusions identified in the brains of patients. Currently, ≤40-50% of FTLD

cases show tau pathology (FTLD-tau), ≤40-50% ubiquitin/TDP-43 pathology (FTLD-TDP), ≤10% FUS positive and tau/TDP-43 negative pathology (FTLD-FUS) and ≤1-2% ubiquitin/p62 pathology (FTLD-UPS) (**Table 1-5**) [28].

1.2.3.1 FTLD-tau

Hyperphosphorylated tau inclusions accumulate in the cytoplasm of neurons and glial cells located, mainly, in the frontal/temporal cortex, hippocampus and subcortical nuclei and, more rarely, in the midbrain, brain stem, cerebellum and spinal cord [17, 29]. Tau hyperphosphorylation seems to be mediated by the activity of several kinases including the glycogen synthase kinases (GSKs) [13] (**Table 1-5**). FTLD-tau is associated with mutations in *MAPT* [28], nevertheless tau pathology can also occur in absence of *MAPT* variations [25].

1.2.3.2 FTLD-TDP

Ubiquitin/TDP-43 positive inclusions accumulate in the cytoplasm (NCI) or nucleus (NII) of neurons and cause dystrophic neurites in brain cortex, subcortex and hippocampus [30]. TDP-43 pathology has recently been sub-classified in the subtypes A, B, C and D [31] (**Table 1-5**) as an updated classification system to replace previous conflicting nomenclature. FTLD-TDP is associated with genetic variability in *GRN*, *C9orf72* and *VCP* (see the genes further below in section ***1.2.4 Genetics***) and each

subtype (A, B, C and D) is characterised by specific signatures (**Table 1-5**). Type A, which associates mainly with bvFTD and PNFA, presents with dystrophic neurites (DN) and oval neuronal cytoplasmic inclusions (NCI) in the neocortical layer and is mainly associated with mutations in *GRN* [31]; type B, primarily seen in bvFTD as well as FTD-MND cases, shows NCI and rather few DN in all cortical layers and is associated with chromosome 9, i.e. with *C9orf72* [31]; type C, described in bvFTD and SD cases, is defined by presence of elongated DN and few NCI in the upper cortical layers and no particular genetic association has been recognised thus far [31], and; type D, which associates with familial inclusion body myopathy with Paget's disease of bone and frontotemporal dementia (IBMPFD), is characterised by short DN and lentiform neuronal intranuclear inclusions (NII) in all layers and is associated with *VCP* mutations [31].

1.2.3.3 FTLD-FUS

FUS immunoreactive inclusions have been described in ~10% of FTD cases (FTLD-FUS) that do not present TDP-43 or tau positive inclusions (**Table 1-5**) [22]. FUS pathology has been consistently reported in rare FTD subtypes such as atypical FTD (aFTD), neuronal intermediate filament inclusion disease (NIFID) and a subgroup of FTLD-U with early onset and caudate atrophy. The topographic distribution of FUS pathology includes cortical layers, the hippocampus, subcortical regions, brainstem and spinal cord (**Table 1-5**) [32] [33]. The molecular pathology may vary based on the affected brain region and/or on the syndrome (NIFID vs. aFTD). In the cortical layers a

common feature between NIFID and aFTD is represented by crescent-shaped NCI; globular or flame-like NCI, and rod-shaped (occasionally) and vermiform NII are more frequently found in NIFID, whereas bean- or annular-shaped NCI, and FUS-positive neurites (in white matter) mainly characterise aFTD [33]. In the hippocampus crescent-shaped NCI, bean-like or globular NCI/NII as well as coiled body-like inclusions in white matter have been described in NIFID; conversely, only bean-shaped NII may be found in aFTD [33]. Globular NCI and skein-like NCI in motor neurons as well as (occasionally) bean-shaped NCI are found in the brainstem of NIFID cases, whereas dense globular NCI and skein-like NCI in motor neurons characterise aFTD [33]. Last, in the spinal cord differential FUS positive aggregates including filamentous skein- or dot-like inclusions define NIFID, whereas mainly skein-like NCI have been described in aFTD [33]. It has been suggested that a minority of cases presenting FUS pathology carry mutations in *FUS* [28] [33].

1.2.3.4 FTLD-UPS

FTLD-UPS defines a very rare subgroup of FTD which is linked to chromosome 3 (FTD3) [34] and is characterised by ubiquitin/p62 positive NCI either in the dentate gyrus or sparse in the frontal or other cortical areas [35] (**Table 1-5**). Based on the fact that these proteins are part of the ubiquitin proteasome system (UPS) this FTD subtype (≤1-2%) is named FTLD-UPS [36] [37]. FTLD-UPS has been associated with variability in *CHMP2B* [35].

In summary, the nomenclatures used to classify FTLD pathology are based on the most frequently observed patterns of atrophy and types of protein inclusions and are being constantly updated in order to uniform the classification of FTLD.

Table 1-5. Classification of pathology and genetics associated with FTD

Pathology and Genetics	FTLD-tau (≤40-50%)	FTLD-TDP (≤40-50%)	FTLD-FUS (≤10%)	FTLD-UPS (≤1-2%)
Inclusion protein	Tau	Ubiquitin/TDP-43	FUS	Ubiquitin/p62
Cellular location	Neurons and glial cells[a1,a2]	Neurons[c]	Neurons and glial cells[e]	Neurons[f]
Brain topography	Frontal cortex[a1,a2] Temporal cortex[a1,a2] Hippocampus[a1,a2] Subcortical nuclei[a1,a2] Midbrain (rarely)[a1,a2] Brain stem (rarely)[a1,a2] Cerebellum (rarely)[a1,a2] Spinal cord (rarely)[a1,a2]	Frontal cortex[c] Subcortex[c] Hippocampus[c]	Cerebral cortex[e] Hippocampus[e] Subcortical regions[e] Branistem[h] Spinal cord[h]	Dentate gyrus[f] Frontal cortex (rarely)[f] Other cortical areas (rarely)[f]
Pathological features	Hyperphosphorylated tau inclusion[b] Paired helical filaments[b] Neurofibrillary tangles[b]	**Type A**: DN or oval NCI in neocortical layer 2[d] **Type B**: Mostly NCI and few DN in all cortical layers[d] **Type C**: Mostly elongated DN and few NCI in upper cortical layers[d] **Type D**: Short DN and lentiform NII in all layers[d]	*FUS inclusions (vermiform NII)*[e]	*p62 inclusions (NCI)*[f]
Genetic features	*MAPT* mutations[g]	**Type A**: *GRN* mutations[d] **Type B**: *C9orf72* expansion[d] **Type C**: genetic component not known **Type D**: *VCP* mutations[d]	*FUS* mutations (very rarely)[g]	*CHMP2B* mutations[g]

Table 1-5. Classification of pathology associated with FTD and their characteristics. **Abbreviations**: FTLD-tau=frontotemporal lobar degeneration with tau pathology; FTLD-TDP=frontotemporal lobar degeneration with ubiquitin/TDP-43 pathology; FTLD-FUS=frontotemporal lobar degeneration with FUS pathology; FTLD-UPS=frontotemporal lobar degeneration with ubiquitin/p62 pathology (which are elements of the Ubiquitin Proteasome System);TDP-43=transactive response DNA-binding protein 43 kDa; FUS=fused in Sarcoma; p62=ubiquitin proteasome system associated sequestosome p62; DN=dystrophic neurites; NCI=neuronal cytoplasmic inclusions; NII=neuronal intranuclear inclusions. References: a1-[17]; a2-[29]; b-[38]; c-[30]; d-[31]; e-[32]; f- [35]; g- [28]; h- [33] . Table taken, or adapted, from [26].

1.2.4 Genetics

Genetics of FTD is heterogeneous. Approximately, 50% of familial cases show association with *MAPT* and progranulin (*GRN*) genes [17]. Recently, a proportion of FTD cases, mainly bvFTD (7-11% for sporadic cases and 12-25% for familial cases [39]), was associated with abnormal repeat expansion in either the promoter region or the first intron of chromosome 9 open reading frame 72 (*C9orf72*). These three genes are currently the main genetic factors associated with FTD. In addition, variations in valosin containing protein (*VCP*), charged multivesicular body protein 2B (*CHMP2B*), *TDP-43*, *FUS*, sequestosome 1 (*SQSTM1*) [40] and genetic variability in the transmembrane protein 106 B (*TMEM106B*) [41] genes contribute, all together, to ≤5% of cases [17].

In the next paragraphs each gene will be described in their most relevant characteristics, with a particular attention for *MAPT*, *GRN* and *C9orf72*.

1.2.4.1 MAPT

MAPT is located on chromosome 17q21.1 and encodes the microtubule associated protein tau. There are 8 different isoforms of tau protein. Six of these are present in the CNS and are produced due to alternative splicing of exons 2, 3, and 10 [42, 43].

The protein holds an amino-terminal projection domain and a carboxy-terminal microtubule binding domain (MT-binding domain). The latter consists of three or four repeats (3R- or 4R-tau) depending on alternative splicing of exon 10 [44] (**Figure 1-1**).

From a functional perspective, tau binds and stabilizes the microtubules (MTs) supporting the maintenance of cell shape and the trafficking of molecules in the axon [45, 46].

Tau seems also to have a role in neuronal polarity, signal transduction [29] and in mediating MTs interaction with other elements of the cytoskeleton as well as with components of the neural plasma membrane [46].

Not least, it has been suggested that tau regulates the spacing between MTs [29].

Currently, more than 70 variations in *MAPT* have been reported [44, 47] (**Figure 1-1**). Of these ~45 lead to deletions, missense and splice site mutations through:

1 – Exonic point mutations [45];

2 – Intronic mutations that affect exon 10 splicing [45, 48], and;

3 – Structural variations [49].

Exonic variants cause mainly missense mutations, most of which are located within the four MT-binding domains and contribute to a decrease the ability of tau in binding MTs and/or an increase in aggregation propensity [50].

Figure 1-1. *MAPT* gene

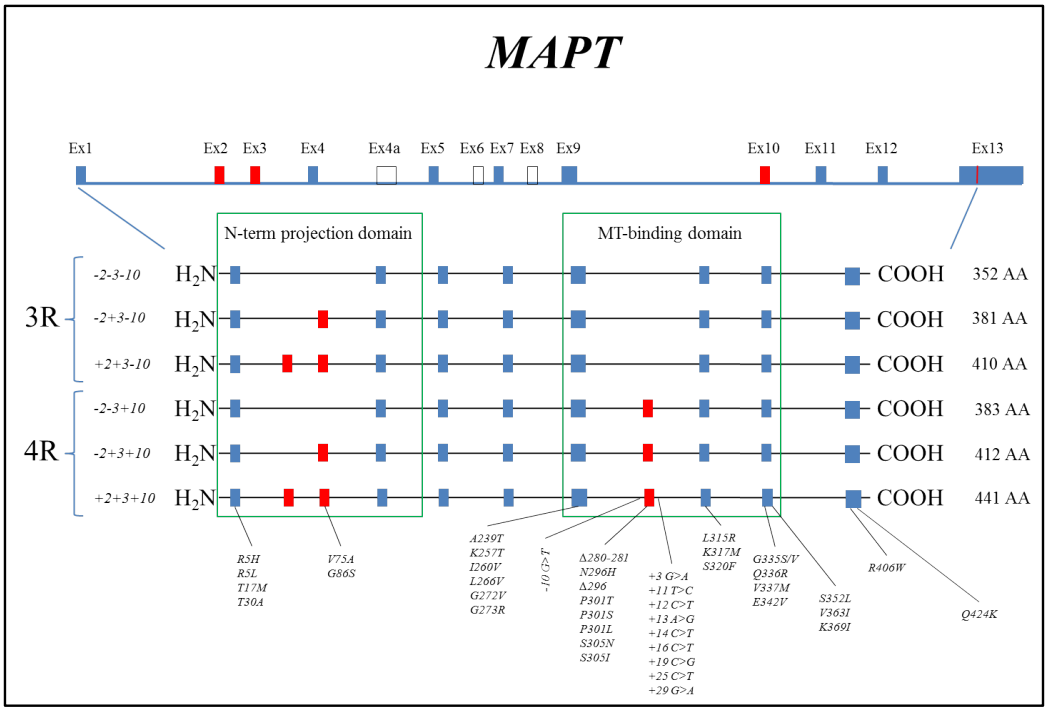


Figure 1-1. The *MAPT* gene consits of 13 exons of which 3 (exons 2, 3 and 10) are alternatively spliced and 3 are rarely transcribed (exons 4a, 6 and 8). The 6 isoforms of MAPT are here represented showing the 3 and 4 repeats (3R and 4R) that compose the MT-binding domain. The most important reported mutations are shown with amino-acid change and position. Figure taken from [51].

Several missense mutations (Ala152Thr, Pro332Ser, Gly366Arg and Pro364Ser) were recently isolated and shown to negatively affect the interaction between tau and the MTs [52-54]. Other missense mutations (Arg5Leu, Lys257Thr, Ile260Val, Pro301Leu, Pro301Ser, Gln336Arg, Val337Met, and Arg406Trp) promote tau's self-aggregation [50], whilst a minority of missense mutations (Asn279Lys, Asn296His, Ser305Asn),

together with deletions (ΔK280, ΔN296) and intronic mutations (intron 9 g(-10)t and intron 10 +3, +11, +12, +13, +14, +16) affect exon 10 splicing and, therefore, the ratio between 3R/4R and between the different isoforms [47, 50].

Features such as mental retardation and learning disabilities in children were reported in presence of structural changes at the *MAPT* locus [55-57]. A more detailed and critical assessment of structural variability at the *MAPT* locus and its association with FTD is discussed in **Chapter 4**, section *4.2.2.4*.

The overall frequency of *MAPT* mutations in FTD is 2-11% [25]. The mean onset age in FTD cases caused by *MAPT* mutations is ~55 years and disease duration is on average 9 years (5-20 years) [17]. *MAPT* variability is mainly associated with bvFTD whilst occasionally (and rather rarely) with SD (Pro301Leu) [58] or PNFA (Val363Ile and Gly304Ser) [59, 60].

1.2.4.2 GRN

The progranulin gene (*GRN*) is located on chromosome 17q21.32 and codes for the 593 amino-acid long protein PGRN, which is a precursor of granulins (GRNs) [61] (**Figure 1-2**). PGRN is expressed in epithelial and hematopoietic tissues and within the cerebral cortex, cerebellum, and hippocampus in the central nervous system (CNS) [62]. From a functional perspective, it is involved in cell cycle progression, cell growth regulation, wound healing and inflammatory processes [63]. Interestingly, PGRN acts as an anti-

inflammatory agent, whilst GRNs have pro-inflammatory effect [64]. Finally, PGRN is known to activate several kinase-dependant signalling cascades, the vascular endothelial growth factor (VEGF), and, not least, to play a role in brain development [65]. The first *GRN* mutations in FTD were reported in 2006 [66, 67]. Currently, more than 140 variants have been reported [47] (**Figure 1-2**). Approximately 65 of these lead, mainly, to non-sense, frameshift, missense, and intronic mutations [47, 61, 65].

Figure 1-2. *GRN* gene

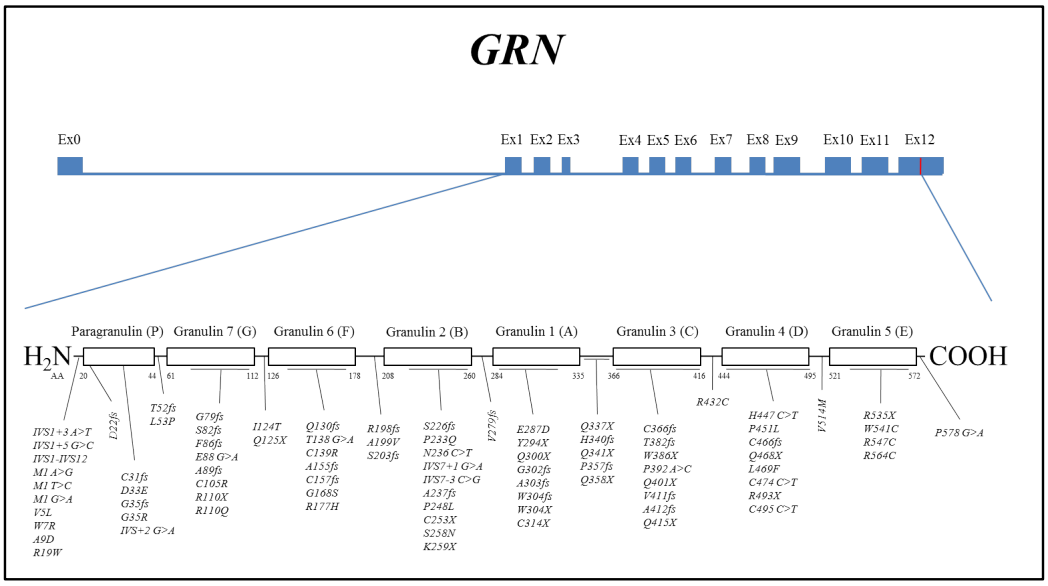


Figure 1-2. The *GRN* gene consists of 12 coding exons and one non-coding exon (exon 0). PGRN protein is precursor of granulins. The most important reported mutations are shown with amino-acid change and position. Figure adapted from [51].

Non-sense, frameshift, and intronic (IVS1 +5G>C) mutations cause haploinsufficiency (a decrease in the amount of translated protein) due to either mRNA non-sense mediated decay (NMD) or nuclear degradation of transcripts retaining the first intron of *GRN* [61]. Also some missense mutations contribute to a decreased secretion of PGRN [68]: Ala9Asp [69], Pro248Leu and Arg432Cys [70] have been reported causing a functional haploinsufficinecy due to cytoplasmic missorting [61, 70]. Other missense mutations can affect PGRN functionality (Cys105Arg, Ile124Thr, Gly168Ser, Pro248Leu, Arg432Cys, Trp541Cys, and Arg556Cys) [61, 68] or its structure and stability (Ile124Thr, Pro248Leu, Ser258Asn, Ala324Thr, and Arg432Cys) [61]. Structural variations such as large deletions (54-69kb comprising *GRN*) [61] or removal of exons 1-11 [71] can also lead to PGRN loss of function. Recent studies reported novel variants/mutations such as IVS9+1delG, identified and characterised in a patient with bvFTD and asymmetrical parkinsonism [72], Ala266Pro associated with reduced PGRN plasma levels [73] and Cys157LfsX97 in a patient diagnosed with PNFA [74].

The overall frequency of *GRN* mutations in FTD is 5-11% [25]. The mean onset age is ~60 years (35-89) and the average duration of disease is 8 years (3-22 years) [17]. Individuals carrying *GRN* mutations manifest mainly apathy and social withdrawal as behavioural symptoms, whilst cases with language impairment are by far less frequent but can also occur [17]. *GRN* mutations have been associated with asymmetrical frontal, temporal, and inferior parietal lobe atrophy [75] and, at the molecular level, mainly with tau-negative and TDP-43 positive inclusions (FTLD-TDP type A).

1.2.4.3 C9orf72

The chromosome 9 open reading frame 72 gene (*C9orf72*) encodes the uncharacterised protein C9orf72. The two isoforms derived from alternative splicing of exons 1a and 1b (481 or 222 amino-acid long protein, respectively) are ubiquitously expressed in kidney, lung, liver, heart, testis and in various brain regions such as frontal cortex and cerebellum. In cells, C9orf72 protein locates to the cytoplasm and the nucleus [76, 77].

Several studies reported a link to chromosome 9, locus 9p21.2, in amyotrophic lateral sclerosis (ALS) and FTD-ALS cases [78-82]. This locus comprises kinase activator MOB 3B (*MOB3B* or *MOBKL2B*), interferon kappa (*IFNK*) and *C9orf72*. No coding variants were identified in these genes [78, 83] until a GGGGCC hexanucleotide repeat expansion in the promoter region of isoform 1a and in intron 1 of isoform 1b in *C9orf72* was reported as the reason for the association with 9p21.2 [76, 77].

To date, the repeat expansion seems more frequent in familial cases, such as familial ALS (fALS), fFTD-ALS or fFTD, and less frequent in sporadic cases with the same diagnoses. Further, it seems that European and American (with European background) populations carry the expansion the most, especially, northern European such as Finnish and Swedish populations [39]. In FTD cases, the vast majority of carriers present with bvFTD (>60%), occasionally with PNFA and very rarely with SD.

Frequencies of the repeat expansion across different studies (and populations) are summarized in **Table 1-6**.

Table 1-6. Summary of features associated with *C9orf72* expansion

Study cohort (ethnicity)	Study size of FTLD cases (n)	Frequency of expansion in FTLD cohort (%)	Diagnosis of positive cases	Status of positive cases (Sporadic/ Familial)	Pathology of positive cases	References (adapted from)
Finnish	75	~29	bvFTD (n=16) , PNFA (n=6)	Familial (n=8)	N.A	[77]
White (92), Asian (1)	93	~18	N.A	N.A	FTLD-TDP	[76]
White (366), Asian (5), Black or African American (3)	374	~3	clinical FTD	Sporadic	N.A	
		~12		Familial		
Flanders-Belgian	305	16	N.A	Familial	N.A	[84]
		~4	N.A	Sporadic	N.A	
UK	398	~8	bvFTD(n=19), PNFA (n=3), SD/FTD (n=1), FTD-MND (n=9)	Familial (n=20) , Sporadic (n=11) , Data N.A (n=1)	FTLD-TDP(n=4), FTLD-tau (n=1)	[85]
Dutch	353	~12	bvFTD(n=34) , PPA (n=8)	Familial (n=37) , Sporadic (n=5)	FTLD-TDP (n=10)	[86]
Finnish	48	~19	N.A	Sporadic	N.A	[87]
UK	543	~6				
Dutch	224	~2				
French	150	~9				
Finnish	27	~48	N.A	Familial	N.A	
UK	170	~16				
German	29	~14				
Dutch	116	~26				
French	50	~44				
Spanish	75	~9	bvFTD	Familial	N.A	[88]
Australian (European descent)	89	~10	bvFTD (n=6), FTD-ALS (n=3)	Familial	N.A	[89]
	22	~41	bvFTD (n=4), AD (n=3), FTD-ALS (n=2)	Familial (n=5), Sporadic (n=4)	FTLD-TDP	
European and North Americans (European descent)	520	~5	bvFTD (n=11), FTLD unspecified (n=8), FTLD-MND (n=8)	Familial (n=22), Sporadic (n=5)	N.A	[90]
Combined European and Flanders-Belgian cohorts	1118	~6	bvFTD (n=22), PNFA (n=1), Others (n=48, Data N.A)	Familial (n=33) , Sporadic (n=22), Others (n=16, Data N.A)	FTLD-TDP (n=11)	[39]
Western European (meta-analysis)	2668	~10	N.A	Familial (n=494), Sporadic (n=167), Others (n=2007, Data N.A)	N.A	[39]

Table 1-6. Frequency, clinical and neuropathological characteristics of C9orf72 expansion carriers from different FTLD studies. **Abbreviations**: N.A= Data not available; C9orf72=chromosome 9 open reading frame 72; FTLD=frontotemporal lobar degeneration; bvFTD=behavioural variant FTD; SD=semantic dementia; PNFA=progressive nonfluent aphasia; SD/FTD = semantic dementia with frontal features; FTD/MND = Frontotemporal dementia with motor neuron disease; FTD-ALS = frontotemporal dementia with amyotrophic lateral sclerosis; AD= Alzheimer's disease; FTLD-TDP=frontotemporal lobar degeneration with ubiquitin/TDP-43 pathology (that is subdivided in Types A, B, C, D); FTLD-tau=frontotemporal lobar degeneration with tau pathology.

The pathology associated with *C9orf72* falls within the FTLD-TDP group, specifically, types A and B [31, 84-86, 89, 91] with some unique features such as the ubiquitin/p62 positive and TDP-43 negative inclusions in cerebellum and hippocampus [92]. The ubiquitin/p62 positive inclusions occur as NCI in cerebellum (molecular and granular layers) and in the hippocampus (pyramidal layer), and as NII in the cerebellum (just granular layer) and hippocampus (pyramidal layer) [86, 93, 94] [95]. Moreover, concomitant ubiquilin positive inclusions [96] or tau pathology have been described in some expansion carriers with typical expansion associated pathology [97, 98].

Mean age at onset is ~55-57 years of age [76, 84-86] (**Table 1-6**) and, among the clinical features, a psychotic component has been constantly described across studies [85, 99]. The neuroimaging profile shows symmetrical atrophy in frontal and temporal neocortex, cerebellum and hippocampus (hippocampal sclerosis) [39, 92, 95]. The pathogenic mechanisms of *C9orf72* expansion remain poorly understood. However, it has been hypothesized that the *C9orf72* expansion may lead to haploinsufficiency by affecting transcription or splicing, or to RNA toxicity by sequestering non-identified RNA binding proteins, or to the aberrant translation and accumulation of poly-GA/GP/GR (glycine-alanine/glycine-proline/glycine-arginine poly-dipeptides) by non ATG-initiated translation [100, 101]. Specifically, the two mechanisms involving RNA toxicity as well as the aberrant di-peptide repeat (DPR) deserve further consideration. The former (RNA toxicity) is supposedly a consequence of aberrant interaction between the expansion and RNA transcripts leading to the accumulation of intranuclear RNA foci in neurons [76, 102-105]. The downstream effect of the formation of RNA foci has been suggested

affecting major RNA processing thus impacting vital cellular physiological processes such as the RNA metabolism [106]. The latter (DPR) has been proposed as a consequence of unconventional repeat-associated non-ATG (RAN) initiated translation of the G_4C_2 repeat. Based on immunohistochemical evidence, DPR accumulates in neurons of the neocortex, hippocampal pyramidal neurons and cerebellum [104]. DPR represents a unique and typical feature of the *C9orf72* expansion carriers, specifically, as part of the NCI-TDP-43-negative inclusions in the cerebellum [104]. Nevertheless, although being undoubtedly a pathological hallmark associated with presence of *C9orf72* expansion, the direct implication of DPR in neurodegeneration, specifically in terms of clinico-pathological correlations, has been, to date, unclear and controversial [107].

In summary, the pathogenic role of the *C9orf72* expansion in FTLD (as well as in ALS and the FTD-ALS spectrum) is currently not elucidated, mainly owing to a limited number of studies assessing correlation between clinical and pathological aspects of the disease, and will thus need to be further studied and characterised.

A critical and updated assessment of the *C9orf72* expansion is detailed in **Chapter 5**, section *5.2.3.3*.

1.2.4.4 VCP

The valosin-containing protein gene (*VCP*) maps to chromosome 9p13.3 and encodes the 806 amino-acid long protein VCP. VCP is expressed in most mammalian tissues [108]

and is involved in maintaining the balance between protein synthesis and degradation [109] as well as in various processes such as cell cycle regulation, post-mitotic Golgi reassembly, suppression of apoptosis, DNA-damage response and protein degradation [110-114].

Mutations in *VCP* cause autosomal dominantly inherited inclusion body myopathy (IBM) associated with Paget's disease of the bone (PDB) and frontotemporal dementia (FTD) or IBMPFD [115, 116]. Currently, more than 15 mutations have been reported in *VCP* [47], being Arg155Cys the first pathogenic missense mutation reported in FTD [117]. Other missense mutations were identified in selective domains of the VCP protein: The amino-terminal domain (Ile27Val, Arg93Cys, Arg95Cys/Gly, Pro137Leu, Gly157Arg and Arg159His/Cys), the L1 linker connecting the amino-terminal and D1 domains (Arg191Gln and Leu198Trp), the D1 domain (Ala232Glu, Thr262Ala, Asn387His) and the L2 linker (Ala439Ser) [115]. *VCP* variability has phenotypically also been associated with early signs of language impairment [118] and ALS [119]. To date functional studies have been partially equivocal, but it seems sensible to consider that *VCP* mutations negatively affect autophagic structures and processes [115].

IBMPFD patients present three main phenotypic signatures: Either disabling weakness (mean onset age 45), or osteolytic lesions with PDB or language and behavioural impairment (mean age of onset 54) [116]. IBMPFD pathology was originally reported being ubiquitin-positive and tau-, β-amyloid- and α-synuclein-negative (beside a few cases with law density staining for the latter proteins) [120]. In IBMPFD the ubiquitin

positive inclusions are abundantly intranuclear (NII), and may also be identified in DN and as NCI [120]. In addition, and relevantly, the ubiquitin-positive inclusions were reported to co-stain, thus co-localise, with TDP-43 predominantly as NII, feature which is consistent with FTLD-TDP Type D [25, 110] [120].

1.2.4.5 CHMP2B

The charged multivesicular body protein 2B gene (*CHMP2B*) is located on chromosome 3p11.2 and codes for the 213 amino-acid long protein CHMP2B that holds a coiled coil domain at the amino-terminus and a microtubule interacting transport (MIT)-interacting region (MIR) at the carboxy-terminus. CHMP2B is a subunit of the endosomal sorting complex required for transport III (ESCRT-III) involved in sorting and trafficking surface receptors or proteins into intraluminal vesicles (ILVs) for lysosomal degradation [35]. CHMP2B is expressed in frontal and temporal lobes, cerebellum and hippocampus [121].

A familial FTD case was linked to chromosome 3 (FTD3) [122] and later two variants, $CHMP2B^{intron5}$ and $CHMP2B^{\Delta 10}$, were reported co-segregating with disease in that family [121]. To date, 12 variants have been reported [47] in FTD, ALS and CBS cases [34, 121, 123-127] [128]. Although the majority of the variants seem non-pathogenic three have been associated with disease or aberrant phenotype [121, 127, 129].

With a mean onset age of 57 years, FTD3 patients present clinically with personality change, hyperorality, dyscalculia, a range of speech disturbances and dystonic postures

[35]. Brain atrophy affects, symmetrically, the frontal and temporal cortex and TDP-43 negative and ubiquitin and/or p62 positive neuronal inclusions are identified in the dentate gyrus or sparse in the frontal cortical areas (FTLD-UPS) [35].

A critical and updated assessment of *CHMP2B* is detailed in **Chapter 5**, section *5.2.1.2*.

1.2.4.6 TDP-43

The TAR DNA binding protein gene (*TARDBP* or *TDP-43*) is located on chromosome 1p36.22 and encodes the 414 amino-acid long protein TDP-43 that holds two RNA recognition motifs and a carboxy-terminal glycine-rich region [130]. Wild-type (WT) TDP-43 is highly conserved and ubiquitously expressed in different tissues, including the central nervous system (CNS) [130]. TDP-43 localises to the nucleus of neurons and glial cells, and is involved in exon skipping of several genes and in the biogenesis of mRNA [130]. Over 40 variants have been reported in *TDP-43* [47] of which >30 map to the glycine-rich region. Genetic screening of *TDP-43* in FTD cases has been controversial. Initially, variants were not found in FTD [131] [132] and it was suggested that TDP-43 pathology may be a consequence rather than a cause of FTLD [132]. Nevertheless, two studies reported a missense change (Lys263Glu) and concomitant TDP-43 pathology in a FTD/PSP/chorea patient [133], and a 3'-UTR (untranslated region) variant in *TDP-43* associated with increased TDP-43 mRNA levels in FTLD-TDP and bvFTD-ALS cases [134]. Also, the missense change Asn267Ser was identified in bvFTD and CBS cases

[135] [136] [137], and Ala382Thr was associated with a unique phenotype, comprising FTD, ALS and parkinsonian features [138].

Functional studies of 3 recurrent missense mutations Ala315Thr, Gly348Cys and Ala382Thr did not show differences with wild-type-TDP-43 in transfected cell lines, whilst a peculiar peri-nuclear localization and aggregation of TDP-43 in motor neurons were observed in living zebra-fish embryos [139]. In addition, investigation of amino- and carboxy-terminal TDP-43 fragments in mammalian cells and cultured neurons showed that, particularly, carboxy-terminal fragments can negatively affect neuronal differentiation and interfere with normal function of full length wild-type-TDP-43 protein, contributing, possibly, to neurodegenerative processes [140]. Taken together, these results support the notion that *TDP-43* abnormalities may associate with toxic gain- or loss-of-function.

In summary, the association of *TDP-43* genetic variability with FTD remains equivocal; in fact most of *TDP-43* mutations seem mainly related to ALS cases [141]. Clearly, further screening is needed in order to shed light on *TDP-43* genetic variability and its role in FTLD.

1.2.4.7 FUS

The fused in sarcoma gene (*FUS*) is located on chromosome 16p11.2 and encodes the 526 amino-acid long protein FUS, which is part of the heterogeneous nuclear

ribonucleoprotein (hnRNP) complex and is involved in pre-mRNA splicing and in the export of fully processed mRNA to the cytoplasm. Mutations in *FUS* were reported in familial ALS (fALS), whilst no variants were found in sporadic ALS (sALS) cases [142, 143]. These reports suggested *FUS* mutations as potential cause of fALS with a frequency of 4%, second only to superoxide dismutase-1 gene (*SOD-1*) mutations for fALS cases. To date over 40 variants have been reported for *FUS* gene [47], but their pathogenicity seems to some extent controversial. Findings support the notion that exon 15 is highly polymorphic and variability in this gene mainly associated with fALS rather than sALS [144]. Further investigation of *FUS* in FTD led to discordant results: Despite the identification of variability in single cases such as a Arg521Cys mutation in one FTD case [144] and a missense mutation, Met254Val (predicted, *in silico*, to be pathogenic), in one FTLD patient [145], other studies mainly rejected the notion of association between *FUS* and FTD as in the case of the screening of exon 15 in a cohort of ~220 FTD cases [146], or a haplotype analysis (tagging the SNPs rs741810 and rs1052352) that failed to identify haplotypes associated with FTLD in sporadic FTD [147]. In addition one of the insertion/deletion (indel) reported as pathogenic in the original reports [142, 143] was later also isolated in normal controls [145]. All these reports taken together may suggest that *FUS* is unlikely to be a susceptibility factor for FTLD and probably caution is needed in ascribing pathogenicity to *FUS* variants in neurodegenerative diseases. However, functional studies on mutations such as Arg521Gly, Arg521Cys and Arg521His showed aberrant retention of the mutated protein in the cytoplasm (Arg521Gly) [142] or increased cytoplasmic localization of the protein (Arg521Cys and Arg521His) [143].

The penetrance and relevance of *FUS* variants in FTD is to date rather unclear. From a functional perspective, despite the identification of FUS pathology in a small minority of FTLD cases (FTLD-FUS pathology) and a correlation with aberrant cellular distribution of the FUS protein for some of the most relevant missense changes (Arg521Gly, Arg521Cys and Arg521His), the exact role of *FUS* in neurodegeneration will need to be further elucidated.

1.2.4.8 TMEM106B

TMEM106B gene is located on chromosome 7p21 and encodes the 274 amino-acid long protein TMEM106B, which has recently been described as a type 2 integral transmembrane protein characterised by an amino-terminal domain located within the cytoplasm and a carboxy-terminal domain that may also locate to the cytosol or protrude to the extracellular microenvironment [148]. Association between *TMEM106B* and FTLD was established through a genome-wide association study (GWAS) on 515 FTLD-TDP pathologically confirmed cases [41] and three top SNPs (rs1020004, rs6966915 and rs1990622) were reported being associated with increased expression levels of TMEM106B in brain [41]. Specifically, an increase of TMEM106B mRNA levels was shown in post-mortem human brain tissue in association with rs1990622, especially, in *GRN* mutation carriers [41, 149]. It has been suggested that increased expression of TMEM106B may modulate GRN plasma levels and, thus, contribute to pathogenic processes [150]. Several studies support the idea that variability in *TMEM106B*

predisposes to FTD, especially, in *GRN* mutation carriers [41, 151-153]. However its exact pathogenic role still remains widely unclear. The three most strongly associated SNPs (rs1020004, rs6966915 and rs1990622) and the missense mutation Thr185Ser (rs3173615) were extensively tested in different cohorts with the aim of identifying correlations between TMEM106B expression levels, GRN protein levels and disease onset age and progression [151, 152, 154]. The missense change T185S (rs3173615), predicted to be possibly damaging (PolyPhen-2, **Chapter 2**, section *2.3.3.1*), was suggested to be possibly pathogenic in FTLD or FTLD-TDP cohorts in presence of *GRN* variability [152]. The SNP rs1990622 showed linkage disequilibrium (LD) with other SNPs within 10kb of the *TMEM106B* gene locus including complete LD with rs3173615. As well, in sporadic FTLD-TDP cases with *GRN* mutations, carriers of the homozygous allele T (rs1990622) showed association with thirteen years earlier onset and lower GRN plasma levels [151]. For SNP rs1020004 an association between genotype and disease course was identified in that the homozygous genotype A seemed to lead to shorter disease duration [41]. TMEM106B was shown to be involved in regulating lysosomal morphology and function in FTLD-TDP cases with *GRN* mutations [150] supporting a previous study suggesting a link between TMEM106B and GRN biology through late endosome and lysosome processes [148].

In summary, variability in *TMEM106B* is associated with FTLD-TDP and *GRN* mutation carriers as well as, to a certain extent, with features such as poor cognitive performance in ALS cases [153] and AD [155]. The main pathological signatures are Neuronal cytoplasmic inclusions and short dystrophic neurites in neocortex/cortical layers,

corresponding to FTLD-TDP Type A pathology [31]. It seems that *TMEM106B* might act as a disease modifier affecting age of onset and modulating PGRN protein levels [151].

The heterogeneous clinical, pathological and genetic features of FTD are summarised in **Figure 1-3**.

Figure 1-3. Clinical, pathologic and genetic features associated with FTD

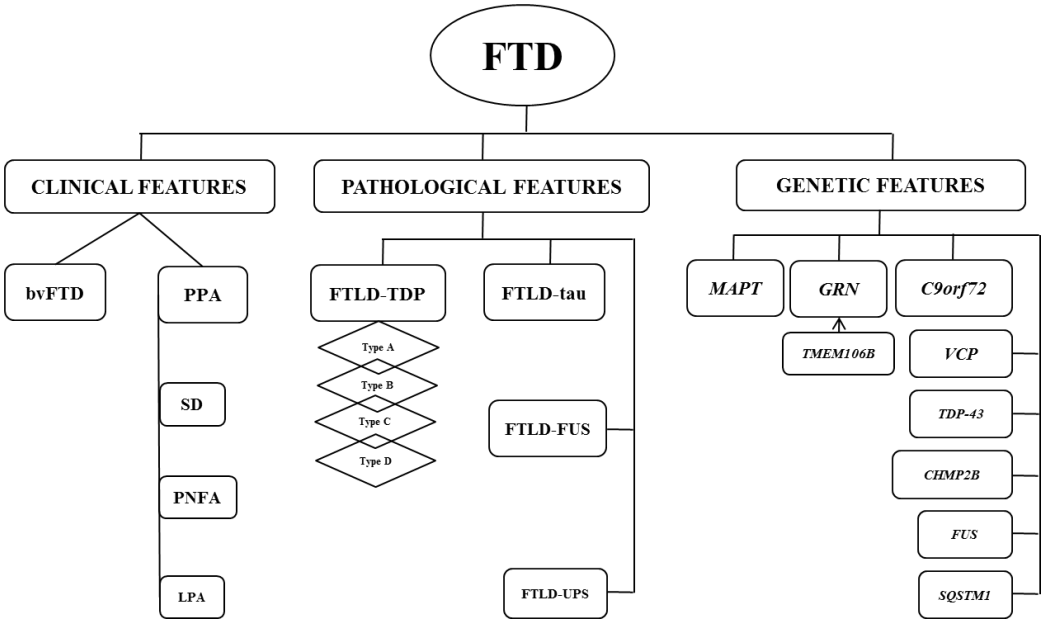


Figure 1-3. The main clinical, pathological and genetic features associated with FTD are summarized. In each section the most common features are located at the top with bigger font size which is decreasing once features become less common or frequent. In the Genetics section *TMEM106B* is shown as a possible modifier of *GRN*. **Abbreviations**: FTD=frontotemporal dementia; bvFTD=behavioural variant FTD; PPA=primary progressive aphasia; SD=semantic dementia; PNFA=progressive nonfluent aphasia; LPA=logopenic progressive aphasia; FTLD-TDP=frontotemporal lobar degeneration with ubiquitin/TDP-43 pathology; FTLD-tau=frontotemporal lobar degeneration with tau pathology; FTLD-FUS=frontotemporal lobar degeneration with FUS pathology; FTLD-UPS=frontotemporal lobar degeneration with ubiquitin/p62 pathology (which are elements of the Ubiquitin Proteasome System); MAPT=microtubule associated protein tau; GRN= progranulin; C9orf72=chromosome 9 open reading frame 72; VCP=valosin containing protein; TDP-43=TAR DNA binding protein 43; CHMP2B=charged multivescicular body protein 2B; FUS=fused in sarcoma; TMEM106B=transmembrane protein 106 B. Figure adapted from [26].

1.3 Progressive supranuclear palsy (PSP)

1.3.1 Relevance and statistics of the disease

Progressive supranuclear palsy (PSP) was first described in 1963-1964 by Steele, Richardson and Olszewski [156].

The disorder is characterised by early postural instability, supranuclear gaze palsy, and cognitive decline and can occur as early as 40 years of age, but the average onset age seems to be around 63 years of age [157]. The incidence of PSP has been estimated being 5.3 or 6.4 per 100,000 [158, 159]. Depending on disease morbidity and severity the average disease duration is about 7 years [160]. Moreover, PSP shows a slight predominance in males (64%) [157].

PSP rather presents as a sporadic form (~85% of cases), whilst few familial forms with autosomal dominant inheritance have been described [161].

1.3.2 Diagnostic criteria and clinical features

The main clinical signatures of PSP are vertical gaze palsy, pseudobulbar palsy, axial rigidity and, in some cases, cognitive impairment [156]. Additional features such as sudden falls and supranuclear palsy are included in the clinical diagnostic criteria. Bradykinesia (slowed ability to start and continue movements, and impaired ability to

adjust the body's position), decreased fine motor skills and hypophonia (a weak voice due to incoordination of the vocal muscles) are among the early signs of PSP [157]. Gait (manner of walking) and visual problems are highly prominent symptoms in PSP representing features that help in diagnosing the disorder with higher specificity and sensitivity. In fact, gait difficulties that eventually result in lurching [156, 162], show a faster progression if compared to Parkinson's disease (PD) representing a distinctive factor. Falls develop within the first year after disease onset and the symptoms range from unexplained falls to visual deficits (e.g. blurred vision), cognitive impairment (e.g. apathy or loss of executive functions or language dysfunction) and parkinsonian symptoms [157].

Given the overall picture of the disease summary the essential diagnostic criteria used to diagnose PSP are as follows:

1 – Progressive akinetic-rigid syndrome appearing after age 40;

2 – Progressive vertical supranuclear gaze palsy;

3 – Slowing of vertical saccades, and;

4 – Progressive and prominent postural instability with falls in the first year of illness [157] and at least 1 and 2 or 1 and 3 criteria need to be met for an individual to be diagnosed with PSP.

Together with these pivotal and relevant symptoms additional features can support clinical diagnosis such as axial and appendicular rigidity, symmetry in symptomatology,

paucity of tremor and poor response to levodopa treatment [157]. It needs to be acknowledged that, although these clinical features are well described and may be easily recognisable by an expert physician, there is normally an average delay of 4 to 5 years between the onset of symptoms and diagnosis [157].

1.3.3 Neuropathology

Atrophy patterns in PSP grossly affect the midbrain, particularly the tectum [160]. In addition, the substantia nigra shows loss of pigmentation, and the subthalamic nucleus as well as the superior cerebellar peduncle, pontine tegmentum and dentate nucleus may be atrophic [160].

The molecular pathology of PSP is strictly correlated with tau pathology, hence PSP belongs to the group of disorders called tauopathies [163]. Topographically, tau pathology distributes in different brain areas such as the pallidum, the subthalamic nucleus, the red nucleus, the substantia nigra, the pontine tegmentum, the striatum, the oculomotor nucleus, the medulla and the cerebellar dentate nucleus [163, 164].

In PSP, tau pathology is defined by the accumulation of hyperphosphorylated tau protein inclusions that lead to the formation of neurofibrillary tangles (NFTs) in the cell body and neuropil threads in the dendrites of neurons [26]. NFTs and neuropil threads are composed of hyperphosphorylated 4R-tau isoforms, making PSP a predominantly 4R-tauopathy, implying to a constitutive predominant transcription and translation of this

isoform that has been associated with the H1 *MAPT* haplotype (specifically the H1c clade) [166]. In addition, tau pathology also affects glial cells (astrocytes and oligodendrocytes) with features that are relatively unique to PSP [160, 165]. The argyrophilic tau-positive inclusions that have been described in astrocytes and oligodendrocytes [160] appear as irregular fibrous tufts in astrocytes (thus called tufted astrocytes) in the motor cortex and striatum; conversely they appear as perinuclear fibers, called coiled bodies, together with thread-like processes in oligodendrocytes in cerebellar white matter [160, 167].

There are several variants that separate classical PSP from atypical PSP (aPSP) or PSP-parkinsonism (PSP-P) due to differences in severity or regions of pathology or clinical features while the common link between them is tau pathology [163]. To date there are no definite diagnostic criteria to distinguish the atypical PSP from the classical PSP, however, it has been established that tau pathology is more severe in classical PSP rather than in atypical or PSP-P [163].

Another variant, known as PSP-pure akinesia with gaze freezing (PAGF) shows characteristics of gait disturbance, micrographia, hypophonia, and eventual gait freezing associated with neuronal loss and atrophy in the globus pallidus, substantia nigra and subthalamic nucleus [168, 169].

While other subgroups of PSP show different topography concerning the molecular pathology, tau pathology is more severe than classical PSP, manifested through asymmetric dystonia, apraxia and cortical sensory loss (PSP-CBS). The other subgroup is

characterised by apraxia of speech (PSP with progressive nonfluent aphasia [PSP-PNFA]) [170, 171].

Due to the improved investigation methods, tau pathology has been identified also in the cortical areas with severe tau pathology in the mid-frontal and inferior parietal cortices in PSP-CBS and in the temporal cortex and superior frontal gyrus of patients with PSP-PNFA when compared to classical PSP [171]. Nevertheless, it was recently shown that, although rarely, also accumulation of TDP-43 is detectable in PSP cases, specifically, within the limbic system [172].

1.3.4 Genetics

Belonging to the group of tauopathies, it is not surprising that PSP has been consistently associated with common variation at the *MAPT* locus [173-175] [176]. This is reinforced by the robust genetic association with the *MAPT* H1 haplotype [173, 174, 177-179] that replicated in the recent GWAS with odds ratio greater than 5 [180].

While there is a strong genetic link between PSP and chromosome 17, specifically the locus encompassing the microtubule associated protein tau gene (*MAPT*), pathogenic coding mutations within the *MAPT* gene seem to be rare. To date, PSP has been associated with only four missense mutations, Arg5Leu, Asn279Lys, Leu284Arg, Gly303Val and the deletion of codon 296 [181-187]. Further, association was reported in *MAPT* with intronic variants such as intron 10+3 [188] and 10+16 [189], and with a silent

mutation Ser305 in exon 10 [190]. Despite the fact that association between *MAPT* H1 haplotype and PSP has been previously described [173, 179], the mechanism leading to disease still needs elucidation as the H1 haplotype is relatively common among populations [160]. The H1 haplotype seems responsible for a 1.5-fold higher 4R-tau expression compared to the H2 haplotype, suggesting that, possibly, the lifelong higher expression of 4R-tau could be associated with the onset of PSP [160]. In view of this, SNPs rs1880753 and rs242557 were shown to associate with PSP by putatively influencing *MAPT* expression in *–cis* [177, 178]. A genetic link for PSP was also identified on chromosome 1q31.1 [186] as well as on chromosome 11p12 although, in the latter study, sample size was relatively small (288 cases vs. 344 controls) [191] and the association has not been replicated to date. Furthermore, a study argued for association between the polymorphism Val380Leu in parkin (*park2*) and PSP in that Val carriers might experience an increase of the risk of developing PSP compared to Leu carriers [192].

Elucidation of the possible functional basis of the H1 haplotype association with PSP has led to suggestions of allele-specific differences in transcription and alternative splicing of *MAPT* and its transcripts, respectively. In the first place, both PSP and CBD tau pathology is predominantly four-repeat tau (4R-tau), i.e. consisting of tau protein isoforms with four microtubule-binding domains as a result of alternative splicing of *MAPT* exon 10. Early work [193, 194] showed relative increases of 4R-tau mRNA and protein [195] in affected regions of the PSP brain and, owing to the surplus of these more fibrillogenic isoforms, they preferentially aggregate to form the characteristic

neurofibrillary inclusions. *In vitro,* and in brain, the H1 allele, compared to its counterpart H2 haplotype, is associated with increased splicing of exon 10 and decreased inclusion of exon 3 [196-199]. Furthermore, analysis of the underlying diversity of the H1 haplotype revealed H1c, a common variant of H1 that specifically drives the association with PSP, with a second common variant, H1b not associated and H2, negatively associated as a protective allele [179, 200]. One study showed both increased transcription and exon 10 splicing of the H1c allele [201] with the A-allele of the SNP rs242557 in the *MAPT* promoter region, that segregates with H1c, associated with increased CSF tau levels [202]. Due to the extended region of linkage disequilibrium (LD) [203] that resulted due to a large chromosomal inversion of the region on chromosome 17q containing *MAPT* [204, 205], several other genes, including N-ethylmaleimide-sensitive factor (*NSF*) and corticotropin releasing hormone receptor 1 (*CRHR1*) are also defined by the H1/H2 dichotomy and are thus likely associated with PSP [176].

H1/H2 allele-specific differences of *MAPT* gene expression have been demonstrated at the transcriptional level *in vivo* and *in vitro* [201, 206], though this has not been consistently replicated and recent work suggested that some expression array-based data have been hampered by errors due to SNPs within the gene-specific probes used to quantify expression levels [207]. In fact, Trabzuni et al failed to show differences in *MAPT* transcription but did show a clear allele-specific exclusion of the alternatively spliced exon 3 on the H1 haplotype [199]. They failed, however, to observe increased splicing of exon 10 on the H1 vs. H2 haplotype as reported in other studies [196, 198, 201] and that was postulated to cause the predominant 4R-tau pathology in PSP.

The majority of studies on brain expression are affected by sampling issues likely to result in significant differences in neuronal and glial cell populations and different outcomes in expression levels. This can only be resolved by analysis of individual cell-types isolated by methods such as laser-capture dissection. Recent work comparing gene expression sets from whole-brain sections versus individual neurons or glia showed significantly different outcomes, whereby predictions of Alzheimer's disease progression based on neuronal gene expression are significantly more accurate than when based on glial or whole-tissue expression [208].

1.4 Genetic studies

1.4.1 Basics of genetic variability

Genetic variability (i.e. difference in genotype) is among the main driving factors at the basis of differences between individuals and populations.

The majority of the variants are benign. However, some can be harmful either in a Mendelian fashion (gene/mutation/disease), e.g. in rare or familial forms of disease as in the case of Huntington's disease [209] or cystic fibrosis [210]; or by contributing to increased risk of developing common/complex forms of disease through an interplay between genetic and environmental factors with variable effect size, e.g. as in the case of late onset Alzheimer's disease (LOAD) [211].

Each disorder has a genetic component (with low, intermediate or full penetrance) and, in the majority of cases, the development of a disease is the result of downstream effects that accumulate with time and, ultimately, manifest in clinical symptoms. A better understanding of the variants in the genome and a better genotype-phenotype correlation are critical for understanding and discriminating the genetic factors that influence the conditions of health and disease.

Today we are at a point in time where we can define and identify almost all types of variations that are present in the genome.

There are 2 categories of variants: The simple nucleotide variations (SNVs) and the structural variations (SVs) (**Figure 1-4**). SNVs, comprising single nucleotide polymorphisms (SNPs) and small insertion/deletions (small indels), affect single or few bases, whilst SVs, comprising copy number polymorphisms (CNPs) that include large indels (100bp – 1Kb) and copy number variations (CNVs) (>1Kb), affect larger genomic regions. Also inversions and translocations are types of SVs.

Figure 1-4. Types of genetic variants

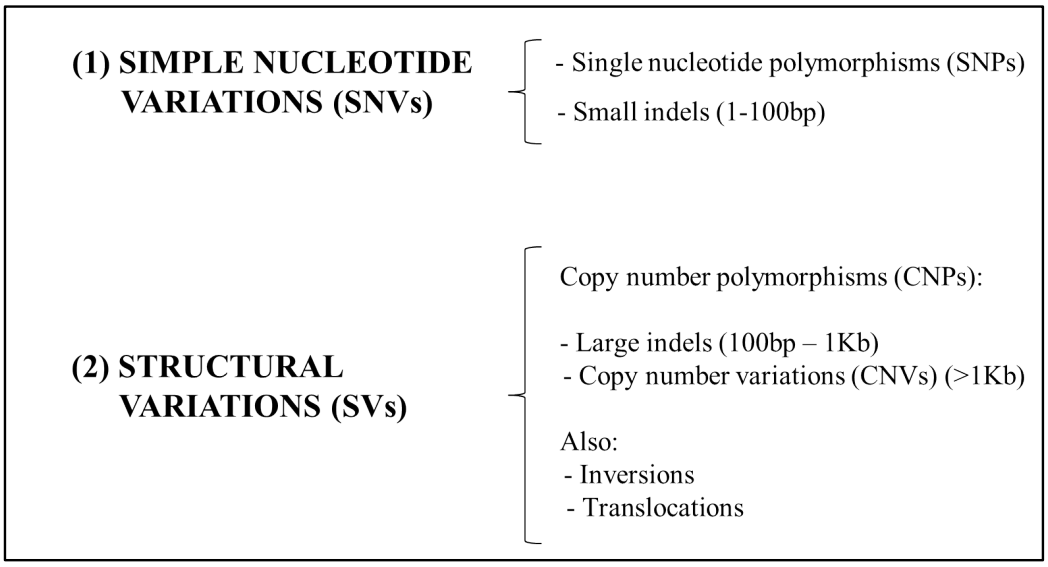


Figure 1-4. Summary of the main types of variants in the genome involving (**1**) single (or few) base pairs (bp) or (**2**) larger areas of the genome.

In the majority of cases, SNVs (**Figure 1-5**) can cause direct changes to proteins (when located in coding regions) or can affect *cis* and/or *trans* gene expression or splicing (when located in intergenic or promoter regions or in introns).

Changes within the coding regions can result in synonymous, missense, non-sense and frameshift mutations. Synonymous changes do not affect the amino-acid sequence of a protein, whilst missense mutations lead to amino-acid change. Non-sense mutations cause a premature truncation of the protein and frameshift mutations cause a shift in the reading frame giving rise to novel translated elements. Missense, non-sense and frameshift mutations can be harmful and exert a pathogenic effect through mechanisms such as loss or toxic gain of function.

Conversely, changes in promoter regions can affect gene expression by negatively modulating the activity of transcription factors, blocking gene expression or causing aberrant gene expression. Changes in introns can affect splicing leading to non-functional or toxic forms of m-RNA and a decrease in protein production (i.e. haploinsufficiency).

Figure 1-5. Characteristics of simple nucleotide variations

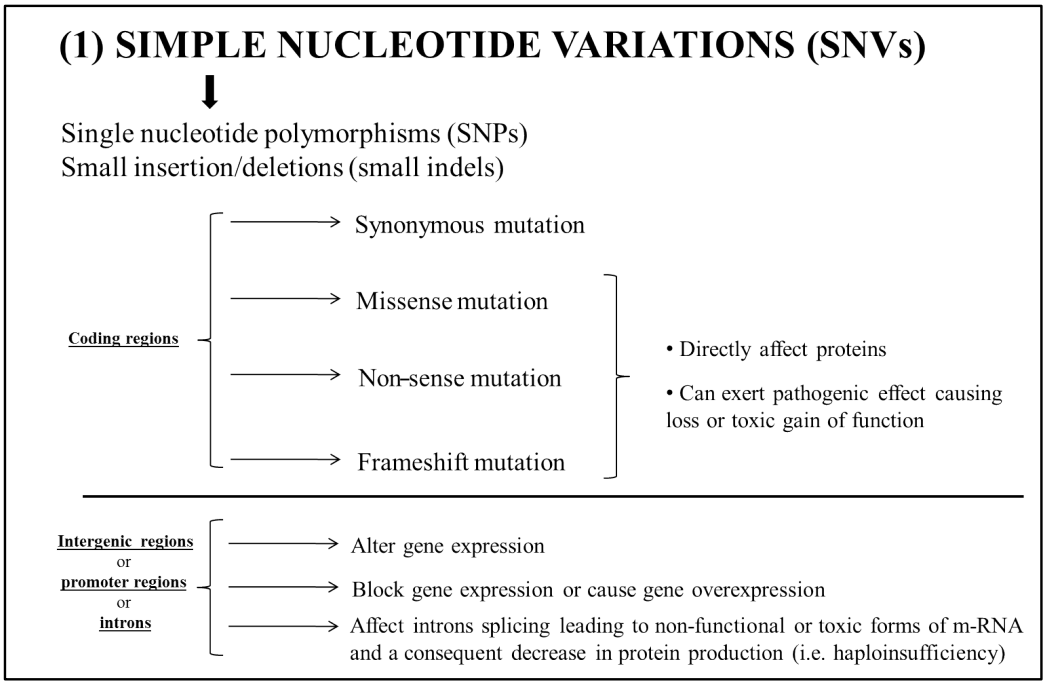


Figure 1-5. Schematic view of the main single nucleotide variations (SNVs) and their related effect.

SVs (**Figure 1-6**) affect larger parts of the genome. These types of variants can cause the loss of portions of DNA (deletions) that, in turn, may lead to haploinsufficiency or aberrant regulation of gene-expression. On the other hand, duplications, which lead to multiple tandem copies of an allele, can cause aberrant phenotypes due to gene over-expression. Duplications can also happen at the level of chromosomes causing over-expression of the set of genes located on that chromosome.

Figure 1-6. Characteristics of structural variations

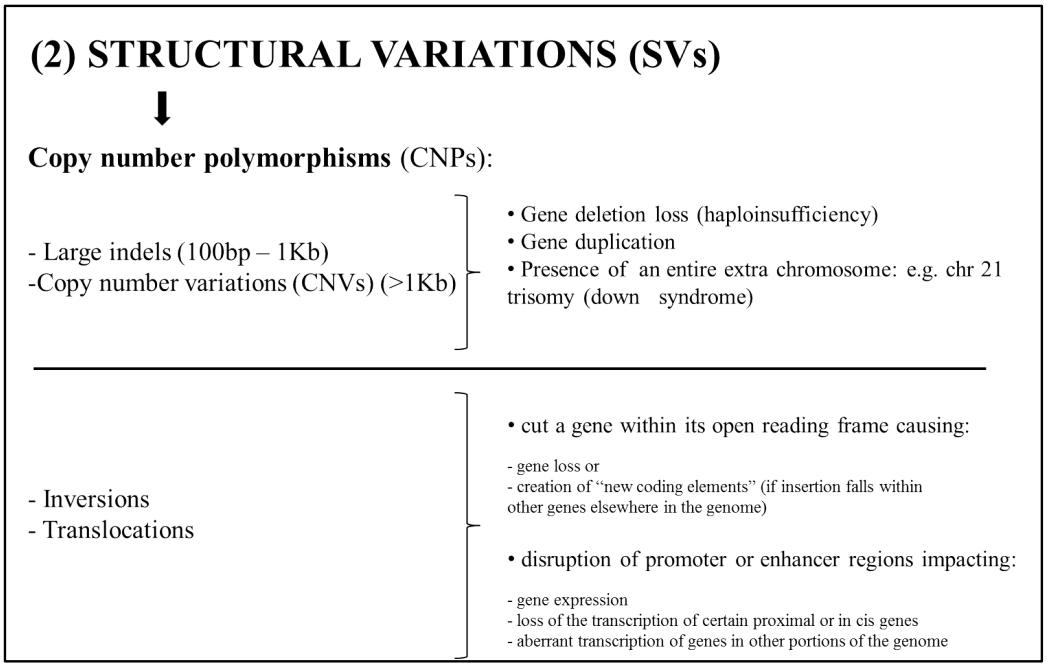


Figure 1-6. Schematic view of the main structural variations (SVs) and their related effect.

1.4.2 The study of genetic variability

Clearly, the first milestone for a better understanding of the human genome was laid in the early 50s' through the biochemical characterisation of the DNA structure [212] (**Figure 1-7**). This discovery provided the basic knowledge for further investigating and understanding how the genetic information is stored, replicated and how it is translated to the next functional level (proteins) [213]. In the late 70s the revolutionary method for sequencing – Sanger Sequencing – represented a paradigm shift in genetic research

because it allowed identification of the exact base sequence of a specific genetic region [214] (**Figure 1-7**). In the early 80s the development of the polymerase chain reaction method (PCR) (**Figure 1-7**) further revolutionized the field of molecular genetics by allowing amplification of fragments of interest (e.g. exons) to perform targeted screening of genetic regions (e.g. sequencing exons to identify coding variants).

Along with the development of these techniques, ground-breaking advances were made in linking traits or diseases to the genome by the study of families (i.e. linkage analysis studies). In the early 80s, gene hunting through positional cloning became a popular approach for identifying specific loci co-segregating only in carriers of a trait/disease in families. This method allowed identification of the causative locus for Huntington's disease (HD), a genetically inherited disease affecting, cognition and psychology [209].

As there was growing awareness in the scientific community that expanding the knowledge on the human genome was fundamental in order to further understand its characteristics as well as its role in health and disease, it was envisioned that sequencing the whole human genome would be the next ambitious endeavour to undertake.

1.4.2.1 Human genome project (HGP) – 1990-2003

The human genome project (HGP) was initiated in the late 80s (http://ornl.gov/sci/techresources/Human_Genome/home.shtml). This international endeavour was driven by the conviction that a deeper knowledge of the genome could, on

one hand, provide clues to better understanding human biology and on the other lead to revolutionary new ways to diagnose, treat, and prevent diseases. HGP accomplished the goal of sequencing the whole human genome providing paramount novel insight into the human genome by:

1 – Revealing that the haploid human genome contains 3.2 billion base pairs;

2 – Revealing that the total number of genes can be estimated being ~25,000;

3 – Revealing that the genome is almost the same (99.9%) across individuals, and;

4 – Mapping 3.7 million SNPs.

These discoveries opened possibilities for entirely new approaches in biological research. In fact, instead of screening one or a few genes at a time it was possible to target the whole genome allowing studying more comprehensively genotype-phenotype correlations and the most fundamental causes of disease.

1.4.2.2 Haplotype map (HapMap) project – 2002-2009

The elucidation of the entire human genome through the HGP provided the basis for the haplotype map (HapMap) project (http://hapmap.ncbi.nlm.nih.gov/). With this international effort DNAs collected from populations of African, Asian, and European ancestry were sequenced to enable the identification of common haplotypes in different populations, worldwide. HapMap revealed that any two people have 99.5% identical

sequences and mapped up to ~10M SNPs with a minor allele frequency (MAF) ≥5%. Moreover, HapMap assessed their allelic and genotypic frequency and distribution among different populations. Considering that sets of nearby SNPs on the same chromosome are inherited in blocks (haplotypes) and that variability within the genome may affect disease risk, by mapping these haplotype blocks and their tag SNPs, HapMap set the basis for the investigation of common/complex diseases by means of genome-wide association studies (GWAS).

Both projects, HGP and HapMap, were fundamental for defining the sequence and the structure of the human genome and, at the same time, for identifying and mapping common polymorphisms. A remarkable advantage inherited from these two pioneering studies was the free availability of the generated data as these were merged and collected in publicly available databases such as dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP/). In addition and finally, HGP and HapMap influenced an unimaginable boost of technology giving rise to the continuous improvement of the platforms and arrays available for genetic screening as well as bioinformatics and biostatistics tools to address the needs associated with data storage, analysis and interpretation. These advancements led to the development of massively parallel sequencing or next generation sequencing (NGS) that enables high coverage and high resolution sequencing of both the whole exome (whole exome sequencing [WES]), and the whole genome (whole genome sequencing [WGS]).

1.4.2.3 1000 Genomes project – 2008-2012

Subsequent to the development of NGS, the 1000 Genomes project was designed to generate an extensive public catalogue of human genetic variation, including rare SNPs (MAF ≤1%), structural variants and their haplotype contexts. Overall, 2,500 unidentified people from about 25 populations around the world were sequenced using NGS technologies. This project was subdivided in three phases each revealing novel and complementary data. Most relevant information came from the first phase that revealed ~15M SNPs and 1.33M indels (>50% of which were novel), and the third that revealed additional ~13K SNPs (70% of which were novel) and 96 indels. In total, the 1000 Genomes project extended the knowledge of common variability up to ~17M variants including SNVs and SVs. The most important outcome of the 1000 Genomes project, besides advancing the pure knowledge of the genome variability and structure, is that the generated data are freely and publicly accessible to provide a robust dataset for the research community to impute genotypes and haplotypes as well as compare the allele frequencies and linkage disequilibrium (LD) patterns each group finds in their own studies.

The HGP, HapMap and 1000 Genome projects have laid the basis of today's deep analysis of the human genome and continuing development of platforms and bioinformatics tools available for the study of genetics of disease (**Figure 1-7**).

Figure 1-7. The study of the genome

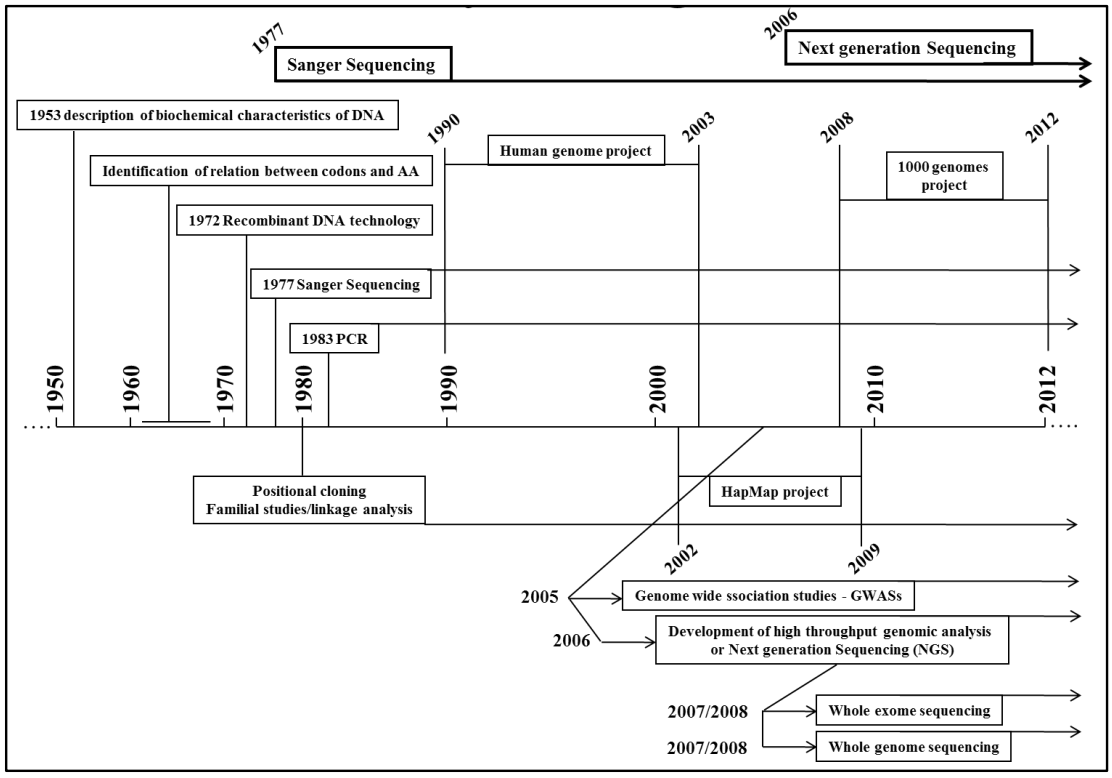


Figure 1-7. Schematic summary of the most important discoveries along with the development of technologies/experimental approaches for the study of the genome and for the study of genetics of disease. AA=amino-acid.

1.4.3 Genome-wide association studies (GWAS)

Although recent (and continuing) improvements of the technologies available for the study of genetics of disease currently provide advanced tools and platforms to scan the exome and the whole genome, the focus of the following sections will be on reviewing the main concepts characterising genome-wide association studies (GWAS) as this investigative approach has been the one on which the project on progressive supranuclear palsy has been based (**Chapter 3**) and the one used to study the genetic underpinnings of frontotemporal dementia (**Chapter 4**) in this thesis.

1.4.3.1 GWAS: Study design

Based on the knowledge acquired through the HGP and the HapMap projects (sections *1.4.2.1* and *1.4.2.2*) along with the progressive development of powerful high throughput genotyping platforms supporting genetic studies of robust size, the design of SNP-based association studies became possible approximately around 2005. As such, the opportunity of high-density genome screening shifted the search for new genes from the traditional candidate gene approach (rare high-risk variants) to the association study approach (common low-risk variants).

GWAS follow the broad hypothesis of "common disease – common variant(s)" and represent a large-scale example of classical cases vs. control studies to assess differences in the allelic frequencies of genotyped (and imputed) genetic markers between the two

study groups. Specifically, differences in the frequencies of the alleles are statistically evaluated for each SNP in order to detect discriminants that may associate with/contribute to disease. Conceptually, GWAS interrogate the genome in an unbiased manner by means of hundreds of thousands of evenly distributed SNPs and allow for the identification of loci that increase susceptibility for disease, i.e. genetic markers within genetic regions with small to moderate effect size. In addition, GWAS is ideal study approach to target common/complex diseases (diseases that are very common among population such as cardiovascular disease, diabetes or Alzheimer's disease) that are the result of the interaction between single or multiple genetic factors and the environment, where both, variably, contribute to the disease.

The workflow of a GWAS is relatively simple and consists of two phases: An exploratory and a replication phase. The exploratory phase (or phase I) is hypothesis free and through massive statistical analyses allows identifying one or more genetic loci that might be associated with the trait under study. Once loci are identified in the discovery phase those that are statistically significant, as well as those that are suggestive of association, need to be selected for replication (phase II) that is to be performed in a novel independent cohort of cases and controls for validation. When and if results of phase I are replicated in phase II it is sensible to infer that most probably the locus/loci that show association contain or are in linkage disequilibrium (LD) with the SNP(s) that is/are responsible for the association.

Generally and finally, after completion of phase II, the loci that are confirmed for association are further investigated through fine mapping, i.e. genotyping a smaller number of SNPs (~10K SNPs) within a smaller region (1-5Mbp) comprising the associated SNP, to identify other associated SNPs and/or, possibly, disease associated haplotypes, or through direct sequencing of all the neighbouring genes implicated by the associated SNP [215].

There is evidence that GWAS are a proven successful method for mapping novel common variants underlying common diseases. Since the beginning of the GWAS era over 1,500 GWASs for over 240 traits or diseases have been published (http://www.genome.gov/gwastudies/, accessed in November 2013).

1.4.3.2 GWAS: Good practice for success

GWAS is a long and complex experimental procedure and there is room for errors and confounding elements to contaminate the final outcome of the study. Errors can occur at different levels, including phenotyping, sample quality, genotyping errors/artefacts, conflicting genetic background within the study cohorts (population stratification) to name the most relevant. However, there are a number of good practices to implement in order to minimise their overall impact. As such, the requirements and the workflow for a successful GWAS can be summarized as follows:

1 – Choice of appropriate genotyping array in order to evenly cover the genome and best target the genetic background of the study population;

2 – Detailed characterisation of the study cohorts through stringent clinical and/or pathological inclusion/exclusion criteria and a well-defined disease phenotype;

3 – Accurate match of cases and controls and large sample and control size in order to increase the power of the study;

4 – Use of stringent quality control (QC) steps prior to and after genotyping. Prior to genotyping excluding poor quality samples is fundamental to avoid genotyping errors/artefacts. After genotyping there are several quality measures to be implemented that target and filter both SNPs and samples to choose those to be included in the association analysis:

4a – SNPs: SNPs with call rates ≥0.95 should be included. This measure can be more stringent (≥0.97-0.99) based, mainly, on study design. To eliminate possible confounding factors, all SNPs with no call, or which are outliers, or that deviate from the Hardy-Weinberg equilibrium law, or that have a MAF<0.01 need to be excluded. These inclusion/exclusion criteria need to be applied for both cases and control sample sets.

4b – Samples: Cases and controls will need to be matched based on ancestry in order to exclude possible false positives simply due to differences in the genetic background of the two cohorts (population

stratification). Samples with missing data for >5% of SNPs, samples that might be related and samples with discordant gender (gender mismatch) need to be excluded because of their high probability of contaminating the results of the association analysis.

After the preliminary QC steps the clean dataset (which means all informative cases, controls and SNPs have been identified and filtered) is used for the association analysis. The latter is performed by means of online free open-source whole genome association analysis toolsets such as Plink (http://pngu.mgh.harvard.edu/~purcell/plink/) and R (http://www.r-project.org/).

In studies evaluating for dichotomous traits (i.e. cases vs. control) the association needs to be tested for its significance and the effect size.

Significance can be assessed through a number of methods including the Chi-squared test with either one degree of freedom (df) (allelic) or two df (genotypic), the Fisher's exact test, or logistic regression. The significance is expressed in p-values for which, currently, an association is considered genome-wide significant when the p-value is $<5x10^{-08}$. Once the association analysis is performed, there is an additional step used to evaluate the impact on results of possible confounding factors. This is the assessment of the distribution of the data through the Quantile-Quantile plots (QQ plots) that allow appreciating inflation/deflation from the expected distribution (**Figure 1-8**). The genomic inflation factor (λ) defines the deviation from the expected distribution under the null model (null hypothesis). Inflation ($\lambda>1$) is generally a signal of possible population

stratification, or an issue of relatedness (sample duplicates), or a technical bias or due to DNA poor quality, whilst deflation (λ>1) is generally a sign of possible phenotype discordance. However, a value of λ~1.05 is considered acceptable in GWAS.

Figure 1-8. Example of quantile-quantile plot

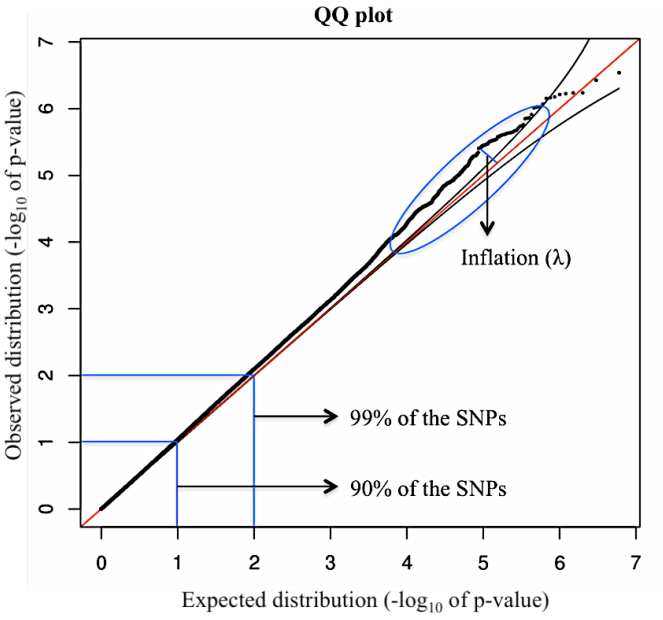


Figure 1-8. Example of a QQ-plot. The red line identifies the expected distribution in concordance with the null hypothesis. The black line represents the observed distribution. The majority of the observed distribution (up to 99%) overlaps with the expected distribution. Only a minority of SNPs (≤1%) deviates from the expected distribution because of presumable association with the trait (disease). However, the deviation underlying true association is expected to be minimal as a major deviation (inflation: λ >1 or deflation λ <1) represents confounding issues (e.g. population stratification).

Effect size is measured in the vast majority of GWAS through odds ratio (OR). An OR greater than 1 generally suggests increased risk, whilst an odds ratio smaller than 1 indicates protection.

1.4.3.3 Key words and definitions

Null hypothesis: The hypothesis that there is no association between genotype and phenotype (i.e. no association between any allelic frequency and disease).

P-value: Probability of finding an association exclusively by chance.

Type I error (α): Probability of identifying an association when there is actually none (false positive also called spurious association).

Type II error (β): Probability of not identifying an association when there is actually one (false negative).

Power: Probability of identifying an association when there is actually one (1-β). The power is function of:

a – Sample size;

b – Allele frequency;

c – Effect size, and;

d – Haplotype structure.

Effect size: Magnitude of risk conferred by a certain allele.

Odds ratio (OR): The measure of association by comparing the odds of an event happening in the presence or absence of a specific variable (e.g. an allele). In the specific case of GWAS the OR is:

$$OR = \frac{\text{presence of allele A and presence of phenotype / presence } of\ allele\ A\ and\ absence\ of\ phenotype}{\text{absence of allele A and presence of phenotype / absence } of\ allele\ A\ and\ absence\ of\ phenotype}$$

1.4.3.4 GWAS: How to interpret the load of data

To date, one of the most important lessons of genome-wide association studies is that the SNPs with the smallest p-values are seldom the real reason of the association.

In the majority of cases the SNPs with highest association, being tag-SNPs and acting as a surrogate for association, indirectly point to neighbouring SNPs in LD for the real cause of association at that locus. In addition, the SNPs showing association are rarely coding, whilst, more frequently, they are intronic or intergenic.

When a GWAS is concluded (discovery + replication phases) caution is warranted in the interpretation of the outcome and following steps are generally recommended (**Figure 1-9**):

1 – Identify all genes in proximity of the top hits and sequence them in the search of possible coding changes. This approach may lead to the identification of pathogenic coding variants and novel genes associated with the disease under study;

2 – Select all identified known polymorphisms to build haplotypes within and around the associated locus/loci to possibly identify disease specific haplotypes and/or particular SNPs that are in linkage disequilibrium with the GWAS top hits to be further studied, and;

3 – Evaluate effects of the associated SNPs (and of those in LD with the associated SNPs) on expression and/or splicing.

For the latter analysis, all SNPs (intergenic, intronic or even synonymous variants) that show association or are in LD with associate SNPs are informative. In fact, when the associated loci do not affect proteins directly it is likely that they exert their effect by [215]:

1 – Altering constitutively transcript levels;

2 – Modulating transcript expression, and;

3 – Affecting splicing.

Figure 1-9. Post-GWAS workflow

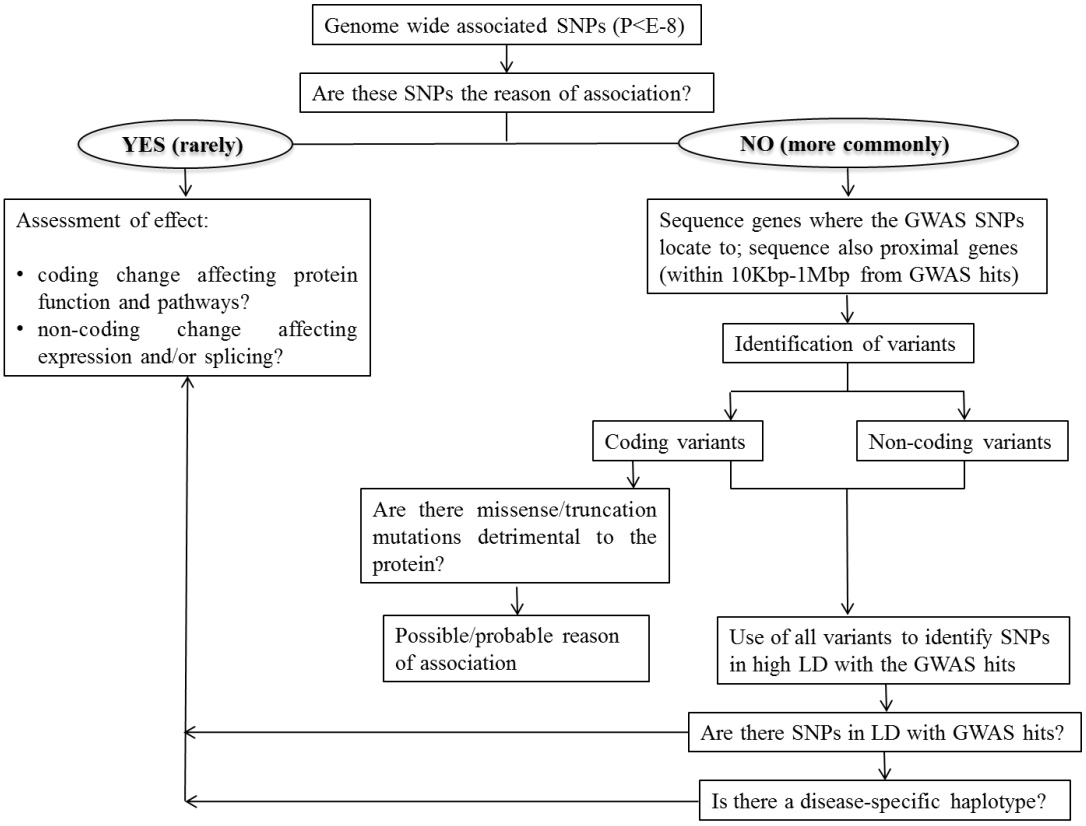


Figure 1-9. Diagram highlighting the recommended workflow subsequent to completion of discovery and replication phases of a GWAS. These are the main (but not exclusive) immediate steps warranted for further interpreting the results of a GWAS.

1.5 Thesis aims and objectives

Exploring the genetics of disease not only allows the identification of genes that may impact disease with differential penetrance (from high to low risk variants) but also may provide clues on expression/splicing patterns that may increase risk or confer protection and, not least, on cellular/biological processes possibly involved in disease pathogenesis.

It is central to understand disease mechanisms, as these will allow the development of preventive as well as therapeutic measures. Within this picture, the genetic approach holds the fundamental role of determining whether and to what extent genetic variability influences pathogenesis and progression of the disease.

To date, despite tremendous progress in the genetics of FTD and PSP, the role of common genetic factors contributing to FTD and PSP is still poorly understood.

The aim of the two main projects presented in this thesis (**Chapters 3** and **4**) was to further characterise these two devastating complex neurological disorders through a follow up study of the recent PSP-GWAS project [180] and the first GWAS on clinical FTD to date.

1.5.1 PSP

PSP is a tauopathy whose strong genetic link to the *MAPT* locus on chromosome 17 has been widely established across multiple studies including the recent GWAS [180]. Nevertheless, to date, the pathogenic mechanism underlying such association has not been fully elucidated, on one hand, because of absence of robust association with coding mutations in *MAPT* and, on the other, because studies of the *MAPT* haplotype only revealed an association with the H1c sub-haplotype without however unequivocally unravelling its pathogenic mode of action. In addition, the *MAPT* association alone does not explain the entire genetic component of PSP suggesting that other genetic loci together with hard-to-determine environmental risk factors must collectively contribute to disease.

The recent GWAS [180] highlighted novel loci in association with PSP that point to specific and plausible biological processes being involved in the pathogenesis of PSP. The aim of the study of PSP presented in this thesis (**Chapter 3**) was to further characterise the newly identified loci by:

1 – Sequencing the genes within the associated loci in the search of relevant coding variants;

2 – Evaluating the presence and, in case, the frequency and distribution of putative disease associated haplotypes, and;

3 – Analysing possible *cis* effects on expression through expression quantitative trait loci (eQTL) analyses of the GWAS-associated SNPs and those in LD.

1.5.2 FTD

Frontotemporal dementia is clinically, pathologically and genetically relatively heterogeneous (**Figure 1-3**). Concerning the genetic component, the screening and characterisation of candidate genes has explained a number of cases (mainly familial) over the past fifteen years, but the vast majority still remain uncharacterised. In addition, and especially, the mechanisms leading to disease are by far not yet understood. The state-of-the-art in the field of FTD research advocates for a better and deeper understanding of genotype-phenotype correlation and, particularly, there is an urgent need for clues about the mechanisms and biological processes that contribute to disease onset and progression. The study presented in this thesis (**Chapter 4**) is the first GWAS on clinical FTD performed to date.

The main aims of this project were to:

1 – Identify novel genetic entities associated with increased risk of FTD, and;

2 – Shed light on possible disease mechanisms to be further explored and transferred to the field of cell biology for functional assessments.

Especially the latter aim is of incommensurable importance as elucidating possible disease mechanisms will allow to better characterise and understand the molecular underpinnings of the disease and, probably even more importantly, establish biological pathways/processes as targets for the development of therapeutic measures, which, to date, are completely lacking in the field of FTD.

In addition, the study of familial and sporadic FTLD cases is also presented in this thesis (**Chapter 5**). The main aim of the latter studies was to further characterise the known candidate genes and gain insight and/or expand on genotype-phenotype correlations through the genetic screening of the main known genes, such as *MAPT*, *GRN*, *C9orf72, TDP-43*, *FUS* and *CHMP2B*, associated with FTLD in population-specific families and/or cohorts.

CHAPTER 2 – Materials and methods

2.1 Materials

2.1.1 PSP cohort

Eighty-four pathologically confirmed progressive supranuclear palsy (PSP) cases (**Figure 2-1**) were analysed in the follow up study (**Chapter 3**) of the PSP-GWAS [180].

Figure 2-1. UK PSP cohort

	1	2	3	4	5	6	7	8	9	10	11	12	
A	P09/97	P88/05	P72/03	P02/01	P42/00	P46/03	P40/05	P66/00	P26/00	P47/01	P29/05		**A**
B	P40/02	P13/05	P58/00	P48/99	P07/03	P31/02	P16/94	P31/97	P13/01	P05/02	P30/05		**B**
C	P11/03	P27/07	P55/02	P57/02	P47/06	P60/02	P60/99	P12/98	P24/05	P16/04	P63/98		**C**
D	P16/03	P18/06	P03/02	P48/04	P27/04	P37/03	P28/01	P22/07	P48/96	P10/04	P16/99		**D**
E	P44/97	P46/99	P21/02	P53/07	P41/06	P70/03	P10/01	P11/01	P65/02	P33/03			**E**
F	P84/05	P31/98	P33/04	P36/05	P19/07	P35/01	P20/99	P08/04	P30/02	P20/03			**F**
G	P20/06	P39/02	P14/00	P50/07	P08/05	P43/04	P11/00	P29/97	P12/07	P01/03			**G**
H	P23/03	P17/06	P44/00	P12/04	P52/01	P07/07	P43/99	P51/01	P14/02	P50/02			**H**
	1	2	3	4	5	6	7	8	9	10	11	12	

Figure 2-1. Code IDs of the 84 PSP cases analysed in the study. The diagnosis of PSP for each case was pathologically confirmed. DNA from each individual has been extracted from brain tissue.

All patients were white western Europeans and were collected over the past 10-15 years, primarily, by Dr Andrew Lees at the Queen Square Brain Bank and Sara Koe PSP Research Centre, Institute of Neurology (UCL). All the PSP cases met modified National Institute of Neurological Disorders and Stroke (NINDS) possible or probable criteria [216] and diagnosis was confirmed pathologically using standardized criteria [160, 217]. The brain pathology was evaluated by neuropathologist at the Queen Square Brain Bank, primarily by Drs Tamas Revesz and Janice Holton, at UCL. Samples were enrolled and collected under approved protocols followed by informed consent and every aspect of the study was approved by the Joint Medical Ethics Committee of the National Hospital of Neurology and Neurosurgery, London.

2.1.2 FTD-GWAS: Cases cohort

Forty-four international research groups from Europe, North America and Australia (**Table 2-1**) participated collectively in the study (**Chapter 4**). Sample collection and DNA extraction was performed at each individual collaborating site where appropriate informed consent was obtained. The DNA material was collected from the participating groups and stored at the three Institutions leading this project: The Department of Molecular Neuroscience at the University College of London (UCL), the Laboratory of Neurogenetics of the National Institute on Aging at the National Institutes of Health (NIH), and the Laboratory of Neurogenetics at the Texas Tech University Health Sciences Center (TTUHSC).

Table 2-1. FTD-GWAS: Collaborative groups

Site PI	Main Institutions	Country
Parastoo Momeni*	Texas Tech University Health Sciences Center, Lubbock, Texas, USA	USA/UK
Caroline Graff	Karolinska Institute, Department of NVS, Stockholm, Sweden	SWEDEN
Isabelle Leber*	Unit of Research of Neurology and experimental Therapy, Paris, France	FRANCE
Jorgen Nielsen	Section of Neurogenetics, University of Copenhagen, Denmark	DENMARK
Adrian Danek	Department of Neurology, Ludwig-Maximilian University of Munich, Germany	GERMANY
Robert Perneczky	Department of Psychiatry and Psychotherapy, Tech University Munich, Germany	GERMANY
Matthias Riemenschneider*	Saarland University Hospital, Department for Psychiatry and Psychotherapy, Homburg/Saar, Germany	GERMANY
Johannes Schlachetzki	Memory Clinic, University Hospital, Freiburg, Germany	GERMANY
Barbara Borroni	Department of Medical Sciences, Neurological Clinic, University of Brescia, Italy	ITALY
Annibale Puca*	Gruppo Multimedica, Milan, Italy	ITALY
Giacomina Rossi	Neurological Institute Carlo Besta, Milan, Italy	ITALY
Daniela Galimberti*	Department of Neurological Sciences, Dino Ferrari Institute, University of Milan, Italy	ITALY
Innocenzo Rainero	Department of Neuroscience, University of Torino, Italy	ITALY
Luisa Benussi	NeuroBioGen Lab-Memory Clinic, Fatebenefratelli, Brescia, Italy	ITALY
Benedetta Nacmias	Department of Neurological and Psychiatric Sciences, University of Florence, Italy	ITALY
Amalia Bruni	Regional Center of Neurogenetic, Lamezia Terme, Italy	ITALY
Giancarlo Logroscino	Department of Basic Medical Sciences, Neurosciences and Sense Organs,"Aldo Moro" University of Bari, Italy	ITALY
Jonathan Rohrer	Institute of Neurology, UCL, London, UK	UK
Simon Mead	Institute of Neurology, Prion Unit, UCL, London, UK	UK
James Rowe	Cambridge University Department of Clinical Neurosciences, Cambridge, CB2 0SZ	UK
Stuart Pickering-Brown*	University of Manchester, Clinical Neuroscience, Manchester, UK	UK
Huw Morris	Department of Neurology, Cardiff University School of Medicine, Cardiff, UK	UK
Ekaterina Roageva	Center for research in Neurodegenerative diseases, Toronto, Canada	CANADA
Rosa Rademakers*	Mayo Clinic Jacksonville, Florida, USA	USA
Vivianna Van Deerlin	University of Pennsylvania Health System, Department of Pathology, Philadelphia, PA, USA	USA
Christine Van Broekhoven*	Department of Molecular Genetics, University of Antwerp, Belgium	BELGIUM
Peter Heutink	German Center of Neurodegenerative Diseases-Tübingen, Germany	NETHERLANDS
John van Swieten	Department of Neurology, Erasmus Medical Center, Rotterdam, Netherlands	NETHERLANDS
Jordi Clarimon	Genetics of Neurodegenerative Diseases Unit, IIB Hospital Sant Pau, Barcelona, Spain	SPAIN
Pau Pastor	Center for Applied Medical Research, Division of Neurology, University of Navarra, Pamplona, Spain	SPAIN
Carlos Cruchaga	Department of Psychiatry, HPAN, WU, St. Louis, USA	USA
Ian Mackenzie	Department of Pathology, Vancouver General Hospital, Vancouver, BC, Canada	CANADA
Maria Landqvist	Unit of Geriatric Psychiatry, Department of Clinical Sciences, Lund University, Sweden	SWEDEN
Agustin Ruiz	ACE Foundation. Catalan Institute of Applied Neuroscience, Barcelona, Spain	SPAIN
Peter Schofield	Neuroscience Research Australia, Sydney, Australia	AUSTRALIA

Table 2-1. List of the totality of the PIs of collaborative sites that directly sent DNA samples to the three institutions (UCL, NIH and TTUHSC) leading the project. * PI representing multiple centres.

In total, 5,696 samples were collected considering both the discovery (n=3,657) and the replication phase (n=2,039).

I received and managed the samples at UCL (discovery + replication phases) and TTUHSC (replication phase). In addition, I performed all quality control (QC) steps prior to genotyping (section *2.1.2.1*). I partially participated in the process of genotyping that was performed at the UCL core facility, UCL Genomics, located at the Institute of Child Health (ICH). The samples shipped to NIH were received by Cynthia Crews and Dena Hernandez and then underwent QC steps and genotyping that were performed by laboratory technicians at the Laboratory of Neurogenetics (NIH).

2.1.2.1 Discovery phase

Twenty-seven international research groups contributed samples (n=3,657) for the discovery phase (**Table 2-2**).

Given the heterogeneity of the disease the inclusion/exclusion criteria for enrolment in the study were based on:

1 – The most common and worldwide adopted clinical diagnostic criteria [21];

2 – The brain pathology (if available), and;

3 – The genetic screening of the candidate genes (if available).

This project started in late 2008. Being the Neary criteria [21] the gold standard for clinical diagnosis at the time, all samples collected during discovery phase were diagnosed based on those diagnostic criteria.

Table 2-2. FTD-GWAS: Discovery phase cases

Participating Groups	Sending Institution	Samples sent
		Discovery phase
Momeni (USA)	Texas Tech University Health Sciences Center, Lubbock, Texas, USA	
Graff (SWEDEN)	Karolinska Institute, Department of NVS, Stockholm, Sweden	
French Consortium	Unit of Research of Neurology and experimental Therapy, Paris, France Neurological Institute, University of Salpetriere Hospital, Paris, France Department of Neurology and Memory Center, Lille, France	
Nielsen (DENMARK)	Section of Neurogenetics, University of Copenhagen, Denmark	
Danek (GERMANY)	Department of Neurology, Ludwig-Maximilian University of Munich, Germany	
Perneczky (GERMANY)	Department of Psychiatry and Psychotherapy, Tech University Munich, Germany	
Riemenschneider (GERMANY)	Department of Neuropathology, Heinrich Heine University, Düsseldorf, Germany	
Schlachetzki (GERMANY)	Memory Clinic, University Hospital, Freiburg, Germany	
Borroni (ITALY)	Department of Medical Sciences, Neurological Clinic, University of Brescia, Italy	
Novelli - Puca (ITALY)	Gruppo Multimedica, Milan, Italy	
Rossi (ITALY)	Neurological Institute Carlo Besta, Milan, Italy	
Galimberti (ITALY)	Department of Neurological Sciences, Dino Ferrari Institute, University of Milan, Italy	
Rainero (ITALY)	Department of Neuroscience, University of Torino, Italy	**TOTAL 3,657**
Benussi (ITALY)	NeuroBioGen Lab-Memory Clinic, Fatebenefratelli, Brescia, Italy	
Nacmias (ITALY)	Department of Neurological and Psychiatric Sciences, University of Florence, Italy	
Bruni (ITALY)	Regional Center of Neurogenetic, Lamezia Terme, Italy	
Rohrer - Rossor (UK)	Institute of Neurology, UCL and Division of Neuroscience and Psychological medicine, Imperial College, London, UK	
Mead (UK)	Institute of Neurology, Prion Unit, UCL, London, UK	
Pickering-Brown (UK)	University of Manchester, Clinical Neuroscience, Manchester, UK	
Morris (UK)	Department of Neurology, Cardiff University School of Medicine, Cardiff, UK	
Roageva (CANADA)	Center for research in Neurodegenerative diseases, Toronto, Canada	
Rademakers (USA)	Mayo Clinic Jacksonville, Florida, USA	
Van deerlin (USA)	University of Pennsylvania Health System, Department of Pathology, Philadelphia, PA, USA	
Van Broekhoven (BELGIUM)	Department of Molecular Genetics, University of Antwerp, Belgium	
Heutink (NETHERLANDS)	Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, Netherlands	
van Swieten (NETHERLANDS)	Department of Neurology, Erasmus Medical Center, Rotterdam, Netherlands	
Pastor (SPAIN)	Center for Applied Medical Research, Division of Neurology, University of Navarra, Pamplona, Spain	

Table 2-2. List of the research groups (identified by the PI's name) that contributed samples for the discovery phase.

To address and cover the main syndromes associated with FTD, samples clinically diagnosed with behavioural variant FTD (bvFTD), semantic dementia (SD), progressive non-fluent aphasia (PNFA) [20, 21], and FTD with motor neuron disease (FTD-MND) [218] were included in the study. I paid particular attention in identifying cases diagnosed with logopenic aphasia (LPA), because the majority of these cases are associated with Alzheimer's disease pathology [25], reason for which they were eventually excluded from the study. If the analysis of the brain pathology was available it was given priority, in terms of final diagnosis, over the clinical diagnosis. Samples with tau as well as FTLD-U/FTLD-TDP pathology where included in the study. Finally, samples known to carry mutations in the *MAPT* and *GRN* genes were excluded from the study. An ideal requirement was the characterisation of each sample for the *MAPT* haplotype, *ApoE* and *MAPT*/*GRN* sequencing in pre-genotyping phase. In addition, individuals with *C9orf72* expansions were not excluded from the discovery phase, because this locus was identified subsequently (almost 3 years) to sample collection. Of note, whole genome amplified samples were not included in the study to precautionary exclude possible errors in genotyping (**Appendix 2-1**).

Eligible sporadic cases were all included in the study whilst, for familial cases, exclusively index patient/probands were included in the study.

The samples, all of European-North American extraction and of confirmed European ancestry, were received either in tubes or 96 well plates each with a unique alpha-numerical ID specific to the sample and the site. For each sample, a total amount of 2μg

of DNA extracted either from blood or brain at each collaborative site was collected. A description of the main clinical, pathological and genetic known features was integrated through a spread sheet summarizing a panel of information useful in further characterising the collected samples and identifying those eligible for inclusion in the study (**Appendix 2-2**).

Each DNA sample was evaluated for integrity by means of gel electrophoresis on 1% agarose gel (section *2.2.2.2*). Material purity and concentration were analysed by spectrophotometric (Nanodrop, Wilmington, DE, USA) quantification (section *2.2.1.3*). Each sample was brought to a final concentration of 250ng as recommended by manufacturer (Illumina, San Diego, CA, USA) prior to being transferred onto chip for whole genome genotyping.

2.1.2.2 Replication phase

Twenty-eight research groups contributed samples (n=2,039) and matching neurologically normal controls (n=1,600) for the replication phase from Sweden, England, Italy and Spain (**Table 2-3**). Further cases and matching controls from North America (USA and Canada), France, Germany, England and Netherlands (section *2.1.6.3*) were collected at NIH.

Table 2-3. FTD-GWAS: Replication phase cases and controls

Participating Groups	Sending Institution	Samples sent	
		Replication phase	
		Cases	*Controls*
Momeni (USA)	Texas Tech University Health Sciences Center, Lubbock, Texas, USA		
Baborie (UK)	Liverpool - Newcastle		
Graff (SWEDEN)	Karolinska Institute, Department of NVS, Stockholm, Sweden		
French Consortium	Unit of Research of Neurology and experimental Therapy, Paris, France Neurological Institute, University of Salpetriere Hospital, Paris, France Department of Neurology and Memory Center, Lille, France		
Perneczky (GERMANY)	Department of Psychiatry and Psychotherapy, Tech University Munich, Germany		
Borroni (ITALY)	Department of Medical Sciences, Neurological Clinic, University of Brescia, Italy		
Rossi (ITALY)	Neurological Institute Carlo Besta, Milan, Italy		
Galimberti (ITALY)	Department of Neurological Sciences, Dino Ferrari Institute, University of Milan, Italy		
Rainero (ITALY)	Department of Neuroscience, University of Torino, Italy		
Benussi (ITALY)	NeuroBioGen Lab-Memory Clinic, Fatebenefratelli, Brescia, Italy		
Nacmias (ITALY)	Department of Neurological and Psychiatric Sciences, University of Florence, Italy		
Bruni (ITALY)	Regional Center of Neurogenetic, Lamezia Terme, Italy		
Rohrer - Rossor (UK)	Institute of Neurology, UCL and Division of Neuroscience and Psychological medicine, Imperial College, London, UK	**TOTAL 2,039**	**TOTAL 1,600**
Mead (UK)	Institute of Neurology, Prion Unit, UCL, London, UK		
Rowe (UK)	Cambridge		
Pickering-Brown (UK)	University of Manchester, Clinical Neuroscience, Manchester, UK		
Rademakers (USA)	Mayo Clinic Jacksonville, Florida, USA		
Van deerlin (USA)	University of Pennsylvania Health System, Department of Pathology, Philadelphia, PA, USA		
Van Broekhoven (BELGIUM)	Department of Molecular Genetics, University of Antwerp, Belgium		
Heutink (NETHERLANDS)	Department of Clinical Genetics, Erasmus Medical Center, Rotterdam, Netherlands		
van Swieten (NETHERLANDS)	Department of Neurology, Erasmus Medical Center, Rotterdam, Netherlands		
Clarimon (SPAIN)	Genetics of Neurodegenerative Diseases Unit | IIB Hospital Sant Pau, Barcelona, Spain		
Pastor (SPAIN)	Center for Applied Medical Research, Division of Neurology, University of Navarra, Pamplona, Spain		
Cruchaga (USA)	Department of Psychiatry, HPAN, WU, St. Louis, USA		
Mackenzie (CANADA)	Department of Pathology, Vancouver General Hospital, Vancouver, BC, Canada		
Landqvist (SWEDEN)	Unit of Geriatric Psychiatry ,Department of Clinical Sciences, Lund University, Sweden		
Ruiz (SPAIN)	ACE Foundation. Catalan Institute of Applied Neuroscience, Barcelona, Spain		
Schofield (AUSTRALIA)	Neuroscience Research Australia ,Sydney ,Australia		

Table 2-3. List of the research groups that contributed samples for the replication phase. The third column is subdivided into two sub-columns which show the number of cases and neurologically normal controls that were received.

It needs to be acknowledged that during the course of the project the diagnostic criteria for FTD underwent revisions that resulted in two updated versions addressing specifically the behavioural variant [27] and the language variant syndromes [20]. The updated diagnostic criteria were adopted for a part of the samples collected during replication phase. However, inclusion/exclusion criteria for the replication phase were almost entirely the same as in the case of the discovery phase. The differences were only minor:

1 – In addition to samples also matching neurologically normal controls from each site (if available) were collected to best match cases and controls in the association study;

2 – Mutation carriers in the *MAPT* and *GRN* genes were excluded from the study, whilst cases harbouring the *C9orf72* expansion were asked to be identified but were included (**Appendix 2-3**).

The cases (and the controls) collected for the replication phase were a new set of samples, all of European-North American extraction and of confirmed European ancestry each with a unique alpha-numerical ID specific to the sample and the site. For each sample, a total amount of 2μg of DNA extracted either from blood or brain at each collaborative site was collected. The main clinical, pathological and genetic known characteristics were integrated through a spreadsheet that was slightly revised compared to that used in discovery phase (**Appendix 2-3**).

This step was entirely performed as in discovery phase with the difference that purity and concentration were assessed either by spectrophotometric (Nanodrop, Wilmington, DE, USA) or fluorometric (Qubit, Life Technologies, Grand Island, NY, USA) quantification.

2.1.3 The Afrikaaner family

The Afrikaaner family is a multigenerational kindred from South Africa with Dutch ancestry (**Figure 2-2**).

Members of this family have been diagnosed with FTD and the mode of inheritance strongly seems autosomal dominant. Affected and unaffected family members were collected through a collaborative effort between the Institute for Ageing and Health at Newcastle University and The University of Stellenbosch by Drs. Rajesh Kalaria and Felix Potocnick who are the physicians following longitudinally this familial case of FTD since the past 10 years.

The members of this family were consented prior to blood draw according to the ethics committee of the University of Stellenbosch. The DNA used for genetic screening (**Chapter 5**, section ***5.1.1***) was extracted from blood using the standard procedures and the brain pathology was evaluated at the University of Newcastle in three individuals (numbers 19, 20 and 22 in the pedigree; **Figure 2-2**) revealing typical tau pathology.

Figure 2-2. Afrikaaner family pedigree

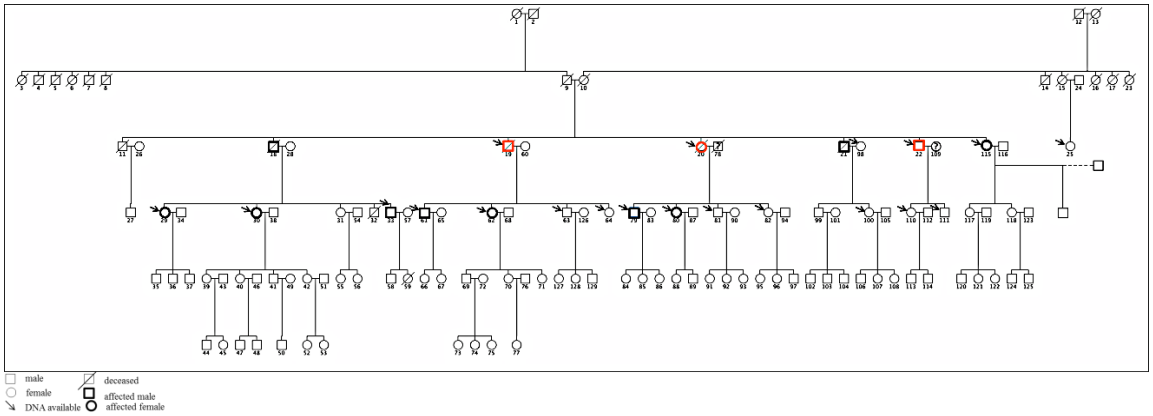


Figure 2-2. Multigenerational pedigree of the Afrikaner family. The individuals for whom there is availability of DNA for genetic screening (arrow), the affected members (bold squares or circles) and those individuals for whom data on the brain pathology are available (bold red squares or circles) are shown.

2.1.4 The Finnish kindred

The Finnish FTD kindred is a two generational family (**Figure 2-3**) of which 5 siblings of the second generation have been followed by a team of neurologists and pathologists at the University of Helsinki over the past 20-30 years.

Two of the five siblings (1 and 2) were diagnosed with FTD at the Department of Neurology at the Helsinki Central University Hospital. Blood samples for DNA analysis were collected from the 2 index patients (1 and 2) and the 3 further siblings (3, 4 and 5) for genetic screening (**Chapter 5**, section ***5.1.2***) after all subjects gave informed consent for research.

Figure 2-3. Finnish family pedigree

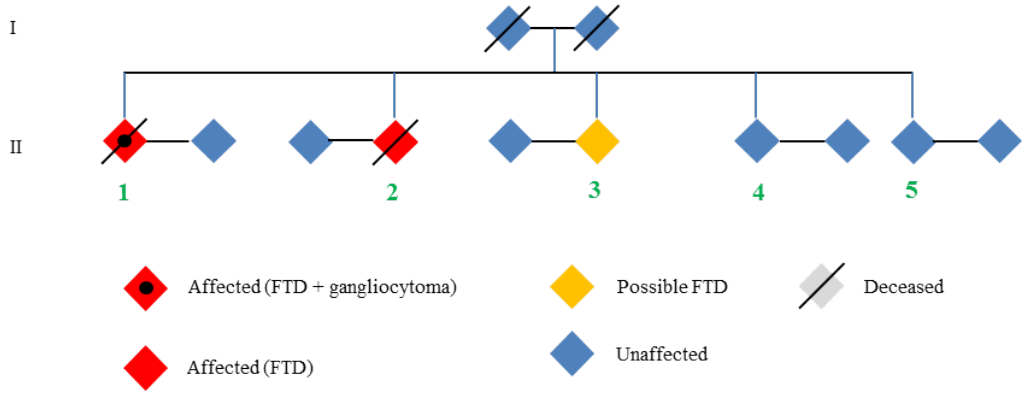


Figure 2-3. Pedigree of the Finnish FTD kindred. First generation: Parents of the five siblings who are depicted in the second generation. Individuals 1 and 2 of the second generation were diagnosed with FTD. Both are deceased. Patient 1 also developed gangliocytoma. This figure has been adapted from Ferrari et al, 2013 [219].

2.1.5 The NINDS sample set

Sporadic cases diagnosed with FTD (n=158) or CBS (n=70) were collected at the Cognitive Neuroscience Division of National Institute of Neurological Disorders and Stroke (NINDS) at the National Institutes of Health (NIH) starting in the early 2000s'.

This cohort included patients with a diagnosis of behavioural variant FTD (bvFTD), progressive non-fluent aphasia (PNFA) [21], and CBS [220] and five of the FTD patients had concomitant motor neuron disease. These subjects were seen as part of an ongoing research study on FTD and CBS at the Cognitive Neuroscience Section of the NINDS of the NIH, Bethesda, MD.

The patients were either self-referred or referred by outside physicians, psychiatrists or neurologists. After arrival to NIH accompanied by a caregiver, they were diagnosed based on an initial clinical evaluation and examination performed by a team of neurologists and psychiatrists, led by Dr Wassermann, based on standard clinical criteria [21, 220]. The patients then spent ~7-10 days participating in extensive neuropsychological and neurologic testing and imaging studies. Their diagnoses were re-evaluated by a neuropsychologist (Dr Jordan Grafman) based on the results of the testing performed at the NIH. All subjects were required to have an assigned research durable power of attorney prior to admission to the protocol and the assigned individuals gave written informed consent for the study.

After the patients gave assent for the study, blood samples for a total of 228 index patients was collected. DNA was extracted for genetic screening (**Chapter 5**, extended section ***5.2***) either at NIH or, later, at TTUHSC. All aspects of the study and the consent procedures were approved by the NINDS, NIA and TTUHSC Institutional Review Boards.

2.1.6 Normal controls

Different types of neurologically normal controls or normal/healthy population samples were used based on the kind of study or on the purpose of the projects.

2.1.6.1 Familial cases

For the two familial cases, the Afrikaaner family (section ***2.1.3***) and the Finnish kindred (section ***2.1.4***), unaffected family members of each family were used as normal controls. These individuals are identified in **Figure 2-2** and **Figure 2-3**, respectively.

2.1.6.2 Sporadic cases

For the sporadic cases and, in general, to assess the presence and distribution of putatively pathogenic missense variants identified during the course of genetic screening of the dementia candidate genes (**Table 5-1**, in **Chapter 5**), DNA samples of neurologically normal controls/healthy individuals (n=672) were used, all obtained from the Coriell repository (http://www.ccr.coriell.org) (**Table 2-4**).

Table 2-4. Normal controls from the Coriell repository

Plate ID	n	Type of control	Ancestry	Web link
NDPT096	94	Neurologically normal controls	Caucasian	http://ccr.coriell.org/Sections/Search/Panel_Detail.aspx?Ref=NDPT096&PgId=202
NDPT098	92	Neurologically normal controls	Caucasian	Discontinued; **Appendix 2-4**
NDPT099	92	Neurologically normal controls	Caucasian	http://ccr.coriell.org/Sections/Search/Panel_Detail.aspx?Ref=NDPT099&PgId=202
NDPT031	94	Population/convenience control	African American	http://www.ccr.coriell.org/Sections/Search/Search.aspx?PgId=165&q=NDPT031
HD200CAU	200	Human variation panel	Caucasian	http://www.ccr.coriell.org/Sections/Search/Panel_Detail.aspx?Ref=HD200CAU&PgId=202
HD100MEX-2	100	Human variation panel	Mexican	http://www.ccr.coriell.org/Sections/Search/Panel_Detail.aspx?Ref=HD100MEX-2&PgId=202

2.1.6.3 Genome-wide association study

Discovery phase

The samples (cases) included in the discovery phase of FTD-GWAS were obtained from different European countries (France, England, Germany, Netherlands, Belgium, Denmark, Sweden and Italy) and from North America (USA and Canada).

To match all the cases, genotyping data of normal controls generated in studies previously conducted at either the Laboratory of Neurogenetics of the National Institute on Aging at the National Institutes of Health or from the Department of Molecular Neuroscience at the University College London were used [221-223]. These control samples had been genotyped on a variety of chips such as Illumina 370K and 550K, and were matched as best as possible based on population ancestry and genotyping platform. Specifically, aggregate data for control samples were merged based on overlapping SNPs from the different chips. In total 7,444 control samples were selected after QC and filtering steps (section *2.3.5.1*) for the association analysis of the discovery phase cohort.

The controls originated from the USA, England, Italy, Germany, France, Sweden and the Netherlands. For each case, at least 2 controls were matched based on compatibility of genetic ancestry estimated by principal components analysis (PCA) (section *2.3.5.1*).

Replication phase

In replication phase the genotyping of controls was performed in concomitance with the genotyping of the cases. The cases that needed to be matched to neurologically normal controls were obtained from Europe (Spain, France, England, Netherlands, Belgium, Germany, Sweden and Italy), from North America (USA and Canada) and from Australia (with British ancestry). All samples were ultimately of European ancestry.

The genotyping of controls was performed at the Laboratory of Neurogenetics of the National Institute on Aging at the NIH (90% of controls) and at the core facility at the Institute of Child Health at UCL (10% of controls). The total number of normal controls that had been collected exceeded the number of controls reported in **Table 2-3**. In fact, **Table 2-3** reports only normal control samples that were collected at UCL, whilst more control cases were genotyped at NIH. All controls (n>7,000) used in the replication phase were genotyped using the NeuroX custom chip (section *2.2.2.8*).

Eventually, n=5,094 survived QC and were used for replication and joint analyses. The normal controls for the replication phase were from the following ancestry backgrounds: USA (European/American), England, Italy, France, Germany, Sweden, Spain, and the Netherlands.

2.2 Methods

2.2.1 DNA extraction

2.2.1.1 DNA from blood

Genomic DNA is extracted from 4-5ml of blood using Promega Wizard DNA extraction kit (Promega, Madison, WI, USA), following manufacturer's instructions. The dried DNA pellet is re-suspended in 250µl of DNA Rehydratation Solution (Promega) and kept 2-5 days at 4°C prior to evaluating the concentration.

2.2.1.2 DNA from Brain

Total DNA is extracted from dissected samples (100–200mg) of human post-mortem brain tissue using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, Manchester, UK) and was performed following manufacturer's instructions (Qiagen). The dried DNA pellet is re-suspended in 100-200µl of DNA Rehydratation Solution (Qiagen) and kept 2-5 days at 4°C prior to evaluating the concentration.

2.2.1.3 DNA quantitation

To evaluate the concentration of the DNA either a spectrophotometer or a fluorometer were used.

Nanodrop spectrophotometer

The NanoDrop Spectrophotometer is equipped with the interactive software ND-1000 V3.5.2 that needs to be installed on PC (NanoDrop Technologies Inc, Wilmington, Delaware, USA). The NanoDrop calculates concentration of a variety of molecular probes including DNA, RNA and proteins. As absorbance measurements will measure any molecules absorbing at a specific wavelength, DNA will absorb at 260 nm and contribute to the total absorbance. For quality control, the ratios of absorbance at 260/280 nm and 260/230 nm are used to assess the purity of DNA. Ratios of 260/280=~1.8 and 260/230=~2.0-2.2 denote pure DNA. The NanoDrop spectrophotometer is initiated with distilled water and then calibrated by using the solution/buffer in which the DNA that will be measured is re-suspended as a blank. Then, 2µl of DNA sample are loaded on the column of the NanoDrop that will read the absorbance and express the DNA concentration in ng/µl. Afterwards the DNA is diluted to the working concentration.

Qubit fluorometer

The Qubit fluorometer (Life Technologies, Paisley, UK) does not require a connection to PC and calculates concentration of a variety of molecular probes including DNA, RNA and proteins. To evaluate concentration of DNA the Quant-iTTM dsDNA BR assay (Life Technologies) is used and samples are prepared as suggested by manufacturer. The Qubit

fluorometer will display the concentration in ng/ml; therefore, calculations will need to be adjusted to ng/µl. Afterwards the DNA is diluted to the working concentration.

2.2.2 Genetic screening

2.2.2.1 Polymerase chain reaction (PCR)

PCR is performed using Roche Faststart PCR Master (400 RXN/10 ML) (Roche Applied Science, Indianapolis, IN, USA), primers (designed with Primer 3 software; http://bioinfo.ut.ee/primer3-0.4.0/) at a dilution of 10pmol/µl and genomic DNA is added to the reaction at a final concentration of 15-20ng. The PCR amplification is performed using a program that uses 35 cycles at 95°C for 15" for denaturation, at 55-57°C for 15" to allow primers' annealing and at 72°C for 30" for extension, followed by a final 7' elongation at 72°C.

2.2.2.2 Agarose gel electrophoresis

Agarose gel electrophoresis is used to verify the quality and size of the PCR amplification product. Either 1%, 1.5% or 2% gel is prepared using Ultrapure agarose (Life Technologies, Grand Island, NY, USA) and Tris/Borate/EDTA (TBE) 1X buffer (diluted 1:10 from VWR, Radnor, PA, USA). To do so, either 1g, 1.5g or 2g of Ultrapure agarose are weighted and added to 100ml of TBE 1X, and heated in a microwave for

1':30". The GelRed nucleic acid gel stain (VWR) is added to visualize the DNA bands in the gel and the gel is left solidifying for 10-15'. Afterwards the gel is placed in a chamber and is covered with TBE 1X buffer. One µl of PCR product is mixed with 4µl of loading dye (EMD Millipore – Novagen, Madison, WI, USA). This mix is loaded into the wells of the solidified gel and the DNA ladder (100bp or 1Kbp) (EMD Millipore – Novagen) is loaded at the beginning of each row to have a mean of comparison for the size of the PCR product. After loading the samples and the DNA ladder the gel is exposed to 120-200mV of voltage for ~20-30'. The DNA fragments are then visualized using a UV trans-illuminator.

2.2.2.3 PCR product purification

The PCR product is purified using either 96 well or 384 well PCR filter plates (Millipore, Billerica, MA, USA). Up to 70µl of distilled/purified water are added to the PCR product which is vacuum-filtered at -12 inch of Mercury (in.Hg). Subsequently, the PCR product is re-suspended in 80µl of distilled/purified water and shaken on the thermo-shaker (Eppendorf, Hauppauge, NY, USA) at 4°C at 300rpm for 10-15'. The purified PCR product is then used for the sequencing reaction.

2.2.2.4 Sequencing

The sequencing reaction is prepared using BigDye Terminator v3.1, 5X Sequencing Buffer (Life Technologies), forward or reverse primers at 5pmol/µl and purified PCR product as recommended by manufacturer (Life Technologies). Standard program recommended by Life Technologies is used to run the sequencing reaction in the thermo-cycler: A sequence of 45 cycles at 95°C for 10" for denaturation, at 50°C for 5" to allow primers' annealing and at 60°C for 4' for extension are performed.

The sequencing reaction product is then purified through vacuum-filtering using Millipore filter plates (Millipore) and re-suspended in 30µl purified water. The filtered sequencing reaction product is transferred into 96 or 384 wells Optical Reaction Plates (Life Technologies) and loaded into the 3730 DNA analyzer (Life Technologies). The output data of the 3730 DNA analyzer is analysed and saved with the Sequencing Analysis Software (section ***3.2***). These files are then exported to the Sequencher 4.9 software (section ***3.2***) for alignment with the reference sequence and for visual inspection and analysis.

2.2.2.5 Repeat expansion

The repeat-primed PCR is used to detect the presence of expansion in *C9orf72*.

Briefly, 100ng of genomic DNA are mixed with FastStart Mix (Roche), dimethyl sulfoxide (DMSO) (VWR), Q Solution (Invitrogen), deazaGTP (Fisher Scientific,

Waltham, MA, USA), and the primers mix (Invitrogen). The primers include a fluorescently labelled forward primer: 6-FAM-AGTCGCTAGAGGCGAAAGC; a reverse primer annealing with the repeat: TACGCATCCCAGTTTGAGACGGGGGCCGGGGCCGGGGCCGGGG, and; an anchor primer: TACGCATCCCAGTTTGAGACG. To amplify this region a touch-down 70-56°C PCR program is used. After the repeat-primed PCR is concluded 2µl of the PCR product are mixed with 0.5µl of Liz500 size standard (used as an internal size ladder, Life Technologies) and 7.5µl HiDi Formamide (Life Technologies). This reaction is loaded onto the 3730 DNA Analyzer for electrophoretic separation and fluorescence detection. Subsequently, the amplified fragment length polymorphism (AFLP) is analysed and visualised through the GeneMapper software (section ***2.3.1***).

As a result of the analysis, the fluorescence intensities are shown in the vertical axis: The size marker is identified by orange vertical lines and the hexanucleotide repeats that extend beyond the 300 bp marker are marked in blue (**Figure 2-4**).

Figure 2-4. Repeat expansion analysis

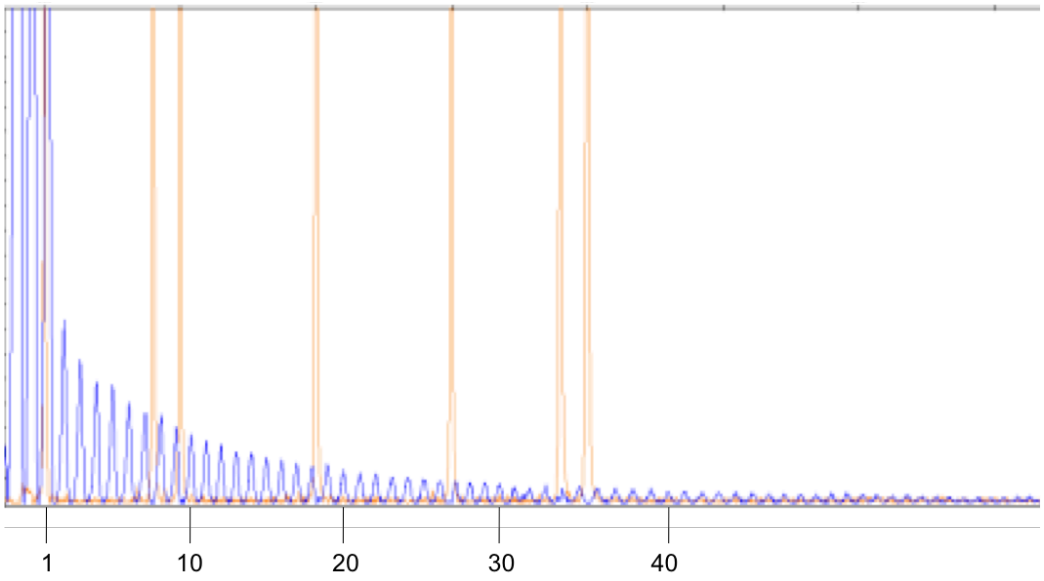


Figure 2-4. Example of software generated graph identifying the standard (orange vertical lines) and the amplified fragment length polymorphism (blue teeth saw pattern).

2.2.2.6 Single SNP genotyping assays

To genotype specific SNPs the TaqMan Genotyping Assay (Life Technologies), a pre-made 5'-nuclease assay containing primers and two allele specific probes is used. Each probe contains a different fluorescent dye (most common fluorescein amidite [FAM; absorbance at 459nm; emission at 521 nm; colour: Green] and [VIC; absorbance at 538nm; emission at 554; colour: Yellow]) for each of the two possible alleles and the primers are specifically designed to amplify the region of interest.

The reactions containing 10ng of DNA, the TaqMan Genotyping Assay (Life Technologies) and the Genotyping Master Mix (Life Technologies) are prepared as recommended by manufacturer (Life Technologies).

The PCR is run on the 7900 HT light cycler (Life Technologies). The setting for the run as well as the post-PCR fluorescent read of the reactions that is performed in the 7900 light cycler (Life Technologies) and the analysis to discriminate the genotypes are performed through the SDS software (section ***2.3***).

This specific type of analysis is called "allelic discrimination" and allows evaluating the frequency and distribution of a certain genotype within a cohort of interest. One SNP at the time, per sample, can be screened and the results are visualized in the form of clusters in the following combinations: 2 types of homozygous calls (i.e. AA or BB) and one heterozygous call (i.e. AB) (**Figure 2-5**).

Figure 2-5. Allelic discrimination analysis

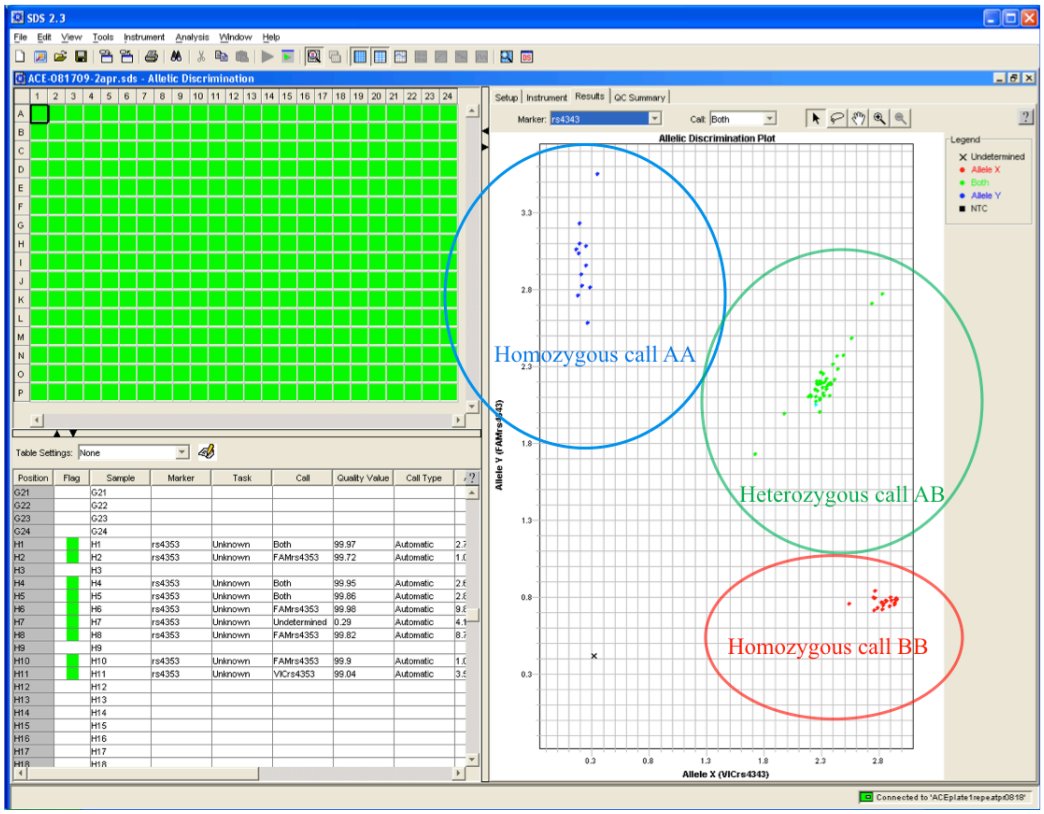


Figure 2-5. Distribution of the allelic distribution analysis. The SNPs are called in either their homozygous or heterozygous combination and data can be exported to exactly match the genotype to the well/sample.

2.2.2.7 Quantitative real-time PCR

To assess specific copy number variations the TaqMan Copy Number Assays (Life Technologies) is used. This assay is run in concomitance with a TaqMan Copy Number Reference Assay (Life Technologies), the Genotyping Master Mix (Life Technologies)

and 10ng of DNA as per manufacturer instructions (Life Technologies). The copy number assay probe is fluorescently labelled with FAM, whilst the reference assay probe is labelled with VIC. The latter provides a standard for normal copy number to which the target probe is compared. The workflow is similar to that of the genotyping assay. The PCR is run on the 7900 HT light cycler (Life Technologies) and the settings for the run as well as the real-time reading are performed and monitored through the SDS software (section ***2.3***).

This approach is called real time PCR because one can follow the process of amplification in real time due to the incorporation of fluorescently labelled probes in the exponentially growing PCR product. The amplification is simultaneously translated into an amplification curve where the threshold line is the level at which a reaction reaches a florescent intensity above background (**Figure 2-6**). The cycle at which the sample reaches this level is called cycle threshold (Ct), which is the value used during data analysis and quantitation.

The SDS software allows saving and exporting the Ct values in a .txt format. These are imported into the Copy Caller software (section ***2.3.2***) where the results of the target probe are standardized to reference genes, such as RNase P.

Figure 2-6. Gene dosage analysis

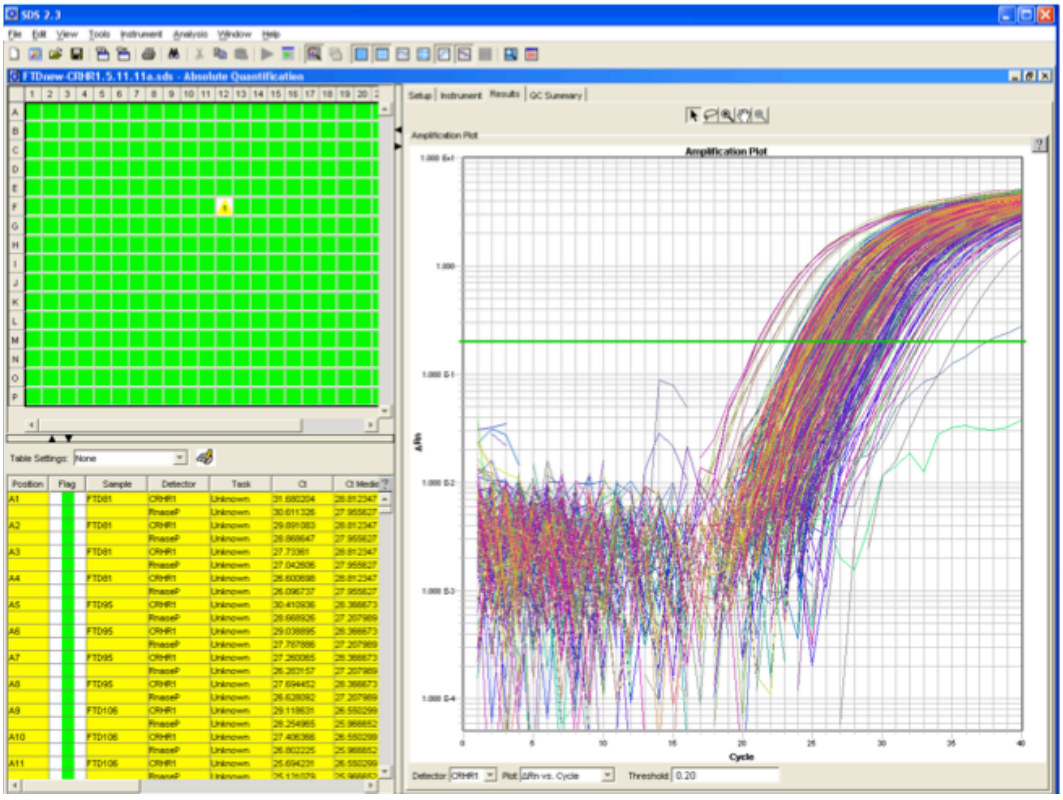


Figure 2-6. Example of a completed CNV/dosage experiment. The square in the top left shows the plate format; the square at the bottom left collects the analysed Ct values and the square on the right displays the final amplification curve and the cycle threshold (green horizontal line).

2.2.2.8 Genome-wide genotyping

Genotyping platform

The Illumina Infinium is an ideal platform for different types of genetic studies including genome-wide association studies (GWAS). This platform allows to genotype from 300K

up to 1M genetic markers in the human genome and the genotyping is based on the BeadArray technology. The Beads are located into micro-wells, within the chip, at a distance of 5.7 microns from each other, and hold several 100,000 oligonucleotide probes (50-mer probes). After having performed quality control of the material (section *2.1.2.1*) the workflow to perform the genome-wide genotyping is as follows:

1 – Genomic DNA (gDNA), in the amount of ~250ng, is amplified through PCR-free whole genome amplification (day 1);

2 – The amplified DNA is then fragmented and added onto the BeadChip in order to allow hybridization to the 50-mer probes (day 2), and;

3 – The oligonucleotide sequence of these probes stops one base before the marker. After hybridization, a fluorescently labelled nucleotide is added through DNA Polymerase mediated enzymatic single-base extension and then the iScan imagining system detects the fluorescent staining discriminating, both, the colour (green or red for discriminating the allele) and the signal intensity (for discriminating the genotype) (day 3).

Genotyping data are exported to Genome Studio (section ***2.3.4***) where, initially, call rates are assessed and the SNPs are manually clustered prior to performing data cleaning and association analysis (section ***2.3.5***).

Array chips

During the discovery phase of the FTD-GWAS (**Chapter 4**) the samples were genotyped either with the Human 660W-Quad BeadChip (88% of the cases), or with the HumanOmniExpress BeadChip (12% of the cases). These chips include tag SNPs that derive from all three HapMap phases specific for the European population and that are optimised to capture large amount of common variation throughout the genome.

The Human 660W-Quad BeadChip provides a relatively dense and uniform coverage of the genome including 550K SNPs and 110K CNV markers. There is the possibility of loading the amplified DNA of 4 samples per chip.

The HumanOmniExpress-12 BeadChip holds more than 730K markers and up to 12 DNA samples can be loaded on this chip. Of note, the normal controls used for analysis in the discovery phase had been previously genotyped on different Illumina array chips (300K and 550K).

For the replication phase all cases and normal controls were genotyped using a custom designed chip called NeuroX. The NeuroX chip specifically targets SNPs and loci associated with different neurological disorders including AD, PD and ALS. The markers included on the NeuroX chip are obtained from published or available GWAS and/or whole exome sequencing data. In total there are ~267K SNPs of which 3,759 are specifically targeting FTD. The latter were selected among those SNPs that reached p-values$<1x10^{-04}$ during the discovery phase of the FTD-GWAS (**Chapter 4**). Of note, the

SNPs included onto the NeuroX chip are tag SNPs based on European ancestry (CEU) linkage disequilibrium (LD) patterns derived from the most recent CEU data from the 1000 Genomes project [224]. For each marker, 5 LD-based proxies with r^2>0.5 or technical replicates were included on the array per locus, tagging associations within +/- 250Kb.

2.2.3 Mutagenesis

The QuikChange II Site-Directed mutagenesis kit (Startagene, CA, USA) was used to introduce point mutations in plasmid constructs (pcDNA3.1+; Invitrogen). For the study of the CHMP2B missense change Ser187Asn, a vector including the wild-type gene CHMP2B was obtained from mRNA extracted from lymphoblasts of a neurologically normal individual. The primers carrying the mutation are designed as per manufacturer's instructions (Startagene). Forward Primer 5'-CCATCAGCTGCTCGAA**A**CTTACCATCTGCCTCTAC-3' and Reverse Primer 5'–GTAGAGGCAGATGGTAAG**T**TTCGAGCAGCTGATGG–3'. The base change G>A is bold and underlined in the forward primer (and in the reverse primer, C>T). PCR is performed following manufacturer's instructions (Startagene) using the plasmid and the primers carrying the base change. Following, XL1-Blue Supercompetent cells (Agilent Technologies, Santa Clara CA, USA) are transformed with the digested PCR product (Dpn1 restriction enzyme) and incubated for 1 hour at 37°C. Afterwards transformed cells are plated on Luria Agar Base (LGA) (VWR) in Petri dishes and left at 37°C for

approximately 16 hours. Bacterial colonies are picked and transferred into Luria Broth (LB) medium (VWR) containing 1:1000 ampicillin (VWR). The plasmid DNA is extracted from the bacteria using of QIAprep spin Miniprep kit (Qiagen, Valencia, CA, USA) following manufacturer's instructions. DNA concentration is measured with the NanoDrop (NanoDrop Technologies Inc). A portion of the DNA is sequenced to confirm the mutation and the rest is stored at -20°C for further experiments.

The following experiments have been performed by Astrid Autier and Dr Hazel Urwin. I did not perform this work and have no access to the protocol. However, in brief, after having isolated the plasmids containing the point mutation these are used to transfect the SK-N-SH cell line (human neural epithelial cells obtained from bone marrow metastasis from the European Collection of Cell Cultures [ECACC, Salisbury, UK]) and fibroblast cell lines with Lipofectamine 2000 (Life Technologies) according to standard protocols. These cells are grown and the expression of mutated vs. wild-type CHMP2B is monitored through immunocytochemistry (immunofluorescence) to label CHMP2B markers to evaluate its location within the cytoplasm. The results of immunocytochemical experiments are visualised at the confocal microscope.

2.3 Analysis

Different software packages for analysis have been used to either generate or analyse genetic data.

2.3.1 3730 DNA Analyzer

For the Sanger Sequencing analysis the Foundation Data Collection Version 3.0 software (Life Technologies) is used to manage instrument/runs setup, real-time data visualization, and collection of raw data. The latter is imported into the Sequencing Analysis Version 5.2 software (Life Technologies) used to analyse, visualise, and save files generated from the 3730 DNA Analyzer. Specifically, the Sequencing Analysis software performs base-call using the KB base-caller program. These files are then exported to the Sequencher 4.9 software (Gene Codes Corporation, Ann Arbor, MI, USA) used to align all sequences to a reference sequence and analyse the data for the identification and annotation of single base pair changes and/or small indels.

For the fragment or expansion analyses, the GeneMapper Version 4.0 software (Life Technologies) is used to assess DNA sizing and quality allele calls. Among other applications, this software is used for analysing and visualising amplified fragment length polymorphism (AFLP).

2.3.2 7900HT fast real-time PCR system

For the SNP genotyping and the CNV assays the Sequence Detection System (SDS) software (Life Technologies) is used. SDS uses an advanced algorithm to calculate the intensities and the type of signal of each fluorescent marker from each well during the plate read to determine and assign a genotype call (allelic discrimination). Conversely, SDS creates and saves analysis templates on a per detector (fluorescent label) basis for the gene expression quantitation, providing the opportunity of visualizing the amplification plot and saving the threshold cycles (Ct) values (fractional cycle number at which the amount of amplified target reaches a fixed threshold) for the copy number and reference assay. The Ct values are exported to the CopyCaller v2.0 software (Life Technologies) that performs a comparative Ct (ΔΔCt) relative quantitation analysis of the real-time data. The comparative Ct (ΔΔCt) method calculates the difference between the threshold cycles of the target and reference assay sequences determining the number of copies of the target sequence in each test genomic DNA sample. The CopyCaller software displays the results of such computations in a bar graph format (**Figure 2-7**).

Figure 2-7. CNV analysis

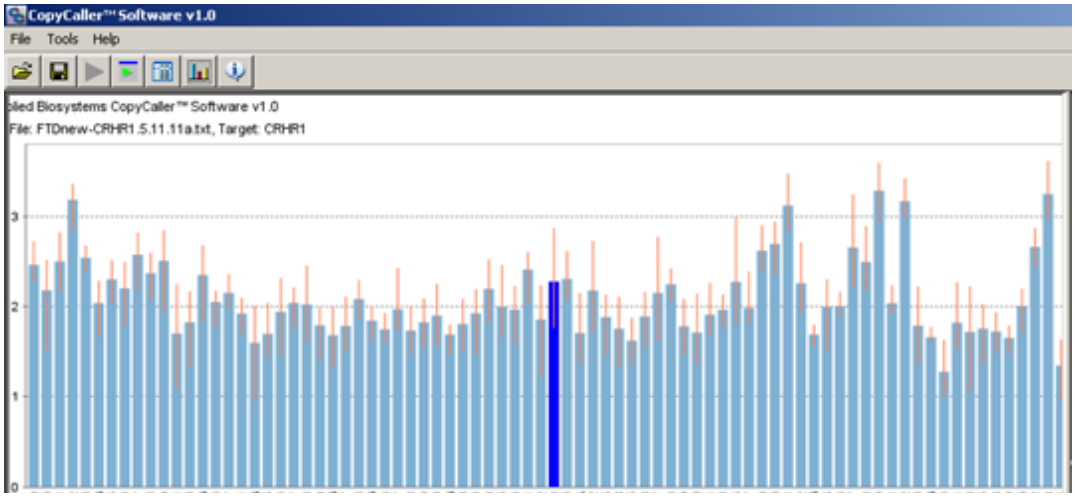


Figure 2-7. The output of the Copy Caller software allows visualising the copy numbers in a bar graph. The value can be exact (as in the figure) or normalised where the copy number is shown as an entire number (e.g. 2.2 is normalised to 2). The copy number is shown on the y axis and samples are displayed on the x axis.

2.3.3 Other tools for analysis (free web resources)

2.3.3.1 PolyPhen-2

When identifying a single base pair transition that leads to a missense change there are a couple of preferable methods to assess pathogenicity. First, one should screen that specific variant in an elevated number of normal control individuals, and, second, one should possibly perform *in vitro* and *in vivo* functional studies. When the latter is not an immediate possibility, a preliminary assessment can be performed *in silico*, through the online free tool PolyPhen-2 (Polymorphism Phenotyping version 2)

(http://genetics.bwh.harvard.edu/pph2/). PolyPhen-2 predicts the possible effect of amino-acid substitutions on protein structure and function by performing functional annotation of SNPs, mapping coding SNPs to gene transcripts and by comparing protein sequence, structural attributes, and conservation profiles [225]. The analysis of these properties results in an estimate of effect: Benign (green/yellow area, 0.0 to 0.50), possibly damaging (0.50 to 0.80) and probably damaging with a score (0.80 to 1.0) and sensitivity/specificity estimates (**Figure 2-8**).

Figure 2-8. Missense mutation effect: *In silico* analysis

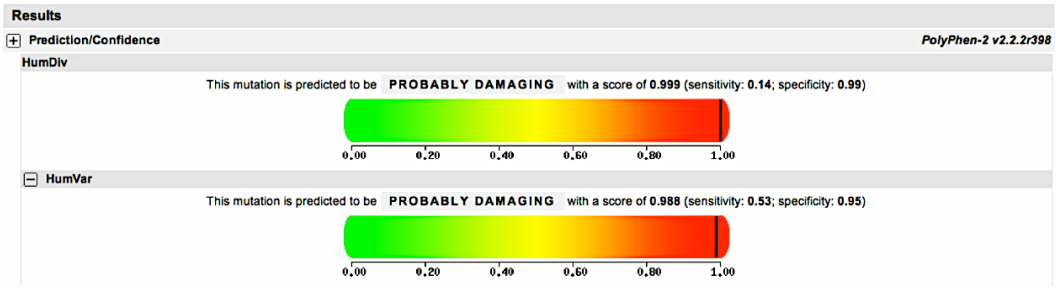


Figure 2-8. Example of the output of the analysis performed through the PolyPhen2 software. It predicts a score for the probability of the missense change of being benign (0.0-0.50), possibly damaging (0.50-0.80) or probably damaging (0.80-1.0) including an estimate of the specificity and sensitivity of the analysis. In the plot showing the results, the "HumDiv" annotation is the model for evaluating rare alleles or dense mapping of regions identified in GWAS, whilst the "HumVar" annotation is the model for Mendelian disease (assessing drastic vs. mild effect).

2.3.3.2 Haploview

When performing genetic screening by sequencing or genotyping (either multiplexed or genome-wide), the availability of a number of markers that cover a large region within and around a gene (or locus) allows evaluating the presence and frequency of specific haplotypes segregating with a certain phenotype. To perform such kind of analysis the Haploview software is used (http://www.broadinstitute.org/scientific-community/science/programs/medical-and-population-genetics/haploview/haploview). Haploview is an online free tool designed to create haplotype plots and evaluate haplotype frequencies (among other applications). The SNPs rs numbers, their chromosomal position as well as the allele frequencies need to be entered in the software that will calculate pairwise LD within a maximum of 500Kb surrounding each SNP of interest. Three coefficients provide an estimate of the strength of the association between two SNPs:

1 – D' is the LD coefficient that varies between 0–1 where the maximum value (1) implies that there is no definite evidence for recombination between two markers but does not entirely guarantees that there is absence of recombination between the two markers [226];

2 – r^2 is the correlation coefficient that indicates the predictive power between two loci [226] indicating strong LD when $r^2=1$, most probable LD when $0.8< r^2<1$ and decreasingly weaker LD when $0<r^2<0.8$, and;

3 – LOD (logarithm of the odds) score indicates whether linkage is suggested (LOD>2), established (LOD>3) or, weak or absent (LOD<2). The reason why LOD>3 indicates LD between two markers is due to the fact that in that specific case, the odds are a thousand (or more) to one that the two markers are linked/inherited together [226].

The haplotype plot visually shows the level of LD between SNPs under investigation (**Figure 2-9**).

Figure 2-9. LD analysis

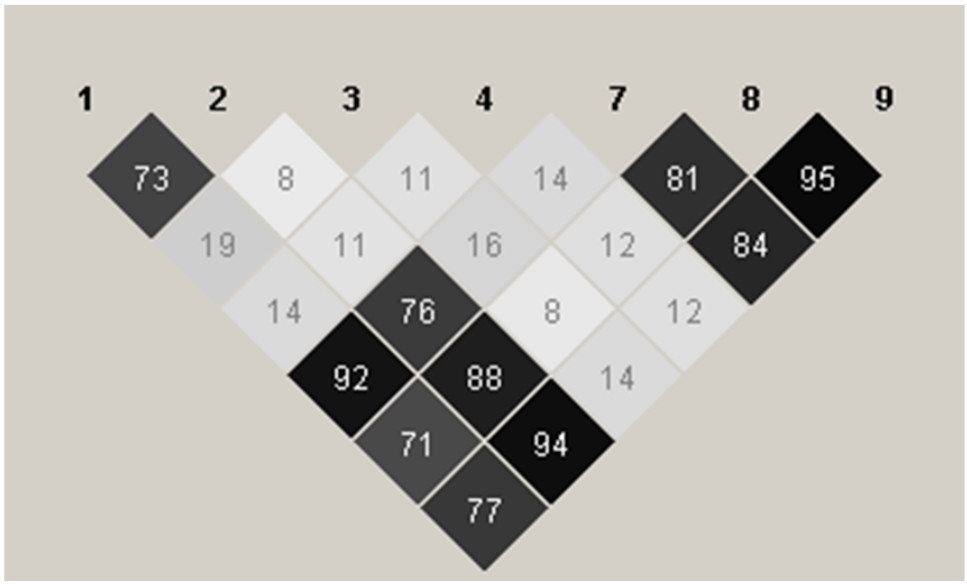


Figure 2-9. In each box the r^2 value between two SNPs is defined by numbers that range between 0 and 0.99. Black boxes without a number indicate r^2=1, i.e. complete LD. All three coefficients provide important information about the LD between two SNPs. For example, strong LD is identified between SNPs 8 and 9 with an r^2=0.95, D'=1 and LOD>3.

2.3.3.3 National Center for Biotechnology information (NCBI)

The ncbi website (http://www.ncbi.nlm.nih.gov) represents a remarkable source of data and information for genetic studies. The website services and tools have mainly been used to search through the literature (http://www.ncbi.nlm.nih.gov/pubmed/), to collect information on genes (http://www.ncbi.nlm.nih.gov/gene/), SNPs (http://www.ncbi.nlm.nih.gov/snp/), proteins (http://www.ncbi.nlm.nih.gov/protein/ and, eventually, diseases through the Online Mendelian Inheritance in Men (OMIM) encyclopedia (http://www.ncbi.nlm.nih.gov/omim/).

2.3.4 Whole genome-wide data

After generating whole genome genotyping data these are imported in the Genome Studio (GS) Data Analysis software (Illumina). This powerful informatics tool allows primary analysis of microarray-based data generated by the iScan System (or BeadXpress reader). The Genome Studio Genotyping (GT) Module (Illumina) is a specific supplementary software that enables a variety of features such as normalization of genotyping data, genotype calling, SNPs clustering, data intensity analysis, preliminary gender-mismatch check, loss of heterozygosity (LOH) calculation, and CNV analysis. Among other relevant features provided by GS and GT, most relevant are:

1 – The calculation of LogR ratio and B Allele Frequency to estimate and visualise CNV;

2 – The availability of multiple CNV algorithms and copy number variation analysis tools (such as cnv partition), and;

3 – The generation of a chromosomal browser heat map for examining and visualising SNPs distribution and CNV across the entire genome.

Report data created through GS are then exported to third party software packages for further analysis:

1 – Plink (http://pngu.mgh.harvard.edu/~purcell/plink/index.shtml) is a sophisticated whole genome association analysis bioinformatics toolset that provides all relevant applications to perform genotyping data cleaning and association analyses, and;

2 – R software (http://www.r-project.org) provides different statistical and graphical techniques including linear and nonlinear modelling and classical statistical tests to perform association analysis on genome-wide data and efficiently visualise results.

2.3.5 GWAS data analyses

This part of the work has been performed by biostatistician Dr Michael Nalls at NIH.

2.3.5.1 Quality control (QC) and data filtering for the FTD-GWAS

FTD-GWAS discovery phase data filtering

The GWAS on frontotemporal dementia included a large number of samples with variable genetic background (although all of overall European ancestry) as they were collected across different countries in Europe and the USA. In addition, genotypes for cases and controls in the discovery phase were mainly generated using different genotyping arrays. Therefore population substructure and limited overlap between the different genotyping arrays represented a challenge at first.

Low quality samples and poorly genotyped SNPs were excluded preliminarily based on the following criteria: Call rate (=percentage of successfully genotyped SNPs) <0.95, MAF<0.01, missing data >0.05, Hardy-Weinberg equilibrium (HWE) with p-value>$1x10^{-06}$ for cases and >$1x10^{-04}$ for controls and gender mismatch (evaluated through chromosome X heterozygosity). Palindrome SNPs (i.e. AT or GC transitions) were excluded; SNPs were flipped to the forward strand position and re-mapped based on HapMap 2 and 23 releases. Then all cases and controls were merged with all HapMap samples to confirm European ancestry that was verified using multidimensional scaling (outliers, >6SD, from the European reference population mean for component vectors 1 and 2 were excluded) and/or by Principal Component Analysis (PCA) as implemented in EIGENSTRAT [227] (cases and controls were matched based on similarity of the first two eigenvectors from PCA). Finally, the level of relatedness was evaluated by identifying and excluding 1st degree relatives through identity by descent (IBD) for any

two samples sharing >12.5% of the genome. To improve genome-wide coverage and statistical power surviving SNPs (~230K) were imputed to the 1000 Genomes August 2010 release, excluding SNPs with MAF<0.01 or imputation quality (Rsq)<0.30. All samples were haplotyped using MACH1 (http://www.sph.umich.edu/csg/abecasis/MACH/index.html) and these haplotype blocks were used for imputation through miniMac (http://genome.sph.umich.edu/wiki/Minimac). Each imputed region included 6,000 SNPs with ~500 SNPs overlapping with each neighbouring block. After regions were merged and the overlapping SNPs of lower quality were excluded 6M SNPs were used for association analysis.

FTD-GWAS replication phase data filtering

Standard QC was performed during replication phase as in the case of discovery phase. The NeuroX chip (section *2.2.2.8*) includes candidate (standard) neurological/neurodegenerative disease SNPs and exonic content. The SNPs were called using a publicly available cluster file based on over 60,000 samples [228]. All the variants with GenTrain scores >0.70 from GS, indicative of high quality genotype clusters, were extracted first to calculate call rates. As in discovery phase, samples with call rates <95%, samples with conflicting gender (matching clinically reported vs. genetically determined sex) and samples exhibiting excess of heterozygosity were excluded. In addition, SNPs with MAF<5%, HWE p-values<$1x10^{-05}$, and per SNP missingness rates>5% were excluded. Also, pairwise IBD filtering was used to remove

cryptically related samples and PCA was used to identify samples to be excluded due to genetic ancestry not consistent with European descent (based on comparisons with all HapMap reference populations).

2.3.5.2 Association analysis

Logistic regression was used based on genotyped and imputed SNP calls to assess the association between each SNP and any of the FTD subtypes first, adjusting for eigenvectors 1 and 2 from PCA as covariates to correct for population substructure. Second, a fixed effects meta-analysis was performed to combine results of each subtype and quantify heterogeneity across subtypes. Genomic inflation was minimal across subtypes and in the meta-analysis across subtypes (λ<1.05) supporting the notion that observed associations are overall trustworthy.

2.3.6 Quantitative trait loci (QTL) analyses

To evaluate *cis*-effects on quantitative traits such as expression and methylation patterns, QTL data were generated by the UK Brain Expression (UKBEC) and North American Brain Expression (NABEC) consortia (**Appendix 2-5**). Specifics about the generation of data and the analyses are included in the appropriate sections in **Chapters 3** and **4**.

2.4 Reagents and equipment

2.4.1 Reagents

The main reagents used to perform the experiments are summarised in the following paragraph. The provider name only is shown as their location has been previously reported (whole section ***2.2***). Promega Wizard DNA extraction kit (Promega), Cell Lysis Solution (Promega), Nuclei Lysis Solution (Promega), RNAse solution (Promega), Protein Precipitation Solution (Promega), isopropanol (VWR), ethanol (VWR), DNA Rehydratation Solution (Promega), Qiagen DNeasy Blood and Tissue Kit (Qiagen), proteinase K, Buffer AL (Qiagen), Buffer AW1 (Qiagen), Buffer AW2 (Qiagen), Buffer AE (Qiagen), Quant-iTTM dsDNA BR assay (Life Technologies), Quant-iT working solution (Life Technologies), DNA standards (Life Technologies), Roche Faststart PCR Master (400 RXN/10 ML) (Roche Applied Science), primers (Invitrogen), Ultrapure agarose (Life Technologies), TBE 1X buffer (VWR), GelRed nucleic acid gel stain (VWR), loading dye (EMD Millipore – Novagen), DNA ladder (EMD Millipore – Novagen), PCR filter plates (Millipore), BigDye Terminator v3.1 (Life Technologies), 5X Sequencing Buffer (Life Technologies), dimethyl sulfoxide (DMSO) (VWR), Q Solution (Invitrogen), deazaGTP (Fisher Scientific), Liz500 size standard (Life Technologies), HiDi Formamide (Life Technologies), TaqMan Genotyping Assay (Life Technologies), Genotyping Master Mix (Life Technologies), TaqMan Copy Number Assays (Life Technologies), TaqMan Copy Number Reference Assay (Life Technologies), QuikChange II Site-Directed mutagenesis kit (Startagene, CA, USA),

plasmid constructs (pcDNA3.1+; Invitrogen), XL1-Blue Supercompetent cells (Agilent Technologies), Dpn1 restriction enzyme (New Englan Biolabs), Luria Agar Base (VWR), Luria Broth (VWR), Ampicillin (VWR), SK-N-SH cell line (European Collection of Cell Cultures [ECACC]), Lipofectamine 2000 (Life Technologies).

All the reagents included in the Illumina Infinium assay used to perform the genome-wide genotyping are summarised in the Infinium HD Super protocol guide, Chapter 2 (http://supportres.illumina.com/documents/myillumina/05340b1f-c179-495d-b790-fa91ecbb6ff2/inf_hd_super_assay_ug_11322427_revc.pdf).

2.4.2 Equipment

The following instruments have been used to perform as well as analyse PCR, sequencing reaction, genotyping (single SNP, multiplexed and high coverage), and gene dosage experiments.

Veriti® 96-Well Thermal Cycler (Life Technologies): A thermal cycler for 96-well PCR plates used to perform high-throughput PCR and cycle sequencing applications.

GeneAmp® PCR System 9700 (Life Technologies): A thermal cycler for two 384-well PCR plates used to perform high-throughput PCR and cycle sequencing applications.

3730 DNA Analyzer (Life Technologies): A 48-capillary DNA Analyzer with medium-to-high throughput genetic analysis used for traditional DNA sequencing and fragment

analysis.

7900HT Fast Real-Time PCR (Life Technologies): A high-throughput real-time PCR system that detects and quantitates nucleic acid sequences. This instrument, that reads wavelength from 500-660 nm, is used to perform single SNP allelic discrimination (TaqMan SNP genotyping assay, section *2.2.2.6*) and gene expression quantitation (TaqMan CNV assay, section *2.2.2.7*).

iScan Reader (Illumina): This instrument is a highly precise and sensitive scanner that supports an extensive range of array applications, including genetic analysis assays for genotyping. This system is used for scanning of high density genotyping assay (Illumina Infinium genotyping assay).

CHAPTER 3 – PSP-GWAS

3.1 PSP-GWAS: The published study

Recently, Drs. Höglinger, Schellenberg and colleagues led the first international genome-wide association study (GWAS) of progressive supranuclear palsy (PSP) [180].

This study was undertaken to identify novel risk loci associated with PSP, beyond the known association with the *MAPT* locus and the H1 haplotype. It was a conventional two-stage case/control association study of a relatively large number of PSP cases provided and shared by Western European and American research groups (n>30). Stage 1 included ~1,100 pathologically proven PSP samples, whilst in stage 2 ~1,000 new PSP samples (~89% clinical and ~11% autopsy confirmed) were analysed.

The samples were genotyped on the Illumina Infinium platform using the 660W-Quad Infinium BeadChips for the cases and the HapMap550 Infinium BeadChip for the controls during stage 1. For stage 2 approximately 5,300 SNPs, that reached p-values$<1x10^{-3}$ during stage 1, were genotyped by means of the Infinium HD iSelect Custom BeadChip. Young individuals recruited at the Children's Hospital of Philadelphia (CHOP) Health Care Network were utilised as controls for stages 1 and 2. Different sets of controls were used for stage 1 and 2.

3.1.1 Discovery phase analysis

After data quality control, non-Caucasian samples and duplicates were dropped. Cases and controls were matched based on principal component analysis (PCA) and similarities of eigenvectors. The genomic inflation factor (λ=1.01) reflected good matching. After quality control filtering, 1,082 PSP cases and 3,317 controls were included in primary analysis. The genome-wide significance threshold was set at the p-value<$1x10^{-7.5}$. The analysis revealed suggestive to significant association at 5 loci (**Figure 3-1**), including the expected *MAPT* locus (**Table 3-1**). More than 5,400 SNPs reached p-values<$1x10^{-3}$.

Figure 3-1. PSP-GWAS discovery phase: Manhattan plot

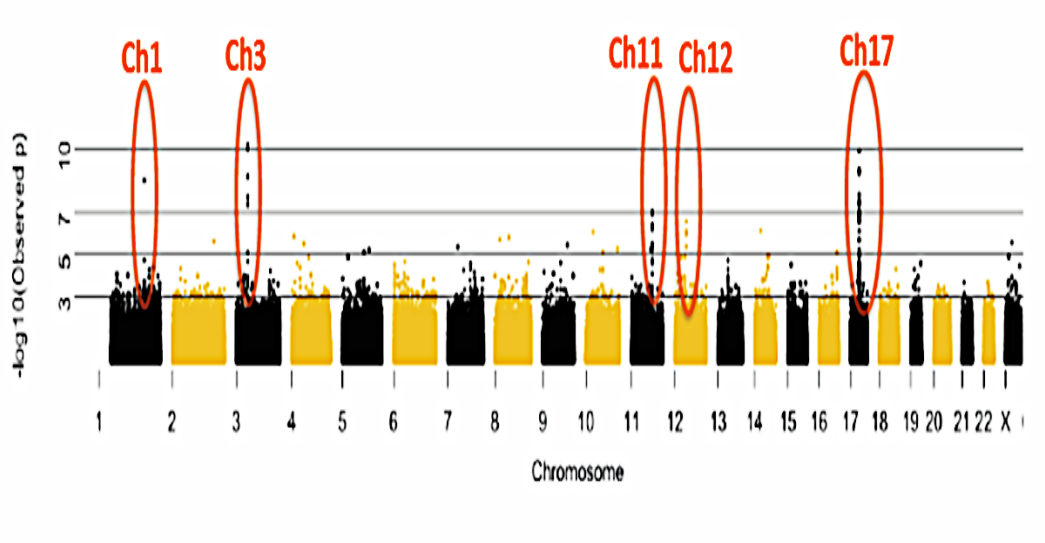


Figure 3-1. Manhattan plot of $-\log_{10}$ (observed p-value) show suggestive to significant association on chromosomes 1, 3, 11, 12 and 17. Derived from files shared by Dr Schellenberg's group.

Table 3-1. PSP-GWAS: Discovery phase associated loci

Chr	Gene	SNP	P-value; OR	Function
		Stage 1		
1	*STX6* (syntaxin 6)	rs12754041 rs1411478	$1.91x10^{-11}$; 1.403 $7.33x10^{-12}$; 1.413	A vesicle transporter protein. Modulator of P53 family. Member of the soluble N-ethylmaleimide-sensitive factor-attachment protein receptor (SNARE) complex for transport from early endosomes to Golgi complex. Induced by P53 and DNA damage. Knockdown leads to cell adhesion, cell cycle arrest and apoptosis
1	*MR1* (major histocompatibility complex, class I-related)	rs1411479	$5.91x10^{-09}$; 1.36	Has antigen presentation function. Involved in the development and expansion of a small population of T cells expressing an invariant T cell receptor alpha chain called mucosal-associated invariant T cells (MAIT)
3	*MOBP* (myelin-associated oligodendrocytic basic protein)	rs1768208 rs1464047	$4.66x10^{-13}$; 1.481 $3.80x10^{-12}$; 0.7018	Relatively abundant central nervous system (CNS)-specific myelin protein, which plays a role in stabilizing the myelin sheath in the CNS, has recently been implicated in the pathogenesis of multiple sclerosis (MS). Shares characteristics with myelin basic protein. Essential for normal arrangement of the radial component of CNS myelin. Target antigen in MS. Multiple splice variants. No myelination defects in MOBP$^{-/-}$ mice
11	*CEP57* (centrosomal protein 57kDa)	rs3017756	$1.66x10^{-06}$; 1.277	Links microtubules (MTs) to kinetochores and centrosome (centriole or centromatrix). Amino-terminus binds to γ-tubulin. Carboxy-terminus binds to MTs (nucleates and bundles MTs *in vitro*)
11	*MTMR2* (myotubularin related protein 2)	rs530914	$3.14x10^{-07}$; 0.77	Small integral membrane protein of lysosome/late endosome. Member of broad family of phosphatases (eg 3-phosphatase activity towards PtdIns3P and PtdIns(3,5) 2polyphosphoinositides. Regulates intracellular trafficking and membrane homeostasis in Schwann cell myelination. Interacts with NFL (low molecular weight neurofilament subunit. Mutations in *MTMR2* cause subtype of Charcot Mary Tooth (CMT) disease (CMT4B1; autosomal recessive). *MTMR 13* mutations cause CMT4B2. MTMR2 interacts with MTMR13. NFL assembly defects due to *MTMR2* mutations could contribute to pathogenesis of CMT4B
12	*NELL2* (NEL-like protein 2*)*	rs7299371	$6.47x10^{-08}$; 1.373	NELL2, a neuron-specific epidermal growth factor (EGF)-like glycoprotein. Specifically expressed during the peak periods of both sensory and motor neuron differentiation. Alternative splicing results in two isoforms, one secretable, located in the ER/Golgi compartments, one non-secretable, cytoplasmic protein. Enriched in glutamatergic neurons. Involved in a major, glutamate-dependent, process of neuroendocrine regulation. May play an important role in the development of the CNS as well as maintenance of neural functions, by regulating the intracellular machinery involving Ca2+ signaling, synaptic transport and/or release of vesicles

Table 3-1. List of the genes mapping to the loci showing association after primary analysis. The known functions of each gene are shown in the third column (information was gathered from the NCBI website [http://www.ncbi.nlm.nih.gov/gene] and shared files by Dr Schellenberg's group). Chr=chromosome; OR=odds ratio.

3.1.2 Replication phase and joint analysis

After genotyping ~5,300 SNPs chosen among those that reached p-values<$1x10^{-3}$ in stage 1, and performing standard quality controls, replication analysis was carried out for 4,099 SNPs in 1,051 PSP cases and 3,560 controls, all of European ancestry.

Of the 1,051 cases assessed in the replication phase 117 (~11%) were autopsy confirmed whilst the remaining were clinically diagnosed (~89%).

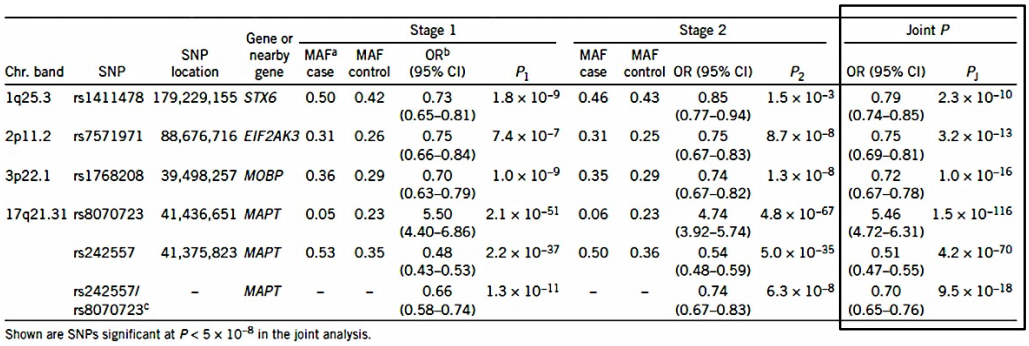
After completion of the joint analysis of stages 1 and 2, besides the expected association with chromosome 17 and the *MAPT* locus, this study identified three novel loci associated with PSP (**Figure 3-2**):

1 – Chr1q25.3 at the gene syntaxin 6 (*STX6*);

2 – Chr2p11.2 at the eukaryotic translation initiation factor 2-alpha kinase 3 gene (*EIF2AK3*), and;

3 – Chr3p22.1 at the myelin-associated oligodendrocyte basic protein gene (*MOBP*) [180].

Figure 3-2. PSP-GWAS: Association analysis



				Stage 1				Stage 2				Joint *P*	
Chr. band	SNP	SNP location	Gene or nearby gene	MAF[a] case	MAF control	OR[b] (95% CI)	P_1	MAF case	MAF control	OR (95% CI)	P_2	OR (95% CI)	P_j
1q25.3	rs1411478	179,229,155	*STX6*	0.50	0.42	0.73 (0.65–0.81)	1.8×10^{-9}	0.46	0.43	0.85 (0.77–0.94)	1.5×10^{-3}	0.79 (0.74–0.85)	2.3×10^{-10}
2p11.2	rs7571971	88,676,716	*EIF2AK3*	0.31	0.26	0.75 (0.66–0.84)	7.4×10^{-7}	0.31	0.25	0.75 (0.67–0.83)	8.7×10^{-8}	0.75 (0.69–0.81)	3.2×10^{-13}
3p22.1	rs1768208	39,498,257	*MOBP*	0.36	0.29	0.70 (0.63–0.79)	1.0×10^{-9}	0.35	0.29	0.74 (0.67–0.82)	1.3×10^{-8}	0.72 (0.67–0.78)	1.0×10^{-16}
17q21.31	rs8070723	41,436,651	*MAPT*	0.05	0.23	5.50 (4.40–6.86)	2.1×10^{-51}	0.06	0.23	4.74 (3.92–5.74)	4.8×10^{-67}	5.46 (4.72–6.31)	1.5×10^{-116}
	rs242557	41,375,823	*MAPT*	0.53	0.35	0.48 (0.43–0.53)	2.2×10^{-37}	0.50	0.36	0.54 (0.48–0.59)	5.0×10^{-35}	0.51 (0.47–0.55)	4.2×10^{-70}
	rs242557/ rs8070723[c]	–	*MAPT*	–	–	0.66 (0.58–0.74)	1.3×10^{-11}	–	–	0.74 (0.67–0.83)	6.3×10^{-8}	0.70 (0.65–0.76)	9.5×10^{-18}

Shown are SNPs significant at $P < 5 \times 10^{-8}$ in the joint analysis.

Figure 3-2. Summary of the results of the PSP-GWAS. P-values and OR for the associated loci are shown for respectively stage 1, stage 2 and the final meta (joint) analysis. The joint analysis shows the association at 1q25.3, 2p11.2, 3p22.1 and 17q21.31. OR was calculated based on the major allele. This Figure is taken form the published manuscript [180]

3.1.3 Lessons learned from the PSP-GWAS

The newly associated loci [180] point to specific pathways and processes that might be involved in the pathogenesis of PSP including:

1 – Intracellular vesicular trafficking;

2 – The endoplasmic reticulum (ER)-mediated cellular response to stress and abnormally unfolded proteins, and;

3 – Central nervous system (CNS) myelination.

More specifically, the associated SNP, rs1411478 (1q25.3; $P_{(joint)}$: $2.3x10^{-10}$), locates to the intronic region of *STX6,* between exons 4 and 5. *STX6* encodes the protein, syntaxin 6

(STX6) that belongs to the SNARE (soluble N-ethylmaleimide-sensitive factor-attachment protein receptor) proteins family and localizes to the trans-Golgi network (TGN) and endosomes, and is involved in intracellular endocytic and secretory trafficking processes [229, 230]. STX6 malfunction was recently associated with different human diseases such as cancer where STX6 seems to induce angiogenesis [231], cystic fibrosis (CF) [229] and Parkinson's and Alzheimer's disease (PD and AD) where STX6 seems to impact neurite outgrowth through a nerve growth factor (NGF)-dependent mechanism [232].

The associated SNP, rs7571971 (2p11.2; $P_{(joint)}$: $3.2x10^{-13}$), locates to the intronic region of *EIF2AK3* between exons 2 and 3. *EIF2AK3* encodes the eukaryotic translation initiation factor 2-alpha kinase 3 protein (EIF2AK3) that locates to the membrane of the endoplasmic reticulum (ER) and is involved in the ER stress response via the unfolded protein response (UPR) [233]. When unfolded proteins accumulate in the ER, EIF2AK3 phosphorylates the eukaryotic translation-initiation factor 2 (EIF-2) leading to a downstream effect that results in a decrease of translation processes [234]. Moreover, mutations in this gene have been associated with Wolcott-Rallison syndrome (WRS), an autosomal recessive disease, characterised by neonatal diabetes, liver and pancreas dysfunctions accompanied, in some cases, by signs of mental retardation [235].

The associated SNP, rs1768208 (3p22.1; $P_{(joint)}$: $1.0x10^{-16}$), locates to the intronic region of *MOBP* between exons 2 and 3. *MOBP* encodes the myelin-associated oligodendrocyte basic protein (MOBP) a member of the central nervous system (CNS) myelin-

constituting proteins [236]. MOBP is abundantly expressed specifically in oligodendrocytes in the white matter of the CNS [237]. It was previously reported that MOBP is involved in stabilizing the myelin constituting the myelin sheath and that it may play a role in multiple sclerosis (MS) as an auto-antigen [238].

Recently, the three PSP-GWAS associated SNPs were genotyped in a Chinese cohort diagnosed with AD and association could not be identified other than for the risk alleles in *EIF2AK3* and *MOBP* in *ApoE*-allele 4 carriers only [239]. Further, the GWAS associated SNP in *STX6* did not show association with Parkinson's disease (PD) in small series of diverse populations [240].

3.2 PSP-GWAS: Follow up study

3.2.1 The design

Our group participated and contributed to the PSP-GWAS with a total of 141 pathologically confirmed PSP cases collected at the Queen Square Brain Bank [180]. For this follow up study, we selected 84 of these 141 cases to further investigate the associated loci in order to possibly uncover the functional basis of the associations analysing coding variants (affecting protein structure and/or function) and non-coding variants (affecting gene expression and/or exon splicing) in our cohort. Specifically, we directly sequenced the coding and intronic flanking regions of the implicated genes with the following aims:

1 – Identifying coding pathogenic polymorphisms;

2 – Assessing genetic variability in the risk genes, the frequency of the variants, their allele frequencies and compare them to the normal population;

3 – Assessing the presence of possible disease specific haplotypes encompassing the risk genes (and the GWAS associated SNPs) and;

4 – Carrying out a regional genome-wide expression quantitative trait locus (eQTL) analysis in control brains to determine if any of the associated SNPs may influence allelic differences in gene expression.

3.2.2 Materials and Methods

3.2.2.1 Patients

We screened 84 pathologically confirmed PSP cases for which genome-wide data were available from the PSP-GWAS [180]. All our cases were of white western European ancestry. The patients diagnosis was assessed through the modified NINDS possible or probable criteria [216] and then confirmed pathologically by standardised criteria [160], [216], [217]. All patients gave informed consent and the samples were collected under protocols approved by the Joint Medical Ethics Committee of the National Hospital of Neurology and Neurosurgery, London. For this paragraph complementary information can be found in **Chapter 2** section ***2.1.1***.

3.2.2.2 Genetic analysis

Genomic DNA was extracted from human post-mortem brain tissue using the Qiagen DNeasy Blood and Tissue Kit (Qiagen, UK). We sequenced all the coding exons and flanking intronic regions of the following genes: *STX6* (exons 2-8), the major histocompatibility complex class I-related (*MR1*) (due to its close proximity to *STX6* [10.6 kb]; exons 2-7), *EIF2AK3* (exons 2-17) and *MOBP* (exons 3-5). Primers for PCR amplification were designed using Primer3 (http://frodo.wi.mit.edu/primer3/). Sequencing reactions of purified PCR amplicons were carried out in a single direction or in both directions if variants needed to be confirmed. Sanger sequencing was performed

through the Big Dye Terminator kit (ABI, Foster City, CA, USA) as recommended by manufacturer, and run on a 3730 DNA Analyzer (ABI), followed by analysis with Sequencher 4.9 software (Gene Codes Corporation, Ann Arbor, MI, USA). The potential damaging effect of novel missense variants on protein structure and function was examined using PolyPhen-2 software [225]. Identified common coding and non-coding variants were used to build haplotypes spanning ~70kb at 1q25.3, ~42.6kb at 2p11.2, and ~32kb at 3p22.1. LD was analysed using Haploview (www.broadinstitute.org/haploview) [241].

3.2.2.3 Expression quantitative trait loci (eQTL) analysis

This analysis was possible through the gene expression/genotyping data generated by the UK Brain Expression Consortium (UKBEC) [199], [207], [242], whose members' names are acknowledged in **Appendix 2-5**. Frozen brains from 134 control individuals were collected by the Medical Research Council (MRC) Sudden Death Brain and Tissue Bank, Edinburgh, UK [243], and the Sun Health Research Institute (SHRI) [244]. All samples were consented and authorized for ethically approved scientific investigation. Mean age at death was 58 years and [male:female] ratio was [1:2.8]. The brain regions analysed included those commonly affected in PSP such as the substantia nigra, putamen, hippocampus and frontal cortex [245] [165]. Total RNA was isolated using the miRNeasy 96 kit (Qiagen) and its quality was evaluated through the 2100 Bioanalyzer and RNA 6000 Nano Kit (Agilent). Total RNA was processed with the Ambion WT

Expression Kit and Affymetrix GeneChip Whole Transcript Sense Target Labeling Assay and hybridized to the Affymetrix Exon 1.0 ST Arrays as recommended by manufacturer. Hybridized arrays were scanned on an Affymetrix GeneChip Scanner 3000 7G and visually inspected for hybridization artifacts. The Affymetrix probe sets were mapped to human genome build 19 (GRCh37). Analysis was restricted to 292K probe sets annotated to NCBI Reference Sequence build 36 through Netaffx annotation file Release 31 (HuEx-1_0-st-v2 Probeset Annotations). Expression data was adjusted for brain bank, gender and batch effects in Partek's Genomics Suite v6.6 (Partek Inc., USA). Samples were genotyped on the Illumina Infinium Omni1-Quad BeadChip and Immunochip [246] [221], scanned using an iScan (Illumina, USA). Genome Studio v.1.8.X (Illumina, USA) was used for generating SNP calls. After standard quality controls (exclusion of samples with non-European descent, call rate <95% and checks on gender, cryptic relatedness, autosomal heterozygosity rate, and HWE [p-value<0.0001]), imputation was performed using MaCH [247] and minimac (http://genome.sph.umich.edu/wiki/Minimac) using the European panel of the 1000 Genomes Project (March 2012: Integrated Phase I haplotype release version 3) leading to ~5.88 million SNPs and ~577K indels with good post-imputation quality (Rsq>0.50) and MAF≥5%. QTL analysis was run for each expression profile (either exon- or gene-level) against either SNP or indel in MatrixEQTL [248]. Subsequent analyses were conducted in R (http://www.r-project.org).

3.2.3 Results

After sequencing the exons and flanking intronic regions of *STX6*, *MR1*, *EIF2AK3* and *MOBP*, we compared the minor allele frequencies (MAFs) of the known SNPs in our PSP cohort with data available from the 1000 Genomes (http://browser.1000genomes.org/index.html) and dbSNP (http://www.ncbi.nlm.nih.gov/SNP/). Furthermore, we used the sequencing results and the allele frequencies to build and possibly identify disease-associated haplotypes in linkage disequilibrium (LD) with the GWAS associated SNPs. The haplotype structures were assessed using Haploview [241]. In addition, of the 26 genetic variants of interest identified in this analysis (including the GWAS associated SNPs), 16 had a MAF of >5% and could be analysed within the available eQTL data set. Each genetic variant was assessed for evidence of regulation of any gene within 1 Mb with particular focus on the effects on *STX6*, *MR1*, *EIF2AK3* and *MOBP*.

3.2.3.1 STX6/MR1

The sequencing of *STX6* resulted in the isolation of three coding variants one of which was novel. The two known variants were synonymous changes, Glu13Glu (rs12125196) and Asn217Asn (rs3747957), whilst the novel variant was a missense mutation located to exon 8 of *STX6*, Cys236Gly (1/58, heterozygous change). *In silico* analysis through PolyPhen-2 software predicted this variant being probably damaging (score: 0.999;

sensitivity: 0.14; specificity: 0.99) [225] (**Figure 3-3**). The carrier of this mutation did not have a reported family history of PSP or other neurodegenerative disorders.

Figure 3-3. *STX6*; missense change Cys236Gly. PolyPhen-2 analysis

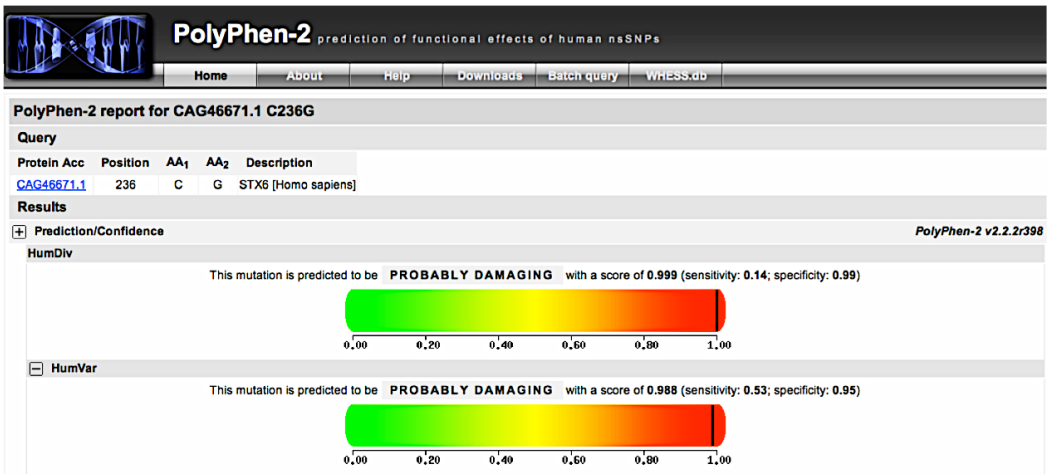


Figure 3-3. *In silico* analysis of the effect of the missense change Cys236Gly

The sequencing of *MR1* resulted in the identification of several variants, both, coding and non-coding. Four variants were intronic two of which very proximal to exon 3 (rs75150495, 8 bp from the start of the exon) and exon 7 (rs3747956, 3 bp after the end of exon 7). The remaining 5 variants were coding. Three of them were missense mutations (rs41268456, Arg31His; rs2236410, His39Ala; rs149433107, Pro203Ser), whilst 2 were synonymous changes (rs3863720, Ser46Ser; rs35223984, Ans239Asn). Of

the missense mutations only Arg31His was predicted to be probably damaging after *in silico* analysis (score: 0.996; sensitivity: 0.55; specificity: 0.98) (**Figure 3-4**) [225].

Figure 3-4. *MR1*; missense change Arg31His. PolyPhen-2 analysis

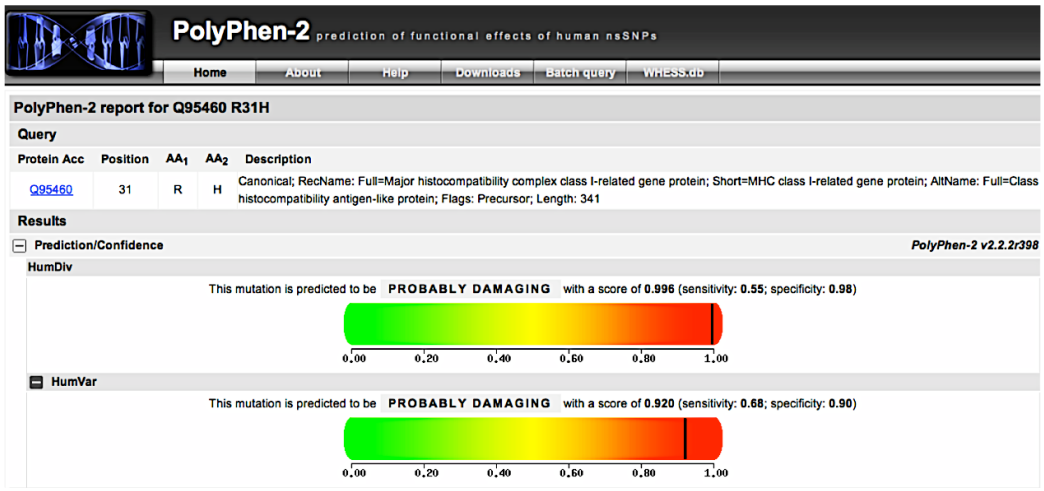


Figure 3-4. *In silico* analysis of the effect of the missense change Arg31His

The results of genetic screening of *STX6* and *MR1* are summarized in **Table 3-2**. Amino-acid numbering for STX6 and MR1 proteins refers to GenBank Accession CAG46671.1 and CAB77667.1, respectively.

Table 3-2. UK PSP study cohort: Genetic screening

Gene	Exon	Intron	rs number	Amino-acid change	Alleles/Frequency			Minor allele	MAF		
									Our samples	*1000 Genomes*	*Caucasian/european (dbSNP)*
STX6	2		rs12125196	Glu13Glu	GG (58/60)	GA (2/60)	AA (0/60)	A	0.017	0.011	0.017-0.025
		4	*rs1411478	/	GG (21/82)	GA (44/82)	AA (17/82)	A	0.475	0.426	0.375-0.473-0.483
	7		rs3747957	Asn217Asn	TT (14/62)	TC (33/62)	CC (15/62)	T	0.492	0.427	0.473-0.483
	8		**novel**	**Cys236Gly**	**TT (57/58)**	**TG (1/58)**	**GG (0/58)**	**G**	**0.009**	**NA**	**NA**
MR1		2	rs2236412	/	TT (64/70)	TC (5/70)	CC (1/70)	C	0.050	0.050	NA
			rs75150495	/	CC (63/70)	CG (4/70)	GG (3/70)	G	0.070	NA	NA
			rs41268456	Arg31His	GG (69/70)	GA (1/70)	AA (0/70)	A	0.007	0.004	NA
	3		rs2236410	His39Ala	AA (54/70)	AG (16/70)	GG (0/70)	G	0.114	0.206	0.150
			rs3863720	Ser46Ser	GG (69/70)	GA (0/70)	AA (1/70)	A	0.014	0.017	0.042
	5		rs149433107	Pro203Ser	CC (78/79)	CT (1/79)	TT (0/79)	T	0.006	0.001	NA
			rs35223984	Asn239Asn	CC (74/78)	CT (4/78)	TT (0/78)	T	0.026	0.014	0.033
		7	rs3747956	/	GG (35/64)	GA (23/64)	AA (6/64)	A	0.273	0.494	0.267-0.302
			rs75073791	/	GG (62/64)	GA (2/64)	AA (0/64)	A	0.016	0.011	0.050
EIF2AK3	2		rs867529	Ser136Cys	CC (31/63)	CG (28/63)	GG (4/63)	G	0.286	0.288	0.267-0.292
		2	*rs7571971	/	CC (36/80)	CT (40/80)	TT (4/80)	T	0.300	0.287	0.267-0.317
	3		rs13045	Gln166Arg	AA (6/59)	AG (25/59)	GG (28/59)	A	0.313	0.347	0.300-0.343
	9		rs141901506	Asp502Asn	GG (57/58)	GA (1/58)	AA (0/58)	A	0.009	NA	NA
	10		rs55791823	Asp566Val	AA (59/61)	AT (2/61)	TT (0/61)	T	0.016	0.001	0.002
		10	rs6750998	/	AA (39/59)	AT (15/59)	TT (5/59)	T	0.212	0.196	0.242-0.258
	11		rs1805164	Gln597Gln	AA (33/65)	AG (26/65)	GG (6/65)	G	0.292	0.293	0.270-0.325-0.350
	13		rs1805165	Ala704Ser	GG (3/43)	GT (17/43)	TT (23/43)	G	0.267	0.287	0.267-0.274-0.280-0.292
		13	rs4972221	/	AA (30/64)	AT (25/64)	TT (9/64)	T	0.336	0.347	0.300-0.317
MOBP		2	*rs1768208	/	CC (32/82)	CT (43/82)	TT (7/82)	T	0.347	0.349	0.292-0.295-0.317
	3		**novel**	**Gln82Lys**	**CC (63/64)**	**CA (1/64)**	**AA (0/64)**	**A**	**0.008**	**NA**	**NA**
		3	rs2233204	/	CC (30/68)	CT (27/68)	TT (11/68)	T	0.360	0.196	0.283-0.317-0.327
		4	rs552724	/	CC (54/69)	CT (8/69)	TT (6/69)	T	0.145	0.210	0.183-0.195

Table 3-2. Summary of common and novel variants identified at the loci associated with PSP [180]. All SNPs with their allele frequencies identified in our sample population are shown and compared with those obtained from the 1000 Genome and dbSNP datasets. Novel variants are shown in bold font. * PSP-GWAS associated SNPs.

In our sample population the minor allele frequency (MAF) of the variants identified in *STX6* and *MR1* did not differ substantially from the 1000 Genomes or dbSNP releases. The haplotype analysis did not reveal any evident LD pattern besides a complete LD (*D*'=1; r^2=1; LOD=25.58) between the GWAS associated SNP rs1411478 and rs3747957 (Asn217Asn in *STX6*) as seen in **Figure 3-5**.

Figure 3-5. LD pattern at the Chr1q25.3 locus

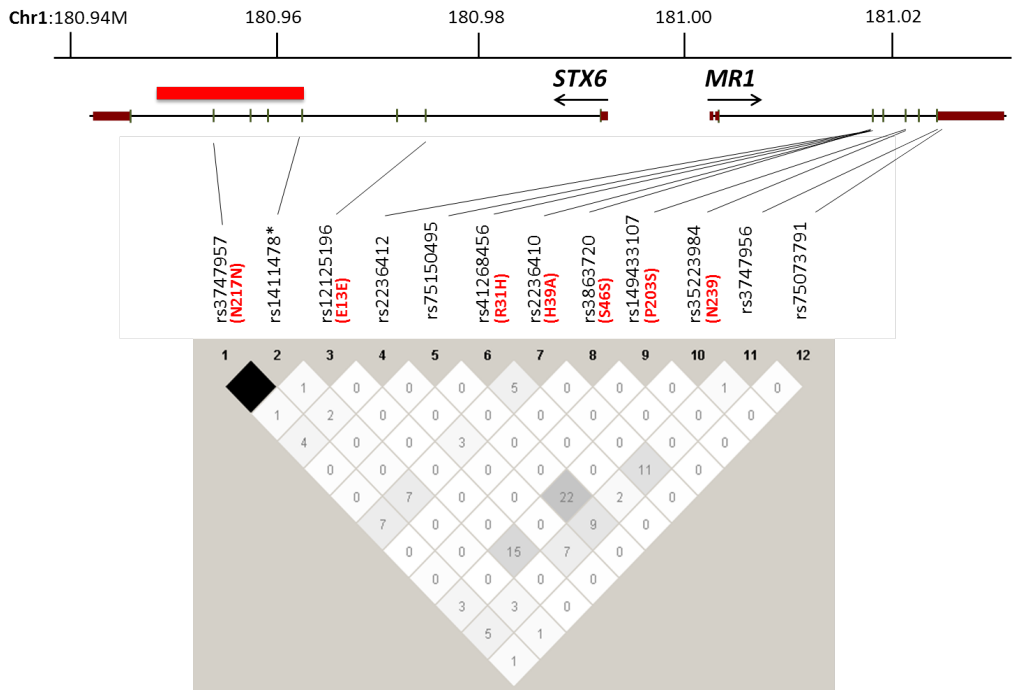


Figure 3-5. The relative location of *STX6* and *MR1* with chromosomal coordinates (Mb; Genome build hg19/GRCh37) is shown at the top. Arrows below gene symbols indicate direction of transcription. In each box the r^2 value between two SNPs is shown with ranges that vary between 0 and 0.99. The black box with no number indicates r^2=1, i.e. complete LD. Complete LD was evident only for SNPs rs3747957 and rs1411478 (* GWAS associated SNP).

Such trend, however, can be seen also in the normal population for both the 1000 Genome release (**Table 3-3A**) and the HapMap 22 release (**Table 3-3B**), where it can be seen that LD between rs1411478 and rs3747957 is complete, with *D*'=1 and r^2=1.

Table 3-3. Pairwise LD in CEU population

A

1000G pilot	SNP	Proxy	Distance	RSquared	DPrime	Chr	Coordinate_HG18
	rs1411478	rs12754041	4320	1	1	chr1	179224585
	rs1411478	**rs3747957**	8429	**1**	1	chr1	179220476
	rs3747957	rs12754041	4109	1	1	chr1	179224585
	rs3747957	**rs1411478**	8429	**1**	1	chr1	179228905

B

HapMap22 release	SNP	Proxy	Distance	RSquared	DPrime	Chr	Coordinate_HG18
	rs1411478	rs12754041	4320	1	1	chr1	179224585
	rs1411478	**rs3747957**	8429	**1**	1	chr1	179220476
	rs3747957	rs12754041	4109	1	1	chr1	179224585
	rs3747957	**rs1411478**	8429	**1**	1	chr1	179228905

Table 3-3. Data extracted from the "Proxy Search" tool for the SNPs identified at the *STX6* locus (http://www.broadinstitute.org/mpg/snap/ldsearch.php). Complete LD (*D*'=1; r^2=1) can be seen for SNPs rs1411478 and rs3747957 in population with European ancestry from the 1000 Genome pilot study (**A**) and HapMap 22 (**B**) releases (as it was the case for our PSP population as per **Figure 3-5**).

The eQTL analysis showed interesting results for rs1411478 as stratifying mRNA levels by genotype of this SNP provided evidence for association between the risk allele A and

decreased expression of *STX6* ($p=1.80x10^{-09}$; Affymetrix exon arrays, transcript ID 2446567) in white matter but in no other brain region (**Figure 3-6A** and **B**).

Figure 3-6. Rs1411478: Expression quantitative trait loci analysis

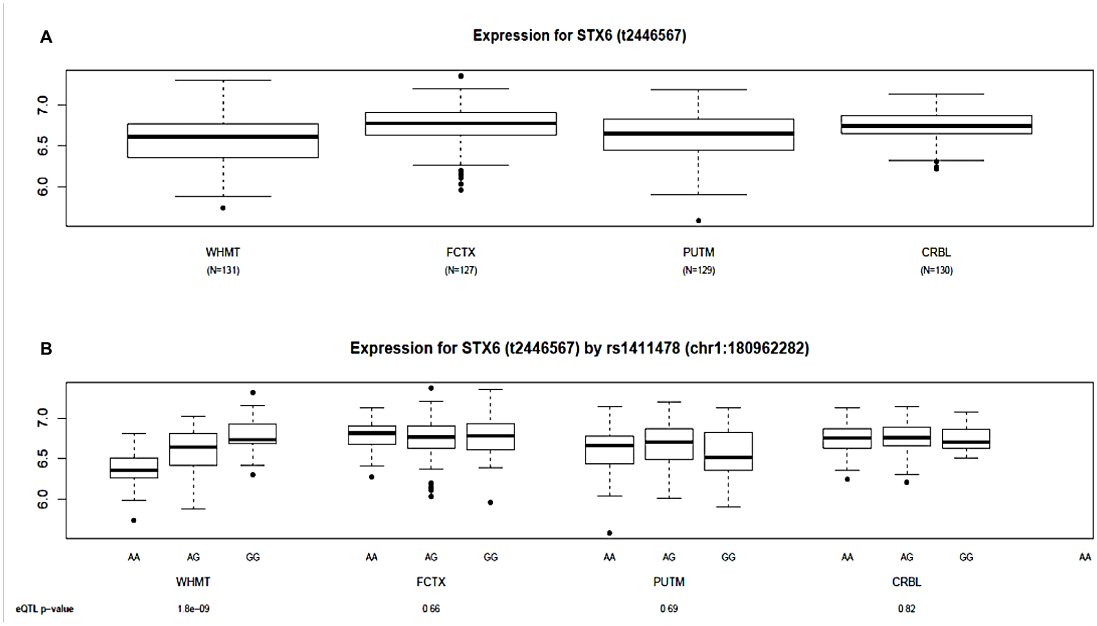


Figure 3-6. eQTL signal associated with rs1411478. (**A**) Boxplot showing STX6 mRNA expression across four brain regions: White matter (WHMT), frontal cortex (FCTX), putamen (PUTM) and cerebellar cortex (CRBL). Whiskers extend from the box to 1.5 times the inter-quartile range. (**B**) STX6 mRNA expression stratified by the genotypes of the PSP risk SNP rs1411478 in WHMT, FCTX, PUTM and CRBL. The p-values are shown on the x-axis.

By investigating all SNPs (genotyped and imputed) within 1Mb of the transcription start and stop site for *STX6* it was confirmed that rs1411478 was the only eQTL signal in this genomic region. Therefore it is difficult to assess whether the SNP rs3747957, which

corresponds to the silent change Asn217Asn, has any implication with the diseased status, considering that:

1 – Strong LD with rs1411478 can be seen in both our PSP population (**Figure 3-5**) and in the normal population with European ancestry from both the 1000 Genomes and HapMap 22 releases (**Table 3-3A** and **B**), and;

2 – the silent change does not affect the protein structure or function and our own QTL data don't support the notion that the base change at rs3747957 may underlie effects on expression or splicing such as in the case for other silent mutations [190].

In addition, there was seemingly no *cis* effect on expression exerted by the intronic polymorphisms rs75150495 and rs3747956 located at 8bp from the start of the exon 3 and at 3bp after the end of exon 7, respectively, in *MR1*.

3.2.3.2 EIF2AK3

The sequencing of *EIF2AK3* resulted in the identification of several known coding and non-coding variants (**Table 3-2**). Two variants were intronic, one of which, rs6750998, locates just 6bp downstream of exon 13. Of the 6 coding variants 5 were missense changes (rs867529, Ser136Cys; rs13045, Gln166Arg; rs141901506, Asp502Asn; rs55791823, Asp566Val; rs1805165, Ala704Ser) and one was a synonymous change (rs1805164, Gln597Gln). After *in silico* analysis, all missense variants were predicted to be benign besides Asp566Val that was shown to be probably damaging (score: 0.993;

sensitivity: 0.7; specificity: 0.97) (**Figure 3-7**) [225]. Amino-acid numbering for EIF2AK3 refers to the NCBI Reference Sequence: NP_004827.4.

Figure 3-7. *EIF2AK3*; missense change Asp566Val. PolyPhen-2 analysis

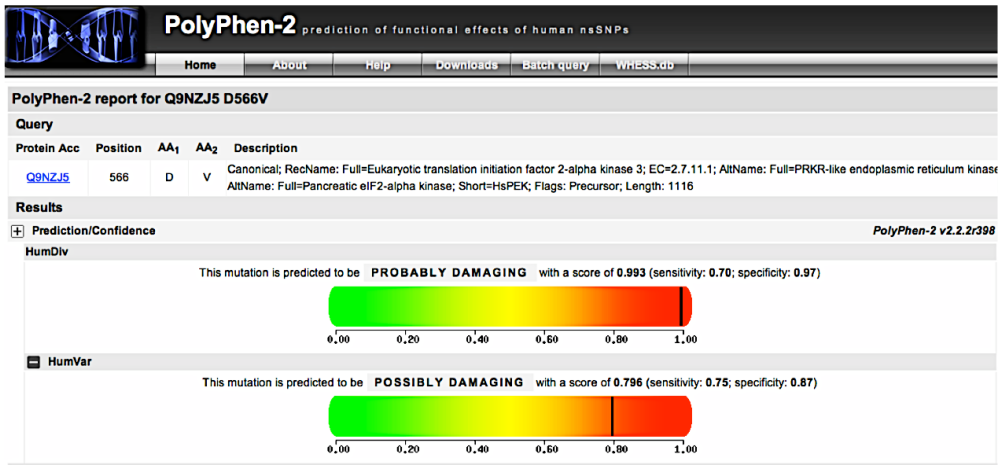


Figure 3-7. *In silico* analysis of the effect of the missense change Asp566Val

It is interesting to note that the MAF for rs55791823 (Asp566Val) was increased in our PSP cohort compared to the 1000 Genomes and dbSNP data for a Caucasian-European population (CEU) (0.016 vs. 0.001 and 0.002; **Table 3-2**), whereas MAFs for all the other *EIF2AK3* SNPs did not differ substantially from the 1000 Genomes or dbSNP data.

The haplotype block analysis identified LD associations (**Figure 3-8**).

Figure 3-8. LD pattern at the Chr2p11.2 locus

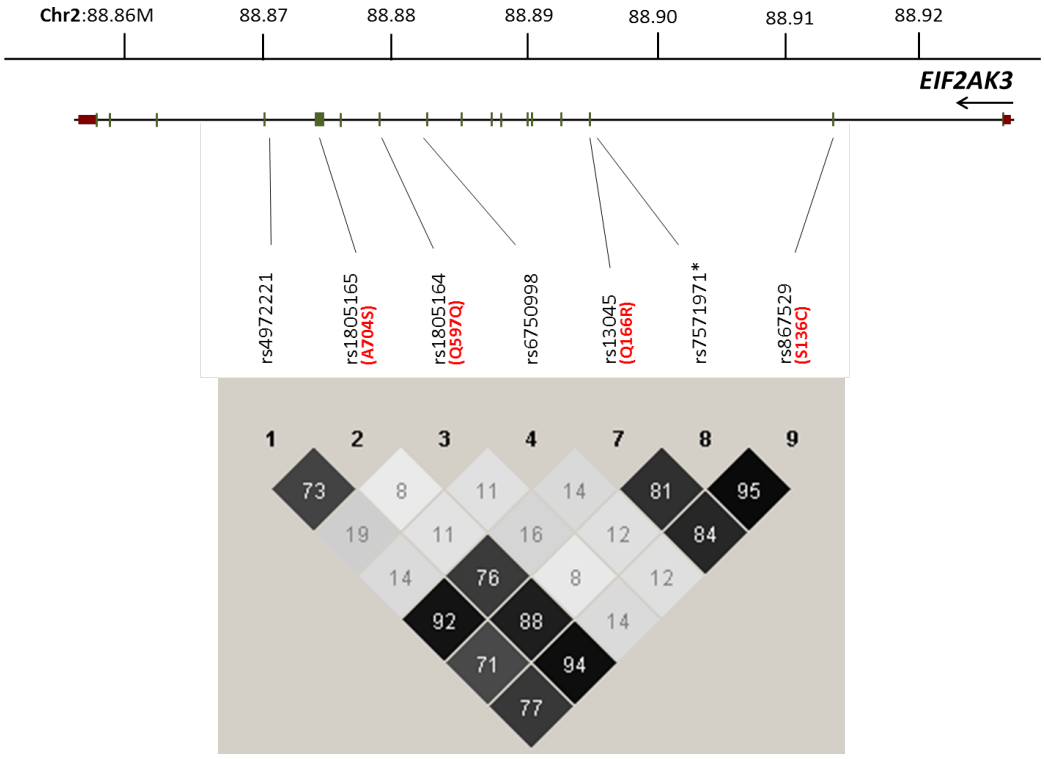


Figure 3-8. The relative location of *EIF2AK3* with chromosomal coordinates (Mb; Genome build hg19/GRCh37) is shown at the top. The arrow below the gene symbol indicates the direction of transcription. In each box the r^2 value between two SNPs is shown with ranges that vary between 0 and 0.99. Any value $0.8<r^2<1$ is suggestive of LD. * GWAS associated SNP

From a careful analysis of the LD patterns in the PSP cohort the following can be derived:

1 – The SNPs rs867529 (Ser136Cys) and rs13045 (Gln166Arg) are in almost complete LD with each other and, both, with the GWAS associated SNP rs7571971;

2 – rs867529 (Ser136Cys) and the GWAS associated SNP rs7571971 are both in almost complete LD with rs1805165 (Ala704Ser), and;

3 – the SNP rs4972221 is in almost complete LD with rs13045 (Gln166Arg). All these observations were based on the notion that complete LD is defined by the following values: *D*'=1; r^2=1; LOD>2.

Of note, we compared the haplotype architecture and LD scores of our PSP cohort with those available for the control population with European ancestry obtained from the "Pairwise LD" analysis (http://www.broadinstitute.org/mpg/snap/ldsearchpw.php, from the HapMap 22 release), to possibly identify differences that might hint for a disease associated haplotype.

However, the LD blocks observed in our study population did not differ from those of the control population as seen in **Table 3-4**.

In addition, our QTL analysis did not reveal significant effects on *cis* expression or splicing for any of the SNPs at this locus.

Table 3-4. Comparison of pairwise LD in CEU population and UK PSP cohort

Pairwise LD/HapMap 22 release							Our study		
SNP	**Proxy**	**Distance**	**RSquared**	**DPrime**	**Chr**	**Coordinate_HG18**	**RSquared**	**DPrime**	**LOD**
rs867529	**rs7571971**	**17922**	**1**	**1**	chr2	88676466	**0.958**	**1**	**20.18**
rs867529	**rs1805165**	**38382**	**1**	**1**	chr2	88656006	**0.942**	**1**	**14.33**
rs867529	**rs13045**	**18150**	**0.848**	**1**	chr2	88676238	**0.844**	**1**	**16.62**
rs867529	rs1805164	34142	0.196	1	chr2	88660246	0.144	1	1.89
rs867529	rs6750998	30331	0.116	1	chr2	88664057	0.125	1	**3.07**
rs7571971	**rs867529**	**17922**	**1**	**1**	chr2	88694388	**0.958**	**1**	**20.18**
rs7571971	**rs1805165**	**20460**	**1**	**1**	chr2	88656006	**0.881**	**1**	**12.60**
rs7571971	**rs13045**	**228**	**0.848**	**1**	chr2	88676238	**0.81**	**1**	**16.32**
rs7571971	rs1805164	16220	0.196	1	chr2	88660246	0.089	0.754	1.00
rs7571971	rs6750998	12409	0.116	1	chr2	88664057	0.124	1	**3.42**
rs13045	**rs7571971**	**228**	**0.848**	**1**	chr2	88676466	**0.81**	**1**	**16.32**
rs13045	**rs867529**	**18150**	**0.848**	**1**	chr2	88694388	**0.844**	**1**	**16.62**
rs13045	**rs1805165**	**20232**	**0.848**	**1**	chr2	88656006	0.768	0.932	**9.93**
rs13045	rs1805164	15992	0.231	1	chr2	88660246	0.164	1	**2.74**
rs13045	rs6750998	12181	0.137	1	chr2	88664057	0.149	1	**3.72**
rs6750998	rs1805164	3811	0.172	1	chr2	88660246	0.119	1	1.75
rs6750998	rs13045	12181	0.137	1	chr2	88676238	0.149	1	**3.72**
rs6750998	rs1805165	8051	0.116	1	chr2	88656006	0.11	1	1.78
rs6750998	rs7571971	12409	0.116	1	chr2	88676466	0.124	1	**3.42**
rs6750998	rs867529	30331	0.116	1	chr2	88694388	0.125	1	**3.07**
rs1805164	rs13045	15992	0.231	1	chr2	88676238	0.164	1	**2.74**
rs1805164	rs1805165	4240	0.196	1	chr2	88656006	0.08	0.731	0.92
rs1805164	rs7571971	16220	0.196	1	chr2	88676466	0.089	0.754	1.00
rs1805164	rs867529	34142	0.196	1	chr2	88694388	0.144	1	1.89
rs1805164	rs6750998	3811	0.172	1	chr2	88664057	0.119	1	1.75
rs1805165	**rs7571971**	**20460**	**1**	**1**	chr2	88676466	**0.881**	**1**	**12.60**
rs1805165	**rs867529**	**38382**	**1**	**1**	chr2	88694388	**0.942**	**1**	**14.33**
rs1805165	**rs13045**	**20232**	**0.848**	**1**	chr2	88676238	0.768	0.932	**9.93**
rs1805165	rs1805164	4240	0.196	1	chr2	88660246	0.08	0.731	0.92
rs1805165	rs6750998	8051	0.116	1	chr2	88664057	0.11	1	1.78
rs4972221	**rs13045**	**24491**	**1**	**1**	chr2	88676238	**0.922**	**0.96**	**20.30**
rs4972221	**rs1805165**	**4259**	**0.848**	**1**	chr2	88656006	0.736	0.935	**10.44**
rs4972221	**rs7571971**	**24719**	**0.848**	**1**	chr2	88676466	0.71	0.911	**13.84**
rs4972221	**rs867529**	**42641**	**0.848**	**1**	chr2	88694388	0.777	0.955	**15.11**
rs4972221	rs1805164	8499	0.231	1	chr2	88660246	0.19	1	**3.52**
rs4972221	rs6750998	12310	0.137	1	chr2	88664057	0.145	1	**3.84**

Table 3-4. *D'* and r^2 values in the normal population with European ancestry (HapMap 22 release) are compared with those of our PSP cohort. All the SNPs with *D'*=1 or >0.95 and $0.8<r^2<1$ are bolded and indicate the overlap between the two populations being compared. No substantial differences can be seen suggesting that there are no obvious divergent LD patterns between the PSP vs. the normal population.

3.2.3.3 MOBP

The sequencing of *MOBP* resulted in the isolation of three variants: One novel missense mutation (Gln82Lys) and two known intronic variants (rs2233204 and rs552724). The missense mutation (1/64, heterozygous) resulted being a benign change after performing *in silico* analysis (**Figure 3-9**). The amino-acid numbering for MOBP is with reference to GenBank Accession BAA05660.1.

Figure 3-9. *MOBP*; missense change Gln82Lys. PolyPhen-2 analysis

Figure 3-9. *In silico* analysis of the effect of the missense change Gln82Lys

MAFs in our cohort were overall comparable to those from the 1000 Genomes releases or Caucasian/European populations (**Table 3-2**).

No relevant data could be obtained from the haplotype analysis (**Figure 3-10**), leaving the association with this locus open for questions as it does not seem, at least in our study population, that there is genetic variability within *MOBP* that justifies the association neither from a coding nor an expressional perspective. For the latter case, in fact, there was no evidence of *cis* gene or exon level QTLs.

Figure 3-10. LD pattern at the Chr3p22.1 locus

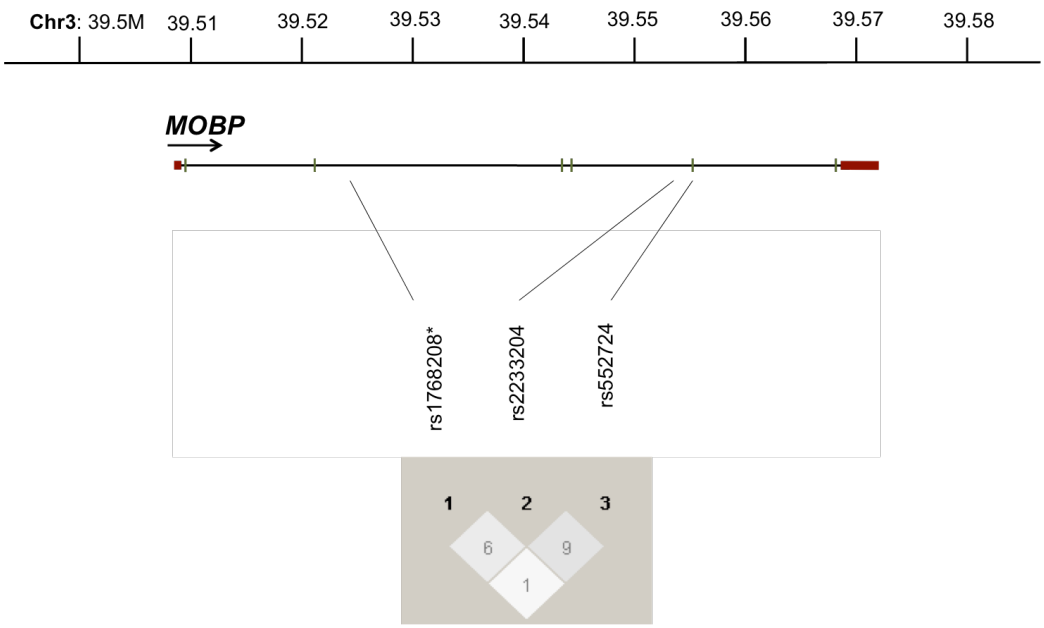


Figure 3-10. LD pattern at 3p22.1. The relative location of *MOBP* with chromosomal coordinates (Mb; Genome build hg19/GRCh37) is shown at the top. The arrow below the gene symbol indicates the direction of transcription. In each box the r^2 value between two SNPs is shown with ranges that vary between 0 and 0.99. Any value $0.8<r^2<1$ is suggestive of LD. * GWAS associated SNP.

3.2.4 Discussion

3.2.4.1 PSP and the MAPT locus

MAPT has long held a distinctive contribution to progressive supranuclear palsy and its definition of primary tauopathy. PSP is, in fact, characterised by:

1 – Hyperphosphorylated tau inclusions as its pathological hallmark, and;

2 – Robust genetic association with the *MAPT* H1 haplotype on chromosome 17 [173], [179], [177], [178].

The latter has been repeatedly reported across different studies and was further confirmed in the recent PSP-GWAS that showed extensive genome-wide significance at rs8070723 (P_j=1.5x10^{-116}) and rs242557 (P_j=4.2x10^{-70}) (**Figure 3-2**) [180]. Some expression studies suggested the role of allele-specific differences in *MAPT* gene expression as the functional basis of this association [201] and as the reason for an increased production of tau isoforms with four microtubule-binding repeat domains (4R-tau), ultimately resulting in the 4R-tau dominant neuronal and glial inclusions in PSP brains [160]. To date, the majority of the variants identified in *MAPT* are point mutations mainly located within and/or around *MAPT* exon 10, which affect tau's 4R/3R ratio [184], [182], [185], [186], [181], [183], [187], [188], [189], [190]. However, *MAPT* variability in this fashion seems too rare to justify the genetic link between PSP, *MAPT* and its extended haplotype. Several studies over the past 5-6 years have tried to untangle the functional basis of the *MAPT* haplotype and its association with PSP. The *MAPT* H1c sub-haplotype was shown

having strong association with PSP [201] [179] but, more recently, results rather seemed discordant as some studies reported a direct link between H1c and *MAPT* expression [201], [178], [249] whilst other studies could not establish significant association, for example, between the H1c tagging SNP rs242557 and an increase in *MAPT* mRNA expression [198], [199]. It was therefore speculated that H1c might affect splicing rather than expression levels of *MAPT* [199], but this hypothesis still needs to be fully investigated. These data suggest that, to date, the exact mechanism by which the *MAPT* H1 haplotype exerts its pathogenic effect in PSP still remains, for the most part, unclear and that the *MAPT* H1 haplotype alone is most probably not sufficient to cause PSP. A yet unexplored possibility to consider might be epistasis. Some genes or genetic loci may, in fact, modulate and/or influence, in *–trans*, either *MAPT* expression/splicing or the expression of the extended *MAPT* haplotype. Conversely, *MAPT*, through the SNPs rs1880753 and rs242557, might influence, in *–trans*, other genes expression. All these mechanisms are possible but, clearly, whether they happen and, especially, whether they have pathogenic effect will need to be elucidated in the future.

Taking all this into consideration, it is reasonable to infer that the biology of PSP likely goes beyond the genetic association with the *MAPT* H1 haplotype, suggesting that other hitherto unidentified environmental and genetic risk factors are most probably at play and contribute to disease pathogenesis.

3.2.4.2 New avenues

The recent PSP-GWAS revealed that novel loci may be involved in the mechanisms leading to disease [180]. With a combination of direct sequencing, and haplotype and eQTL analyses we aimed at fine mapping and validating these loci to possibly better understand their contribution to the pathogenesis of PSP.

For each of the new loci, their possible implication in PSP is discussed in separate paragraphs hereafter.

STX6

STX6 encodes the 255 amino-acid long protein STX6 that is expressed in different cell types including inflammatory/immune, endocrine, endothelial and neuronal cells. STX6 is primarily involved in trafficking of proteins and lipids from the Golgi to the plasma membrane but also in other cellular processes such as endocytosis, exocytosis and recycling [230]. STX6 contains a functional SNARE (soluble N-ethylmaleimide-sensitive factor-attachment protein receptor) domain, which is located between amino-acids 163-225. This domain is responsible for binding to homologue SNARE domains of other SNARE-proteins to direct the transfer of cargo from vesicles to target membranes. In addition, a tyrosine-based sorting motif (YGRL), located between amino-acids 140-143, mediates the retrograde transport of STX6 from the plasma membrane to the Golgi [230] and the carboxy-terminus, between amino-acids 235-255, presents a hydrophobic

transmembrane domain through which STX6 anchors to the Golgi, vesicles or plasma membranes [230]. STX6 is involved in different cellular mechanisms some of which are cell type specific. For example, STX6 mediates exocytosis of inflammatory granules and cytokines in inflammatory/immune cells or allows angiogenesis induced by vascular endothelial growth factor (VEGF) signalling in endothelial cells. More interestingly, STX6 has been suggested being involved in neurite outgrowth as a response to nerve growth factor (NGF) in neuronal cells [232], and impairment of such mechanism might link STX6 to neurological disorders as it was shown in PD and AD [230]. Clearly the recent results of the PSP-GWAS also suggest that STX6 might also be involved in the pathogenesis of PSP [180] supporting the notion of a possible functional link between *STX6* and neurodegenerative diseases.

In our genetic screening we did identify a novel missense mutation (Cys236Gly) in one individual. Its frequency in our cohort is clearly too low (1/58) to justify the genome-wide association. The frequency and distribution of this mutation would need to be verified in bigger cohorts, i.e., ideally, in the totality of the samples included in the PSP-GWAS (stage 1 + stage 2 = ~2,200 samples) to better assess its prevalence in the PSP population. This amino-acid change is located within the carboxy-terminal transmembrane domain of STX6, specifically, the portion of the protein that binds to the Golgi/vesicles membranes. Although to date there is no functional evidence at the cellular level, *in silico* analysis predicted this amino-acid change to be probably damaging (**Figure 3-3**). If this prediction is true and translates its putative damaging

effect to the protein, this mutation might impair the binding of STX6 to the membranes and possibly affect cytoplasmic vesicular trafficking [230].

Moreover, we identified complete LD between the GWAS associated rs1411478 and rs3747957 (Asn217Asn in *STX6*), pattern that could however also be seen for the normal CEU population (**Figure 3-5** and **Tables 3-3A** and **B**). Nevertheless, as a possible reason for association at this locus in our cohort, eQTL data revealed that the GWAS associated rs1411478 seems to influence *cis* expression of *STX6* in white matter ($p=1.80x10^{-09}$). Rs1411478 and rs3747957 are in complete LD, and rs3747957, within *STX6*, locates to exon 7. Interestingly, the encyclopaedia of DNA elements (ENCODE) tracks [250] for intron 7 indicate a conserved region in mammals that is featured by DNAseI hypersensitivity clusters, histone 3 acetylated at lysine 27 (H3K27Ac) and binding of multiple transcription factors from ChIP-seq data. Such epigenetic marks are indicative of gene enhancers [251] that could influence the distant promoter by the formation of chromatin loops and program tissue-specific expression during development [252]. As such, SNPs in LD with the GWAS SNP in this domain may exert an allele/brain region specific effect on *STX6* expression in white matter.

The link with white matter is of extreme interest as the involvement of white matter has been previously described in PSP. Tau lesions, for example, have been identified in white matter of pathologically confirmed PSP cases [253]. The 4R-tau isoform is overexpressed at the protein level in the subcortical regions and the brain stem of these cases of which >90% were H1 homozygous for the *MAPT* haplotype [253]. These two pieces of

information are particularly important as they suggest a possible link between tau pathology, the *MAPT* H1 haplotype and white matter degeneration. Considering our eQTL data, the *cis*–modulation of *STX6* expression in white matter could be an additional factor involved in the pathogenesis of PSP. Another study suggested that the level and topography of degeneration of white matter tracts is likely to correlate with some of the PSP specific symptoms [254]. Specifically, abnormal diffusivity through diffusion tensor imaging was observed in:

1 – Superior cerebellar peduncles;

2 – Corpus callosum and;

3 – Superior and inferior longitudinal fasciculus in patients diagnosed with PSP [254].

Degeneration of superior cerebellar peduncles correlated with disease severity and tau pathology, suggesting that imaging of this tract might serve as a biomarker to monitor PSP severity [254]. Further, degeneration of the superior and inferior longitudinal fasciculus correlated with saccadic impairment and motor functions respectively, whilst no correlations between degeneration of the corpus callosum and PSP specific symptomatology could be established [254]. More recently it was shown that:

1 – White matter of PSP brains have enlarged vacuoles, especially in the alveus, the frontopontine fibers and the central tegmental tract and;

2 – Glial inclusions can be found in basal ganglia, frontopontine fibers, cerebellum and superior frontal gyrus [255].

These features, vacuolation and glial inclusions, were suggested being a consequence of adjacent grey cellular loss and downstream effects of tau pathology, respectively [255].

All this taken together suggests that *STX6* expression might correlate with tau pathology and white matter degeneration in PSP. Although our own eQTL data did not support any effect of rs3747957 (Asn217Asn in *STX6*) on *cis* expression or splicing and its complete LD with the GWAS associated SNP rs1411478 seems reflected also in the normal population, it is reasonable to invest in further investigations to assess whether the effect of rs1411478 on STX6 expression is replicated and confirmed in other PSP cohorts. Not least, the interplay between *MAPT*, *STX6* and white matter degeneration is an intriguing possibility that deserves further studies to establish whether there is evidence of correlation and synergistic effects on disease-specific mechanisms eventually leading to PSP.

MR1

MR1 encodes the 341 amino-acid long transmembrane protein MR1, which belongs to the family of the major histocompatibility (MHC) class I molecules. MR1 is ubiquitously expressed and is implicated in immune response mechanisms such as the antigen presenting function [256].

We performed genetic screening on *MR1* because of its close proximity to *STX6* (10.6 kb) in order to identify possible genetic variants that might further explain the genome-

wide association at this locus. We did identify several known coding and non-coding variants. Among the coding variants, one missense change (Arg31His, rs41268456) was predicted, *in silico*, to be probably damaging (**Figure 3-4**). We identified this as a heterozygous change in 1/70 cases. Based on the results in our cohort, and in line with the 1000 Genomes releases, this variant seems to be extremely rare. The amino-acid change, Arg to His, locates to the extracellular compartment of the MR1 protein and is part of the α1 ligand-binding domain that is highly conserved between different species [257]. It is not clear whether this amino-acid change may cause a severe biochemical change in the α1 subunit that could, in turn, impact the ability of MR1 to bind ligands in the extracellular environment.

In summary, due to lack of LD between the GWAS SNP (rs1411478) and variability in *MR1* as well as absence of significant gene-specific eQTL, it is less plausible that *MR1* is responsible for the observed association at this locus. All the more, MR1 is a protein of the immune system for which there was no evidence of association in the PSP-GWAS [180], unlike in the case of other neurological disorders such as Parkinson's disease (PD) [258] or multiple sclerosis (MS) [259]. It is therefore difficult to expect a possible role for *MR1* in PSP.

EIF2AK3

EIF2AK3 encodes the 1116 amino-acid long protein EIF2AK3 which is ubiquitously expressed in several cell types and tissues, including neurons and the brain (http://www.ebi.ac.uk/gxa/gene/ENSG00000135823). EIF2AK3 is a kinase protein located at the membrane level of the endoplasmic reticulum (ER). The first 514 amino-acids of the protein locate to the lumen of the ER. These are followed by ~20 amino-acids that identify the transmembrane domain. The cytoplasmic portion of the protein presents a structural polypeptide sequence of ~70 amino-acids followed by the ATP-binding domain (amino-acids 599-607) and the kinase domains (amino-acids 593-1077) (http://www.uniprot.org/uniprot/Q9NZJ5). ER stress in eukaryotic cells is handled via the unfolded protein response (UPR), a relatively complicated mechanism that assures for:

1 – Attenuation of translation in the cell;

2 – Repair of unfolded proteins accumulated in the ER lumen, and;

3 – Degradation of the unfolded proteins via the ER associated degradation (ERAD) mechanism [260], [234], [233].

EIF2AK3 is specifically involved in the first process. In steady conditions the lumenal portion of EIF2AK3 oligomerizes with the heat shock-binding protein (BiP). When unfolded proteins start accumulating in the lumen of the ER the BiP chaperone detaches from the amino-terminus of EIF2AK3 and binds to the unfolded proteins. This activates EIF2AK3 which dimerizes and phosphorylates itself and the α-subunit of the eukaryotic

translation-initiation factor 2 (EIF2) that, in turn, promotes the translation of the activating transcription factor 4 (ATF4). The downstream effect of such activation is a general decrease in cellular translation processes [260], [234], [233].

Our genetic screening of *EIF2AK3* resulted in the isolation of several coding and non-coding known variants. Asp566Val (rs5579182) was identified as a heterozygous change in 2/61 cases and was predicted being probably damaging (**Figure 3-7**). This amino-acid change resides in the cytoplasmic region of the protein, approximately 30 amino-acids before the ATP-binding and the kinase domains. The amino-acid Asp is conserved among mammals and the effect of the change from Asp to Val is not clear. Functional studies may elucidate whether such change could, for example, impact the tertiary structure (protein conformation). Comparing the MAF of this variant in our cohort with the 1000 Genomes releases and dbSNP data for a Caucasian-European population (CEU) its frequency seemed higher in our study population (0.016 vs. 0.001 and 0.002; **Table 3-2**).

It is interesting to note that Asp566Val (rs55791823) holds both features of being a probably damaging amino-acid change (**Figure 3-7**) and having a MAF that differs from the 1000 Genomes and dbSNP releases. Nevertheless, our cohort might be too small to consider these two co-occurring characteristics significant and/or strongly associated with disease. In addition, no significant LD pattern could be identified for this specific variant (rs55791823), therefore, its possible pathogenic impact remains hard to interpret in our cohort. The LD analysis identified a significant LD block of ~42 kb that included the

GWAS associated SNP rs7571971 and 4 SNPs, 3 of which were coding (rs1805165, rs13045 and rs867529) and 1 was intronic (rs4972221). The coding variants are all missense and are all in almost complete LD with the GWAS SNP (**Figure 3-8**). Nevertheless, all 3 are predicted to be benign and their MAFs reflect the expected values as of the 1000 Genomes releases and the Caucasian/European general populations. In addition, by comparing the haplotype structure of our PSP cohort with that of the normal CEU population obtained from the HapMap 22 release (and analysed through the online tool http://www.broadinstitute.org/mpg/snap/ldsearchpw.php) no outstanding differences could be identified. However, specific haplotype substructures at this locus had been previously reported [261, 262]. Data for our PSP population did reflect findings in [261, 262] as shown in **Table 3-5**.

Table 3-5. *EIF2AK3* locus associated haplotypes

Haplotype	**SNP/Amino-acid change**				**Haplotype frequency per Study**			
	rs867529/(Ser136Cys)	rs7571971*	rs13045/(Gln166Arg)	rs1805165/(Ala704Ser)	HapMap CEU	[261]	[262]	Our study
A	C/(Ser)	C	G/(Arg)	T/(Ser)	0.646	0.676	0.645	0.661
B	**G/(Cys)**	**T**	**A/(Gln)**	**G/(Ala)**	**0.294**	**0.311**	**0.288**	**0.287**
C	C/(Ser)	C	G/(Arg)	G/(Ala)	/	0.013	0.001	0.011
D	C/(Ser)	C	A/(Gln)	T/(Ser)	0.029/0.016	/	0.061	0.033

Table 3-5. Comparison of the 4 different and most frequent haplotypes (A, B, C and D) associated with the *EIF2AK3* locus. * PSP-GWAS associated SNP

Haplotype B has been suggested to be associated with an increased activation of *EIF2AK3* and the UPR process in response to ER stress [261]. Haplotype B, that includes the minor and risk allele T of the PSP-GWAS associated SNP rs7571971, may be implicated in an increase in risk of developing PSP [262]. Interestingly, the frequency identified for the B haplotype in the Stutzbach study and ours, which are reporting data on PSP cases, are basically the same (0.288 vs. 0.287) regardless from the sample size (~1,000 samples in [262] and ~80 in our study).

Furthermore, considering that our eQTL data did not suggest a possible role in expression or splicing for any of these SNPs it is difficult, at the current status, to assess what the reason of the association at this locus might be. Clearly, neither genetic screening, nor haplotype or eQTL analyses could explain the association with this locus suggesting that possibly trans-active elements might affect this locus and that the PSP-GWAS results rather point to a cellular process as possibly implicated in pathogenesis with yet unclear molecular underlying processes.

In this view, from a functional perspective, it had been shown that efficient synaptic plasticity and memory are related with proper phosphorylation of the EIF-2 protein [235]. Functional EIF2AK3 is necessary for correct phosphorylation of the α-subunit of EIF-2 which consequently suggests not only that EIF2AK3 may have a critical role in overall general cellular homeostasis by controlling protein translation processes but also that these mechanisms may play a role in neurological conditions [235]. It was shown that disrupted EIF2AK3 activity may be involved in the pathophysiology of schizophrenia

[235] and that impaired interaction between EIF2AK3, EIF-2 and ATF4 may contribute to conditions such as impaired cognitive functions and information processing [235]. Several neurological disorders, including AD and PD, have previously been linked to protein translation impairment [235] and EIF2AK3 may be one of the factors at play in this arena. In addition, activation of UPR was previously shown in presence of phosphorylated tau (p-tau) in AD and it was suggested that UPR coincides with early events of tau pathology [262, 263]. Stutzbach and colleagues however also specified that *EIF2AK3* activation may precede tau accumulation rather than follow it [262]. This may indicate a link between the UPR and tau pathology leading to the assumption that UPR processes may be involved in a larger spectrum of tauopathies. In fact, not only this trend was observed in AD [263], but also, more recently, in FTLD-tau, whilst it seemed absent in FTLD-TDP or FTLD-FUS cases or in neurologically normal controls [264]. Another study suggested also a different mechanism by which UPR might be involved in the development of tauopathies in that activated EIF2AK3 may enhance phosphorylation processes with downstream effects on tau contributing, to a certain extent, to the formation of hyperphosphorylated tau and neurofibrillary tangles [264].

In summary, *EIF2AK*3 may be involved in PSP by, possibly, exerting a functional pathogenic role enhanced by an effect driven by the B haplotype [261, 262]. Such effect may happen in concert with *MAPT* and tau pathology. It remains to be established whether the UPR activates because of early tau aggregation (event that was recently dismissed as Tau does not locate to the ER [262]) and, consequently, the B haplotype at the *EIF2AK3* locus increases risk for PSP for the carriers or downstream effects of

EIK2AK3 activation lead to hyper-phosphorylation of Tau [264]. Either possibility needs to be further investigated.

MOBP

MOBP encodes the 183 amino-acid long protein MOBP that locates to the cytoplasm of oligodendrocytes, which are brain cells involved in myelination processes [238]. It is the third most abundant protein in the CNS-myelin [238] and is expressed in humans in the form of two different isoforms resulting from alternative splicing of a single gene [237]. MOBP is expressed in white matter [237] and contains proline-rich tandem repeats which is common in proteins with structural functions [238]. It follows that MOBP is likely to be an important element stabilizing the multi-layered structure of the myelin sheath in the CNS [238]. It was suggested that MOBP may be involved in the pathogenesis of multiple sclerosis (MS) as a result of autoimmune reactivity against MOBP leading to destabilization and destruction of myelin [238].

Our genetic screening revealed a novel missense variant, Gln82Lys (1/63, heterozygous change) that resulted being a benign change after *in silico* analysis (**Figure 3-9**) and two common intronic variants. The missense variant locates to the proline-rich domain that spans between amino-acids 77-136. There was no evidence of LD for the two intronic variants with the GWAS associated SNP as well as no evidence of effect on either *cis* expression or splicing. Based on our data it is difficult to identify a genetic link between

locus and the disease in our cohort. Nevertheless, it was shown that tau pathology in PSP affects neurons, astrocytes and oligodendrocytes [253]. The latter provide the myelin sheath surrounding the axons of neurons. The association at this locus may be rather the result of interplay between MOBP and tau or a downstream effect triggered by tau pathology. Although this hypothesis is clearly intriguing and interesting, it needs to be further investigated before any assumption can be made. In addition, two MOBP isoforms are known. The two proteins are substantially different as one is 183 amino-acids long (referred to as the "canonical"), whilst the other is 81 amino-acids long and is missing the carboxy terminal proline-rich repeat domain. A better understanding of the MOBP isoforms, their expression patterns and their functions, may help in locating this protein within the biological context, in health and disease.

3.2.5 Conclusion

It is worth noting that the PSP-GWAS [180] has, on one hand, confirmed the *MAPT* H1 haplotype being strongly associated with PSP, and, on the other, given rise to novel possible perspectives on the pathogenesis of PSP.

After having concluded our follow up study that aimed at further investigating the newly identified loci [180], our main findings can be summarized as follows:

1 – We identified complete LD between the GWAS associated SNP rs1411478 and rs3747957 (Asn217Asn) in *STX6*, and our data show correlation between *STX6* risk allele and *cis* expression levels in white matter;

2 – The *EIF2AK3* locus showed a ~42kb that included the GWAS associated SNP rs7571971, the 3 coding SNPs rs1805165, rs13045 and rs867529, and the intronic SNP rs4972221. This LD block however did not seem disease specific as it almost equally reflects LD patterns that can be found in the normal population. In addition, no effects on expression/splicing could be identified for any of the SNPs at this locus, and;

3 – There was no direct evidence from our genetic screening of *MOPB* or from our eQTL data for a reason that might underlie the GWAS association at this locus. In addition, although we identified a couple of novel missense changes (Cys236Gly in *STX6* and Gln82Lys in *MOBP*) none seemed convincingly pathogenic, neither they were suggestive of underpinning association at their respective loci, given their extremely low frequency rate of occurrence.

In addition, there are some limitations to this study that need to be acknowledged and kept in mind:

1 – The limited sample size (n=84, although all cases were pathologically confirmed);

2 – The use of neurologically normal control brains to assess expression that may not entirely represent conditions related to the diseased status (i.e. PSP);

3 – The presence of different cell populations (microglial cells, astrocytes, oligodendrocytes, etc.) in addition to neurons in the samples used to produce eQTL data (that might partially confound expression results) and;

4 – The lack of availability of *trans* eQTL and splicing data.

However, considering all of the above, our findings together with those of the PSP-GWAS [180] may suggest a number of ambitious, but, at the same time, intriguing assumptions about possible disease related mechanisms. In fact, the notions that:

1 – Tau pathology and white matter involvement are characteristics of PSP, and;

2 – The newly identified genes/loci which relate to tau pathology and white matter, both offer a reasonable background to base upon the design of future genetic and functional studies in order to verify whether these processes are involved and, if so, to what extent in the pathogenesis of PSP.

In conclusion, these speculations, which are highlighted and summarized in **Figure 3-11**, will clearly need to be further investigated in the near future, in order to be possibly confirmed and validated across studies.

Figure 3-11. Possible PSP-associated disease mechanisms: Speculations

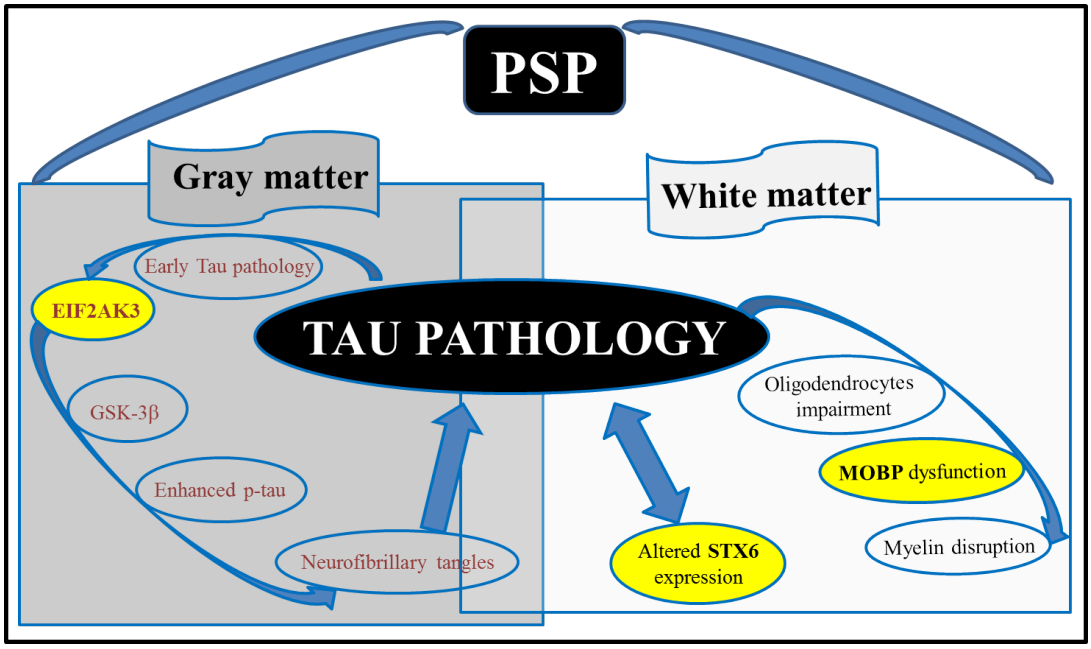


Figure 3-11. The novel loci highlighted in [180] may contribute collectively to the pathogenesis of PSP. Specifically:

1 – Interplay between *MAPT*, the H1 haplotype, tau pathology, white matter degeneration and differential expression patterns of *STX6* may influence each other and have a role in the pathogenesis of PSP;

2 – Activated EIF2AK3 may either be a consequence of early tau pathology or induce phosphorylation of tau by activating the glycogen synthase kinase 3β (GSK-3β) which is involved in tau phosphorylation, implying that the *EIF2AK3* may have a role in the process that leads to the formation of neurofibrillary tangles, and;

3 – MOBP is mainly expressed in white matter and there is evidence of white matter involvement in PSP as well as tau pathology affecting white matter. It may be a possibility that the interaction between *MAPT* and *MOBP* through epistasis, or *trans* regulation of expression or splicing, or downstream effects of tau pathology may impact and impair the biology of MOBP and contribute to myelin disruption.

The work presented in this chapter was performed by the following individuals (or groups):

Raffaele Ferrari: Performed genetic screening (sequencing) and the entire haplotype analyses; **UK Brain Expression Consortium** (UKBEC) (**Appendix 2-5**): Generated expression quantitative trait loci data; **Mina Ryten**: provided expression quantitative trait loci data.

A manuscript derived from the work presented in this chapter has been published in Neurobiology of Aging:

Raffaele Ferrari, Mina Ryten, Roberto Simone, Daniah Trabzuni, Naiya Nicolaou, Geshanthi Hondhamuni, Adaikalavan Ramasamy, UK Brain Expression Consortium, Michael E. Weale, Andrew J Lees, Parastoo Momeni, John Hardy and Rohan de Silva. ***Assessment of common variability and expression quantitative trait loci for GWAS associations for progressive supranuclear palsy***. Neurobiol Aging, 2014 Jun;35(6):1514.e1-1514.e12.

CHAPTER 4 – FTD-GWAS

4.1 FTD-GWAS chronicle

4.1.1 The rationale

Over the past 15 years, the genetic study of frontotemporal dementia has brought to light a handful of mendelian genes/loci (*MAPT*, *GRN* and *C9orf72*) [26, 45, 66, 67, 76, 77] and one possible disease modifying factor (*TMEM106B*) [26, 41]. In addition, a limited number of other genes have been associated with the FTD spectrum although the occurrence of pathogenic mutations has been rare (*VCP*, *SQSTM1* and *UBQLN2*), or equivocal (*TDP-43*, *FUS* and *CHMP2B*) [26, 265]. All the more, a clear disease mechanism has not yet been established and this is for the most part due to the fact that FTD is a complex disorder characterised by heterogeneous features in the clinical, pathological and genetic settings. In particular, the current genetics of FTD, besides explaining approximately 20% or less of cases, do not point to a clear and specific disease mechanism. Only variability in *MAPT*, by either directly affecting exons 9, 10, 11, 12 and 13 or splicing of exon 10, has been proven to lead to an impairment of tau capabilities in binding the microtubules fact that, in turn, affects neuronal stability and partially also contributes to tau pathology [29, 50]. Taking all this together it follows that the molecular underpinnings of FTD are, to date, still not clear.

Advancements of the technologies available to researchers offer increasingly sophisticated tools to study diseases, especially, in the field of genetics. The quantum leap over the past 10-20 years has been unimaginable and unprecedented in this field, crossing over from targeted and limited (in size) Sanger sequencing to genome-wide association (GWA) or, the most recent, exome and whole genome sequencing. Such tools allow the investigation of the genome in a more comprehensive fashion and at a faster pace. Each of these methods facilitates the study different functional and structural characteristics of the genome: The GWA type of approach, for example, identifies genetic loci that are associated with disease [266] and are likely to indicate pathways and/or processes that play a substantial role in the pathogenesis of a certain diseased trait.

With this background in mind, the rationale of the FTD-GWAS was to identify novel genes/loci associated with this neurodegenerative condition that could possibly also shed light on disease associated mechanisms.

4.1.2 The development and its phases

A detailed timeline of the FTD-GWAS is summarized in **Figure 4-1**.

Figure 4-1. FTD-GWAS timeline

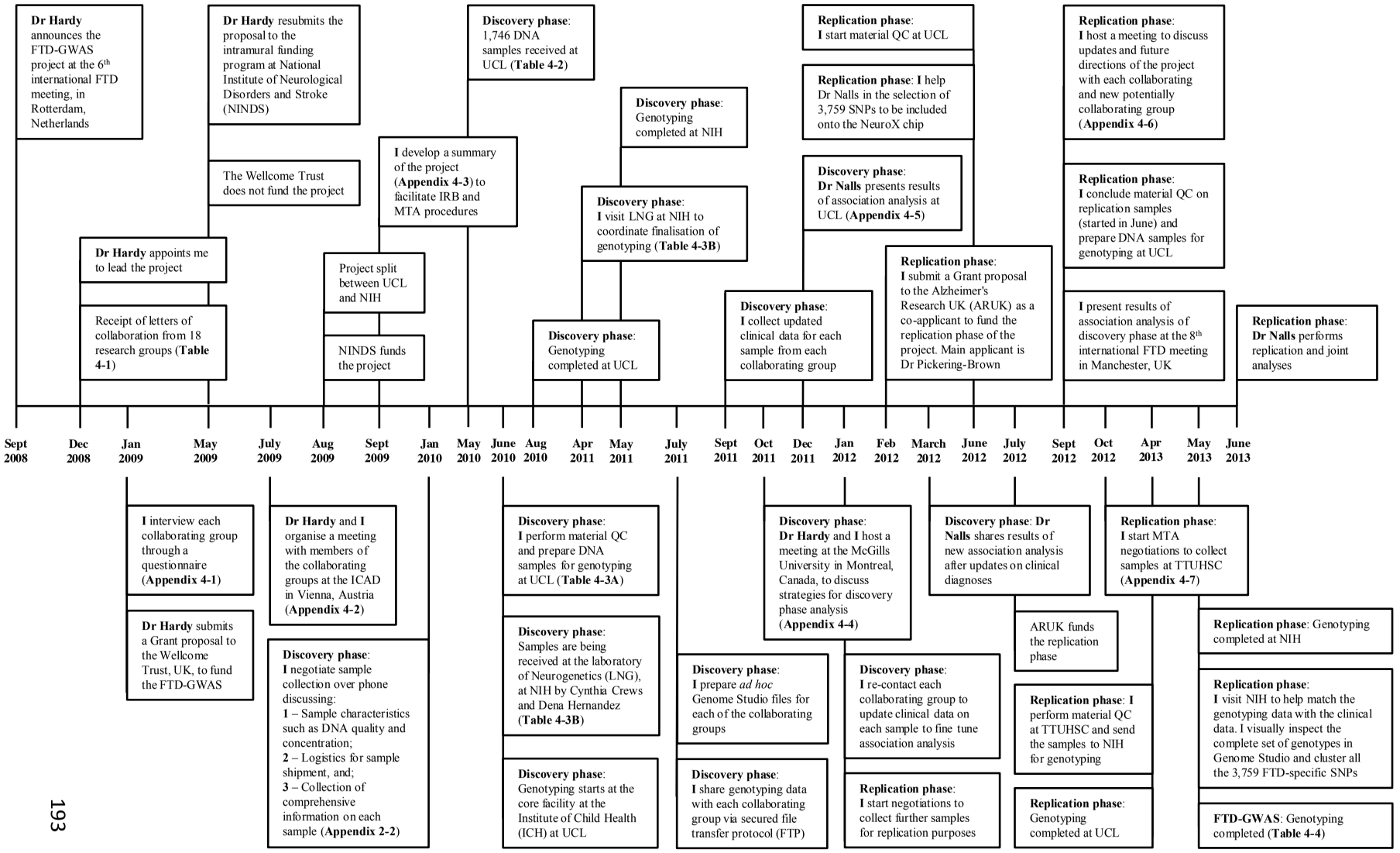


Table 4-1. FTD-GWAS: Collaborative groups as of December 2008

FTD-GWAS Investigators
Martin Rossor/Jonathan Rohrer (UK)
Stuart Pickering-Brown (UK)
Huw Morris (UK)
Rosa Rademakers (USA)
Parastoo Momeni (USA)
Bruce L Miller/Anna Karyadas (USA)
Daniel Geschwind (USA)
Vivianna Van deerlin (USA)
Christine Van Broekhoven (BELGIUM)
Caroline Graff (SWEDEN)
Alexis Brice/Isabelle Leber (FRANCE)
Jorgen Erik Nielsen (DENMARK)
Johannes Schlachetzki (GERMANY)
J van Swieten (NETHERLANDS)
Barbara Borroni (ITALY)
Annibale Puca/Valeria Novelli (ITALY)
Giacomina Rossi (ITALY)
Amalia C Bruni (ITALY)
Pau Pastor (SPAIN)

Table 4-1. List of the investigator(s) or contact people per country who agreed to participate in the FTD-GWAS as of December 2008 after the announcement at the international FTD meeting in Rotterdam, Netherlands.

Table 4-2. FTD-GWAS: Sample receipt at UCL

GROUP 1 = samples received at UCL before May 2010	**Samples (n)**
Schlachetzki J (Germany)	
Borroni B (Italy)	
Novelli V (Italy)	**TOTAL 448**
Momeni P (USA)	
Rossi G (Italy)	
GROUP 2 = samples received at UCL in May 2010	**Samples (n)**
Perneczky R (Germany)	
Danek A/Bader B (Germany)	
Nielssen J (Denmark)	
Riemenschneider M (Germany)	
Galimberti D/Scarpini E (Italy)	
French Consortium	**TOTAL 1,298**
Rainero I (Italy)	
Binetti G (Italy)	
Nacmias B (Italy)	
Bruni AC (Italy)	
Graff C (Sweden)	
TOTAL	**1,746**

Table 4-2. List of the investigator(s) or contact people per country who had sent samples to UCL by May 2010.

Table 4-3. FTD-GWAS: Summary of received and genotyped samples

A

Participating Groups	Samples received	Samples genotyped
Momeni (USA) Graff (SWEDEN) French Consortium Nielsen (DENMARK) Danek (GERMANY) Perneczky (GERMANY) Riemenschneider (GERMANY) Schlachetzki (GERMANY) Borroni (ITALY) Novelli - Puca (ITALY) Rossi (ITALY) Galimberti (ITALY) Rainero (ITALY) Benussi (ITALY) Nacmias (ITALY) Bruni (ITALY)	**TOTAL 1,746**	**TOTAL 1,303**

Table 4-3. (**A**) Summary of samples that have been collected at UCL during discovery phase. After material QC a total of ~450 samples were excluded. Genotyped samples were 1,303.

B

Participating Groups	Samples received	Samples genotyped
Rohrer - Rossor (UK) Mead (UK) Pickering-Brown (UK) Morris (UK) Roageva (CANADA) Rademakers (USA) Van deerlin (USA) Van Broekhoven (BELGIUM) Heutink (NETHERLANDS) van Swieten (NETHERLANDS) Pastor (SPAIN)	**TOTAL 1,911**	**TOTAL 1,655**

Table 4-3. (**B**) Summary of samples that have been collected at NIH during discovery phase. After material QC a total of ~300 samples were excluded. Genotyped samples were 1,655.

Table 4-4. FTD-GWAS: Total of samples and controls received

Participating Groups	Samples sent		
	Discovery phase	*Replication phase*	
		Cases	*Controls*
Momeni (USA) Baborie (UK) Graff (SWEDEN) French Consortium Nielsen (DENMARK) Danek (GERMANY) Perneczky (GERMANY) Riemenschneider (GERMANY) Schlachetzki (GERMANY) Borroni (ITALY) Novelli - Puca (ITALY) Rossi (ITALY) Galimberti (ITALY) Rainero (ITALY) Benussi (ITALY) Nacmias (ITALY) Bruni (ITALY) Rohrer - Rossor (UK) Mead (UK) Rowe (UK) Pickering-Brown (UK) Morris (UK) Roageva (CANADA) Rademakers (USA) Van deerlin (USA) Van Broekhoven (BELGIUM) Heutink (NETHERLANDS) van Swieten (NETHERLANDS) Clarimon (SPAIN) Pastor (SPAIN) Cruchaga (USA) Mackenzie (CANADA) Landqvist (SWEDEN) Ruiz (SPAIN) Schofield	**TOTAL 3,657**	**TOTAL 2,039**	**TOTAL 1,600**

Table 4-4. Summary of the samples sent to the Institutions leading the FTD-GWAS project. The samples are subdivided by stage (discovery and replication phase) and by status (case vs. control).

4.2 FTD-GWAS: The project

The materials and methods as well as the procedures adopted to accomplish the FTD-GWAS project have been extensively reported and explained in **Chapter 2**, sections ***2.1.2***, *2.1.6.3*, *2.2.1.3*, *2.2.2.8*, ***2.3.4*** and ***2.3.5***.

In the following sections, prior to the report of the results of the FTD-GWAS, the materials and methods used to perform methylation and expression quantitative trait loci (m/eQTL) analyses are described.

4.2.1 Expression and methylation quantitative trait loci (QTL)

The work to generate the expression and methylation data was performed by members of the UK Brain Expression Consortium (UKBEC) and North American Brain Expression Consortium (NABEC) whose names and affiliations are listed and acknowledged in **Appendix 2-5**.

4.2.1.1 UK Brain Expression Consortium (UKBEC)

Sample Collection

One hundred and thirty four brain samples of control individuals were collected by the Medical Research Council (MRC) Sudden Death Brain and Tissue Bank, Edinburgh, UK

(n=100), and the Sun Health Research Institute (SHRI) Brain Donation Program, USA (n=34) [243, 244]. All samples were pathologically confirmed to be normal and were fully authorized for scientific investigation. Ten brain regions (cerebellar cortex (CRBL), frontal cortex (FCTX), hippocampus (HIPP), the inferior olivary nucleus (sub-dissected from the medulla, MEDU), occipital cortex (BA17, OCTX), putamen (at the level of the anterior commissure, PUTM), substantia nigra (SNIG), temporal cortex (TCTX), thalamus (at the level of the lateral geniculate nucleus, THAL) and intralobular white matter (WHMT) were sampled per individual for RNA analyses.

RNA isolation and processing of samples using Affymetrix Exon 1.0 ST Arrays

Total RNA was isolated using the miRNeasy 96 well kit (Qiagen, UK) and its quality was assessed through 2100 Bioanalyzer and RNA 6000 Nano Kit (Agilent, UK). Total RNA was then processed with the Ambion® WT Expression Kit and Affymetrix GeneChip Whole Transcript Sense Target Labeling Assay, and hybridized to the Affymetrix Exon 1.0 ST Arrays as recommended by manufacturer. Hybridized arrays were scanned on an Affymetrix GeneChip® Scanner 3000 7G and visually inspected for hybridization artefacts. Arrays were processed using Robust Multi-array Average (quantile normalisation) algorithm [267] in Affymetrix Power Tools 1.14.3 software (http://www.affymetrix.com/partners_programs/programs/developer/tools/powertools.affx). After re-mapping the Affymetrix probe sets onto human genome build 19 (GRCh37), analysis was restricted to probe sets with gene annotation and containing at least 3

uniquely hybridized probes free from polymorphisms (MAF >1% in the European panel of the March 2012 release of 1000 Genomes) [242].

DNA extraction, genotyping and imputation

Genomic DNA was extracted from human post-mortem brain tissue using Qiagen's DNeasy Blood and Tissue Kit (Qiagen,UK) (100–200 mg). Samples from every individual were run on the Illumina Infinium Omni1-Quad BeadChip and the ImmunoChip [221, 222]. The BeadChips were scanned using an iScan (Illumina, USA) with an AutoLoader (Illumina, USA) and Genome Studio (Illumina) was used for generating SNP calls and analysing the data. Individuals with non-European ancestry were identified through PCA and, subsequently excluded from analysis. Prior to imputation, standard QC was performed to exclude samples with call rate<95%, p-value of deviation from Hardy-Weinberg Equilibrium<0.0001 and CNV and indel markers. Then SNPs from both arrays were combined and imputed using MaCH [247] and minimac (http://genome.sph.umich.edu/wiki/Minimac) using the 1000 Genome release (March 2012). More than 5.5M SNPs and ~570K indels with good post-imputation quality (r^2>0.50) and MAF of at least 5% were used in subsequent analyses.

4.2.1.2 North American Brain Expression Consortium (NABEC)

Sample collection

Cerebellar and frontal cortex samples originating from 399 neuropathologically confirmed controls were collected [222]. Briefly the samples originated from the University of Maryland Brain Bank, Baltimore (n=207), Sun Health Research Institute (n=52, non-overlapping with UKBEC), Baltimore Longitudinal Study of Aging (n=20), University of Miami (n=16), the Department of Neuropathology of John Hopkins University (n=9) and the Medical Research Council (MRC) Sudden Death Brain and Tissue Bank (n=95, subset of UKBEC). This study was approved by the appropriate institutional research ethics board.

RNA isolation and processing of samples using Illumina Human HT12-v3 Arrays

Total RNA was extracted from (100–200 mg) human post-mortem brain tissue using a glass-Teflon homogenizer and 1mL TRIzol (Invitrogen, Carlsbad, CA) as recommended by manufacturers (100–200 mg). RNA was biotinylated and amplified using the Illumina® TotalPrep-96 RNA Amplification Kit and directly hybridized onto Human HT12 Expression BeadChips (Illumina Inc., USA). Expression data were analysed using the Gene Expression Module 3.2.7 within Illumina® BeadStudio. Raw intensity values for each probe were transformed using the cubic spline normalization method and then

log2 transformed for mRNA analysis. We re-mapped the annotation for probes according to ReMOAT [268] on the human genome build 19 and then restricted the analysis to genes that were uniquely hybridized and associated with gene descriptions.

DNA isolation and processing of samples using Illumina Human Methylation27 arrays

The CpG methylation in cerebellum and frontal cortex was determined in a subset of the NABEC dataset (n=292). Genomic DNA was phenol–chloroform extracted and quantified on the Nanodrop1000 spectrophotometer prior to bisulfite conversion. Bisulfite conversion of 1 μg of genomic DNA was performed using Zymo EZ-96 DNA Methylation Kit as recommended by manufacturer. CpG methylation status of at 27,578 CpG dinucleotides at 14,495 genes was determined using Illumina Infinium HumanMethylation27 BeadChip, as per the manufacturer's protocol. Data were analyzed in BeadStudio software (Illumina Beadstudio v.3.0) with threshold call rate>95%. QC of sample handling included comparison of genders reported by the brain banks and determined by analyzing methylation levels of CpG sites on the X chromosome. Samples where genders did not match between brain bank and methylation data were excluded from our analyses. Individuals not overlapping with UKBEC were genotyped on Illumina Infinium HumanHap550 v3 (Illumina, USA). The SNPs that passed QC and were common to both UKBEC and non-UKBEC individuals were extracted for imputation. Approximately 5.3M SNPs were available after imputation and QC.

eQTL analyses

SNP dosage was analysed through an additive genetic model using linear regression adjusting for covariates of gender, age at death, post mortem interval (PMI), brain bank, batch in which preparation or hybridization were performed (and for the first two PCA of population stratification in NABEC dataset). The analyses were conducted using MACH2QTLv1.11 (http://www.sph.umich.edu/csg/abecasis/MaCH/download/) for NABEC and MatrixEQTL [248] and R (http://www.R-project.org/) for UKBEC. Association was tested for all probe sets located within +/- 1Mb of each SNP. Gene expression and methylation data are available from Gene Expression Omnibus under the accession number GSE36192.

4.2.2 Results

In total (discovery + replication phases), 3,526 FTD samples that survived QC were analysed in this study. These were divided in 2,154 cases (discovery phase) and 1,372 cases (replication phase) (**Table 4-5**).

The QQ plots for all subtypes and the whole cohort are shown in **Figure 4-2**.

Table 4-5. FTD-GWAS: demographics of FTD samples used in the study

Country	*Samples*									
	Collected			*Included in analysis*			*Females % (n)*		*Mean age at onset*	
	Discovery phase	*Replication phase*	*Total*	*Discovery phase*	*Replication phase*	*Total*	*Discovery phase*	*Replication phase*	*Discovery phase*	*Replication phase*
USA	706	209	915	579	175	754	44·4 (n=257/579)	48·9 (n=85/174)	60 (23-85; n=520)	63 (24-93; n=120)
Canada	25	37	62	24	29	53	52·2 (n=12/23)	57·1 (n=8/14)	64 (43-85; n=15)	59 (43-75; n=9)
UK	494	372	866	401	284	685	42·8 (n=171/400)	39·7 (n=108/272)	60 (23-83; n=372)	61 (35-86; n=167)
Spain	100	330	430	0	309	309	NA	43 (n=133/309)	NA	65 (32-89; n=308)
France	238	54	292	205	42	247	44·4 (n=91/205)	47·6 (n=20/42)	62 (39-79; n=190)	NA
Belgium	240	51	291	191	42	233	46·1 (n=88/191)	28·6 (n=12/42)	63 (29-90; n=191)	64 (43-84; n=42)
Netherlands	333	93	426	250	77	327	51·6 (n=129/250)	40·3 (n=31/77)	58 (29-76; n=250)	61 (51-69; n=59)
Denmark	35	0	35	7	0	7	71·4 (n=5/7)	NA	57 (40-62; n=7)	NA
Germany	349	34	383	320	33	353	NA	50 (n=15/30)	61 (36-83; n=243)	57 (29-72; n=30)
Sweden	26	112	138	18	98	116	55·6 (n=10/18)	61·2 (n=60/98)	57 (38-75; n=16)	62 (28-78; n=93)
Italy	1035	563	1598	564	371	935	52·9 (n=297/561)	45·3 (n=168/371)	64 (31-83; n=429)	65 (31-87; n=353)
Australia	0	138	138	0	121	121	NA	36·4 (n=44/121)	NA	59 (32-77; n=112)
Meta	**3,581**	**1,993**	**5,574**	**2,559 (2,154°)**	**1,581 (1,372°)**	**4,140 (3,526°)**	**46·5 (n=1,186/2,552)**	**44·1 (n=684/1,550)**	**61 (23-90; n=2,233)**	**62 (24-93; n=1,293)**

° = after quality control filtering

Table 4-5. Samples that have been collected during the discovery and the replication phases and their characteristics are shown. The numbers of samples that have been included in analysis after the material QC are depicted; ° identifies the samples that survived genotyping data QC and that were used to perform actual association analyses. Also female percentages as well as mean age at onset for both discovery and replication cohorts are shown (per country and in total).

Figure 4-2. QQ-plots for all FTD-GWAS association analyses

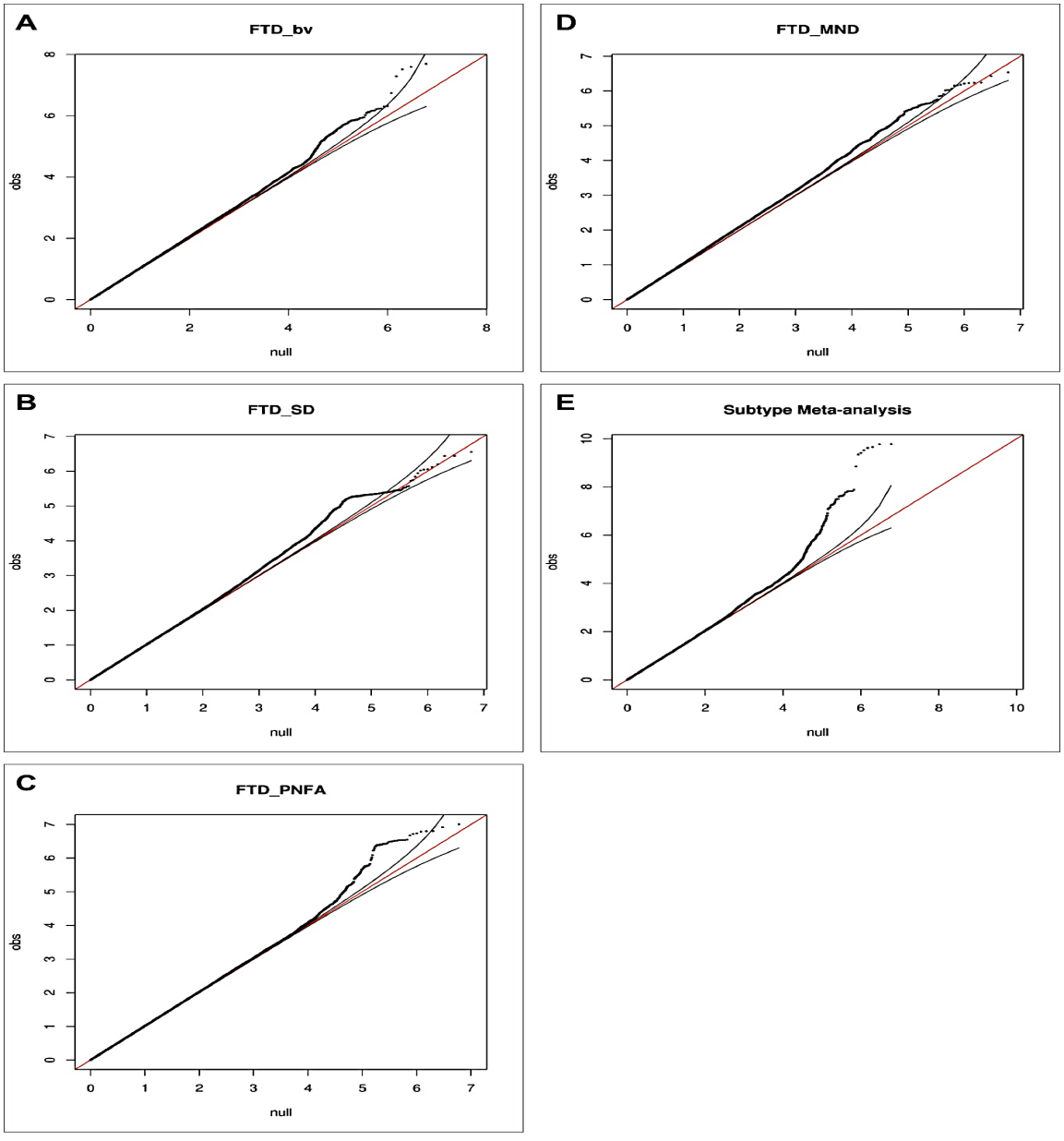


Figure 4-2. QQ plots for each subtype (**A-D**) and for the entire cohort (**E**). In each and every subtype there is almost no departure from the expected distribution. An inflation can be seen in the QQ plot of the entire cohort, however raw lambda value was 1.054, thus acceptable for GWAS standards.

4.2.2.1 Discovery phase

Primary analysis for the discovery phase was performed on a total of n=2,154 cases (**Table 4-5**) and n=7,444 control samples (after quality control filtering).

Based on the different subtypes that had been collected (and that contribute to the FTD spectrum), the primary association analyses were accordingly strategized leading to four separate analyses each targeting every single FTD subtype (bvFTD, SD, PNFA and FTD-MND) individually (**Table 4-6**) followed by a final meta-analysis on the entire cohort.

Table 4-6. FTD-GWAS: subtypes

Country	*Subtypes*														
	bvFTD			***SD***			***PNFA***			***FTD-MND***			***FTLD-U***		
	Discovery phase	*Replication phase*	*Total*	*Discovery phase*	*Replication phase*	*Total*	*Discovery phase*	*Replication phase*	*Total*	*Discovery phase*	*Replication phase*	*Total*	*Discovery phase*	*Replication phase*	*Total*
USA	315	25	340	147	12	159	81	15	96	36	21	57	0	102	102
Canada	22	5	27	1	1	2	0	5	5	1	7	8	0	11	11
UK"	207	152	359	75	53	128	69	44	113	50	16	66	0	19	19
Spain	NA	194	194	NA	41	41	NA	51	51	NA	13	13	NA	10	10
France	135	30	165	3	0	3	8	3	11	59	8	67	0	1	1
Belgium*	135	27*	162	13	1*	14	22	2*	24	21	2*	23	0	10*	10
Netherlands	159	37	196	47	31	78	24	6	30	20	3	23	0	0	0
Denmark	2	NA	2	0	NA	0	1	NA	1	4	NA	4	0	NA	0
Germany	209	18	227	45	8	53	55	6	61	11	1	12	0	0	0
Sweden	7	53	60	2	20	22	6	10	16	3	8	11	0	7	7
Italy	443	186	629	28	22	50	69	86	155	24	16	40	0	61	61
Australia"	NA	56	56	NA	26	26	NA	19	19	NA	20	20	NA	0	0
Meta	**1,634 (1,377°)**	**783 (690°)**	**2,417 (2,061°)**	**361 (308°)**	**215 (190°)**	**576 (495°)**	**335 (269°)**	**247 (221°)**	**582 (486°)**	**229 (200°)**	**115 (94°)**	**344 (294°)**	**0 (0)**	**221 (177°)**	**221 (177°)**

° = after quality control filtering; * = from discovery phase; " = sharing same controls

Table 4-6. The number of samples collected per subtype and per country are shown.

BvFTD subtype analysis

During the discovery phase, the association analysis for the bvFTD subtype was performed on n=1,377 cases (**Table 4-6**) and n=2,754 controls (after quality control filtering). As a result of the statistical analyses two non-coding SNPs passed the genome-wide significance threshold (p-value<$5x10^{-08}$) in this subtype (**Table 4-7**).

Table 4-7. Discovery phase: SNPs associated with the bvFTD subtype

	Discovery phase										
Trait	***Marker***	***Chr***	***BP***	***Candidate gene***	***Effect allele***	***Alternate allele***	***Frequency of Effect allele***	***Imputation quality***	***OR***	***SE***	***P***
bvFTD	rs302652	11	87894831	*RAB38*	T	A	0.741	0.9296	1.369	0.057	$2.02x10^{-08}$
	rs74977128	11	87936874	*RAB38/CTSC*	T	C	0.882	0.4182	0.551	0.107	$3.06x10^{-08}$

Table 4-7. SNPs exceeding genome-wide significance in discovery phase for the bvFTD subtype are listed along with their chromosomal position and nearby genes. The OR is calculated based on the effect allele whilst the risk allele can be derived.

Both SNPs map to chromosome 11 at the locus 11q14: The first associated SNP, rs302652, had a p-value=$2.02x10^{-08}$ and an OR=1.369, whilst the second, rs74977128, had a p-value=$3.06 x10^{-08}$ and OR=0.551 (**Table 4-7**).

It is noteworthy that rs302652 is located in intron 1 of the gene RAB38, member RAS oncogene family (*RAB38*), whilst rs74977128 maps to the intergenic region between *RAB38* and cathepsin C (*CTSC*), however closer to the gene *RAB38*, being ~25 kb upstream from its open reading frame (ORF) (**Figure 4-3**).

Figure 4-3. BvFTD: Manhattan plot

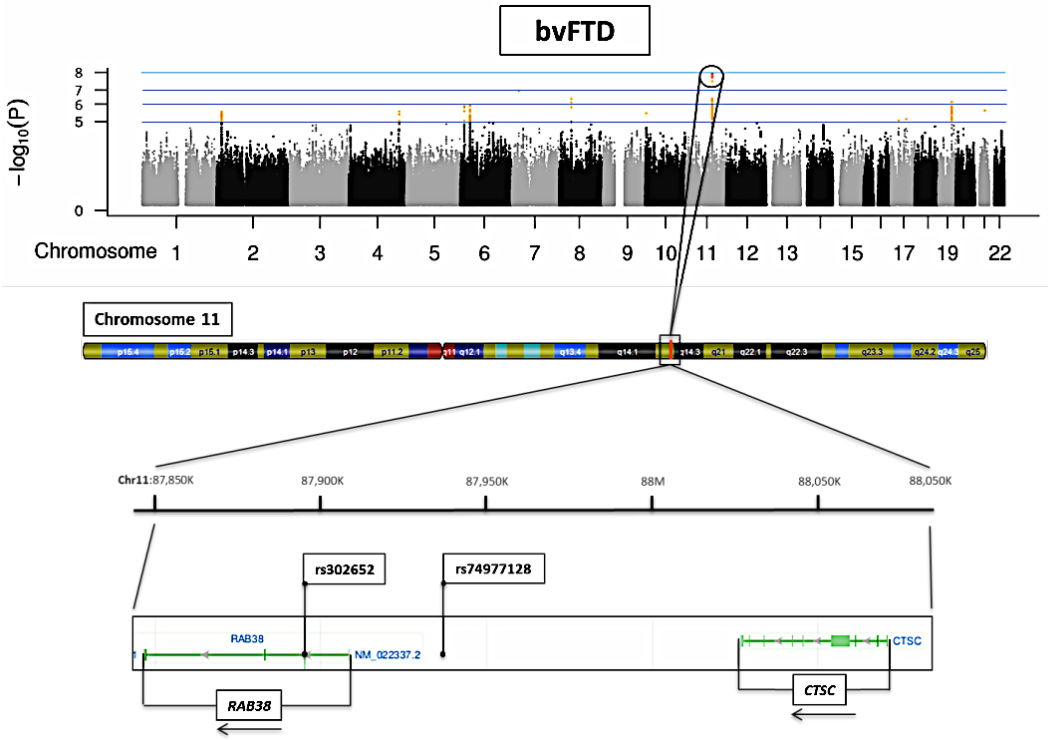


Figure 4-3. The Manhattan plot of −log10 p-values across genome for the bvFTD subtype and the locus bearing genome-wide significant associations are shown. The Manhattan plot depicts the associated region at 11q14. The red dots identify the SNPs that reached genome-wide significance; the orange dots identify the SNPs suggestive of association.

SD, PNFA and FTD-MND subtypes analyses

Similarly, association analysis was carried out on the other three FTD subtypes. After quality control filtering statistical analysis was performed on 308 SD cases (compared to 616 controls), 269 PNFA cases (compared to 538 controls) and 200 FTD-MND cases (compared to 400 controls) (**Table 4-6**).

There were no SNPs that reached genome-wide significance for any of these three subtypes. This outcome can probably be ascribed to the fact that for each of the subtypes the sample size was relatively small and, therefore, most likely not sufficient to confer enough power to the association analysis.

However, several SNPs showed suggestive associations (p-values between 10^{-06} and 10^{-07}) in each subtype (**Figure 4-4**), fact that suggests all such SNPs being a robust subset of markers of interest, which can be further investigated as they may hint subtype-specific associations and may be valuable candidates to be further investigated and followed up in fine mapping types of studies.

Figure 4-4. SD, PNFA and FTD-MND: Manhattan plots

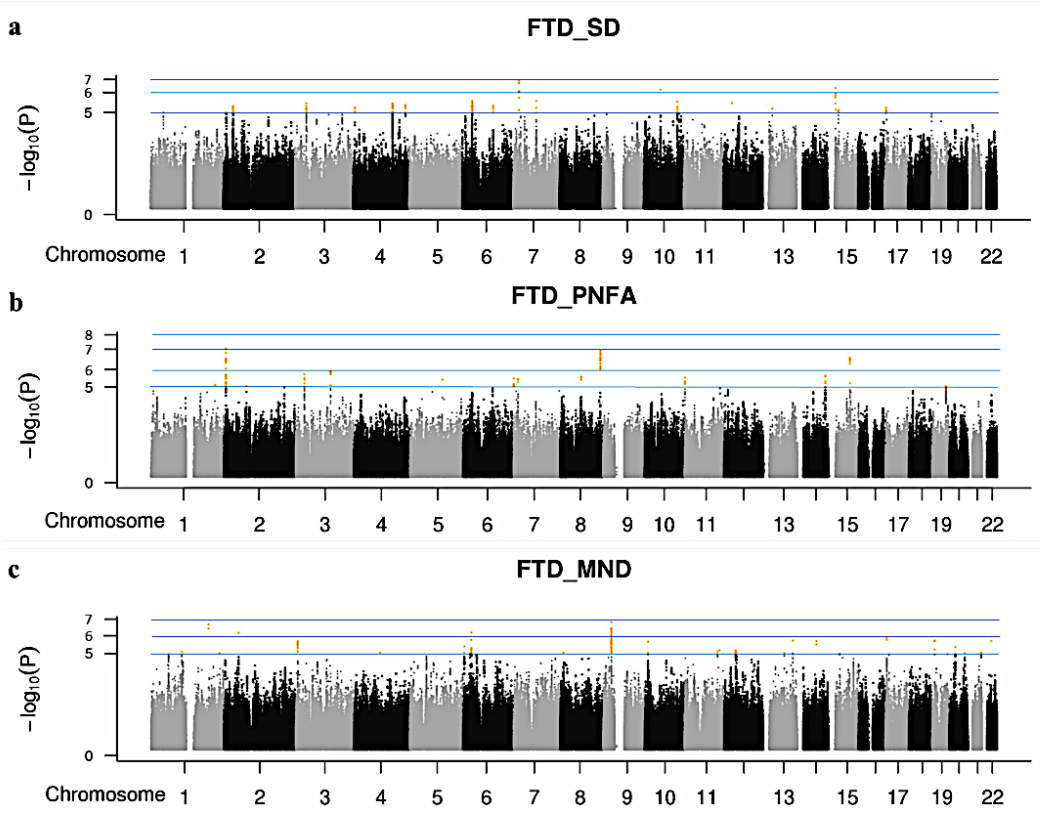


Figure 4-4. The Manhattan plots for each of the other subtypes are shown: SD (**a**), PNFA (**b**) and FTD-MND (**c**). SNPs that achieved p-values<$1x10^{-05}$ are depicted by the orange dots. SNPs with associated 10^{-07}≤p-values≤10^{-06} are suggestive of associations at loci to be further investigated as possibly relevant to each specific subtype.

Whole cohort meta-analysis

Next, the interest was to establish whether there was association with specific loci for the entire cohort. After performing association analysis for the totality of the samples included in the discovery phase (n=2,154 cases and n=7,444 controls after quality control

filtering), 29 SNPs resulted exceeding the threshold of genome-wide significance (p-value<$5x10^{-08}$) at the 6p21.3 locus, on chromosome 6 (**Table 4-8**).

Table 4-8. Discovery phase: SNPs associated with the entire cohort

SNP	Chr	Position (BP)	Effect allele	Alternate allele	Frequency of effect allele	Beta	OR	Standard error	P-value	Directionality across subtypes	Heterogeneity p-value (Cochrane's Q)
rs9268877	6	32431147	A	G	0.440	0.286	1.331	0.045	1.65E-10	++++	0.8184
rs9268852	6	32429594	A	G	0.439	0.280	1.323	0.044	1.68E-10	++++	0.8242
rs9268863	6	32430289	A	G	0.442	0.279	1.321	0.044	2.20E-10	++++	0.8295
rs9268881	6	32431606	A	T	0.560	-0.288	0.750	0.045	2.39E-10	----	0.8102
rs9268888	6	32431867	T	G	0.556	-0.302	0.740	0.048	2.99E-10	----	0.645
rs9268912	6	32432509	A	C	0.559	-0.300	0.741	0.048	3.88E-10	----	0.7369
rs9268883	6	32431638	A	T	0.553	-0.289	0.749	0.046	4.49E-10	----	0.7834
rs9268893	6	32431927	T	C	0.539	-0.290	0.748	0.048	1.42E-09	----	0.721
rs9268856	6	32429719	A	C	0.251	-0.285	0.752	0.050	1.30E-08	----	0.9437
rs9268854	6	32429672	A	G	0.752	0.288	1.334	0.051	1.47E-08	++++	0.943
rs9268855	6	32429675	A	G	0.248	-0.288	0.750	0.051	1.47E-08	----	0.943
rs9268862	6	32430167	A	C	0.750	0.284	1.328	0.050	1.48E-08	++++	0.9426
rs9268845	6	32429204	T	C	0.240	-0.295	0.744	0.052	1.48E-08	----	0.97
rs9268857	6	32429739	A	G	0.750	0.283	1.327	0.050	1.62E-08	++++	0.9431
rs9268850	6	32429477	A	G	0.248	-0.286	0.751	0.051	1.69E-08	----	0.9431
rs9268840	6	32428804	T	C	0.246	-0.284	0.753	0.051	2.06E-08	----	0.9504
rs7747521	6	32431105	A	G	0.761	0.303	1.353	0.054	2.26E-08	++++	0.9374
rs4434496	6	32430508	T	C	0.249	-0.289	0.749	0.052	2.31E-08	----	0.941
rs4428528	6	32430362	C	G	0.752	0.289	1.334	0.052	2.36E-08	++++	0.9406
rs7766843	6	32430729	T	C	0.245	-0.293	0.746	0.053	2.37E-08	----	0.9421
rs6940440	6	32429087	A	C	0.747	0.274	1.315	0.049	2.41E-08	++++	0.9543
rs7766854	6	32430752	T	C	0.245	-0.293	0.746	0.053	2.47E-08	----	0.9425
rs7747010	6	32430800	A	G	0.759	0.299	1.348	0.054	3.11E-08	++++	0.9409
rs7747025	6	32430814	A	G	0.759	0.299	1.348	0.054	3.11E-08	++++	0.9409
rs4280993	6	32430604	A	C	0.249	-0.289	0.749	0.052	3.20E-08	----	0.9399
rs7746751	6	32430867	A	G	0.249	-0.289	0.749	0.052	3.21E-08	----	0.9398
rs7746922	6	32430975	A	C	0.245	-0.292	0.747	0.053	3.82E-08	----	0.9374
rs9268885	6	32431705	T	C	0.245	-0.305	0.737	0.056	4.43E-08	----	0.9275
rs1980493	6	32363215	T	C	0.853	0.329	1.389	0.060	4.94E-08	++++	0.793

Table 4-8. All the 29 SNPs exceeding the genome-wide significance threshold at 6p21.3 are shown. The OR is calculated based on the effect allele.

Among the associated SNPs three were chosen as reference SNPs having in mind the need to address the potential issue of basing the replication analyses on surrogate/proxy SNPs. The selected SNP rs1980493 (p-value=$4.94x10^{-08}$, OR=1.389) locates in intron 5 of the butyrophilin-like 2 (MHC class II associated) gene (*BTNL2*), whilst the other two selected SNPs rs9268877 (p-value=$1.65x10^{-10}$, OR=1.331) and rs9268856 (p-value=$1{\cdot}30x10^{-08}$, OR=0.752) are both located between the major histocompatibility complex class II, DR alpha and DR beta 5 genes (*HLA-DRA*; *HLA-DRB5*), specifically, ~18.5–20 kb downstream from *HLA-DRA* (**Table 4-9** and **Figure 4-5**).

Table 4-9. Discovery phase: SNPs associated with the entire cohort (chosen for replication and joint analysis)

	Discovery phase										
Trait	***Marker***	***Chr***	***BP***	***Candidate gene***	***Effect allele***	***Alternate allele***	***Frequency of Effect allele***	***Imputation quality***	***OR***	***SE***	***P***
All FTD*	rs9268877	6	32431147	*HLA-DRA/HLA-DRB5*	A	G	0.440	0.7783	1.331	0.045	$1.65x10^{-10}$
	rs9268856	6	32429719	*HLA-DRA/HLA-DRB5*	A	C	0.251	0.8563	0.752	0.050	$1.30x10^{-08}$
	rs1980493	6	32363215	*BTNL2*	T	C	0.853	0.9642	1.389	0.060	$4.94x10^{-08}$

Table 4-9. SNPs exceeding genome-wide significance in discovery phase for the entire cohort are listed along with their chromosomal position (BP) and nearby genes. The OR is calculated based on the effect allele whilst the risk allele can be derived. * denotes only minimal cross subtype heterogeneity, with heterogeneity p-values ranging from 0.793 – 0.944 based on Cochran's Q.

Of note, the SNPs with smallest p-values ($1x10^{-10} \leq$ p-value $<5x10^{-08}$) (**Table 4-8**) all encompassed the genetic region containing the *BTNL2* and the *HLA-DRA* and *HLA-DRB5* genes (**Figure 4-5**).

Figure 4-5. Meta-analysis: Manhattan plot

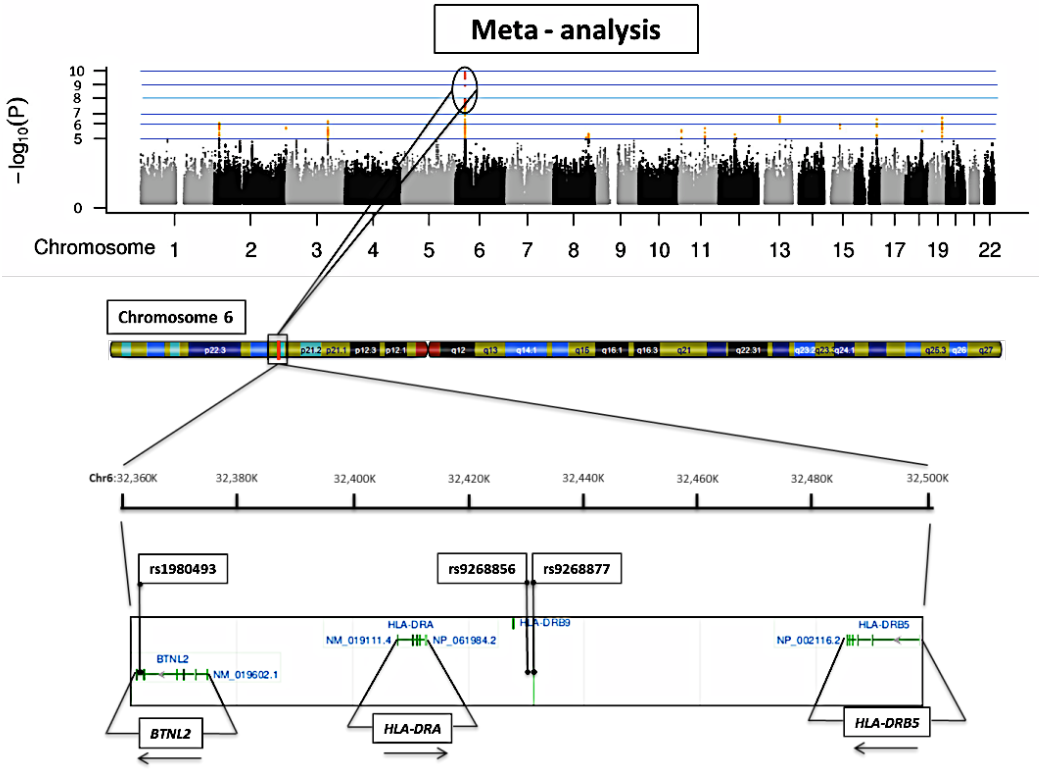


Figure 4-5. Manhattan plots of −log10 p-values across genome identifying the regions with genome-wide significant associations for the entire cohort. The Manhattan plot depicts the associated region at 6p21.3. The red dots identify the SNPs that reached genome-wide significance; the orange dots identify the SNPs suggestive of association.

In summary, after completion of all the planned association analyses for the discovery phase two novel loci were evident.

The first locus of interest was on chromosome 11, at 11q14, and was relevant for the bvFTD subtype only. The associated region included the *RAB38*/*CTSC* genes (**Table 4-7**

and **Figure 4-3**). The second relevant locus was on chromosome 6, at 6p21.3, and indicated a probable link between the FTD spectrum and the *BTNL2* and *HLA-DRA*/*HLA-DRB5* genes (**Table 4-9** and **Figure 4-5**).

On the other hand, possibly due to sample size, no genome-wide significant association could be identified in the SD, PNFA and FTD-MND subtypes.

Interestingly, there was no association at the expected or known loci such as those encompassing *MAPT* and *PGRN* genes on chromosome 17, or *C9orf72* on chromosome 9. The most likely reasons that explain the lack of association at these loci are:

1 – The fact that all known chromosome 17 mutation carriers were excluded from analysis, and;

2 – The frequency of *C9orf72* expansion carriers within the whole discovery cohort was ~8% as revealed by *post hoc* analysis (**Table 4-10**), therefore probably too low and insufficient for contributing to a genome-wide significant signal.

Of note, the discovery phase dataset included samples collected from the USA and Europe and revealed a frequency of only 8% of expansion carriers (n=194/2412). First, this possibly suggests that the expansion may be more frequent in specific geographical areas (such as central/northern Europe, specifically, France, Belgium, Sweden and partially UK as reflected in the discovery cohort) and be more characteristic of a specific subtype, i.e. FTD-MND (=52/221, 23.5%), rather than the bvFTD subtype or the language variant syndromes (**Table 4-10**). A closer look at the Manhattan plots (**Figures**

4-3 – 4-5) shows complete absence of any signal at the *C9orf72* locus in the bvFTD, SD and PNFA subtypes, as well as in the whole cohort, whilst a signal is only detectable in the FTD-MND subtype (**Figure 4-4c**), which, however, does not reach genome-wide significance, being the p-value=$2.12x10^{-06}$ (**Table 4-17**).

Table 4-10. Frequency of *C9orf72* positive cases within the discovery cohort

	Discovery phase									
	bvFDT		**SD**		**PNFA**		**FTD-MND**		**Total**	
C9orf 72+	n+/nS	%	n+/nS	%	n+/nS	%	n+/nS	%	n+/nS	%
	121/1537	**7.9**	*12/350*	**3.4**	*9/304*	**3.0**	*52/221*	**23.5**	*194/2412*	**8.0**

Table 4-10. Summary of the frequency of *C9orf72* positive cases within the discovery cohort shown for each subtype separately and for the totality of samples. n+ = number of cases positive for the repeat expansion; nS = number of cases screened.

It is also noteworthy that the association analyses performed during the discovery phase of the FTD-GWAS generated a robust number of SNPs with p-values between $1x10^{-07}$ and $1x10^{-06}$. These SNPs are all suggestive of possible association and are all valuable candidates for follow up studies. It is in fact a possibility that specific loci that might act as disease modifiers are hidden within these data and, all the more, that they may highlight genetic differences between the different subtypes and point to specific genetic markers (and pathways) as factors underlying the development of one rather than the other phenotype.

4.2.2.2 Replication phase

During the replication phase, n=1,372 cases were analysed (**Table 4-5**) along with n=5,094 controls. At this stage, the aim was to replicate the loci identified during the discovery phase that were associated with genome-wide significance with the bvFTD subtype and the entire cohort.

Surrogate/proxy SNPs were used to assess replication for the bvFTD subtype, whilst the same associated SNPs identified during the discovery phase were evaluated for the entire cohort.

Replication and joint analysis for the bvFTD subtype

The surrogate/proxy SNPs rs302668 and rs16913634 were assessed at the 11q14 locus in n=690 bvFTD cases (**Table 4-6**).

The immediate replication analyses revealed a moderately significant p-value in the case of SNP rs302668 (p-value=$4.1x10^{-02}$, OR=1.139), whilst significance seemed weaker for rs16913634 (p-value=$7.1x10^{-01}$, OR=1.037) (**Table 4-11**).

Table 4-11. Replication phase: SNPs associated with the bvFTD subtype

Replication phase									
Trait	***Marker***	***Chr***	***BP***	***Candidate gene***	***Effect allele***	***Alternate allele***	***OR***	***SE***	***P***
bvFTD	rs302668 (proxy)	11	87876911	*RAB38*	T	C	1.139	0.064	0.041
	rs16913634 (proxy)	11	87934068	*RAB38/CTSC*	G	A	1.037	0.098	0.710

Table 4-11. SNPs assessed to verify replication at the 11q14 locus. The markers along with their chromosomal positions (BP) and p-values are shown. The OR is calculated based on the effect allele whilst the risk allele can be derived.

Following, the combination of datasets relative to the discovery and the replication phases showed joint p-values that were attenuated in the proximal SNPs rs302668 (p-value=$2.44x10^{-07}$, OR= 1.229) and rs16913634 (p-value=$8.15x10^{-04}$, OR=0.801) (**Table 4-12**) when compared with the p-values reached during discovery phase only.

Table 4-12. Joint analysis: SNPs associated with the bvFTD subtype

Discovery and replication combined									
Trait	***Marker***	***Chr***	***BP***	***Candidate gene***	***Effect allele***	***Alternate allele***	***OR***	***SE***	***P***
bvFTD	rs302668 (proxy)	11	87876911	*RAB38*	T	C	1.229	0.064	$2.44x10^{-07}$
	rs16913634 (proxy)$	11	87934068	*RAB38/CTSC*	G	A	0.801	0.049	$8.15x10^{-04}$

Table 4-12. SNPs assessed for the joint analysis at the 11q14 locus. The markers along with their chromosomal positions (BP) and p-values are shown. The OR is calculated based on the effect allele whilst the risk allele can be derived. $ denotes heterogeneity p-value < 0.01 in the meta-analysis of the discovery and replication phases combined.

These results most probably reflect a decrease in power due to proxy-based replication. However, given these circumstances, rs302668 closely approached genome-wide significance suggesting that this marker is, most probably, truly linked to the bvFTD phenotype and that variability at this locus may really contribute to disease, although, possibly, with small to moderate effect.

Replication and joint analysis for the entire cohort

The three SNPs rs9268877, rs9268856 and rs1980493 could be verified at the 6p21.3 locus in the whole replication cohort for a total of n=1,372 cases (**Table 4-5**) and n=5,094 controls after quality control filtering.

The replication analysis revealed moderate significance for 2 of the 3 assessed SNPs, in that rs9268877 showed a p-value=$1.04x10^{-01}$ with OR=1.080, whilst rs9268856 showed a p-value=$1.4x10^{-02}$ with OR=0.878 and rs1980493 showed a p-value=$2.0x10^{-02}$ with OR=1.172 (**Table 4-13**).

Table 4-13. Replication phase: SNPs associated with the entire cohort

	Replication phase								
Trait	***Marker***	***Chr***	***BP***	***Candidate gene***	***Effect allele***	***Alternate allele***	***OR***	***SE***	***P***
	rs9268877	6	32431147	*HLA-DRA/HLA-DRB5*	A	G	1.080	0.047	0.104
All FTD*	rs9268856	6	32429719	*HLA-DRA/HLA-DRB5*	A	C	0.878	0.053	0.014
	rs1980493	6	32363215	*BTNL2*	T	C	1.172	0.068	0.020

Table 4-13. SNPs assessed to verify replication at the 6p21.3 locus. The markers along with their chromosomal positions (BP) and p-values are shown. The OR is calculated based on the effect allele whilst the risk allele can be derived. * denotes only minimal cross subtype heterogeneity, with heterogeneity p-values ranging from 0.793 – 0.944 based on Cochran's Q.

Combined analyses of discovery and replication phases confirmed the strong association seen in the discovery phase. In fact the p-values for each SNP passed the genome-wide threshold: rs9268877 (p-value=1.05×10^{-08}, OR=1.204), rs9268856 (p-value=5.51×10^{-09}, OR=0.809), and rs1980493 (p-value=1.57×10^{-08}, OR=1.290) (**Table 4-14**).

Table 4-14. Joint analysis: SNPs associated with the entire cohort

	Discovery and replication combined								
Trait	***Marker***	***Chr***	***BP***	***Candidate gene***	***Effect allele***	***Alternate allele***	***OR***	***SE***	***P***
	rs9268877$	6	32431147	*HLA-DRA/HLA-DRB5*	A	G	1.204	0.039	1.05×10^{-08}
All FTD*	rs9268856	6	32429719	*HLA-DRA/HLA-DRB5*	A	C	0.809	0.029	5.51×10^{-09}
	rs1980493	6	32363215	*BTNL2*	T	C	1.290	0.058	1.57×10^{-08}

Table 4-14. SNPs assessed for the joint analysis at the 6p21.3 locus. The markers along with their chromosomal positions (BP) and p-values are shown. The OR is calculated based on the effect allele whilst the risk allele can be derived. $ denotes heterogeneity p-value <0.01 in the meta-analysis of the discovery and replication phases combined. * denotes only minimal cross subtype heterogeneity, with heterogeneity p-values ranging from 0.793 – 0.944 based on Cochran's Q. $ denotes heterogeneity p-value<0.01 in the meta-analysis of the discovery and replication phases combined.

The results for the replication and joint analyses in the entire cohort confirmed the association of the 6p21.3 locus with the FTD spectrum strongly arguing for a remarkable link between FTD and variability at the human *HLA* locus.

4.2.2.3 Quantitative trait loci (QTL) analyses

The GWAS hits identified in this study highlight two novel loci associated with the disease, namely the 11q14 locus, encompassing the genes *RAB38*/*CTSC* for the bvFTD subtype, and the 6p21.3 locus, encompassing the genes *BTNL2* and *HLA-DRA*/*DRB5* for the entire FTD cohort.

The next step was to evaluate whether any of the associated SNPs held a possible biological relevance, e.g. by influencing methylation and/or expression levels. These aspects could be investigated in brain tissues that were assayed to assess genome-wide *cis* methylation and expression patterns.

Methylation quantitative trait loci (mQTL)

The Methylation27 dataset (section ***4.2.1***) was used to assess effects on methylation exerted by any of the associated SNPs.

SNPs identified in the discovery phase at the 6p21.3 locus showed significant associations with proximal CpG methylation probes after multiple test correction.

Of particular interest were two SNPs that were closely associated with changes in methylation levels related to the *HLA-DRA* gene in both frontal cortex and cerebellar tissue.

The first SNP, rs1980493 (located in intron 5 of *BTNL2*, risk allele T), showed p-value=$3.34x10^{-08}$ in the cerebellum and p-value=$2.17x10^{-08}$ in the frontal cortex, whilst the second SNP, rs9268856 (located downstream from *HLA-DRA*, risk allele C), revealed p-value=$1.16x10^{-06}$ in the frontal cortex (**Table 4-15**).

Data suggest that the risk alleles of both SNPs are associated with changes in the methylation levels related to *HLA-DRA*.

Comparatively, there was no evidence of *cis*-mQTL at 11q14.

Table 4-15. FTD-GWAS: Methylation quantitative trait loci

Dataset	CpG probe	SNP	Chr	Position (bp)	Reference allele	Alternate allele	Frequency of reference allele	Imputation quality	Effect estimate of alternate allele (in Z units)	Standard error	P-value	FDR adjusted p-value	Probe start (BP)	Probe end (BP)	Symbol
Cerebellum - CpG methylation	cg18055007	rs1980493	6	32363215	T	C	0.8361	0.9888	0.456	0.116	0.0000893	0.0106267	31698226	31698276	DDAH2
Cerebellum - CpG methylation	cg21415604	rs1980493	6	32363215	T	C	0.8361	0.9888	-0.44	0.116	0.0001588	0.01133832	31948433	31948483	C4B
Cerebellum - CpG methylation	cg03543593	rs1980493	6	32363215	T	C	0.8361	0.9888	-0.447	0.116	0.000123	0.01097775	32076417	32076467	TNXB
Cerebellum - CpG methylation	cg25764570	rs1980493	6	32363215	T	C	0.8361	0.9888	-0.643	0.116	$3.34x10^{-08}$	0.0000119	32407239	32407289	HLA-DRA
Cerebellum - CpG methylation	cg19248557	rs1980493	6	32363215	T	C	0.8361	0.9888	-0.381	0.116	0.00106	0.0473025	32407838	32407888	HLA-DRA
Cerebellum - CpG methylation	cg14833385	rs1980493	6	32363215	T	C	0.8361	0.9888	0.394	0.116	0.0007042	0.0418999	32921142	32921192	HLA-DMA
Cerebellum - CpG methylation	cg25764570	rs9268856	6	32429719	C	A	0.748	0.9687	-0.422	0.1	0.0000228	0.004064445	32407239	32407289	HLA-DRA
Cerebellum - CpG methylation	cg15982308	rs9268877	6	32431147	G	A	0.5785	0.9532	0.281	0.087	0.001253	0.049702333	32145780	32145830	AGPAT1
Cerebellum - CpG methylation	cg10714284	rs9268877	6	32431147	G	A	0.5785	0.9532	-0.287	0.087	0.0009995	0.0473025	32908851	32908901	HLA-DMB
Frontal cortex - CpG methylation	cg21415604	rs1980493	6	32363215	T	C	0.8361	0.9888	-0.463	0.116	0.0000701	0.00834666	31948433	31948483	C4B
Frontal cortex - CpG methylation	cg25764570	rs1980493	6	32363215	T	C	0.8361	0.9888	-0.652	0.116	$2.17x10^{-08}$	0.00000773	32407239	32407289	HLA-DRA
Frontal cortex - CpG methylation	cg25764570	rs9268856	6	32429719	C	A	0.748	0.9687	-0.484	0.1	$1.16x10^{-06}$	0.000207417	32407239	32407289	HLA-DRA

Table 4-15. Summary of association of top hits with -*cis* methylation levels at the 6p21.3 locus. Association is shown for SNPs rs1980493 and rs9268877 suggestive of their implication in methylation processes and patterns in relation to *HLA-DRA*

Expression quantitative trait loci (eQTL)

To assess effects on expression exerted by the associated SNPs, several different datasets were scrutinized (section ***4.2.1***).

There were no significant associations in the HT12 dataset after adjusting for multiple test correction, whilst in the ExonArray dataset there was only one *cis*-eQTL at the 6p21.3 locus that passed Bonferroni correction of 0.05 ($p=4.8x10^{-07}$) involving rs1980493 (intron 5 of *BTNL2*, risk allele T) and exon 7 (Affymetrix ID 2903485) (**Table 4-16a**) of the solute carrier family 39 (zinc transporter), member 7 gene (*SLC39A7*) that encodes a protein involved in transport of zinc from both the extracellular environment and the intracellular storage (OMIM: #601416).

Table 4-16. FTD-GWAS: Expression quantitative trait loci

a

rsid	chr:pos.hg19	exprID	tID	Symbol.NA31	CRBL	FCTX	HIPP	MEDU	OCTX	PUTM	SNIG	TCTX	THAL	WHMT
rs1980493	chr6:32363215	2903485	2903470	SLC39A7	6.9E-01	3.1E-01	5.0E-01	4.8E-07	9.9E-01	8.5E-01	1.8E-02	3.9E-01	2.4E-01	2.2E-01

b

SNP	probe_id	ILMN_Gene	snp_probe_distance	p-value
rs302652	ILMN_2134974	RAB38	312,176	5.05E-32

Table 4-16. Summary and statistics of effects on expression exerted by rs1980493 (S6a) and rs302652 (S6b).

This eQTL appears to be tissue-specific (medulla) (**Figure 4-6**), and the risk allele is associated with increased expression of *SLC39A7* (trans-membrane zinc transporter that transports zinc from Golgi and endoplasmic reticulum within the cytoplasm). A possible link between elevated expression of SLC39A7 and pathogenesis of neurodegenerative diseases needs to be further investigated.

Figure 4-6. Regional distribution of *SLC39A7* mRNA expression

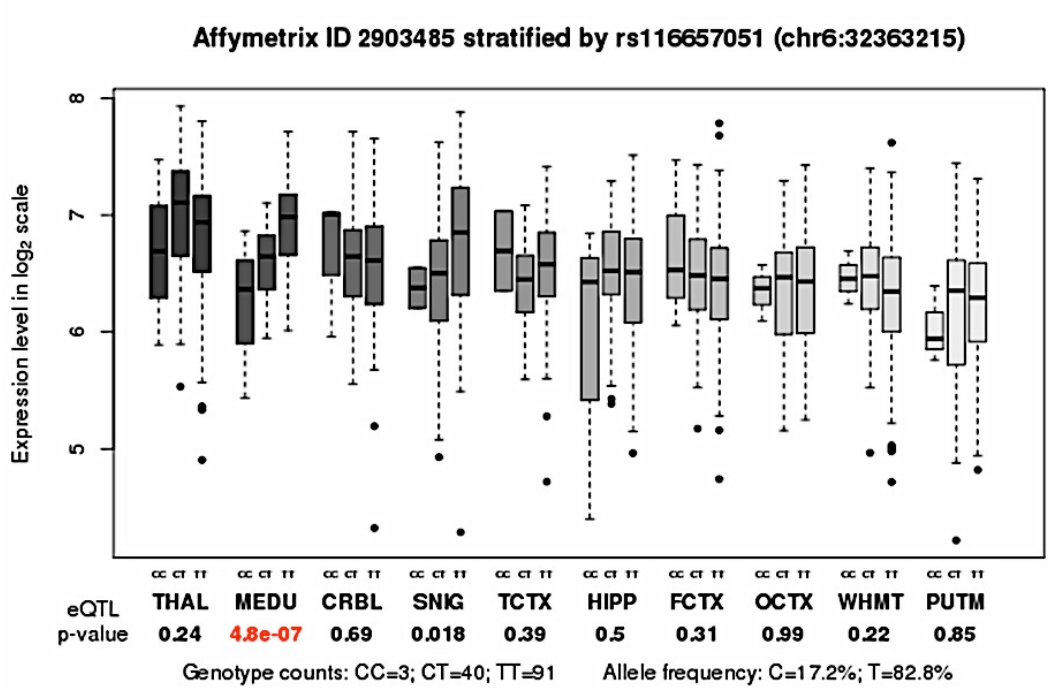


Figure 4-6. Box plots of mRNA expression levels for 10 brain regions are shown, based on microarray experiments and plotted on a log2 scale (y-axis). **Abbreviations**: Thalamus (THAL), medulla (MEDU), cerebellum (CRBL), substantia nigra (SNIG), temporal cortex (TCTX), hippocampus (HIPP), frontal cortex (FCTX), occipital cortex (OCTX), white matter (WHMT) and putamen (PUTM). Whiskers extend from the box to 1.5 times the inter-quartile range.

In addition, the Zeller et al dataset [269] revealed a remarkable *cis*-eQTL ($p=5.05x10^{-32}$) at the 11q14 locus between SNP rs302652 (chr11: 87894881, risk allele T) (**Table 4-16b**) and decreased expression of *RAB38* (Illumina ILMN_2134974 located in chr11: 87846656-87846705) in monocytes suggesting a direct role in transcriptional processes in –*cis* for this SNP.

In summary, both methylation and expression quantitative trait loci (m/eQTL) data did support the notion that risk at the associated loci may involve *cis* changes of methylation and expression levels pointing to the regulation of transcription as the likely mechanism by which variability at the associated loci exerts its effect.

4.2.2.4 Copy number polymorphisms (CNPs) and FTD

Background

There is evidence in the literature of a link between some rare cases of FTD and the occurrence of copy number polymorphisms (CNPs) on chromosome 17, specifically, at the *MAPT* and *GRN* loci [49]. However, CNPs at these loci had also been described in a number of other phenotypes. For example, in 2006, several research groups reported aberrant rearrangements on chromosome 17 associated with mental retardation, developmental impairment, hypotonia and facial dysmorphism: This was named "the chromosome 17q21.31 microdeletion syndrome" with an estimated frequency of 1 in 16,000 [55], [56], [57]. The microdeletion was suggested to span a region of 0.5 Mb on

chromosome 17, encompassing the 4 genes, *MAPT*, intramembrane protease 5 (*IMP5*), corticotropin releasing hormone receptor 1 (*CRHR1*) and saitohin (*STH*) [105]. Moreover, a *de-novo* duplication at the same locus, reciprocal to the 17q21.31 microdeletion, was reported in an individual affected by psychomotor developmental delay, facial dysmorphism and microcephaly [102]. Finally, a microduplication on chromosome 17q21.31 was reported in a patient with behavioral problems as well as poor social interactions with features of the autism spectrum disorder [103].

In the FTD arena, two studies failed to identify association between CNPs at the *MAPT* and *GRN* loci and FTD phenotypes [270], [271]. However, two other studies reported a heterozygous 17.3 kb deletion responsible for the removal of exons 6 to 9 of *MAPT* in one FTD patient [272] and a 439 kb duplication in the region encompassing *CRHR1*, *IMP5*, *MAPT* and *STH*, in a patient diagnosed with bvFTD and amnestic syndrome [273].

Although CNVs at these loci seem to associate with a variety of phenotypes including developmental and mental impairments, autism spectrum, social/behavioural issues, and FTD, this implied to possible occurrence of CNVs among the FTD-GWAS discovery cohort. To investigate such possibility I visually inspected the chromosome 17 genotypes in the search for copy number polymorphisms (CNPs) within the extended discovery cohort assessing B Allele Frequency and LogR Ratio plots the chromosome browser tool from the Genome Studio software.

The analysis

This preliminary analysis targeted exclusively chromosome 17 with the specific aim of evaluating presence of possible deletions and/or duplications at either the *MAPT* (chr17:43,960kb–44,120kb +/- 2Mb) or *GRN* (chr17: 42,422kb–42,431kb +/- 2Mb) loci. After having inspected a sample size of ~1,100 from the discovery phase cohort, which included FTD samples form the USA, France, Denmark, Germany, Sweden and Italy, the outcome could be interpreted based on three fundamental observations:

1 – One group of samples showed LOH regions at the *MAPT*/*GRN* loci with sizes varying from a minimum of ~400 kb to a remarkable maximum of 22 Mb (**Figure 4-7a-f**);

2 – One group of samples revealed duplications at other loci (not *MAPT*/*GRN*) (**Figure 4-8a-h**), and;

3 – One group of samples showed CNVs at other loci (not *MAPT*/*GRN*) (**Figure 4-9a-h**).

LOH at the MAPT/GRN loci

The smallest LOH region was ~400 kb (**Figure 4-7a**), and was encompassing *MAPT*. It still needs to be established whether this LOH causes a change in the gene dosage. In the case of the larger LOH regions of sizes of ~22 Mb (**Figure 4-7b**), 11.3 Mb (**Figure 4-7c**) and 9 Mb (**Figure 4-7d**), two points need to be mentioned:

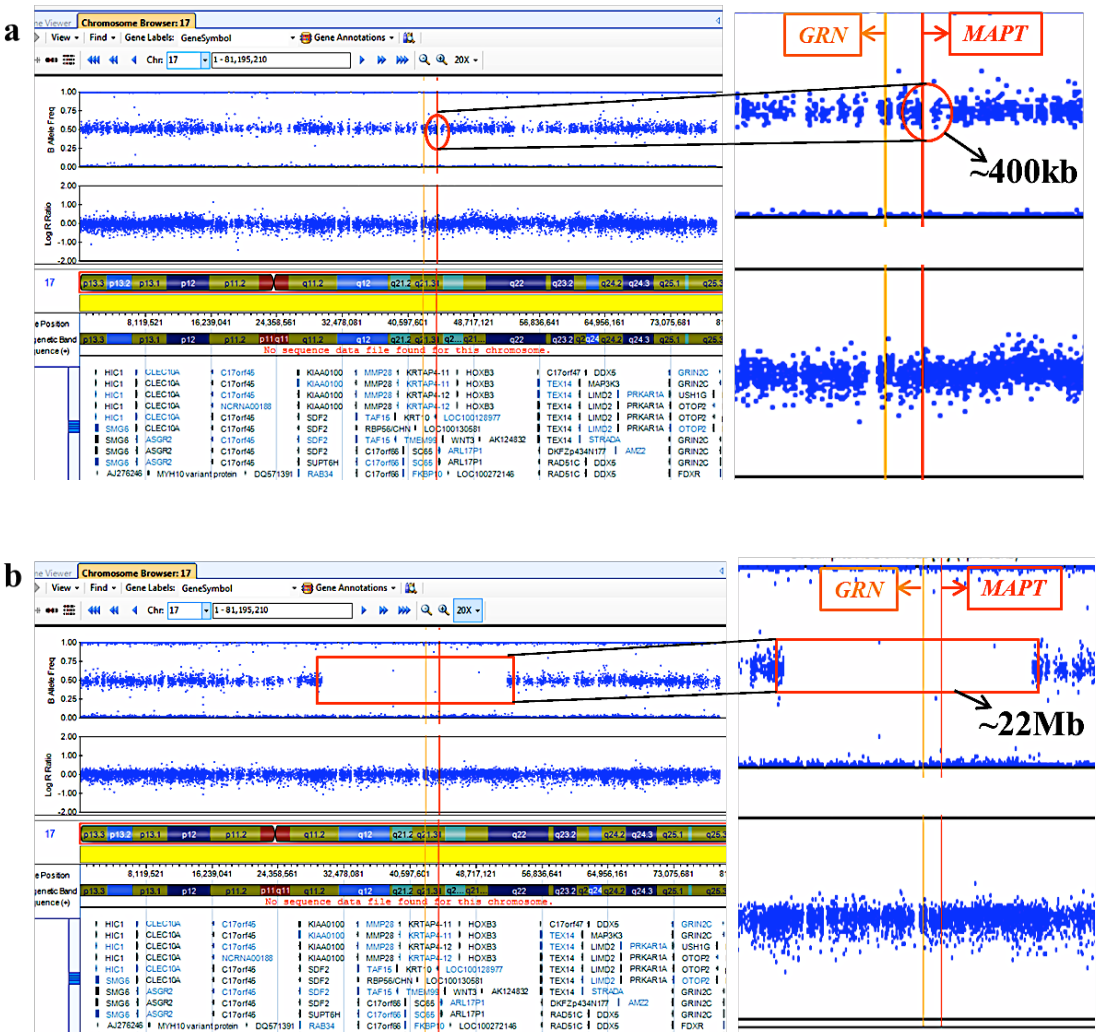
1 – These LOH regions are extraordinarily large and encompass ~15-30 genes including both *GRN* and *MAPT*, and;

2 – Although the LogR Ratio graph does not show an inflation, the LOH displayed in the B Allele Frequency graph is clearly evident.

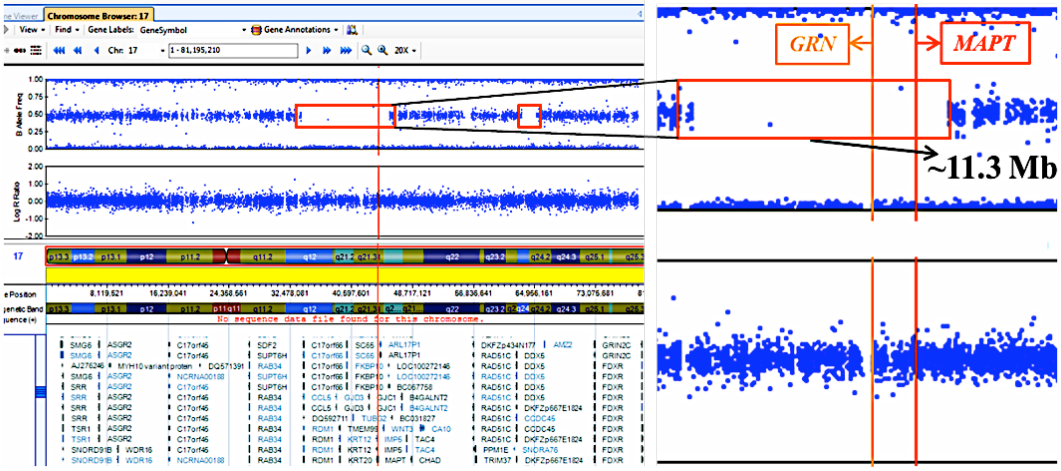
Provided further proof this may mean that either there is loss of one allele (that affects a remarkable number of genes collectively), or else, there is an extended and consecutive series of homozygous SNPs in this region that are unusual. It needs to be established whether either of these possibilities may influence the phenotype.

Among the six cases presented in **Figure 4-7**, 3 were diagnosed with bvFTD (**Figure 4-7c**, **e** and **f**), 2 with PNFA (**Figure 4-7a** and **b**) and one with SD (**Figure 4-7d**).

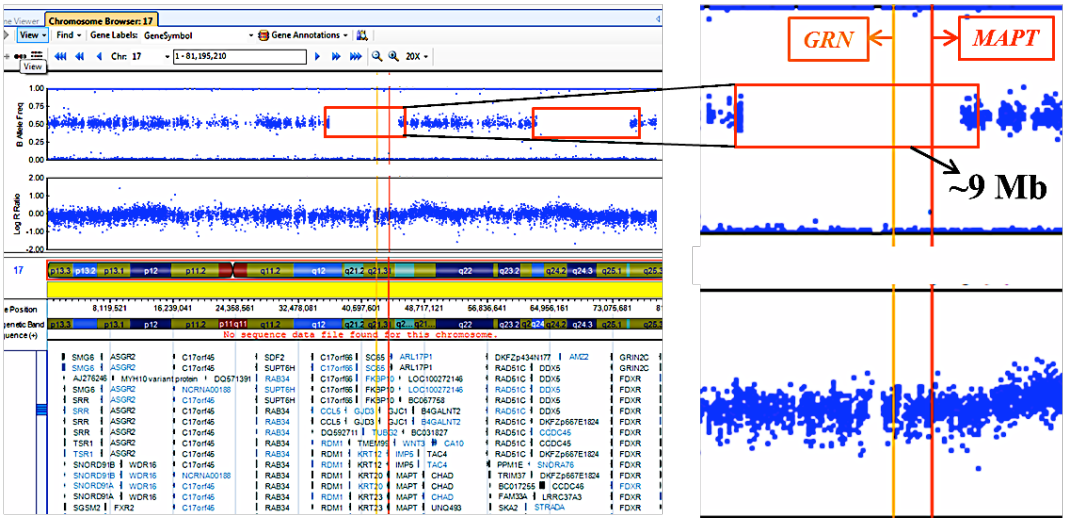
Figure 4-7. LOH analysis at the *MAPT* and *GRN* loci



c


Chromosome Browser: 17
GRN
MAPT
~11.3 Mb

d


Chromosome Browser: 17
GRN
MAPT
~9 Mb

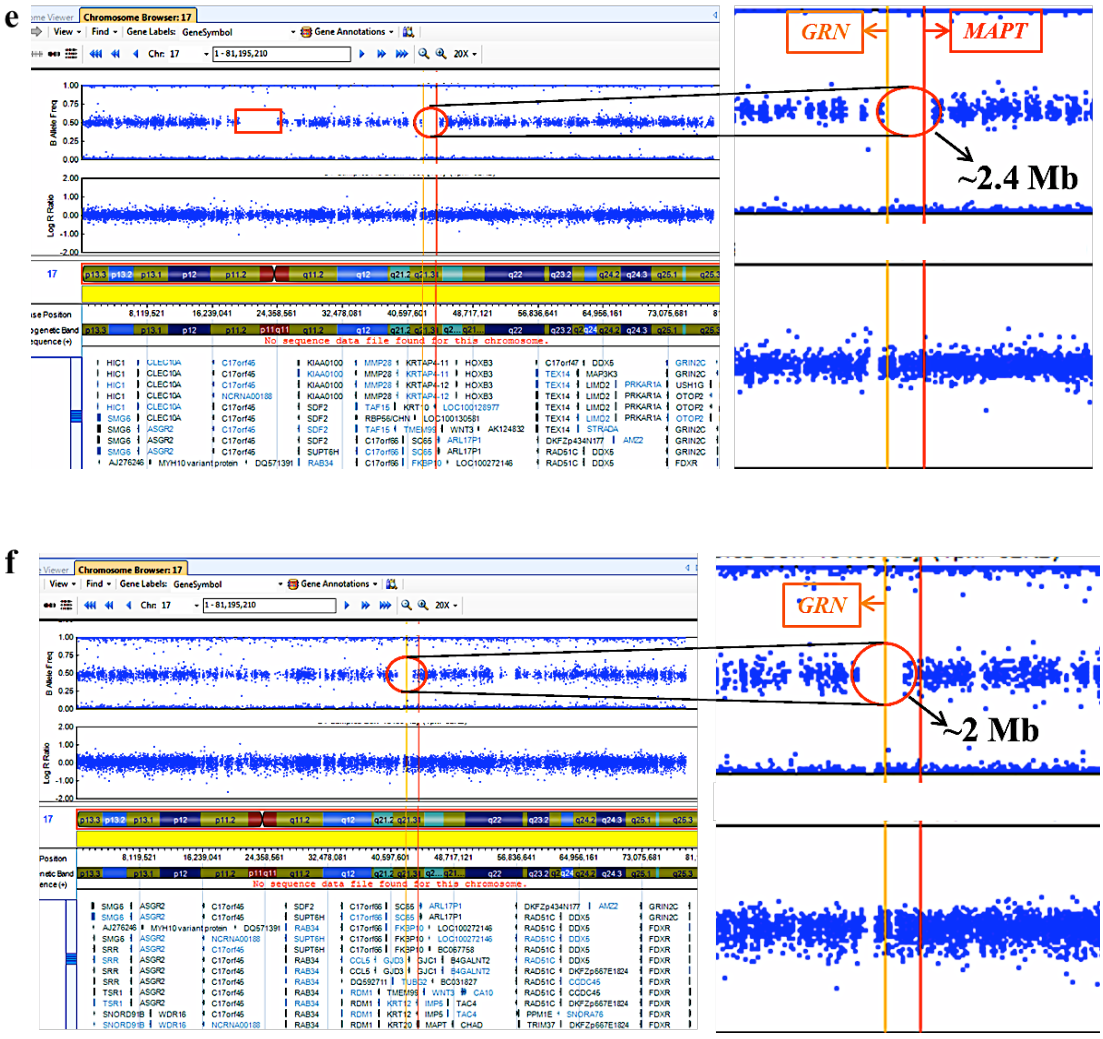


Figure 4-7. Six different examples of LOH regions encompassing *MAPT*, or *GRN* or both, identified among ~1,100 samples from the discovery cohort. (**a**) Small LOH region of ~400 kb encompassing *MAPT*; this LOH pattern belongs to a patient diagnosed with PNFA. (**b**) Largest LOH region of ~22 Mb, encompassing ~30 genes including *MAPT*/*GRN*; this LOH pattern belongs to a patient diagnosed with PNFA. (**c**) Large LOH region of ~11.3 Mb, encompassing ~20 genes including *MAPT*/*GRN*; this LOH pattern belongs to a patient diagnosed with bvFTD. (**d**) Large LOH region of ~9 Mb, encompassing ~15 genes including *MAPT*/*GRN*; this LOH pattern belongs to a patient diagnosed with SD. (**e**) Intermediate LOH region of ~2.4 Mb encompassing *MAPT*/*GRN*; this LOH pattern belongs to a patient diagnosed with bvFTD. (**f**) Intermediate LOH region of ~2 Mb encompassing *GRN* only; this LOH pattern belongs to a patient diagnosed with bvFTD.

Duplications on chromosome 17

Duplications seemed evident in a number of cases. Keeping in mind that some may represent artefacts rather than indicate a true CNV and need therefore to be further investigated and confirmed, this analysis seemed interesting for three main reasons:

1 – All putative duplications were detectable in various regions of chromosome 17 but not at the *MAPT*/*GRN* loci;

2 – Six of the 8 of the presented cases (**Figure 4-8a-h**) were diagnosed with bvFTD, and;

3 – Three putative duplications occurred in more than one sample and two out of three shared the same diagnosis (bvFTD).

These CNV patterns need to be confirmed, first, in more/other patient cohorts; second, possibly, in familial cases where a clear pattern of inheritance is evident. Third, such CNVs need to be absent from the normal population to support the argument that duplications at these loci (**Figure 4-8a-h**) are associated with the diagnosis of FTD.

Furthermore and very interestingly, there was one case (diagnosed with SD) that showed a duplication spanning almost the whole short arm of chromosome 17 (**Figure 4-8a**).

Figure 4-8. Duplications on chromosome 17

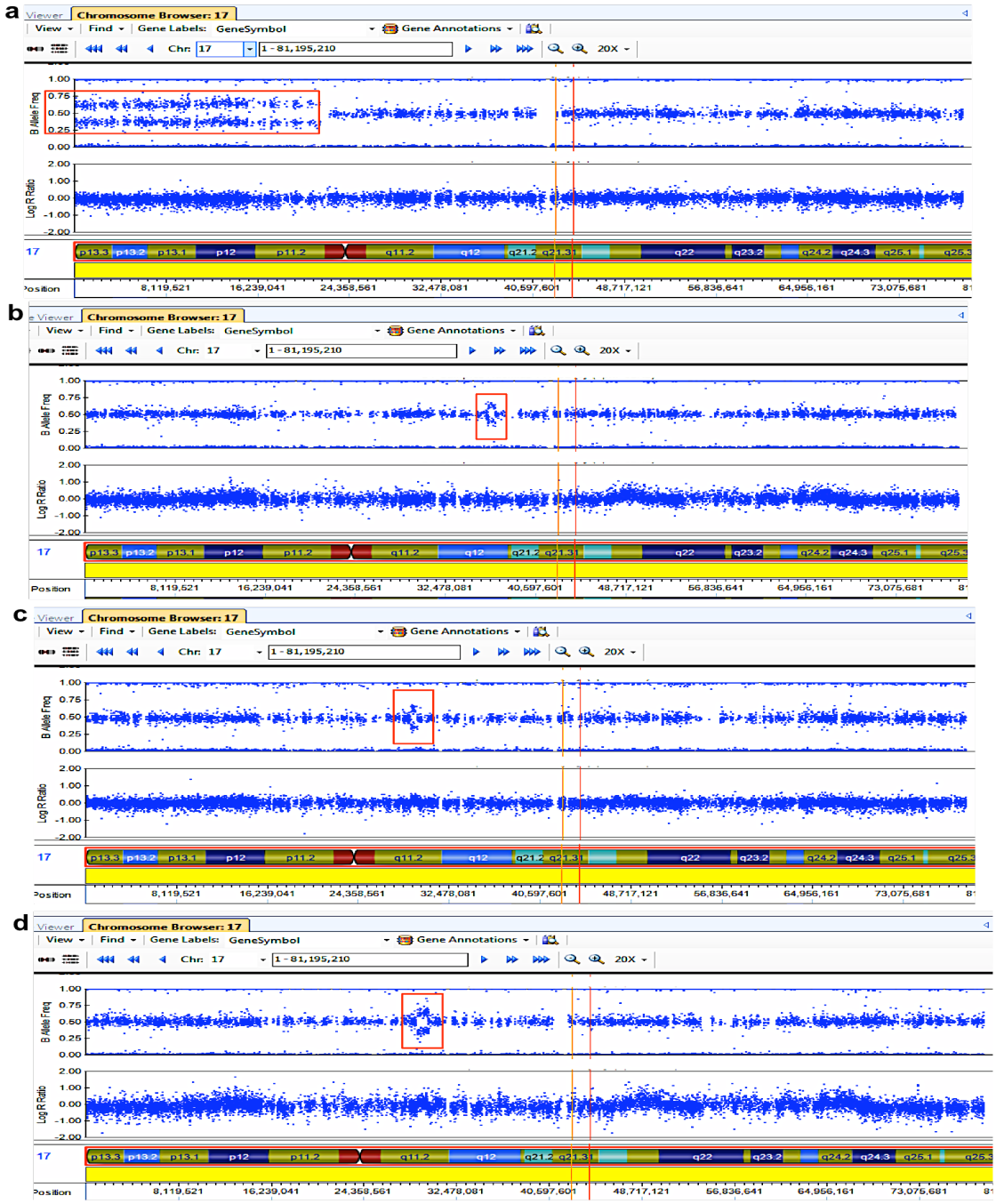


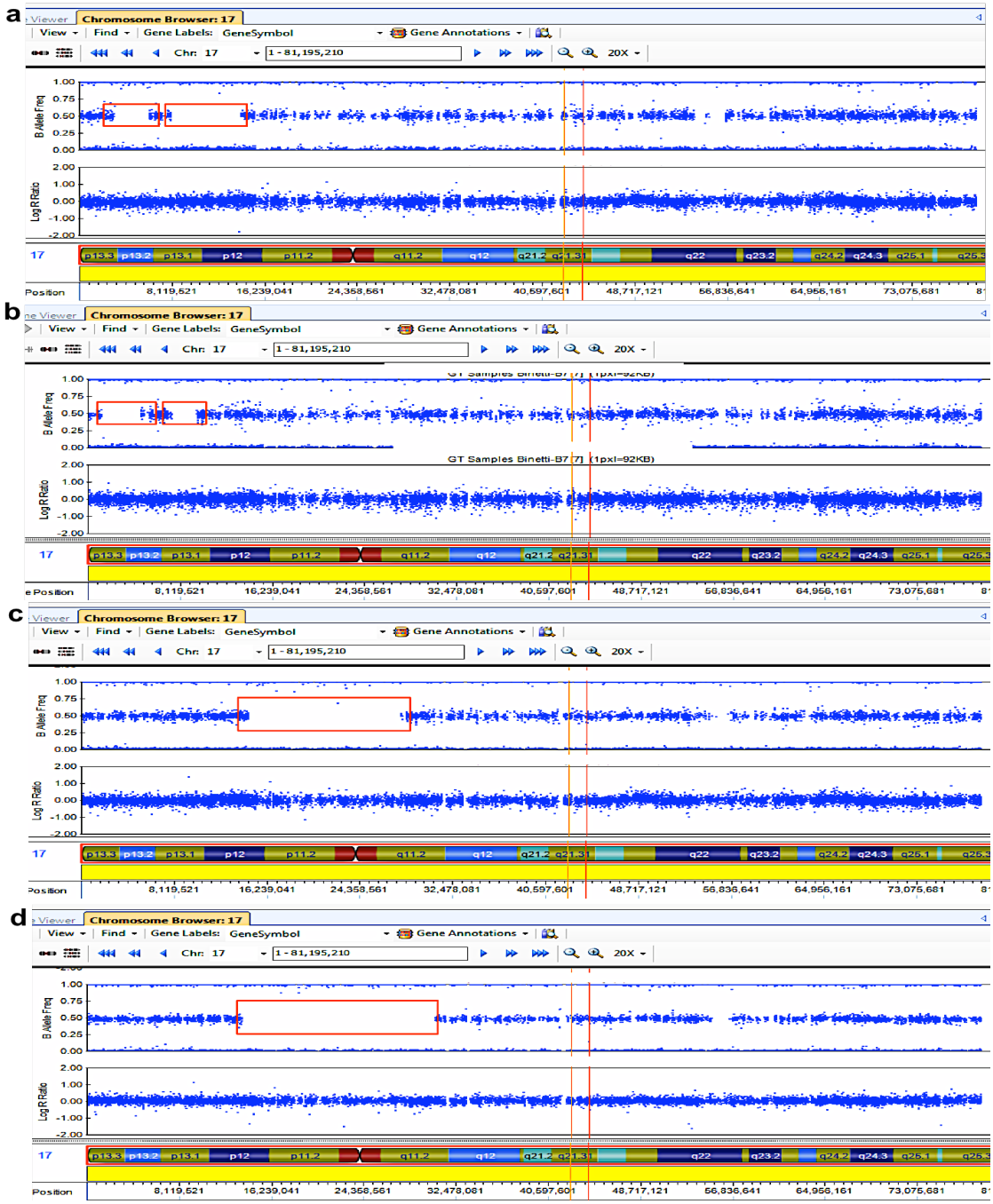
Figure 4-8. Eight different examples of duplications on chromosome 17 at loci different from *MAPT*/*GRN* identified among ~1,100 samples from the discovery cohort. (**a**) Duplication encompassing almost the complete short arm of chromosome 17 in one SD patient. (**b**) Duplication at q12 in one bvFTD patient. (**c**–**d**) Duplication at q11.2 in two bvFTD patients. (**e**–**f**) Duplication at p11.2 in two bvFTD patients. (**g**–**h**) Duplication at q24.3 in one PNFA and one bvFTD patient.

CNVs at other loci (not MAPT/GRN)

There were several cases that showed relatively large LOH regions spanning from ~2–2.5Mb (**Figure 4-9a** and **b**), to 5.7–9.7Mb (**Figure 4-9c, e**, **f** and **g**), up to 13Mb (**Figure 4-9d** and **h**) in different regions of chromosome 17 none of which were encompassing the *MAPT*/*GRN* loci. The majority of cases (6/8) were diagnosed with bvFTD (**Figure 4-9a**, **c**, **d**, **f**, **g** and **h**), SD (**Figure 4-9e**), and one case was PNFA (**Figure 4-9b**). The B Allele Frequency plots showed LOH, however the LogR Ratio plots did not clearly show an according inflation. This may mean that there is either a heterozygous deletion, with a hard to detect inflation, or that there are large portions of exclusive homozygous calls. The real nature of these CNVs needs to be further investigated through gene dosage experiments. If confirmed, haploinsufficiency of the genes in the LOH region may influence disease.

In summary it is important to note that these are only preliminary results. In addition, there is room for some of the heterozygous deletions or duplications to be artefacts (e.g. **Figure 4-8a** or **Figure 4-9d**). At the same time, some CNVs seem real, as in the case of the duplications seen in **Figure 4-8c**, **e**–**h**. At this stage, the next crucial steps will be validating these preliminary results through gene dosage experiments followed by a scrupulous ascertainment of the genes involved at either deleted or duplicated regions. Only then it will be possible to correlate the actual CNVs to the genes functions and, eventually, to the phenotype.

Figure 4-9. Deletions on chromosome 17



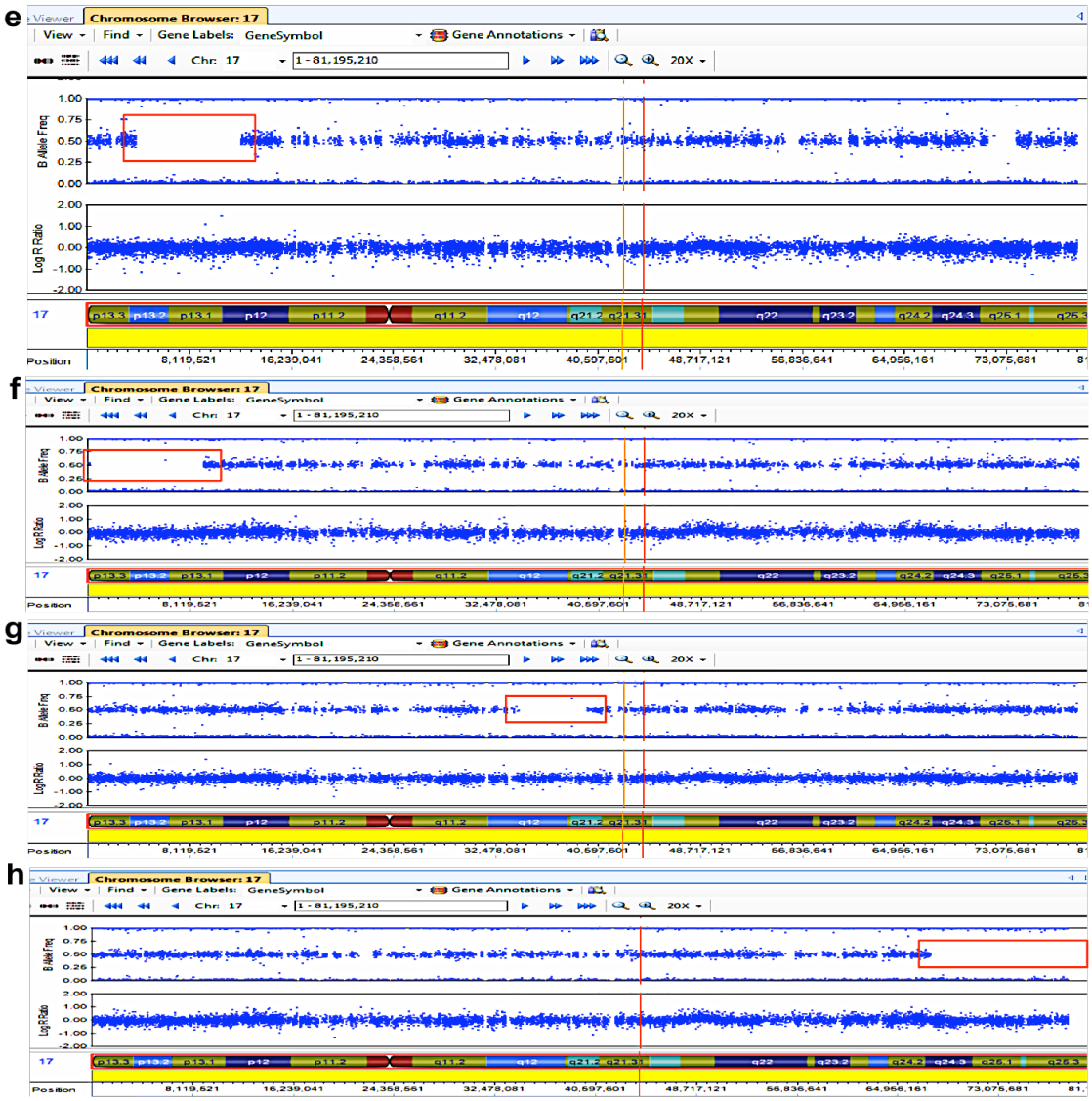


Figure 4-9. Eight different examples of CNVs on chromosome 17 at loci different from *MAPT*/*GRN* identified among ~1,100 samples from the discovery cohort. (**a**–**b**) LOH encompassing two telomeric regions of the short arm of chromosome 17 in one bvFTD and one PNFA patient. (**c**–**d**) LOH encompassing part of both the short and the long arms of chromosome 17 in two bvFTD patients. (**e**–**f**) LOH encompassing two telomeric regions of the short arm of chromosome 17 in one SD and one bvFTD patient. (**g**–**h**) LOH encompassing two different regions of the long arm of chromosome 17 for two bvFTD patients.

4.2.2.5 Candidate loci

This study provided the opportunity to explore possible genetic overlap among different neurological disorders. It is known that several neurodegenerative diseases share, to some extent, a number of features that can fall within the clinical, pathological or genetic categories. For example:

1 – Features such as broad cognitive impairment, memory deficits, changes in the behavioural habits of an individual and language dysfunction can represent a subtle clinical overlap between AD and FTD [274];

2 – Features such as tau pathology or TDP-43 pathology represent a common factor characterising FTD, AD, CBS and PSP, or FTD and ALS, respectively [275-277], and;

3 – Features such as the *C9orf72* repeat expansion represent a genetic overlap for FTD and ALS or, more generally, for the FTD-MND spectrum [76, 77, 87].

To further explore possible genetic overlap among neurodegenerative diseases, a subset of candidate loci known to be associated either with FTD or other closely related conditions such as ALS [81], PSP/CBD [180] and AD [278] were investigated and compared for association (**Table 4-17**).

Overlap between FTD and PSP/CBD

Two SNPs (rs242557 and rs8070723) that were associated with the PSP/CBD phenotypes at the *MAPT* locus [180] reached p-values between 10^{-03} and 10^{-04} in the entire FTD

discovery cohort, whilst rs8070723 only in the bvFTD as well as in the PNFA subtypes, in the FTD-GWAS (**Table 4-17**). The risk associated with the effect allele was in the same direction as in the GWAS for progressive supranuclear palsy, but the noticeable difference in OR for rs8070723 (5.46 [180] vs. ~1.2-1.4 [FTD-GWAS]) (**Table 4-17**) may reflect a smaller proportion of subjects with tau pathology or association with the *MAPT* locus within the discovery cohort, most probably due to the exclusion of *MAPT* mutation carriers from the FTD-GWAS.

Overlap between FTD and AD

One SNP, rs2075650, defining the association between AD and the *TOMM40*/*ApoE* locus approached genome-wide significance in the entire dataset (p-value=8.81×10^{-07}) and in the bvFTD subtype (p-value=1.37×10^{-06}), but not in the SD, PNFA, and FTD-MND subtypes (**Table 4-17**). Several AD-GWAS reported association with the minor allele of this SNP with OR>2.5 [278], whilst in the FTD-GWAS the OR was ~1.3 for most subtypes and in the entire cohort (**Table 4-17**). This association may reflect subtle clinical overlap of clinically diagnosed FTD and AD cases [279], or may indicate that there is a link between this locus and FTD. Based on some studies that suggest that there may be an involvement for *ApoE* also in the FTD spectrum [280-283], this hypothesis will need to be further investigated to establish whether the association of this locus with FTD is actually phenotype specific and, if so, how it exerts its detrimental effect.

Table 4-17. Candidate loci comparison

						OTHER STUDIES				OUR STUDY (discovery phase)									
Chr	**Disease**	**Gene/nearby gene**	**SNP (bp)**	**Effect allele**	**Alternate allele**	**Frequency of effect allele**	**Reported association**		**Ref**	**Frequency of effect allele**	**Imputation quality (RSQ)**	**Meta (All FTD)**		**bvFTD**		**SD**		**PNFA**		**FTD-MND**	
							P (joint)	**OR**				**P**	**OR**	**P**	**OR**	**P**	**OR**	**P**	**OR**	**P**	**OR**
17	PSP/CBD	*MAPT*	rs8070723 (44081064)	A	G	0·950 (Stage 1); 0·940 (Stage 2)	**$1{\cdot}50x10^{-116}$**	**5·46**	27	0·765	0·8400	**$2{\cdot}80x10^{-04}$**	**1·201**	**$3{\cdot}14x10^{-03}$**	**1·201**	**$4{\cdot}34x10^{-01}$**	**1·103**	**$8{\cdot}72x10^{-03}$**	**1·471**	**$5{\cdot}82x10^{-01}$**	**1·091**
			rs242557 (44019712)	G	A	0·470 (Stage 1); 0·500 (Stage 2)	**$4{\cdot}20x10^{-70}$**	**0·51**		0·634	0·5246	**$4{\cdot}82x10^{-03}$**	**0·853**	**$1{\cdot}27x10^{-02}$**	**0·841**	**$3{\cdot}20x10^{-01}$**	**0·867**	**$2{\cdot}02x10^{-01}$**	**0·815**	**$8{\cdot}23x10^{-01}$**	**0·961**
19	AD	*TOMM40/APOE*	rs2075650 (45395619)	G	A	0·150	**$1{\cdot}80x10^{-157}$**	**2·53**	28	0·141	0·9978	**$8{\cdot}81x10^{-07}$**	**1·304**	**$1{\cdot}37x10^{-06}$**	**1·383**	**$3{\cdot}64x10^{-02}$**	**1·326**	**$8{\cdot}69x10^{-01}$**	**0·976**	**$2{\cdot}06x10^{-01}$**	**1·252**
9	ALS	*C9orf72/MOB3B*	rs3849942 (27543281)	A	G	0·260	**$1{\cdot}01x10^{-08}$**	**1·20**	26	0·253	0·9996	**$4{\cdot}38x10^{-04}$**	**1·166**	**$7{\cdot}38x10^{-03}$**	**1·155**	**$9{\cdot}89x10^{-01}$**	**1·010**	**$9{\cdot}03x10^{-01}$**	**0·990**	**$2{\cdot}12x10^{-06}$**	**1·957**
7	FTLD-TDP	*TMEM106B*	rs1990622 (12283787)	A	G	0·679	**$1{\cdot}08x10^{-11}$**	**1·64**	17	0·600	0·9588	**$7{\cdot}88x10^{-02}$**	**1·073**	**$5{\cdot}85x10^{-03}$**	**1·144**	**$8{\cdot}36x10^{-01}$**	**0·978**	**$8{\cdot}98x10^{-01}$**	**0·985**	**$3{\cdot}11x10^{-01}$**	**0·876**
11	MS	*RAB38*	rs1386330 (87819427)	C	T	0·130	**$2{\cdot}00x10^{-06}$**	**NA**	30	0·141	0·9694	**$3{\cdot}35x10^{-01}$**	**1·050**	**$6{\cdot}09x10^{-01}$**	**1·040**	**$7{\cdot}60x10^{-01}$**	**1·040**	**$6{\cdot}97x10^{-01}$**	**1·060**	**$3{\cdot}00x10^{-01}$**	**0·829**
6	Vitiligo	*BTNL2*	rs3806156 (32373698)	T	G	0·474	**$7{\cdot}22x10^{-19}$**	**1·42**	31	0·346	0·9734	**$1{\cdot}26x10^{-01}$**	**0·939**	**$6{\cdot}19x10^{-01}$**	**0·975**	**$4{\cdot}43x10^{-01}$**	**0·923**	**$6{\cdot}25x10^{-01}$**	**0·944**	**$2{\cdot}08x10^{-02}$**	**0·736**
	Rheumatoid arthritis	*HLA-DRA/HLA-DRB5*	rs9268853 (32429643)	C	T	NA	**$4{\cdot}56x10^{-109}$**	**2·4**	32	0·308	0·8475	**$1{\cdot}43x10^{-01}$**	**0·936**	**$3{\cdot}77x10^{-01}$**	**0·951**	**$6{\cdot}15x10^{-01}$**	**0·45**	**$8{\cdot}68x10^{-01}$**	**0·978**	**$7{\cdot}86x10^{-02}$**	**0·770**
	MS	*HLA-DRA*	rs3135388 (32413051)	A	G	0·230	**$8{\cdot}94x10^{-81}$**	**1·99**	34	0·131	0·9734	**$4{\cdot}80x10^{-02}$**	**1·122**	**$2{\cdot}10x10^{-01}$**	**1·095**	**$1{\cdot}25x10^{-01}$**	**1·254**	**$5{\cdot}02x10^{-01}$**	**1·120**	**$6{\cdot}10x10^{-01}$**	**1·105**
			rs3129871 (32406342)	A	C	0·504	**$5{\cdot}70x10^{-15}$**	**1·72**	35	0·337	0·9379	**$3{\cdot}43x10^{-01}$**	**0·961**	**$3{\cdot}15x10^{-01}$**	**0·949**	**$4{\cdot}94x10^{-01}$**	**1·078**	**$8{\cdot}24x10^{-01}$**	**0·974**	**$2{\cdot}72x10^{-01}$**	**0·859**
	PD	*HLA-DRA*	rs3129882 (32409530)	G	A	0·450	**$1{\cdot}90x10^{-10}$**	**1·26**	36	0·456	0·9992	**$3{\cdot}36x10^{-02}$**	**1·086**	**$3{\cdot}27x10^{-02}$**	**1·106**	**$7{\cdot}52x10^{-01}$**	**1·033**	**$5{\cdot}74x10^{-01}$**	**1·065**	**$7{\cdot}07x10^{-01}$**	**1·049**

Table 4-17. Comparison between SNPs and loci associated with other conditions than FTD or other closely related conditions such as ALS, PSP/CBD and AD, and our study. P-values are shown for each different locus and OR is calculated for the same effect allele in all conditions and our study.

Overlap between FTD and ALS

The marker for the *C9orf72*/*MOB3B* locus, rs3849942, (*C9orf72*/*MOB3B* locus) [81] achieved a p-value of $2.12x10^{-06}$ and an OR of 1.957 in the FTD-MND subtype consistent with *post hoc* analyses (~23% of expansion carriers among this subtype) (**Tables 4-10** and **4-18**). In addition, rs3849942 reached p-values between 10^{-03} and 10^{-04} in bvFTD (p-value=$7.38x10^{-03}$, OR=1.155) and the entire discovery cohort (p-value=$4.38x10^{-04}$, OR=1.166), whilst in the SD or PNFA subtypes p-values were not relevant (**Table 4-17**). Overall, association at this locus in the FTD-GWAS seemed rather surprising. Considering that the expansion carriers were included in the analysis because this genetic variant was not yet known at the time analysis on the discovery cohort was performed, and given the high frequency that had been reported in FTD-ALS cases across different studies [76, 87], one would have expected a robust signal at least in the bvFTD and FTD-MND subtypes and, possibly, also in the meta-analysis. This was not the case if we exclude the FTD-MND subtype where a relatively high frequency among cases (~23%) and a sensible but not extraordinary large sample size (n=200) were not sufficient to lead to a signal with genome-wide significance. However, it is a possibility that if the FTD-MND subtype would have included a larger number of samples, increasing therefore power of analysis, association at this locus might have reached genome-wide significance. At the same time there is another possibility that should not be dismissed *a priori*, i.e. the presence of *C9orf72* expansion in neurologically normal controls [76, 77, 86, 284, 285] that may have lowered the signal.

Overlap between FTD-GWAS and previous FTLD-TDP-GWAS

The previous GWAS on FTLD-TDP [41] highlighted the *TMEM106B* locus as associated with the TDP-43 pathologically confirmed FTLD-TDP cases; such association was even stronger in the presence of concomitant *GRN* mutations [41]. Evaluating rs1990622 in the FTD-GWAS discovery cohort revealed that this SNP achieved p-values between 10^{-02} and 10^{-03} in the entire dataset (p-value=$7.88x10^{-02}$, OR=1.073) and in the bvFTD subtype (p-value=$5.85x^{-03}$, OR=1.144) (**Table 4-17**). It has to be noted however that the original study [41] was performed with autopsy-confirmed FTLD-TDP cases and the current FTD-GWAS was with clinically defined cases. In addition, the previous study included a number of *GRN* mutation carriers [41], whilst in our study these are excluded. Considering these differences and the fact that recent studies suggested *TMEM106B* to be directly related to *GRN* metabolism [26], data for our FTD discovery cohort may suggest our results being a partial or limited replication of the original finding.

4.2.2.6 Overlap with putative pleiotropic loci

The results of the association, replication and joint analyses indicated two clear and novel susceptibility loci at 11q14 and 6p21.3 for FTD. At this point, one question that arose was whether association at these loci had previously been identified in other conditions and if there was any possible relation between those conditions and FTD. All comparisons discussed in this section are summarized in (**Table 4-17**).

Chromosome 11q14

A suggestive association with the *RAB38* locus had been reported for rs1386330 (p-value=2·00x10^{-06}) in multiple sclerosis (MS) [286]. Although the same locus points in this case to another neurological disorder it is hard to interpret this as a definite overlap between FTD and MS. In fact, rs1386330 was not statistically significant in the FTD discovery dataset, despite an OR>1 in almost all categories. The SNPs that reached statistical significance in the FTD-GWAS locate in intron 1 (rs302652) and upstream (rs74977128) of *RAB38* whilst the SNP associated with MS (rs1386330) locates downstream of *RAB38* gene. This could possibly indicate that the two different disorders associate with the same locus but are characterised by two different haplotype substructures, which are tagged by rs1386330 in the case of MS and rs302652/rs74977128 in the case of FTD.

Chromosome 6p21.3

Compared to the *RAB38* locus, association with the *HLA* locus seemed to be more frequent and ubiquitous among other conditions. As such association with *HLA* was evident in a higher number of studies than in the case of *RAB38* (http://www.genome.gov/gwastudies/ accessed in the period June – September 2013). However, after careful assessment of the reported associated SNPs, only a handful of markers seemed potentially relevant.

First, SNP rs3806156 (p-value=$7x10^{-19}$, OR=1.42), located in intron 1 of *BTNL2*, was found associated with vitiligo, an autoimmune condition that causes depigmentation in melanocytes [287]. Second, SNP rs9268853 (p-value=$5x10^{-109}$, OR=2.4), intergenic between *HLA-DRA* and *HLA-DRB5*, was identified in a GWAS targeting rheumatoid arthritis [288]. Although these two forms of autoimmune disease were recently shown to be occasional co-morbidities in some FTLD cases [289], the two associated SNPs (rs3806156 and rs9268853) were not significant in the FTD discovery cohort.

Third, rs3135388 (p-value=$9x10^{-81}$, OR=1.99), located downstream of *HLA-DRA*, and rs3129871 (p-value=$6x10^{-15}$, OR=1.72), located upstream of *HLA-DRA*, were associated with MS [259, 290]. These SNPs were not significant in the FTD-GWAS dataset, however, it is interesting to note that again MS showed association with a locus that also is closely related to FTD. It follows that by critically considering that for both loci, *RAB38* and *HLA*, association was evident in two clinically distinct conditions (MS and FTD), the SNPs that did reflect the association were different. Especially, those that showed significance in one condition did not in the other and vice versa. This could mean that both loci are implicated in the pathogenic processes that lead to disease, but, at the same time, that the association is ultimately defined by different haplotype sub-structures that, eventually, exert their effect through diverse pathomechanisms.

Fourth and finally, rs3129882 (p-value=$2x10^{-10}$, OR=1.26), located in intron 1 of *HLA-DRA*, was shown being associated with PD [291]. In our dataset, rs3129882 did not show significant association (p-values<0.05, OR>1) again suggesting that the same locus may

underpin two different neurodegenerative diseases such as PD and FTD but through different markers that, in turn, lead to different pathomechanisms.

In summary, this section highlights that the extended loci that reached genome-wide significant association in the FTD discovery cohort had also been reported in a number of other conditions. Some of these fall under the category of neurodegenerative diseases (MS and PD) and others are autoimmune diseases (vitiligo and rheumatoid arthritis) that, interestingly, have also been shown to be occasional co-morbidities of FTD [289]. However, it is important to note that the markers driving the genetic association were different across the various conditions, for which it may be sensible to infer that variability for each marker ultimately contributes to the pathogenic effects leading to different diseased phenotypes through differential processes, pathways or mechanisms.

4.2.3 Discussion

Frontotemporal dementia is a complex disorder characterised by a broad range of clinical manifestations, differential pathological signatures and considerable genetic variability [17, 26, 166]. These features clearly imply to complex disease mechanisms. As such, many elements, including interplay between environmental and genetic risk factors, are likely to contribute to onset and progression.

To date there have been few studies of environmental factors in FTD. These pointed mainly to head trauma and thyroid disease [292] as well as traumatic brain injury (TBI)

[293] as factors that may possibly influence or be related to an increased risk of developing FTD, whereas cerebrovascular disease seemed not to influence or sufficiently associate with FTD [293]. However, despite a handful of reports, environmental factors are still understudied and, as such, not too well defined in FTD. On the other hand, genetic investigation has led to the identification of a number of candidate genes [26, 45, 67] [76], but ultimately only a number of mutations in these genes have proven to be pathogenic in FTD [47]. Currently, besides those cases carrying *MAPT* mutations that lead to tau pathology, there is still no clear and linear correlation between pathology, clinical syndromes and genetic variability. In addition, apart from a few highly penetrant pathological mutations in *MAPT* and *GRN* the other susceptibility genes seem to have lower penetrance and effect size, regardless of their frequency [26, 145, 294-296].

With this in mind, we performed the largest and most comprehensive genome-wide association study on frontotemporal dementia, to date, in the search for novel disease risk loci and modifiers associated with the disorder that could possibly also highlight pathways and mechanisms relevant to disease development and progression, as still not much is known about the underpinning molecular processes that influence the pathogenesis of FTD. After the completion of this project, we identified two novel loci associated with disease, specifically, the Chr11q14 locus, encompassing *RAB38*/*CTSC* for the bvFTD subtype, and the Chr6p21.3 locus, encompassing *BTNL2* and *HLA-DRA*/*DRB5* for the entire cohort, suggesting that FTD pathogenesis likely involves lysosomal/phagosomal pathways (link to Chr11q14) and immune system processes (link to Chr6p21.3).

All the SNPs that met genome-wide significance (p-values<$5x10^{-08}$) were non-coding. Although it is well known that the SNPs reflecting the association signal may not be the actual risk causing variants, it is relevant to note that either intronic or intergenic SNPs may underlie the reason of association especially by influencing expression or splicing events. In our study for example, expression and methylation quantitative trait loci data (e/mQTL) did support the notion that risk at these loci may be driven by *cis* changes of expression and methylation levels and, therefore, likely through regulation of transcription processes.

Perhaps, however, the most important outcome of our study is that the association with these two loci provides, prospectively, novel and unprecedented insight into the pathobiology of FTD.

4.2.3.1 The chromosome 11q14 locus

The first gene of interest in this associated locus is *RAB38*, an oncogene that was reported being mutated in melanoma [297] and that encodes the transmembrane protein RAB38 which is expressed in thyroid, elements of the immune system such as whole blood, lymph nodes and immune cells (dentritic cells, NK cells, T/B cells), and, most importantly, the brain (http://www.genecards.org/cgi-bin/carddisp.pl?gene=RAB38). From a functional perspective, RAB38 has been suggested to be involved in two main cellular processes:

1 – Mediation of protein trafficking to lysosomal-related organelles within the trans-Golgi network (TGN) [298, 299], and;

2 – Maturation of phagosomes that envelop pathogens [300].

It was speculated that impairment of these pathways may lead to cargo accumulation in the early endosomes with downstream effects on other pathways such as the recycling pathway (toward the plasma membrane) and/or the degradative pathway (affecting late endosome/multivesicular body [MVB]) [298].

The second gene of interest in this locus is *CTSC*, a lysosomal cysteine-proteinase that participates in the activation of serine proteinases in immune/inflammatory cells that are involved in immune and inflammatory processes including phagocytosis of pathogens and local activation/deactivation of inflammatory factors (e.g. cytokines) (OMIM: #602365).

The associated SNP at the *RAB38/CTSC* locus, rs302652, showed a remarkable eQTL in monocytes ($p=5.05x10^{-32}$ [269]) for the Illumina probe ILMN_2134974 with the risk allele T being associated with decreased expression of *RAB38*. These data suggest that a decreased function of RAB38 might contribute to pathogenesis and may be the mechanism by which the association at this locus is mediated. Despite the fact that one should be cautious at interpreting expression data generated in a tissue which is different from the tissue of interest, in this case the data are extremely relevant since monocytes and microglial cells are both cells of the innate immune system and both derive from

myeloid progenitors [301]. It is interesting to note that both RAB38 and CTSC, based on their intracellular location and functions, correlate with lysosomal as well as phagosomal biology. An association with lysosomal processes in FTD was hinted by two studies on the *GRN* [302] and *TMEM106B* [150] genes. In fact, PGRN not only is ubiquitously expressed [62], but also it is involved in a variety of diverse biological processes [63-65]. Among these, PGRN is involved in regulating lysosome functions, and TMEM106B, which locates intracellularly to the late endosome/lysosome compartments, appears to regulate PGRN levels [150]. Considering that endolysosomal homeostasis is essential for the health of neurons, the link between the two proteins, RAB38 and CTSC, and FTD (in addition to the one of TMEM106B and PGRN) suggests that autophagosomal/lysosomal dysfunctions may play a critical role and be implicated in the onset and/or progression of the disease.

4.2.3.2 The chromosome 6p21.3 locus

The first genes of interest in this associated locus are *HLA-DRA* and *HLA-DRB* that, respectively, encode monomorphic and polymorphic class II HLA-DR transmembrane chains, which are expressed on the surface of antigen presenting immune cells. Microglial cells are scavenger cells that are involved in several processes in the brain such as modulation of immune responses (including phagocytosis) as well as cytokines and growth factor secretion [303]. The HLA-DR molecules are known to be expressed on the surface of microglia, and it has been suggested that increased expression of HLA-DR

molecules on microglia may reflect pathological activity, as previously reported, for example, in AD and PD [304]. From a functional perspective, the HLA-DRA/DRB molecules promote interaction between microglial cells and CD4+ T-cell receptors allowing for modulation and regulation of immune responses (OMIM#142860).

The second gene of interest in this locus is *BTNL2*, which encodes a membrane protein that is ubiquitously expressed across different tissues, including the brain, and is involved in repressing T-cells proliferation [305].

Our own mQTL data and a couple of previous reports support the possibility of a link between this locus and effects on transcription [221, 291]. For example, our data reveal an association between the risk allele T of SNP rs1980493 (located in *BTNL2*), and increased methylation at *HLA-DRA* in cerebellum (p-value=$3.34x10^{-08}$) and frontal cortex (p-value= $2.17x10^{-08}$), as well as a suggestive association (p-value=$4.8x10^{-07}$) with increased *cis* expression of the gene *SLC39A7*. The latter gene encodes the protein SLC39A7, which is a cytoplasmic zinc transporter. The potential pathogenic effect of an increased expression of the SLC39A7 protein in neurodegenerative conditions is not known and needs to be fully elucidated before any speculation could be made. However, these transcription-related mechanisms highlight the likely functional reason for association at this locus. Besides the fact that the exact mode of action by which variability at this locus might exert its pathogenic effect needs to be further investigated, the valuable message here is that the association with *HLA* strongly supports a tie between FTD and the immune system.

Taking one step back, the immune system is known to be fundamental in modulating several processes in the central nervous system (CNS). For example, in normal conditions, microglial cells:

1 – Play an important role during brain development by pruning neurons, and;

2 – Maintain CNS homeostasis through removal of either debris and apoptotic cells or pathogens via phagocytosis [306, 307].

Considering that:

1 – The class II HLA-DR molecules are expressed on the surface of microglial cells, and;

2 – Microglia interact with CD4+ T cells, it follows that microglia themselves play a key role in regulating immune responses and homeostasis in brain.

As such, these notions offer insight into a possible role for the *HLA* locus in FTD pointing to the fact that variability at this locus may lead to aberrant/detrimental immune responses in the brain and, ultimately, to neurodegeneration.

4.2.3.3 The immune system and neurodegeneration

Ageing causes changes in both the CNS and the immune system. There is evidence in the literature for a link between changes in immune responses (that happen during the process of aging) and an increased incidence of neurodegenerative diseases [306, 307].

Specifically, the innate and the adaptive immune responses have been suggested to be involved in processes that lead to neurodegenerative diseases [306] including AD [308], PD [291, 309], ALS [310], MS [286, 311] and Huntington disease (HD) [312].

Our current study points to a possible role of the immune system in the pathogenesis of FTD adding this neurodegenerative disease to the above mentioned list.

It is noteworthy that brain homeostasis is guaranteed by the harmonic interplay of microglial cells, regulatory T and B cells and efficient cross talk between neurons and microglia [306]. Ageing, stress as well as infection at site or peripheral inflammation have been reported as factors that can perturb brain homeostasis by possibly leading to an excessive or aberrant immune response in the brain that, in turn, may lead to neuroinflammation and neurodegeneration [306, 312]. Some reports have shown that increased vulnerability to infection in later age may associate with behavioural and cognitive deficits [313]. For example, peripheral infections have been associated with deterioration of cognition in AD patients [306].

However, to date, it is still not clear whether inflammation plays a primary or secondary role in neurodegenerative diseases.

Microglia are activated in areas of pathology as shown in substantia nigra and areas where dopaminergic neurons die in PD [314] or within and around Aβ plaques in AD [312]. In addition, positron emission tomography (PET) imaging studies showed that such activation is visible in pre-manifest gene carriers; therefore, it is a possibility that

such an event (microglia activation) is not just secondary to large scale neuronal death [315], and so microglial chronic activation was suggested having a primary role in triggering neurodegenerative diseases and their progression [307].

Another important aspect is that inflammatory markers accumulate around the pathological lesions especially in the brain areas where synaptic or neuronal loss occur [312]. This was observed in a number of neurodegenerative diseases, including FTD [316]; it was shown that accumulation of inflammatory markers in areas of brain pathology is reflected, partially, by an increase of cytokine levels in the cerebrospinal fluid (CSF) [292] and/or in the plasma [312]. For example, elevated plasma levels of cytokines and chemokines such as Interferon γ (INF-γ) or tumour necrosis factor α (TNF-α) were shown to correlate with cognitive decline in AD [317], and, furthermore, increase of TNF-α in CSF was also observed in FTD [318]. More recently, an elegant study, evaluating correlation between thyroid and non-thyroid autoimmune disorders and two groups of patients either diagnosed with SD or carrying *GRN* mutations, established a correlation between non-thyroid autoimmune diseases – including rheumatoid arthritis and vitiligo – and those two groups of patients, including elevated TNF-α plasma levels [289].

All this taken together leads to the notion that the immune system may act as a disease modifier in neuropathogenesis and that systemic immune changes may affect disease onset and progression in neurodegeneration [312].

4.2.4 Summary and conclusion

The associations identified in our study at Chr11q4 – encompassing *RAB38* and *CTSC* – and at Chr6p21.3 – encompassing *BTNL2*, *HLA-DRA* and *HLA-DRB5* – indicate genes involved in lysosomal/phagosomal as well as inflammatory processes to be likely involved in the pathogenesis of FTD.

After completion of this study, there are, possibly, three important points to be highlighted:

1 – The Chr11q4 locus has previously been associated with another neurodegenerative disorder, namely MS [286], a disease that is also associated with inflammation [319] [259, 290]. This possibly suggests shared pathogenic pathways in FTD and MS. However, the association with this locus for each disorder is driven by two different SNPs, probably implying to two distinctive disease-specific haplotypes;

2 – Association with the Chr6p21.3 locus had been previously reported in MS and PD [291, 319] as well as in the autoimmune diseases vitiligo and rheumatoid arthritis [287, 288]. Interestingly, the latter conditions were recently shown to correlate with either the diagnosis of SD or the presence of *GRN* mutations in concomitance with elevated TNF-α plasma levels [289]. It has been previously implied in the literature that deficiencies in the immune system trigger the development of neurodegenerative diseases [306]. The association with FTD that we report in this study provides further genetic evidence and support for this hypothesis, and;

3 – At this stage, considering that both RAB38 and CTSC are involved in lysosomal biology and, to a certain extent, in immune processes one might speculate that lysosomal/phagosomal pathways and immune system processes may individually as well as synergistically play a critical role in the development of FTD.

Undoubtedly, the new loci presented in this work will need to be replicated in other FTD cohorts and further investigated in order to establish a link between the genetic association and biological processes underlying disease.

These findings however hold promise for a better understanding of the pathogenesis of FTD and for the development of tools to be implemented for preventive and therapeutical measures.

The work presented in this chapter was performed by the following individuals (or groups):

Raffaele Ferrari: Contacted all PIs and research groups that participated in the study, coordinated sample collection at UCL, TTUHSC and NIH, performed material QC for discovery & replication phases, coordinated and supervised genotyping of cases and controls at UCL for discovery & replication phases, collected all clinical information for each collected sample (UCL+TTUHSC+NIH), participated in genotyping data QC, assisted Dr Michael Nalls in the analysis phase, performed LOH and candidate loci analyses; **Michael Nalls**: performed genotyping QC, genetic statistical and association analyses for discovery, replication and joint phases; **Jonathan Rohrer**: designed the clinical characterization datasheet (**Appendix 2-2**) and reviewed the clinical phenotypic data from the cohort with Raffaele Ferrari; **UK Brain Expression Consortium** (UKBEC) and **North American Brain Expression Consortium** (NABEC) (**Appendix 2-5**): Generated expression and methylation quantitative trait loci data, respectively.

A manuscript derived from the work presented in this chapter has been accepted for publication in Lancet Neurology:

Raffaele Ferrari, Dena G Hernandez, Michael A Nalls, Jonathan D Rohrer, Adaikalavan Ramasamy, John BJ Kwok, Carol Dobson-Stone, William S Brooks, Peter R Schofield, Glenda M Halliday, John R Hodges, Olivier Piguet, Lauren Bartley, Elizabeth Thompson, Eric Haan, Isabel Hernández, Agustín Ruiz, Mercè Boada, Barbara Borroni, Alessandro Padovani, Carlos Cruchaga, Nigel J Cairns, Luisa Benussi, Giuliano Binetti, Roberta Ghidoni, Gianluigi Forloni, Daniela Galimberti, Chiara Fenoglio, Maria Serpente, Elio Scarpini, Jordi Clarimón, Alberto Lleó, Rafael Blesa, Maria Landqvist Waldö, Karin Nilsson, Christer Nilsson, Ian RA Mackenzie, Ging-Yuek R Hsiung, David MA Mann, Jordan Grafman, Christopher M Morris, Johannes Attems, Timothy D

Griffiths, Ian G McKeith, Alan J Thomas, Pietro Pietrini, Edward D Huey, Eric M Wassermann, Atik Baborie, Evelyn Jaros, Michael C Tierney, Pau Pastor, Cristina Razquin, Sara Ortega-Cubero, Elena Alonso, Robert Perneczky, Janine Diehl-Schmid, Panagiotis Alexopoulos, Alexander Kurz, Innocenzo Rainero, Elisa Rubino, Lorenzo Pinessi, Ekaterina Rogaeva, Peter St George-Hyslop, Giacomina Rossi, Fabrizio Tagliavini, Giorgio Giaccone, James B Rowe, Johannes CM Schlachetzki, James Uphill, John Collinge, Simon Mead, Adrian Danek, Vivianna M Van Deerlin, Murray Grossman, John Q Trojanowski, Julie van der Zee, William Deschamps, Tim Van Langenhove, Marc Cruts, Christine Van Broeckhoven, The Belgian Neurology Consortium and the European Early-Onset Dementia consortium, Stefano F Cappa, Isabelle Le Ber, Didier Hannequin, Véronique Golfier, Martine Vercelletto, Alexis Brice, The French research network on FTLD/FTLD-ALS, Benedetta Nacmias, Sandro Sorbi, Silvia Bagnoli, Irene Piaceri, Jørgen E Nielsen, Lena E Hjermind, Matthias Riemenschneider, Manuel Mayhaus, Bernd Ibach, Gilles Gasparoni, Sabrina Pichler, Wei Gu, Martin N Rossor, Nick C Fox, Jason D Warren, Maria Grazia Spillantini, Huw R Morris, Patrizia Rizzu, Peter Heutink, Julie S Snowden, Sara Rollinson, Anna Richardson, Alexander Gerhard, Amalia C Bruni, Raffaele Maletta, Francesca Frangipane, Chiara Cupidi, Livia Bernardi, Maria Anfossi, Maura Gallo, Maria Elena Conidi, Nicoletta Smirne, Rosa Rademakers, Matt Baker, Dennis W Dickson, Neill R Graff-Radford, Ronald C Petersen, David Knopman, Keith A Josephs, Bradley F Boeve, Joseph E Parisi, William W Seeley, Bruce L Miller, Anna M Karydas, Howard Rosen, John C van Swieten, Elise GP Dopper, Harro Seelaar, Yolande AL Pijnenburg, Philip Scheltens, Giancarlo Logroscino, Rosa Capozzo, Valeria Novelli, Annibale A Puca, Massimo Franceschi, Alfredo Postiglione, Graziella Milan, Paolo Sorrentino, Mark Kristiansen, Huei-Hsin Chiang, Caroline Graff, Florence Pasquier, Adeline Rollin, Vincent Deramecourt, Florence Lebert, Dimitrios Kapogiannis, UK Brain Expression Consortium, North American Brain Expression Consortium, Luigi Ferrucci, Stuart Pickering-Brown, John Hardy, Parastoo Momeni and Andrew B Singleton. ***Genome-wide association study reveals lysosomal and immune system involvement in frontotemporal dementia***. Lancet Neurology (accepted for publication on May 6th, 2014).

CHAPTER 5 – Other FTD related projects

During the course of this research, in addition to the two main projects, i.e. the follow up study on the PSP-GWAS [180] (**Chapter 3**) and the entire FTD-GWAS (**Chapter 4**), I have actively participated in a number of other research projects.

These have mainly, but not exclusively, focused on the study of FTD. In fact, especially, at TTUHSC under my supervisor Dr Momeni, I have also been working on the screening of the known candidate genes for FTD (**Table 5-1**) in cases diagnosed with corticobasal syndrome/degeneration (CBS/CBD) [320]. In addition, I have been involved not only in the screening of the Alzheimer's disease (AD) risk factors and candidate genes (**Table 5-1**) in large AD cohorts [321], but also in an innovative and ongoing study aimed at untangling the gender differences in the incidence and prevalence of AD [322].

In the following sections I will exclusively summarize the study of FTD families and the screening of the FTD candidate genes in our extended cohort of sporadic FTLD and CBS/D cases at TTUHSC, reason being that these research projects are of great interest in that they provide direct insight into the complexity of FTD and its genetic heterogeneity.

Table 5-1. Candidate genes for genetic screening

Gene	Full name	Chromosome	Location	BP	Associated Disease	Reference
MAPT	microtubule-associated protein tau	17	17q21.1	43971702 - 44105700	FTD	[45]
					CBS/D	[180]; [179]; [323]; [324]
					PSP	[180]; [179]
					AD	[325]; [52]; [326]
GRN	granulin	17	17q21.32	42422491 - 42430474	FTD	[66]; [67]
					CBS/D	[327]; [328]; [329]; [330]
C9orf72	chromosome 9 open reading frame 72	9	9p21.2	27546543 - 27573864	FTD-ALS	[76]; [77]
TDP-43	TAR DNA binding protein	1	1p36.22	11072462 - 11085549	ALS	[331]
					FTD	[136]
FUS	fused in sarcoma	16	16p11.2	31191431 - 31206192	ALS	[142]; [143]
CHMP2B	charged multivesicular body protein 2B	3	3p11.2	87276413 - 87304698	FTD	[121]
ApoE	apolipoprotein E	19	19q13.2	45409039 - 45412650	AD	[332]
APP	amyloid beta (A4) precursor protein	21	21q21.3	27252861 - 27543446	AD	[333]
PSEN1	presenilin 1	14	14q24.3	73603143 - 73690399	AD	[334]
PSEN2	presenilin 2	1	1q31-q42	227057885 - 227083804	AD	[335]
CR1	complement component (3b/4b) receptor 1	1	1q32	207669473 - 207815110	AD	[336]
CLU	clusterin	8	8p21-p12	27454434 - 27472328	AD	[336]; [337]
PICALM	phosphatidylinositol binding clathrin assembly protein	11	11q14	85668214 - 85780923	AD	[337]

5.1 Familial cases

5.1.1 A South African FTD familial case

5.1.1.1 Introduction

The Afrikaaner family is a South African kindred of Dutch ancestry whose members affected by FTD appear to inherit the disease in an autosomal dominant fashion. Affected and unaffected family members were collected by Drs. Rajesh Kalaria and Felix Potocnik through a collaborative effort between the Institute for Ageing and Health at the Newcastle University, UK and the University of Stellenbosch in South Africa. All members of the family participating in this longitudinal study were consented according to the ethics committee of the University of Newcastle prior to enrolment.

5.1.1.2 Clinical and pathological features

The Afrikaaner family is a multigenerational pedigree (**Chapter 2**, **Figure 2-2**) with early onset FTD as the affected individuals consistently develop the disease in their late 40s or early 50s. The clinical phenotype is marked by executive dysfunction, social disinhibition and dementia, being alcohol dependency and violent behaviour the main early signs in most of the affected members. Pathological investigation was performed on post-mortem brain tissues of the consented affected individuals that went to autopsy and

revealed widespread tauopathy involving both the grey and white matter in three affected siblings (numbers 19, 20 and 22 in the pedigree).

5.1.1.3 Genetic screening of the candidate genes

The DNA material of the affected and unaffected members of this family was extracted from blood either at the University of Newcastle or at the National Institute on Ageing (NIA) using standard procedures (**Chapter 2**, section ***2.2.1***).

The DNA has been screened for mutations in the candidate genes *MAPT*, *GRN*, *CHMP2B* (at the NIA by Dr Momeni) [125], and *FUS* and *TDP-43* (at TTUHSC by me) [137]. In addition, I performed gene dosage experiments for *MAPT* at TTUHSC.

There was no evidence of abnormal copy number variations (CNVs) in *MAPT*. In addition, sequencing experiments did not reveal any pathogenic variants in *MAPT*, *GRN* or *TDP-43* [125, 137]. On the contrary, sequencing of the *CHMP2B* and *FUS* genes did lead to unexpected findings.

CHMP2B

The screening of *CHMP2B* revealed a truncation mutation, Arg186X due to the base change C>T in exon 6 of *CHMP2B* (**Figure 5-1**), in two unaffected siblings (individuals 110 and 111 on the pedigree) (**Figure 5-2**). Interestingly, this mutation was not carried by

their affected father (number 22 on the pedigree) or by any other affected family member [125].

Figure 5-1. *CHMP2B*: Chromatogram of Arg186X

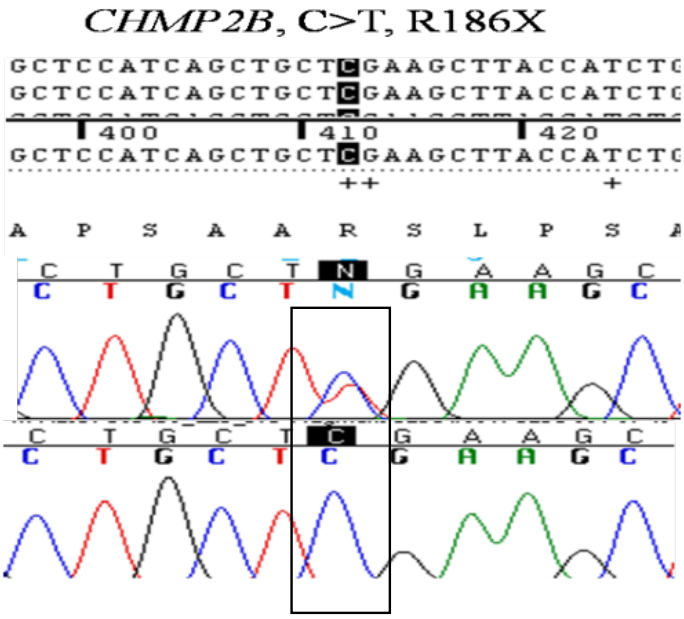


Figure 5-1. The chromatogram shows the base change C>T causing the non-sense change Arg186X in *CHMP2B* [125].

Figure 5-2. *FUS* and *CHMP2B* mutations in the Afrikaaner pedigree

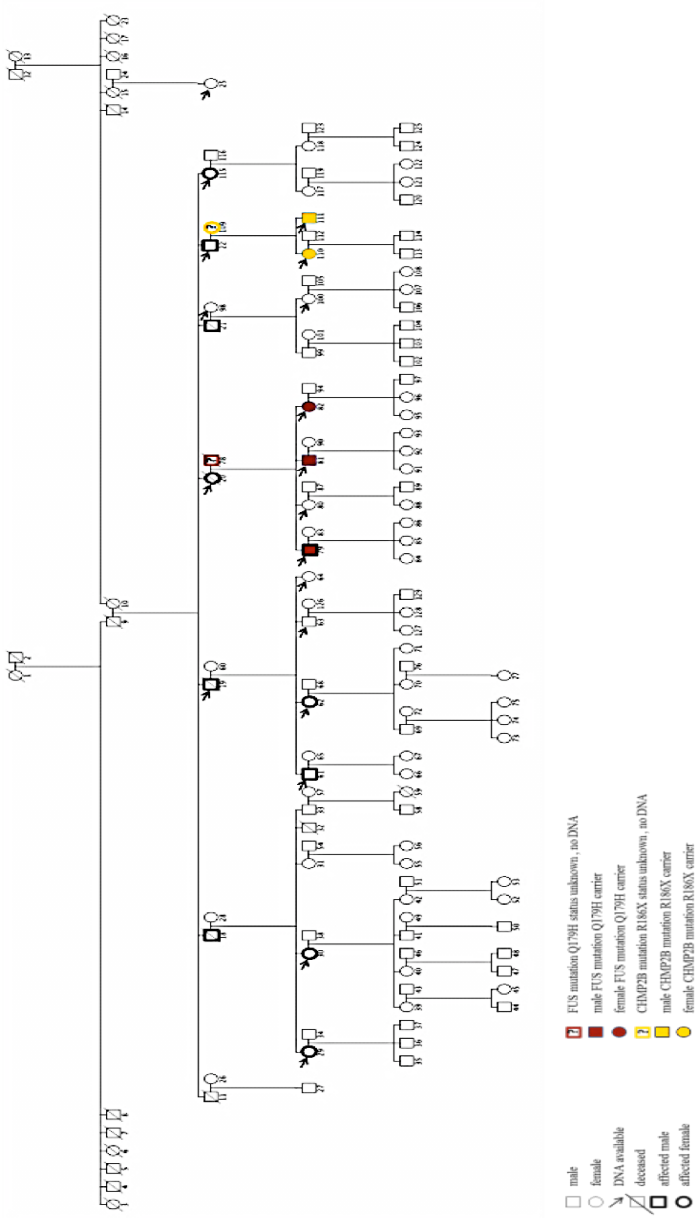


Figure 5-2. Pedigree of the Afrikaner family identifying those individuals for which DNA samples are available for screening (arrows). Carriers of the *FUS* and *CHMP2B* mutations are highlighted in brown and yellow, respectively.

It is important to note that, since the Arg186X mutation was not carried by the affected father, it must have been inherited from the unaffected mother (number 109 in the pedigree, **Figures 5-2**) who died at the age of 68 years without any report of dementia or neurological dysfunctions. In addition, based on the available family history, the parents and 3 siblings of the unaffected mother were not reported having ever had signs of neurological impairment [125]. These findings are of interest when put into context. In fact, a previous study by Skibinski and colleagues [121] reported a G>C transition in the acceptor splice site of exon 6 of *CHMP2B* that co-segregated in 11 affected family members and that was projected to lead to two aberrant transcripts: $CHMP2B^{intron5}$ and $CHMP2B^{\Delta 10}$ [121]. These three variants, intron5, Δ10 and Arg186X (circled in red in **Figure 5-3**), locate at the carboxy-terminus of the protein CHMP2B and cause the disruption of the microtubule interacting and transport (MIT)-interacting motif (MIM).

Figure 5-3. Wild-type and mutated CHMP2B protein primary structure

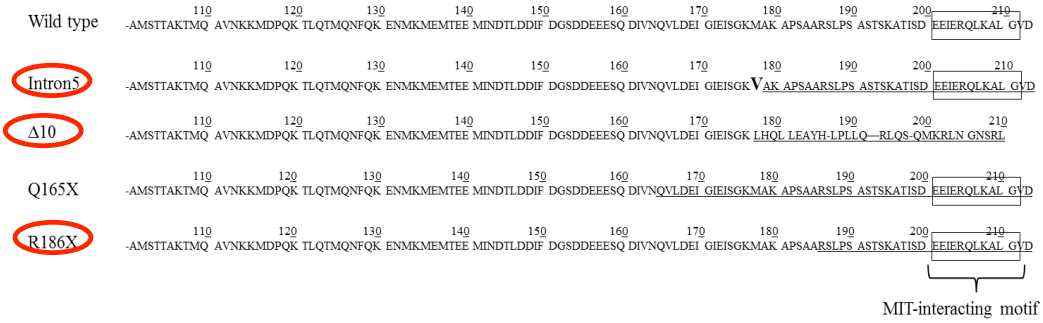


Figure 5-3. All the known mutations leading to a carboxy-terminal truncation phenotype are shown and compared to the wild-type protein sequence of CHMP2B (top). The carboxy-terminal protein sequence that is lost is underlined. These truncation mutations determine the loss of the MIT-interacting motif (MIM).

The MIM is a fundamental functional domain of the CHMP2B protein that recruits the vacuolar protein sorting 4 homolog A protein (VPS4), which is an ATPase and is involved in intracellular protein trafficking [338] (further discussed in section *5.2.1.2*). Clearly, the loss of the MIM domain would be expected to impact the ability of CHMP2B in recruiting VPS4 and, in turn, the cytosolic protein/vesicular trafficking machinery [339]. In the attempt to interpret the relevance of the Arg186X mutation one needs to keep in mind that:

1 – The two Afrikaaner family members who carry this truncation mutation are currently 56 (individual 110) and 54 (individual 111) years old and, still, unaffected, and;

2 – The mutation was inherited from the unaffected mother.

Therefore, it is plausible, at this stage, to infer that Arg186X may be a non-pathogenic variant, and that the phenotype of this family may be independent from genetic variability in *CHMP2B*. In addition, since the mutations intron5, Δ10 and Arg186X are expected to similarly affect the CHMP2B protein because of their location, one might speculate that not all mutations leading to the disruption of the MIM domain are pathogenic or fully penetrant. A more comprehensive and critical assessment of the variability in *CHMP2B* and, especially, of the effects of the carboxy-terminal truncation mutations is elaborated in section *5.2.1.2*.

FUS

Screening of the *FUS* gene in members of the Afrikaaner family also led to surprising results. At first, the novel Gln179Leu missense mutation (**Figure 5-4**) was identified in one affected individual (number 79 in the pedigree) (**Figure 5-2**) [137]. However, this variant was absent from other affected members of the family. In the attempt to better understand its origin and pattern of inheritance, we screened the DNA of the affected mother (number 20 on the pedigree) and the siblings (numbers 80, 81 and 82 on the pedigree) (**Figure 5-2**). The mother who died of FTD at the age of 74 years did not carry the variant, whilst two of the three siblings (numbers 81 and 82 on the pedigree) who are still asymptomatic did carry the variant.

Figure 5-4. *FUS*: Chromatogram of Gln179Leu

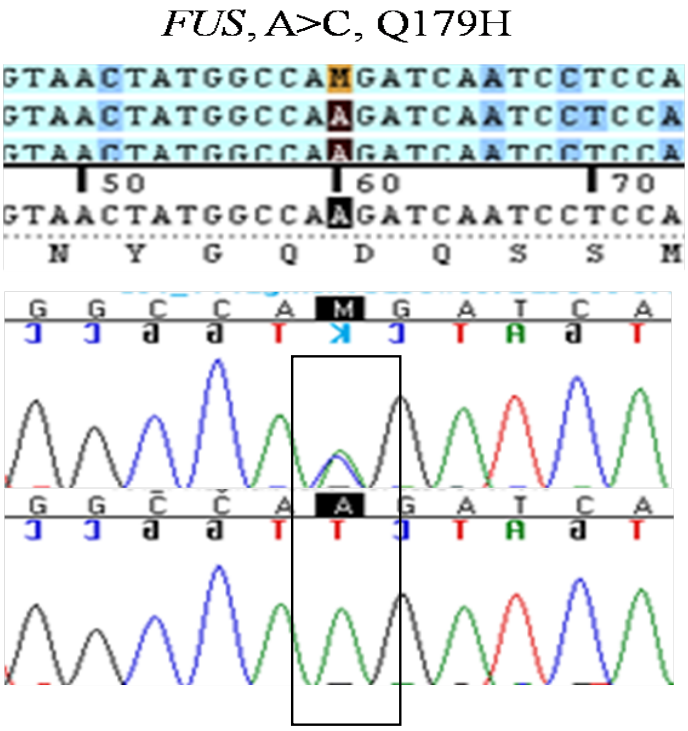


Figure 5-4. The chromatogram shows the base change A>C in exon 6 of the *FUS* gene causing the missense mutation Gln179Leu [137].

The DNA from the unaffected father (number 78 on the pedigree), whose medical history shows no evidence of any neurodegenerative condition, was not available to test for segregation. Nevertheless, since the affected mother did not carry the Gln179Leu missense change, the unaffected father must be the obligate carrier. Therefore, although:

1 – PolyPhen-2 scores this missense change as possibly damaging (**Figure 5-5**) [225], and;

2 – We did not isolate Gln179Leu in 659 neurologically normal controls;

Due to its presence in unaffected family members (unaffected father [78] and unaffected siblings [81 and 82]) and its mode of inheritance, we assume that this is rather a rare non-pathogenic variant [137].

Figure 5-5. *FUS*; missense change Gln179Leu. PolyPhen-2 analysis

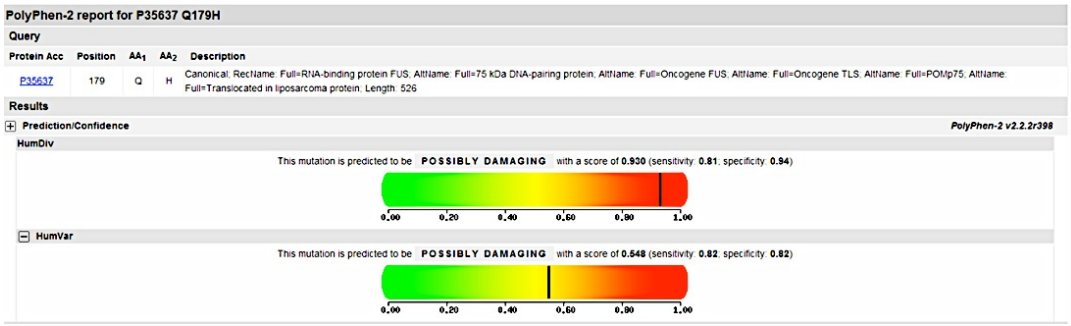


Figure 5-5. Polyphen-2 predicted the variant Gln179Leu being possibly damaging.

Taking these results together, the Afrikaaner family clearly resembles a complicated FTD familial case characterised by:

1 – Early onset;

2 – Autosomal dominant mode of inheritance;

3 – Tau pathology and absence of ubiquitin, TDP-43, FUS or p62 positive inclusions;

4 – Absence of variants in *MAPT, GRN* and *TDP-43*, and;

5 – Controversial variability and inheritance pattern for *CHMP2B* and *FUS*.

5.1.1.4 Summary

In conclusion, the screening of the known candidate genes did not reveal any pathogenic variant associated with the diseased condition, leaving the question open for the genetic component(s) underlying this familial case. To proceed with the investigation of this complex FTD kindred the next steps will entail a linkage analysis and exome sequencing. In the case of the first approach I have already generated genome-wide data (**Chapter 4**, section *4.1.2.3*) and together with the extended team who is working on this project, we have identified the cases that will be used for the linkage study (**Table 5-2**). For the second endeavour a number of samples (either n=4/6 [2/3 affected vs. 2/3 unaffected] will be chosen to be screened by means of next generation sequencing tools.

Table 5-2. List of Afrikaaner family members for linkage analysis

ID	Status	Notes	Age in 2010
Sample 1	A	Dementia	64
Sample 2	A	Cognitive impairment	62
Sample 3	A	Cognitive impairment	55
Sample 4	A	Cognitive impairment	48
Sample 5	U		37
Sample 6	A	Cognitive impairment	60
Sample 7	U		56
Sample 8	A	Dementia	71
Sample 9	U		54
Sample 10	A	Cognitive impairment	57
Sample 11	U		50
Sample 12	U		53
Sample 13	U		49
Sample 14	U		50
Sample 15	A	Dementia	62

Table 5-2. List of the 15 samples of the Afrikaaner family to be included in the linkage analysis. A=affected; U=unaffected.

5.1.2 A Finnish FTD kindred

5.1.2.1 Introduction

During the discovery phase of the FTD-GWAS I genotyped nine DNA samples of a Finnish family (**Chapter 4**) that I received from Dr Myllykangas, a collaborator from the Department of Pathology, Haartman Institute University of Helsinki, Finland. Five members (1–5) are siblings, two of which (1 and 2) were diagnosed with FTD and are now deceased (**Chapter 2**, **Figure 2-3**). Hereafter the clinical, pathological and genetic characterisation of the affected individuals of this interesting FTD kindred is reported.

5.1.2.2 Clinical assessment

Dr Auli Verkkoniemi-Ahola performed the clinical analysis. Patient 1 was diagnosed at an early stage of the disease, whilst patient 2 was seen at a late stage of dementia when already hospitalized (with medical records available from 1979 to 2002). Of the 3 further siblings, one was suspected of being a possible FTD case (3) and two were unaffected (4 and 5) (**Chapter 2**, **Figure 2-3**). Neuropsychological testing was performed for both index patients.

The first signs seen in patient 1 were memory and orientation problems, and cerebellar symptoms such as ataxia, at age 55. Patient 1 showed a decline of cognitive functions (executive functions and attention) as well as visuospatial, memory and language

impairment. Patient 1 also had difficulties in writing, counting and learning, decreased initiative and lack of insight. In addition, hypomimia, reduced and slow speech, and bradykinesia of both upper and lower extremities were also evident. However, the minimental state examination (MMSE) was 28/30. Patient 1 died at 56 years of age due to pneumonia. Conversely, early signs in patient 2 were slowness of movements, reasoning and speech, and social avoidance at age 54. Patient 2 showed broad cognitive decline, disturbances in executive and language functions (aphasia), poor attention, and memory dysfunctions. Also extrapyramidal signs such as bradykinesia and reduced accessory movements were present. In the final phases of disease, the patient was hypomimic and unable to speak. The MMSE score was 14/30 at that stage. The patient died at 62 years of age due to bronchopneumonia.

5.1.2.3 Brain imaging and pathology

Both structural magnetic resonance imaging (MRI) and functional imaging using single-photon emission computed tomography (SPECT) were performed for patient 1, whilst only computed tomography (CT) was available for patient 2. The MRI of patient 1 showed general cerebral atrophy and a calcified tumour in the left cerebellar lobe, diagnosed as a dysplastic gangliocytoma (**Figure 5-6**), whilst SPECT showed hypoperfusion of both frontal lobes.

Figure 5-6. Finnish FTD family index patient: MRI

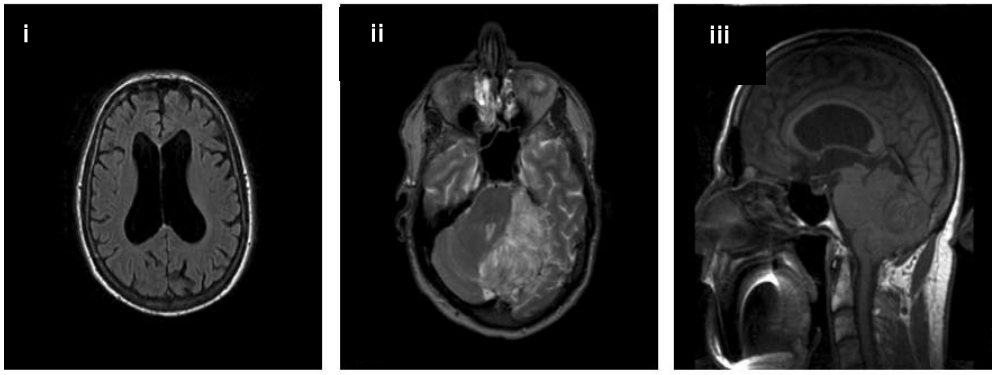


Figure 5-6. The MRI from the index patient 1 shows a slightly thinned frontal cortex and dilated lateral ventricles. (**i**) cerebellar tumor (axial view). Axial (**ii**) and sagittal (**iii**) view of the tumour mass located in the left cerebellar hemisphere [219]. Thanks to Dr Myllykangas' extended research group for this figure.

The CT of patient 2 only revealed general central and cortical atrophy.

Dr Myllykangas' extended research group investigated the brain pathology. In patient 1 the frontal cortex and the hippocampi appeared atrophic and the left lateral and third ventricles were enlarged. The thalamus and the brain stem were reduced in size, and the calcified cerebellar tumour (dysplastic gangliocytoma) obstructed the fourth ventricle. In patient 2 the atrophy was severe in the frontal lobes and in the striatum. It was less severe in the anterior and lateral parts of the temporal lobes, in the hippocampus and brain stem, whilst it was absent from the cerebellum and the spinal cord. Lastly, the substantia nigra was degenerated.

Concerning the molecular pathology, in patient 1 cytoplasmic pTDP-43-positive inclusions were seen in the dentate fascia, pyramidal neurons of the hippocampus and in all layers of the cortical areas, whilst p62-positive and TDP-43-negative cytoplasmic and intranuclear inclusions were found in the granular cells of the cerebellum. In patient 2 there were cytoplasmic pTDP43-positive inclusions and abundant dystrophic neurites in the hippocampus and all layers of the cortical areas. The granular cells of the cerebellum had cytoplasmic and intranuclear p62-positive and TDP-43-negative inclusions, whilst the substantia nigra showed p62- and TDP-43-positive neuronal cytoplasmic inclusions.

5.1.2.4 Genetic screening

Analysis of the DNA was performed on the relevant family members, namely, the 2 index patients (1 and 2) and the 3 further siblings (3, 4 and 5) (**Chapter 2**, **Figure 2-3**). All subjects gave informed consent for research.

First, since these samples were included in the FTD-GWAS (**Chapter 4**) to generate genome-wide genotyping data, Dr Myllykangas and colleagues performed a linkage analysis with the goal of identifying disease specific locus/loci. Second, I performed the sequencing for four FTD candidate genes (*MAPT*, *GRN*, *TDP-43* and *C9orf72*) in all 5 siblings and I screened the phosphatase and tensin homolog gene (*PTEN*) a tumour suppressor associated with the development of hamartomas (including gangliocytomas that was observed in patient 1).

Linkage analysis

Dr Myllykangas and her group performed linkage analysis of this kindred using the Merlin program (http://www.sph.umich.edu/csg/abecasis/Merlin/tour/linkage.html). I did not have access to the actual linkage data. However, I was informed about the outcome of the linkage analysis via personal (email) communication. The linkage analysis revealed several chromosomal areas with low LOD scores. Values of 0.6020 were seen in recessive model and 0.3010 in dominant model, the most interesting of which was found on chromosome 1 in proximity of the *TDP-43* locus. The maximum LOD score in this peak (0.6020 in recessive and 0.3010 in dominant model) was found 9 Mb centromeric from *TDP-43* locus. Also, a peak with max LOD score 0.931 was observed in the region encompassing the *C9ORF72* gene in 9p21.2.

Sequencing

We decided to screen the candidate genes *MAPT*, *GRN* and, especially, *TDP-43* and *C9orf72* based on the linkage results (although the LOD scores were relatively low) and the type of pathology identified in the two index patients.

The candidate genes *MAPT* (exons 1, 9, 10, 11, 12 and 13 and intronic flanking regions), *GRN* and *TDP-43* (all exons and intronic flanking regions) were sequenced following standard methods (**Chapter 2**, section *2.2.2.4*) and the repeat expansion in *C9orf72* was assessed by repeat-primed polymerase chain reaction (RP-PCR) [77] in all 5 siblings.

As a result of the sequencing experiments, there was no evidence of pathogenic variants in the candidate genes *MAPT*, *GRN* and *TDP-43*. However, patients 1 and 2, and sibling 3 did show a number of repeats that was above the presumed pathogenic range in *C9orf72* (≥40 repeats) (**Figure 5-7**).

Figure 5-7. Finnish FTD family: *C9orf72* expansion analysis

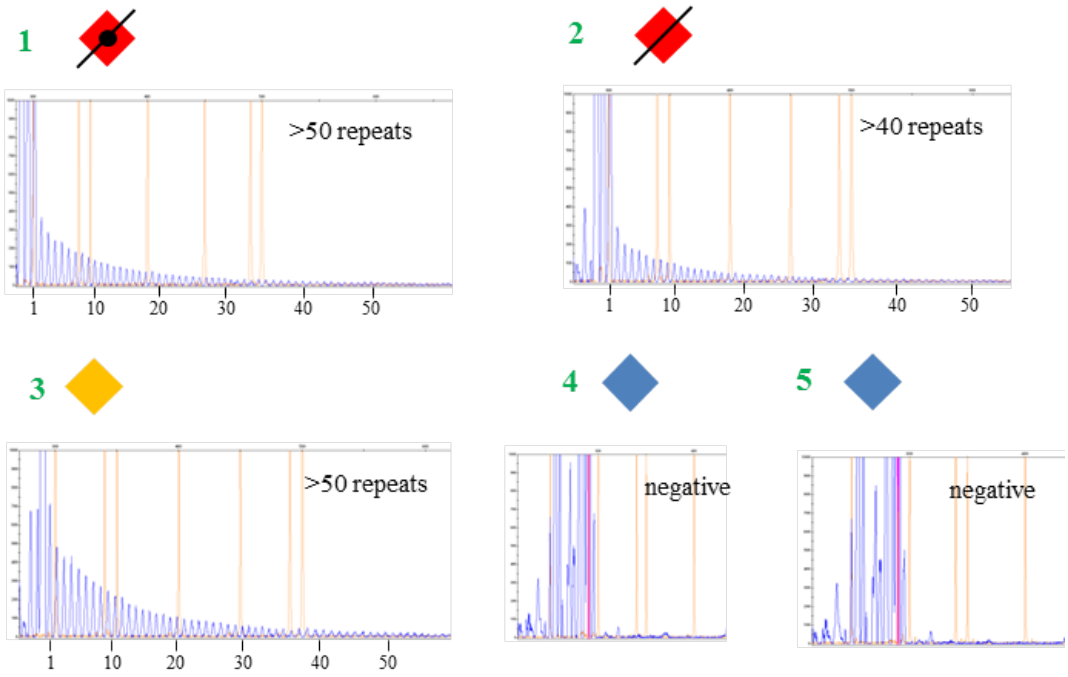


Figure 5-7. Results of *C9orf72* repeat expansion screening through primed PCR method for all five siblings of the second generation.

Finnish chromosome 9 haplotype

After having identified the presence of the expansion and verified its distribution among the family members, I evaluated if the expansion carriers also held the previously reported chromosome 9p risk haplotype [78, 340]. I could derive the genotypes of the risk haplotype directly from the FTD-GWAS data. The expansion carriers did bear at least one of the risk alleles of all 42 SNPs, whilst this was not the case in the non-expansion carriers (**Table 5-3**). These data support the notion that the risk haplotype block and the expansion are likely to be inherited together in the Finnish population [78, 340].

Table 5-3. Finnish *C9orf72*-risk haplotype analysis

SNPs	Risk allele	Patient				
		2	3	5	4	1
rs3849942	A	**AG**	**AG**	GG	GG	**AG**
rs1330921	T	**TT**	**TT**	**TT**	**TT**	**TT**
rs10121765	C	**AC**	**CC**	**CC**	**CC**	**AC**
rs1110264	A	**AG**	**AG**	GG	GG	**AG**
rs1110155	A	**AA**	**AA**	**AC**	**AC**	**AA**
rs2150336	C	**TC**	**TC**	**TC**	**TC**	**TC**
rs2225389	C	**AC**	**AC**	AA	AA	**AC**
rs1161680	G	**AG**	**AG**	AA	AA	**AG**
rs2120718	G	**AG**	**GG**	**AG**	**AG**	**AG**
rs2589054	G	**GG**	**AG**	**AG**	**AG**	**GG**
rs10812596	A	**AA**	**AA**	**AA**	**AA**	**AA**
rs1058326	C	**CC**	**TC**	TT	TT	**CC**
rs944404	T	**TT**	**TT**	**TT**	**TT**	**TT**

Table 5-3 (continued)

SNPs	Risk allele	Patient				
		2	3	5	4	1
rs765709	A	**AA**	**AA**	**AA**	**AA**	**AA**
rs1316679	G	**GG**	**GG**	**GG**	**GG**	**GG**
rs4406503	G	**GG**	**GG**	**GG**	**GG**	**GG**
rs10511817	C	**CC**	**CC**	**CC**	**CC**	**CC**
rs725804	A	**AA**	**AC**	CC	CC	**AA**
rs10511816	T	**TG**	**TG**	**TG**	**TG**	**TG**
rs1444533	A	**AA**	**AG**	**AG**	**AG**	**AA**
rs1822723	C	**CC**	**CC**	**TC**	**TC**	**CC**
rs4879515	T	**TT**	**TC**	CC	CC	**TT**
rs895023	T	**TT**	**TT**	**TT**	**TT**	**TT**
rs868856	T	**TC**	**TC**	CC	CC	**TC**
rs7046653	A	**AG**	**AG**	GG	GG	**AG**
rs2440622	A	**AA**	**AA**	**AA**	**AA**	**AA**
rs1977661	C	**CC**	**CC**	**CC**	**CC**	**CC**
rs903603	C	**CC**	**TC**	**TC**	**TC**	**CC**
rs10812610	C	**CC**	**AC**	**AC**	**AC**	**CC**
rs2814707	A	**AG**	**AG**	GG	GG	**AG**
rs12349820	T	**TT**	**TT**	**TC**	**TC**	**TT**
rs10122902	G	**AG**	**GG**	**GG**	**GG**	**AG**
rs10757665	T	**TT**	**TT**	**TC**	**TC**	**TT**
rs1565948	G	**AG**	**GG**	**AG**	**AG**	**AG**
rs774359	C	**TC**	**TC**	TT	TT	**TC**
rs2282241	G	**TG**	**TG**	**TG**	**TG**	**TG**
rs1948522	C	**TC**	**CC**	**CC**	**CC**	**TC**
rs1982915	G	**GG**	**AG**	AA	AA	**GG**
rs2453556	G	**GG**	**AG**	AA	AA	**GG**
rs702231	A	**AC**	**AA**	**AA**	**AA**	**AC**
rs696826	G	**AG**	**GG**	**GG**	**GG**	**AG**
rs2477518	T	**TC**	**TT**	**TT**	**TT**	**TC**

Table 5-3. The 42 SNPs constituting the Finnish ALS risk haplotype and the distribution of their genotypes among the Finnish FTD kindred are shown. All the expansion carriers (1, 2 and 3) hold at least one risk allele when compared to the non-carriers (highlighted in light brown). When considering only the 21 of 24 SNPs (highlighted in yellow) of the shorter haplotype [78, 340] (due to chip design), at least one of the risk alleles was present in all the expansion carriers whilst this was not the case in the non-expansion carriers.

Screening of PTEN

As previously mentioned (section *5.1.2.3*), patient 1 also had a dysplastic gangliocytoma that was evident from the neuroimaging data (**Figure 5-6**).

The latter condition had been associated in the past with mutations in the phosphatase and tensin homolog gene (*PTEN*) [341, 342]. In the attempt to verify whether patient 1 did carry pathogenic variants in *PTEN* and, therefore, whether mutations in this gene may explain the presence of the tumour, we decided to screen this gene. The sequencing of *PTEN* in patient 1 only was first outsourced to the University of Ghent to assess the presence of coding variants through Multiplex ligation-dependent probe amplification (MLPA), because at the time there was no possibility to perform this experiment at any of the laboratories collaborating in this project. The MLPA screening did not result in the identification of any mutation in the coding regions of *PTEN*. Subsequently, further review of the literature revealed that not only single base pairs transitions leading to missense or non-sense changes [343], but also heterozygous deletions in the *PTEN* gene had been associated with the development of hamartomas [344, 345]. We had the opportunity to use the genotyping data derived from the whole genome-wide screening (FTD-GWAS) for a variety of types of analysis. As such, I thought of visually inspecting the *PTEN* locus using the Chromosome Browser tool in the Genome Studio software in all 5 siblings in the search for possible loss of heterozygosity (LOH) regions encompassing *PTEN*. Surprisingly there seemed to be a pattern suggestive of a possible LOH region spanning ~200Kbp, exclusively in the patient with gangliocytoma (B Allele

Frequency plot in **Figure 5-8**). An obvious inflation of the LogR Ratio plot was not detectable (**Figure 5-8**), but this also still does not entirely excludes the possibility of a true LOH. Taking these observations together, they imply that the putative LOH region could be due either to an extraordinarily long series of homozygous calls in the whole region (chromosome10: 89,586K–89,780K) or to a possible case of uniparental disomy (UPD) that results in a copy-neutral LOH (cnLOH) in Genome Studio (**Figure 5-8**).

Figure 5-8. Finnish FTD family: LOH analysis

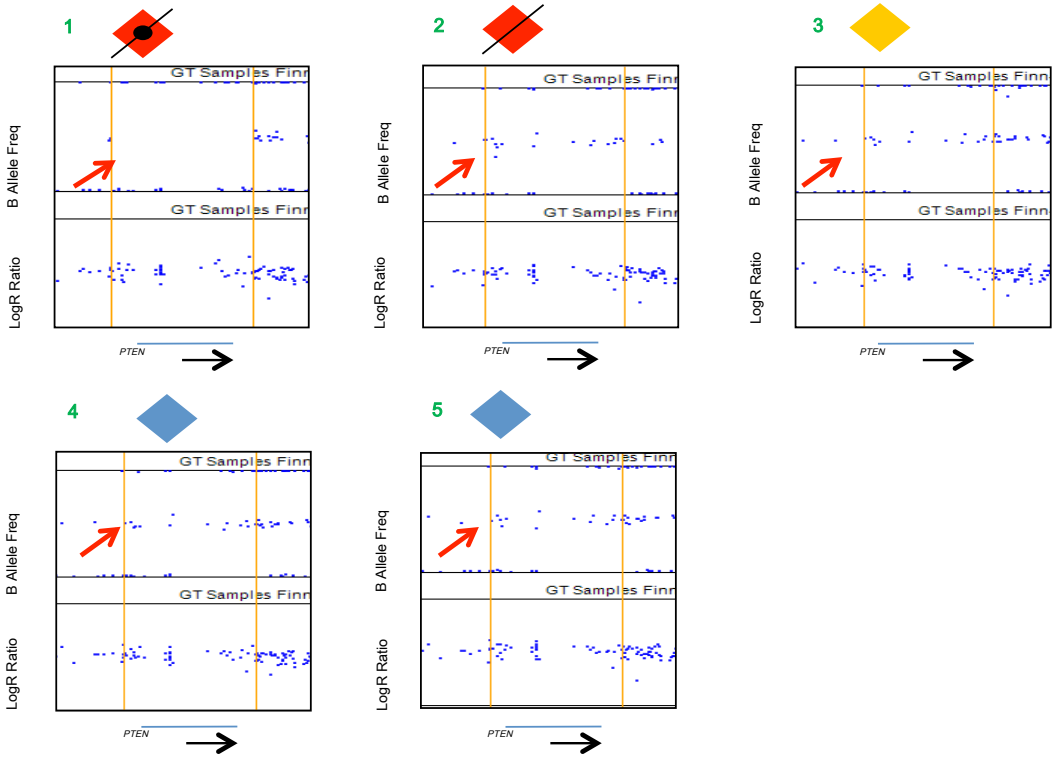


Figure 5-8. Suggestive LOH region (space delimited by the two orange vertical lines and indicated by the red arrow) encompassing the *PTEN* gene in patient 1.

This region included 27 SNPs that encompassed the whole *PTEN* gene (**Table 5-4**).

Table 5-4. *PTEN* locus: Genotyping data

Name	Chr	Position (bp)	Patient					MAF	MA	Location
			3	5	4	**1**	2			
rs1871055	10	89571955	BB	BB	BB	**AB**	BB			
rs6586103	10	89573982	AB	AA	AA	**BB**	AA			
rs11202576	10	89574264	AA	AA	AA	**AB**	AA			
rs1871057	10	89574656	BB	BB	BB	**BB**	BB			
rs17702687	10	89575957	AB (T/C)	AB (T/C)	AB (T/C)	**AA (T/T)**	AB (T/C)	0.349	T	5' from PTEN
rs10887755	10	89578868	AA	AA	AA	**AA**	AA			
rs12775504	10	89585632	AA	AA	AA	**AA**	AA			
rs7078010	10	89586538	AB (T/G)	AB (T/G)	AB (T/G)	**AA (T/T)**	AB (T/G)	0.356	T	5' from PTEN
rs10509412	10	89589334	AB	AA	AA	**BB**	AA			
rs7076964	10	89589556	AB (A/G)	AB (A/G)	AB (A/G)	**BB (G/G)**	AB (A/G)	0.481	A	5' from PTEN
rs10887758	10	89593295	AB (T/C)	AB (T/C)	AB (T/C)	**BB (C/C)**	AB (T/C)	0.23	C	5' from PTEN
rs1022427	10	89596169	BB	BB	BB	**BB**	BB			
rs1234212	10	89598872	AB (T/C)	AB (T/C)	AB (T/C)	**AA (T/T)**	AB (T/C)	0.407	T	5' from PTEN
rs1234221	10	89606459	AA	AA	AA	**AA**	AA			
rs1234220	10	89635453	AA	AA	AA	**AA**	AA			
rs11202596	10	89637134	BB	BB	BB	**BB**	BB			
rs2299939	10	89647130	AB (A/C)	AB (A/C)	AB (A/C)	**BB (C/C)**	AB (A/C)	0.147	A	PTEN intron 2
rs2248293	10	89697245	AB (T/C)	AB (T/C)	AB (T/C)	**BB (C/C)**	AB (T/C)	0.434	C	PTEN intron 5
rs11202607	10	89717394	BB	BB	BB	**BB**	BB			
rs478839	10	89721850	AB (A/G)	AB (A/G)	AB (A/G)	**AA (A/A)**	AB (A/G)	0.342	G	3' after PTEN
rs10509532	10	89727534	AB (T/C)	AB (T/C)	AB (T/C)	**AA (T/T)**	AB (T/C)	0.179	T	3' after PTEN
rs809367	10	89731786	BB	BB	BB	**BB**	BB			
rs644205	10	89735794	BB	BB	BB	**BB**	BB			
rs10887768	10	89745581	AB (T/C)	AB (T/C)	AB (T/C)	**BB (C/C)**	AB (T/C)	0.289	C	3' after PTEN
rs10509533	10	89747929	BB	BB	BB	**BB**	BB			
rs10887774	10	89756023	AB (A/G)	AB (A/G)	AB (A/G)	**AA (A/A)**	AB (A/G)	0.309	A	3' after PTEN
rs10788575	10	89758564	AB (A/G)	AB (A/G)	AB (A/G)	**AA (A/A)**	AB (A/G)	0.197	A	3' after PTEN
rs2785070	10	89764558	AB	BB	BB	**AA**	BB			
rs2244092	10	89767768	AB	BB	BB	**AA**	BB			
rs2673825	10	89768475	BB (G/G)	AB (A/G)	AB (A/G)	**BB (G/G)**	AB (A/G)	0.344	T	3' after PTEN
rs1360950	10	89770130	BB (G/G)	AB (A/G)	AB (A/G)	**BB (G/G)**	AB (A/G)	0.271	A	3' after PTEN

Table 5-4. The genotypes of 27 SNPs that in the 660K Quad chip array cover the region between chromosomal positions 89,586K–89,780K encompassing the *PTEN* gene are shown. Patient 1 (red arrow and genotypes in bold) is the only case where the calls of the region encompassing *PTEN* are homozygous (between SNPs rs17702687 and rs1360950).

Considering that the known SNPs screened for *PTEN* through the MLPA method revealed only homozygous alleles (personal communication from Dr Myllykangas) in patient 1 this was supportive of the inference that there was a possible LOH, as suggested by the analysis through Genome Studio. To verify whether there was a deletion of one of the two alleles of *PTEN* gene which may have resulted in PTEN haploinsufficiency and, possibly, played a role in the development of the gangliocytoma in patient 1, I investigated the presence of CNV at this locus in patient 1 and the other 4 siblings through a TaqMan CNV assay (**Chapter 2**, section *2.2.2.7*). As previous reports had shown that deletions occurring at the *PTEN* locus may affect either exon 1 only, or exons 1-5 or the whole gene [344, 345], I chose an assay that was targeting exon 1 in order to capture either of the possibilities.

The experiment was performed for the 5 siblings at two different concentrations and in quadruplicate each time (**Figure 5-9a** and **b**) and confirmed later using a second DNA extraction from sibling 1, 2 and 3 again at two different concentrations and in quadruplicate each time (**Figure 5-9c** and **d**). Each set of experiments revealed normal copy numbers (n=2) for each of the sibling. *A priori*, patient 1 could have been expected to carry a deletion (copy number=1) but the CNV experiment did not confirm this assumption. Such results pointed out that the LOH pattern seen in **Figure 5-8** could most probably reflect either an extraordinarily extended region of exclusive homozygous calls or, although very unlikely, a cnLOH.

Figure 5-9. Finnish FTD family: CNV analysis

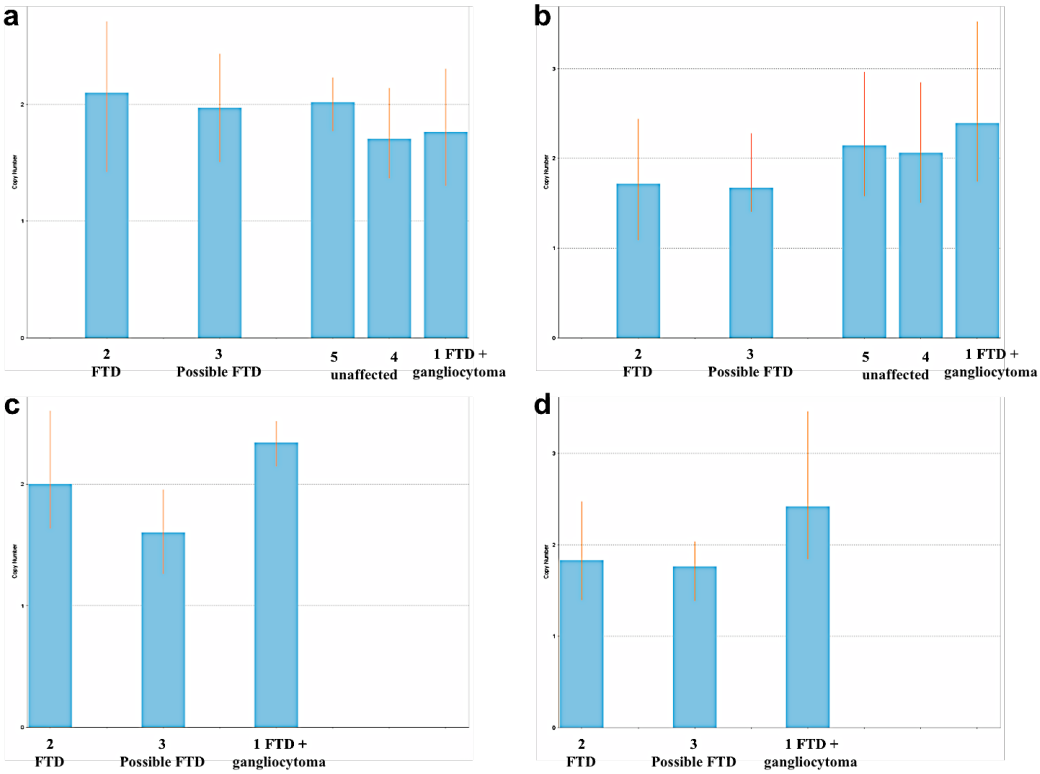


Figure 5-9. CNV assay targeting exon 1 of *PTEN*. (**a**) The bars depict results of the 5 siblings for whom DNA was used at a concentration of 50 ng/ul. Each bar represents 4 replicates. (**b**) The bars depict results of the 5 siblings for whom DNA was used at a concentration of 25 ng/ul. Each bar represents 4 replicates. (**c**) The bars depict results of 3 (1, 2 and 3) siblings for whom new DNA extractions were performed. DNA was used at a concentration of 50 ng/ul. Each bar represents 4 replicates. (**d**) The bars depict results of 3 (1, 2 and 3) siblings for whom new DNA extractions were performed. DNA was used at a concentration of 25 ng/ul. Each bar represents 4 replicates. Normal copy numbers (2) are predicted for all siblings. The bars in this figure are not normalized.

In order to verify homozygosity of *PTEN* gene in this family through a method other than chip or CNV TaqMan assay I sequenced exon 1, which was the target of the CNV assay and exon 5, which is known from the literature to be highly polymorphic. According to

dbSNP at least 8 SNPs (rs149772796, rs145695240, rs57374291, rs145124907, rs139767111, rs370795352, rs144545031 and rs9651492) are harboured in exon 5. Detecting homozygous alleles in patient 1 and, conversely, heterozygous calls in the other family members for the known SNPs could confirm at least the pattern seen in **Figure 5-8**. However, sequencing did reveal homozygous calls for all the known SNPs in all the siblings. At present, provided that the CNV assay showed 2 allele copies in this genetic region in all siblings, including patient 1, the reason underlying the LOH is still unresolved.

5.1.2.5 Summary

Genetics of the Finnish kindred

This study was performed to identify the genetic underpinnings of a Finnish FTD kindred. After I included all family members on the 660K GWAS chip and generated genome-wide genotyping data, Dr Myllykangas performed a linkage analysis that pointed to the *TDP-43* and *C9orf72* loci as possibly associated in this family. Subsequently, the genetic screening was focused on the 5 siblings (**Chapter 2**, **Figure 2-3**) by screening for the candidate genes *MAPT*, *PGRN*, *TDP-43* and *C9orf72*. No variants were detected in the coding regions of the first 3 genes, whilst three of the siblings (1, 2 and 3) carried the *C9orf72* expansion (**Figure 5-7**). Of note, the third sibling (3) recently became symptomatic. All three siblings (1, 2 and 3) also carried the 42-SNP 9p21 risk haplotype

associated with ALS in the Finnish population which was absent from the two non-expansion carriers (siblings 4 and 5) (**Table 5-3**), supporting previous reports specific to the Finnish population [78, 340] and pointing to this locus as associated with the diseased phenotype in this family.

To put these data into context, it is relevant to consider the clinical and pathological features of the index patients.

The early clinical signs in patient 1 involved memory and orientation problems on one hand and ataxia (movement problems) on the other. In addition, although with less emphasis, cognition and speech were impaired, and there was evidence of bradykinesia. The early clinical signs of patients 2 involved psychiatric problems (social avoidance) and movement impairment. As well cognition and language impairments were present accompanied by memory dysfunction and bradykinesia. In both cases the diagnosis was FTD but evidently the broad features were heterogeneous encompassing typical signs of AD (memory and visuospatial), FTD (cognition, executive function and language), and included features that are seen in FTD but less frequently (psychiatric and extrapyramidal/parkinsonian signs). These features on one hand reflect the topographic atrophy patterns observed in the brains of the patients and on the other, to a great extent, the pathological signatures that have been described in expansion carriers [85] [91, 92, 346]. Although, to date, the most frequent diagnoses associated with the expansion fall in the categories of bvFTD and FTD-MND [86, 94, 347], heterogeneity of the clinical features in presence of *C9orf72* expansions is evident [348]. The current work thus

confirms and shows that clinical signs associated with the expansion in *C9orf72* can encompass features resembling AD, bvFTD, PPA and parkinsonism all together. However, it is difficult, at this stage, to correlate the expansion alone to these many co-morbidities or broad clinical manifestations. Clearly, additional studies are needed to further elucidate this aspect. The molecular pathology associated with the index patients revealed FTLD-TDP pathology type B (patient 1) or A/B (patient 2) and included the typical signature of the expansion carriers, i.e., the cytoplasmic and intranuclear p62-positive and TDP-43-negative inclusions in the granular cells of the cerebellum. These features collectively mirror previous reports describing the most common forms of pathology associated with *C9orf72* expansions [31, 84, 85, 89, 91].

PTEN and gangliocytoma

On the other hand, patient 1 had a dysplastic gangliocytoma (Lhermitte-Duclos disease), which is a rare benign brain tumour/hamartoma of cerebellar granular cells [341] that became an interesting, *per se*, study within the main (genetic) study of the Finnish FTD family. It had been hypothesized that the phenomenon of aberrant migration and hypertrophy of granule cells is at the basis of the development of these benign lesions [341]. Adult-onset dysplastic gangliocytomas can occur in individuals with Cowden syndrome, a condition that is associated with mutations in the *PTEN* gene [341, 342]. Although patient 1 did not explicitly show signs of Cowden syndrome we decided to screen *PTEN* to assess possible genetic variability in this gene and its association with

this condition. In normal conditions PTEN is a tumour suppressor that controls cell cycle, cell migration and apoptosis. When the gene presents either coding mutations, mutations in the promoter region or partial/complete gene deletion, PTEN malfunction enhances cell survival, cell growth and decreases apoptosis [349]. From a functional perspective it was shown that *PTEN* mutations lead to activation of Akt/mTOR pathway during the formation of dysplastic gangliocytoma [341, 342]. Interestingly, such activation was also seen in neurodegenerative diseases, including ALS, especially, in those cases associated with *SOD1* mutations [349-353]. The biology of PTEN seems complex; knock-down/inhibition of the protein leads to a robust sprouting of adult sensory neurons [354], which may seem a mechanism that supports neuronal survival. However, based on the nature of PTEN protein, it cannot be dismissed that such positive effect may be only initial due to the fact that the normal function of PTEN as a tumour suppressor is impaired leading, ultimately, to tumourigenesis. In fact, a reduction of PTEN protein levels has been shown to associate with or induce tumourigenesis across studies [342, 354-356]. Targeted analysis, including sequencing and copy number variation (CNV) assessment, of *PTEN* did not support the presence of missense/non-sense mutations or heterozygous deletion in patient 1. However, genome-wide data showed a possible LOH encompassing *PTEN* in patient 1 but in none of the other siblings (**Figure 5-8**). A recurring segmental LOH pattern could be seen throughout chromosome 10 (where *PTEN* resides), but it was not encompassing a large continuous region of the chromosome virtually excluding the UPD as a possibility (**Figure 5-10**). At this stage, having ruled out a heterozygous deletion through the CNV experiments (copy number=2), it is difficult to

interpret the LOH patterns as presented in the B Allele Frequency plot (**Figure 5-10**). It is possible that such heterogeneous LOH pattern in patient 1 just represents regions in which the calls are unusually homozygous or it is an artefact due to a chip defect. However, the possibility of homozygous calls may be further investigated through whole genome sequencing, which would clarify whether the whole region encompassing the *PTEN* gene (and the other LOH regions; **Figure 5-10**) is/are truly homozygous.

Figure 5-10. Finnish FTD family index patient: CNV analysis of whole chr10

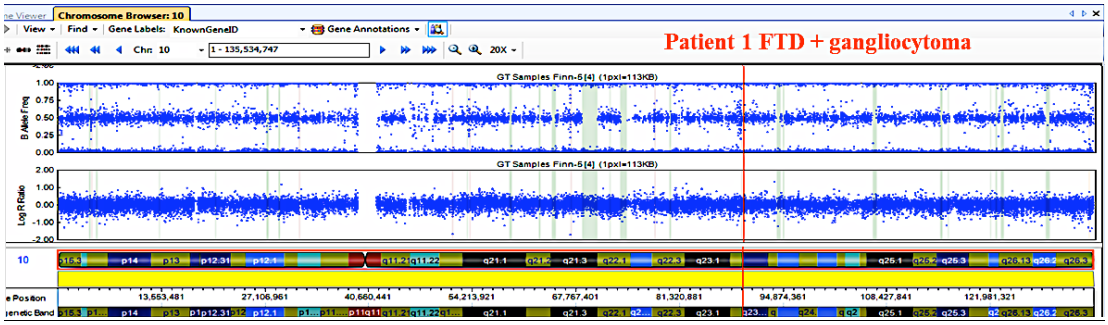


Figure 5-10. The chromosome browser view of the Genome Studio software is shown. The LOH patterns visible throughout chromosome 10 unlikely represent UPD. Putative LOH regions are highlighted by green bars on the B Allele Frequency and LogR Ratio plots. The red vertical line localises *PTEN* gene.

5.2 Sporadic cases

This section summarizes work that targeted the extended FTLD and CBS/CBD cohorts received and stored at TTUHSC by Dr Momeni.

In brief, the DNA material that was screened in these projects was extracted from blood samples of a total of 228 index patients with the broad diagnoses of FTLD (n=158) or CBS/CBD (n=40/30). These cases were originally collected at the Cognitive Neuroscience Division of National Institute of Neurological Disorders and Stroke (NINDS) at the National Institutes of Health (NIH) in collaboration with Drs. Grafman, Wassermann, Pietrini, Kapogiannis and Huey [137] and later sent to and stored at TTUHSC. This cohort included patients with diagnoses of behavioural variant FTD (bvFTD), FTD-MND, progressive nonfluent aphasia (PNFA) [21], and CBS/CBD [220]. The procedure to enrol and diagnose the patients was standardized at NIH by Dr Grafman. The patients were either self-referred or referred by outside physicians or psychiatrists or neurologists. At NIH each patient underwent extensive neurologicical, neuropsychological and imaging examination, and the diagnosis was based on standard clinical criteria [21, 220]. All subjects gave written informed consent for the study and the NINDS Institutional Review Boards, as well as those at TTUHSC, approved all aspects of the studies (**Chapter 2**, section ***2.1.5***). These samples were screened during the past 5–8 years for the main candidate genes (**Table 5-1**).

5.2.1 FTD and CHMP2B

5.2.1.1 Genetic screening and functional study

During the aforementioned extended screening, I identified a novel heterozygous g.26218G>A variant in exon 6 of *CHMP2B* (**Figure 5-11**), predicted to cause the amino-acid change Ser187Asn in the carboxy-terminal region of the CHMP2B protein [123].

Figure 5-11. *CHMP2B*: Chromatogram of Ser187Asn

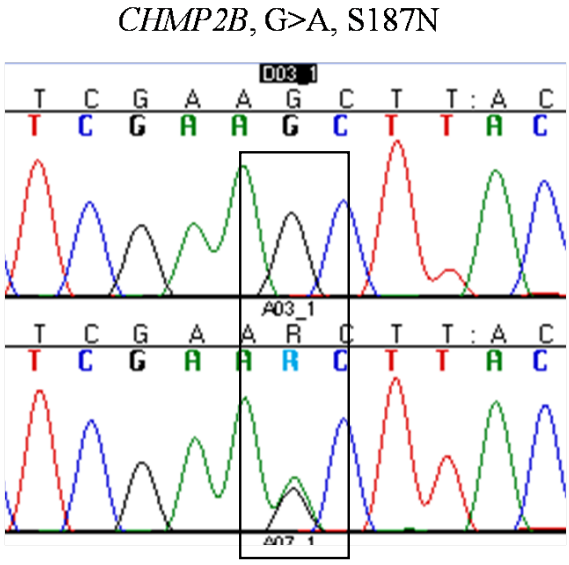


Figure 5-11. The chromatogram shows the base change G>A causing the missense mutation Ser187Asn in *CHMP2B* [123].

This variant was identified in one patient of English/Swedish heritage diagnosed with bvFTD characterised by executive dysfunction and personality change (**Figure 5-12**) [123].

Figure 5-12. Index patient pedigree

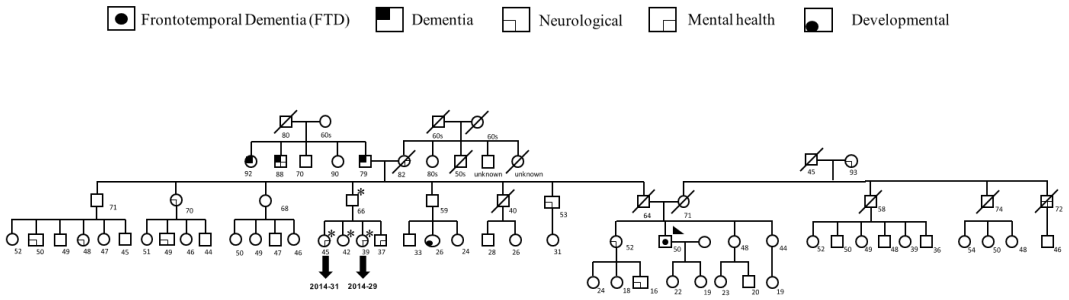


Figure 5-12. The pedigree of the FTD proband is shown. The paternal grandfather and two of his siblings were diagnosed with dementia. Family members screened along with the proband are countersigned by an asterisk [123].

The complete results of the genetic screening performed on the proband are shown in **Table 5-5**. These include characterisation for the *ApoE* genotype, the *MAPT* haplotype (H1 vs. H2), and a summary of the findings in the other candidate genes *MAPT* and *GRN*. Although a number of variants were identified in *MAPT* and *GRN*, none resulted being pathogenic.

Table 5-5. Candidate genes screening

ApoE	Sequencing	MAPT Haplotype
	MAPT	
E3/e3	Exon 1, A>G (+8) from 5' exon 1, homozygous G (rs17650901) Intron 8, G>A (-26) from 5' of exon 9, homozygous A (rs62063850) Exon 9, A>G, p.A227A (silent), homozygous G (rs1052553) Exon 9, T>C, p.N255N (silent), homozygous C (rs17652121) Intron 11, G>A (+34) from 3' of exon 11, homozygous A *	H2/H2
	GRN	
	Intron 3, G>A (+21) from 3' of exon 3 (rs9897526) Intron 4, del/ins GTCA (-47-50) from 5' exon 5, heterozygous deletion (rs34424835) Intron 5, G>A (+24) from 3' exon 5 (rs850713) Intron 12, C>T (+78) from 3' exon 12 (rs5848)	
	CHMP2B	
	Exon 3, T>C, p.T104T (silent), homozygous C (rs11540913) Exon 6, G>A, p.S187N *	

Table 5-5. Sequencing analysis resulted in the identification of single nucleotide polymorphysms (SNPs) in *MAPT, GRN* and *CHMP2B*. Most of the SNPs are known. Novel (non-previously reported) SNPs are indicated by an asterisk.

To evaluate the possible pathogenicity of the newly identified missense change, Ser187Asn, I screened neurologically normal controls of white Caucasian and African American ancestry. The g.26218G>A variant leading to Ser187Asn was not found neither in 273 (NDPT 098, NDPT 099, and NDPT 096: Coriell Cell Repositories, Camden, NJ, USA), nor it was present in additional 400 Caucasian neurologically normal controls that had previously been screened for *CHMP2B* exon 6 [125]. However, the novel *CHMP2B* change was harboured by 6 of 94 (6.4%) normal controls of African American heritage (Coriell Cell Repositories: NDPT 031). Unfortunately, it was not possible to determine

the mode of inheritance of this variant, since the father of the proband is deceased and blood samples from the mother could not be collected. However, samples from an uncle and 3 cousins of the proband were available and were screened (**Figure 5-12**). None of them carried the variant.

In silico analysis of the effect of the Ser187Asn missense change on the protein structure and function was evaluated using the PolyPhen-2 software [225] and it was predicted to be benign (**Figure 5-13**).

Figure 5-13. *CHMP2B*; missense change Ser187Asn. PolyPhen-2 analysis

Figure 5-13. Polyphen-2 predicted the g.26218G>A variant (Ser187Asn) being benign.

To further explore the pathogenicity of Ser187Asn I performed a functional study in the laboratory of Dr Collinge under supervision of Dr Adrian Isaacs at the Prion Unit, at the

Institute of Neurology at UCL. Mrs. Astrid Authier and I generated a plasmid construct containing the mutated base g.26218G>A within *CHMP2B* exon 6 by means of a mutagenesis assay (**Chapter 2**, section ***2.2.3***). To confirm the presence of the base change G>A in the plasmid I sequenced it for *CHMP2B* exon 6 (**Figure 5-14**).

Figure 5-14. *CHMP2B*: Chromatogram of Ser187Asn (plasmid)

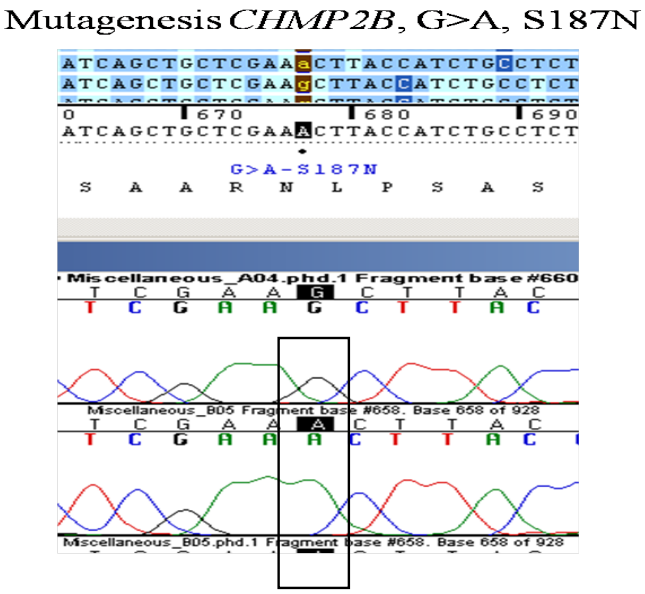


Figure 5-14. The plasmid holding the homozygous G>A base change which causes Ser187Asn.

To progress in the functional study SK-N-SH cell line (human neural epithelial cells obtained from bone marrow metastasis from the European Collection of Cell Cultures

[ECACC]) and fibroblast cell lines were used for transfection of mutated *CHMP2B* (this part of the work was performed by Mrs. Astrid Authier and Dr Hazel Urwin). The expression of *CHMP2B* carrying the missense mutation Ser187Asn was compared to the expression of wild-type *CHMP2B*. Analysis by confocal microscopy microscopy by Dr Urwin did not reveal differences in the cytoplasmic localization and distribution of $CHMP2B^{wild\text{-}type}$ and $CHMP2B^{S187N}$ suggesting and confirming the most probable non-pathogenicity of this missense change confirming the prediction of PolyPhen-2. This outcome was replicated in a further confirmation experiment (Dr H. Urwin, personal communication).

5.2.1.2 Comment on variability in CHMP2B

Mutations in *CHMP2B* seem to cause extremely rare cases of FTD [51].

Although there is evidence of genetic variability in *CHMP2B* in the literature [47] and the original finding of a link to chromosome 3 as well as the segregation in the Danish family seem convincing stories [121, 122], there is also evidence and support for the notion that the role of *CHMP2B* in FTD is still, to date, equivocal.

Based on reports of the past eight years mutations in *CHMP2B* can be divided into two groups:

1 – The carboxy-terminal truncation, and;

2 – The missense mutations (**Table 5-6**).

Table 5-6. Mutations in *CHMP2B*

Truncation mutations	Base change	Disease	Reference
$CHMP2B^{intron5}$	g.26189G>C	FTD	[121]
$CHMP2B^{\Delta 10}$	g.26189G>C	FTD	[121]
p.Gln165X	g.25950C>T	FTD	[127]
p.Arg186X	g.26214C>T	FTD	[125]
Missense mutations			
p.Ile29Val	g.13227A>G	FTD3 ALS	[126, 128]
p.Asn143Ser	g.25885A>G	CBD	[127]
p.Asp148Tyr	g.25899G>T	FTD	[121]
p.Ser187Asn	g.26218G>A	FTD	[123]
p.Gln206His	g.26276A>C	ALS	[128]

Table 5-6. Summary of most important truncation and missense mutations in *CHMP2B*.

In order to better understand the effects of these mutations on the structure and function of the protein CHMP2B and their possible involvement in neurodegeneration, the role of wild-type CHMP2B needs to defined, first. The endosomal sorting complex required for transport (ESCRTs) I, II and III is a multimeric protein complex in charge for sorting proteins into the intraluminal vesicles (ILVs), which constitute multivesicular bodies (MVBs). These are a subset of late endosomes that eventually fuse with lysosomes for cargo proteins degradation. CHMP2B is part of the ESCRT-III complex, which is the portion directly involved in sorting the cargo proteins into ILVs [357, 358]. CHMP2B presents a coiled coil domain between amino-acids 25-55 at the amino-terminus (**Figure**

5-15A) and a microtubule interacting and transport (MIT) – interacting motif (MIM), between the carboxy-terminal amino-acids 201-211 (**Figure 5-15B**).

Figure 5-15. CHMP2B protein domains

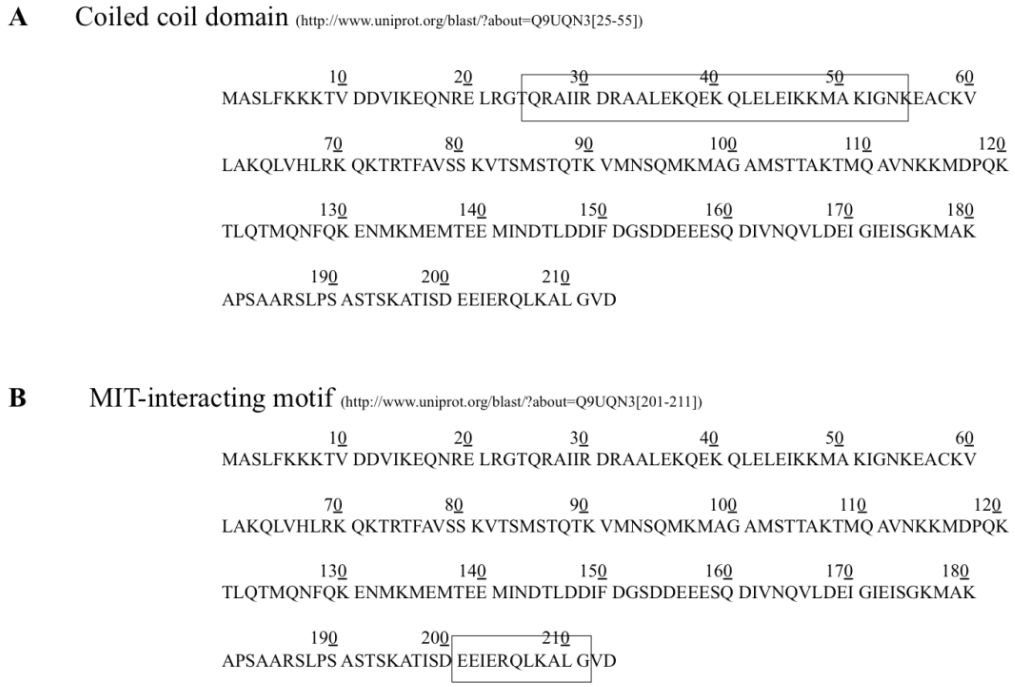


Figure 5-15. (**A**) The coiled coil domain is located close to the amino-terminus of CHMP2B. (**B**) The microtubule-interacting and transport (MIT)-interacting motif (MIM) is located at the carboxy-terminus of CHMP2B.

Within the ESCRT-III complex, the three proteins CHMP2B, CHMP2A and Vps24 (vacuolar protein sorting 24), together, bind the MIT domain of the hexameric protein

Vps4 [359, 360]. The interaction between these components leads to two main functional consequences:

1 – The active dissociation of ESCRTs from the endosomal membrane, and;

2 – The formation and release of ILVs [361, 362].

In this scenario, the carboxy-terminal truncation mutations are predicted to cause loss in CHMP2B of the domain that binds Vsp4.

Carboxy-terminal truncation mutations

Functional analysis of the intron 5 and Gln165X mutations (**Table 5-6**) showed that they lead to aberrant cytoplasmic phenotype if compared to cells transfected with wild-type CHMP2B [35, 121, 127]. However, interestingly, functional analysis of the Δ10 mutation did not reveal clear effects on neurodegeneration [363] suggesting that, possibly, not all reported *CHMP2B* mutations are pathogenic, including some of the carboxy-terminal truncating mutations [364]. In support of this observation, the isolation of the Arg186X truncation mutation in two asymptomatic members of an FTD family (section *5.1.1.3*) also raises questions about the pathogenicity or penetrance of the carboxy-terminal truncating mutations in *CHMP2B* [125]. Of course those individuals from the Afrikaaner kindred who carry the truncation mutation will need to be longitudinally followed to verify whether they will eventually develop FTD. However, at the current stage,

Arg186X may either be a non-pathogenic change, or a pathogenic variant with variable penetrance or a pathogenic mutation associated with differential age of onset.

Missense mutations

Some missense mutations, such as Asp148Tyr and Asn143Ser, that were previously reported were neither associated with pathogenicity nor aberrant endosomal phenotype [121, 127, 364]. The variant leading to Ile29Val was identified in normal controls, weakening its pathogenic implication [34, 126] and Gln206His, which is located in the MIM domain (**Figure 5-15B**), was shown to lead to a cytoplasmic phenotype characterised by enlarged vacuoles [365]. Interestingly, Gln206His was isolated in both FTD [128] and ALS cases [365].

The novel missense mutation that I identified, Ser187Asn, was not found in a total of 673 Caucasian neurologically normal controls [123, 125]; however it was identified in a French control cohort [124]. In addition, it reached a frequency of 6.4% in African American controls suggesting it being a common polymorphism in the normal population. Although for the latter we could not verify the pattern of inheritance, its presence in the normal population, as well as the *in silico* prediction of non-pathogenicity through PolyPhen 2 software and the confirmation through a functional study (data not shown, section *5.2.1.1*), all together, suggest that this variant is most probably not pathogenic. All these data do not entirely support a pathogenic role for *CHMP2B*

missense mutations in FTD. Possibly, due to its location within the MIM, Gln206His may be the only missense change with a detrimental effect.

Functional studies and animal models

Most of the functional work performed to date indicates that impaired trafficking of multivesicular bodies (MVBs) to lysosomes can cause cytoplasmic accumulation of vesicles [366, 367]. In this view, it was shown that:

1 – Depletion of ESCRT subunits can lead to abnormal morphology of the MVBs fact that in turn affects late endosomal trafficking and causes protein or vesicles accumulation [368], and;

2 – Intron 5 was implicated in the misregulation of Toll-like receptor (TLR) leading to abnormal sorting in the endocytic pathway [364].

More recently, a study proposed the possibility of using RNA interference to knockdown the expression of mutated *CHMP2B* [369]. The authors argued that this approach could clear abnormally enlarged endosomes from the cytoplasm of fibroblasts derived from patients carrying the intron 5 and Δ10 mutations [369]. Although a partial decrease of abnormally enlarged endosomes was observed, the long term effects of CHMP2B depletion could not be assessed; therefore it would probably be sensible to verify this approach and replicate these findings in animal models as well as clinical trials to prove their chances of success in humans. Finally, an animal model developed to study the *in*

vivo effects of intron 5 reported a reduced survival of intron 5+ transgenic mice, in addition to lesions such as presence of p62 positive pathology and axonal degeneration [129]. These results are clearly of interest but, to support pathogenicity of carboxy-terminal truncation mutations in *CHMP2B*, it would be illuminating if Δ10, Gln165X and Arg186X would replicate these findings and lead to the same phenotype.

5.2.1.3 Summary

In conclusion, some studies support the idea that variability in *CHMP2B* may lead to impairment of the ESCRT machinery. This may disrupt endosomal trafficking causing, potentially, a number of functional consequences:

1 – Lack of trophic support for the cell;

2 – Aberrant cellular signalling and;

3 – Impairment of autophagy [35].

Interestingly, especially the third mechanism seems supportive of the results of the FTD-GWAS (**Chapter 4**) implying to impaired lysosomal biology in FTD. However, the *CHMP2B* locus did not reach suggestive association confirming that variability in *CHMP2B* is extremely rare and definitely not sufficient to give rise to genome-wide significant signals. All this taken together supports the notion that *CHMP2B* is a very rare

cause of FTD and indicates, all the more, that the pathogenicity of the wide range of the mutations identified in *CHMP2B* to date requires further supportive evidence.

5.2.2 FTLD and CBS/CBD: Variability in TDP-43 and FUS

5.2.2.1 Introduction

In 2009, the fused in sarcoma gene (*FUS*) was suggested to be involved in the pathogenesis of familial amyotrophic lateral sclerosis (fALS) [142, 143]. *FUS* encodes a protein that regulates DNA repair and transcription processes such as RNA splicing. Similarly, the TAR-DNA Binding Protein 43 (TDP-43) acts as a transcription factor and is also involved in RNA alternative splicing processes [370]. Thus, both FUS and TDP-43 play a critical role in the RNA metabolism. In addition, TDP-43 pathology occurs in ~50% of FTD cases (FTLD-TDP pathology) [31] and represents a remarkable overlap between the FTLD and ALS spectrum in that it co-localises with ubiquitin positive deposits in the brains of both FTD and ALS patients [371, 372]. Genetic variability in *FUS* and *TDP-43* was mainly associated with ALS [142, 143, 331, 373], whilst their mutation rate in FTD has been extremely rare and, all the more, yet under-investigated [135, 145].

With study, we intended to further screen them in FTLD as well as in CBS/CBD cases in order to better assess their implication in these disorders.

5.2.2.2 Genetic screening

All exons of the *FUS* (n=15) and *TDP-43* (n=6) genes and their flanking intronic regions were sequenced in 158 FTLD and 70 CBS/D cases.

The variants identified in FUS were further screened in a total of 569 neurologically normal controls of Caucasian, African American and Mexican ancestry that were obtained from the Coriell Institute (http://www.ccr.coriell.org; **Table 5-8**).

FUS

Sequencing analysis of *FUS* resulted in the identification of a limited number of variants within the patient cohort: One missense change, 2 indels, 2 silent changes, and one intronic variant in the 3'UTR (**Table 5-7**).

Table 5-7. Candidate genes screening

Patient ID	Dx	Mutation	Exon	Previously reported	Controls
FUS					
182	FTD	Pro106Leu CCC>CTC	4	No	No
85	FTD	Gly174-Gly175 del GG (g. 4180-4185 delGAGGTG)	5	FALS [142] NC [145]	No
135	CBS	Gly175-Gly176 insGG (g. 4185-4186insGAGGTG)	5	ALS [142]	Yes (1/659)
593-1	FTD	Pro125 CCC>CCG	5		No
568-1	FTD	Arg522 AGG>AGA	15		Yes (1/659)
87	CBS	3'UTR exon 15 STOP +41	3'UTR		Yes (9/659)
109	CBS	3'UTR exon 15 STOP +41	3'UTR		
TDP-43					
125	CBS	Asn267Ser	6	ALS [374] FTLD [135]	Not tested

Table 5-7. Genetic variants in the *FUS* and *TDP-43* genes identified among the FTD and CBS patients. DX=diagnosis.

The novel variant Pro106Leu in exon 4 of *FUS* (**Table 5-7**) was identified in an FTD patient characterised by personality/behavioural changes, executive dysfunction as well as a decrease in language production and apathy. Co-segregation with disease could not be verified in this case due to lack of DNA samples from informative members of the family of the proband. Although this variant was not identified in 569 neurologically normal controls it is difficult to fully assess its pathogenicity in absence of functional data. Nevertheless, *in silico* analysis suggested this missense change being most likely benign (**Figure 5-16**).

Figure 5-16. *FUS*; missense change Pro106Leu. PolyPhen-2 analysis

Figure 5-16. Polyphen-2 predicted the missense change Pro106Leu in *FUS* being benign.

Further, we identified the heterozygous 6 base pair deletion Gly174-Gly175 del GG (g. 4180-4185 delGAGGTG) in exon 5 of *FUS* in an FTD patient whose main symptoms resembled PNFA. Although the same deletion had previously been described in an fALS case [142] it was also, more recently, isolated in 2/638 controls aged 70 and 37 years, hence questioning the pathogenicity of this variant [145]. In addition, we isolated the heterozygous insertion Gly175-Gly176 insGG (g. 4185-4186insGAGGTG) in exon 5 of *FUS* in a patient with the diagnosis of CBS. This variant had also been previously reported in fALS [142] but we identified it in 1/659 of our normal controls population.

The latter is the first report of this variant in a normal control suggesting that its pathogenic role may be questionable. In light of these results, the indel polymorphisms originally identified in familial ALS and thought to be pathogenic [142, 143] may, at this

point, be considered non-pathogenic and excluded from involvement in the pathogenesis of FTD, CBS and, retrospectively, ALS. In addition, we identified a synonymous variant, Pro125Pro, in a FTD patient and in none of our controls whereas we found another synonymous variant, Arg522Arg, in both one FTD patient and one control (**Table 5-7**). Of note, one variant that we isolated in the 3'UTR region of *FUS* (41 base pairs after stop codon) of two CBS patients was also present in nine controls (**Table 5-7**). To better interpret the potential pathogenicity of the *FUS* variants that were found in the cases, particular attention was given to the analysis of *FUS* exons 4, 5, 6, and 15 in 659 neurologically normal controls. A detailed list of the variants identified in the normal controls with their respective DNA ID from Coriell is summarized in **Table 5-8**.

Table 5-8. *FUS* screening in neurologically normal controls

Coriell series ID	Sample ID	Mutation	Exon	Frequency
Coriell NDP031	ND08239	C>G (exon 4 -23)	4	1/659
Caucasian control plate 101-200	NA18010-061702 NA18074-061802	Y66YTAT>TAC	4	2/659
Coriell NDP099	ND09433 B07	GGC>GGT G76G	4	1/659
Coriell NDP031	ND13554	TAC>TAT Tyr91 rs73530286	4	1/659
Caucasian control	NA17229-121201	Ser135Asn AGC>AAC rs61732970	4	1/659
Caucasian control	NA17295- 010902	Gly175-Gly176 insGG (g. 4185-4186insGAGGTG)	5	1/659
Coriell NDP099	ND01689 C03 ND02820 E04	C>T intron 5 +22	Intron 5	2/659
Coriell NDP098	ND06756	T>C intron (start ex6 -21) rs74015090	Intron 5	34/659
Coriell NDP99	ND10775	T>C intron (start ex6 -21) rs74015090	Intron 5	
Coriell NDP96	ND11603 ND11575	T>C intron (start ex6 -21) rs74015090	Intron 5	

Table 5-8 (continued)

Coriell series ID	Sample ID	Mutation	Exon	Frequency
Coriell NDP031	ND04013 ND09688 ND11333 ND14116 ND14118 ND03137 ND05016 ND09724 ND05233 ND09725 ND09828 ND11656 ND04478 ND05278 ND04479 ND09728 ND09947 ND11906 ND02365 ND04011 ND12395 ND08253 homozygous C ND09824 homozygous C ND12149 homozygous C ND11905 homozygous C	T>C intron (start ex6 -21) rs74015090	Intron 5	34/659
Coriell Mexican American controls	NA17452 MX00014 120301 NA17712 MX0099 061002 NA17678 MX0073 061002 NA17709 MX0096 062402 homozygous C	T>C intron (start ex6 -21) rs74015090	Intron 5	
Caucasian control	NA18069-062402	T>C intron (start ex6 -21) rs74015090	Intron 5	
Coriell NDP031	ND09824 ND09724 ND05233 ND09828 ND05278 ND11906 ND02365 ND12395 ND12149 homozygous T	C>T intron ex6 -5 rs73530287	Intron 5	11/659

Table 5-8 (continued)

Coriell series ID	Sample ID	Mutation	Exon	Frequency
Coriell Mexican American controls	NA17709 MX0096 062402	C>T intron ex6 -5 rs73530287	Intron 5	11/659
Caucasian control	NA18069-062402	C>T intron ex6 -5 rs73530287	Intron 5	
Coriell NDP098	ND05808	GGC.GGC.GGC.GGC.G GT to GGC.GGT Gly228_Gly230del	6	1/659
Coriell NDP031	ND14118 ND11656 ND04479 ND04011	GAC>GAT D212D	Exon 6	4/659
Caucasian control	NA17225-040902	GGT>GGC Gly222 rs61732969	Exon 6	1/659
Coriell NDP031	ND00662 ND09723 ND09401 ND09726 ND09658	(GGC)5/6/7 Gly224 rs72550890 11 Gly	Exon 6	6/659
Caucasian control	NA17225-040902	(GGC)5/6/7 Gly224 rs72550890 9 Gly	Exon 6	
Coriell NDP031	ND08253 ND09728	GGC>GGT G225G	Exon 6	1/659
Coriell NDP099	ND03969	GGC>GAC G227D	Exon 6	1/659
Coriell Mexican American controls	NA17465 MX00025 060602	GGT>GGC G229G		1/659
Coriell NDP098	ND05808	GGC.GGC.GGC.GGC.G GT to GGC.GGT Gly228_Gly230del	Exon 6	1/659
Coriell NDP031	ND11333 ND09947	CGC>CGT R234R		2/659
Coriell Mexican American controls	NA17465 MX00025 060602	GGA ins R244	Exon 6	1/659
Caucasian control	NA17278-090601	AGG>AGA Arg522Arg	Exon 15	1/659
Coriell NDP099	ND01689 ND02820	G>A exon 15 STOP +41 3'UTR rs80301724	3'UTR	9/659
Coriell NDP096	ND13769	G>A exon 15 STOP +41 3'UTR rs80301724	3'UTR	
Coriell Mexican American controls	NA17676 MX0071 061002	G>A exon 15 STOP +41 3'UTR rs80301724	3'UTR	
Caucasian control	NA17238-010202 NA17242-040402 NA17255-002 040202 NA18013-061102 NA18059-062602	G>A exon 15 STOP +41 3'UTR rs80301724	3'UTR	

Table 5-8. The complete list of variants found in the exons 4, 5, 6, and 15 of *FUS* gene among 569 Coriell normal control DNA samples is summarized.

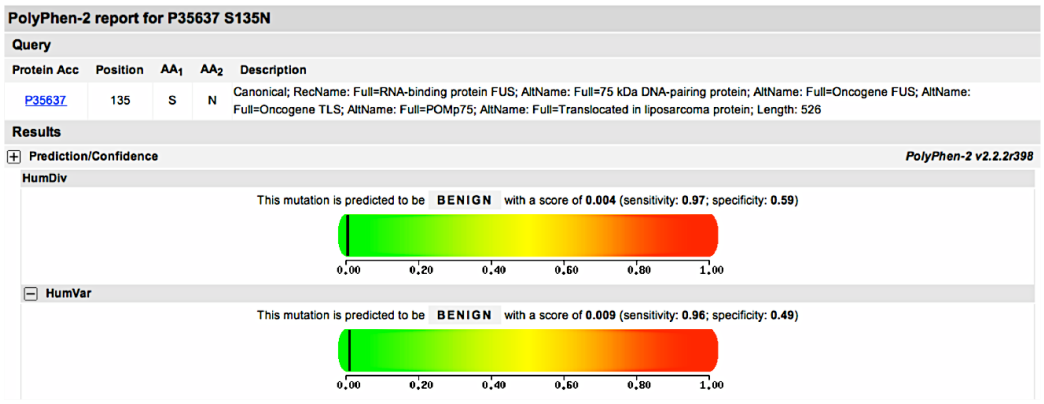
The screening in the normal population resulted in an outstanding outcome in that it did lead to the isolation of a considerable number of variants that were exclusive of our control cohort.

These included the missense variants, Ser135Asn (predicted to be benign, **Figure 5-17A**) (rs61732970) and Gly227Asp (surprisingly predicted to be probably damaging, **Figure 5-17B**), the indels GGC.GGC.GGC.GGC.GGT to GGC.GGT Gly228_Gly230del [375], (GGC) 5/6/7 Gly224 (rs72550890) 11 or 9 Gly, and the GGA ins Arg244.

Results of this kind are unprecedented.

Figure 5-17. *FUS*; missense changes Ser135Asn and Gly227Asp. PolyPhen-2 analysis

A



B

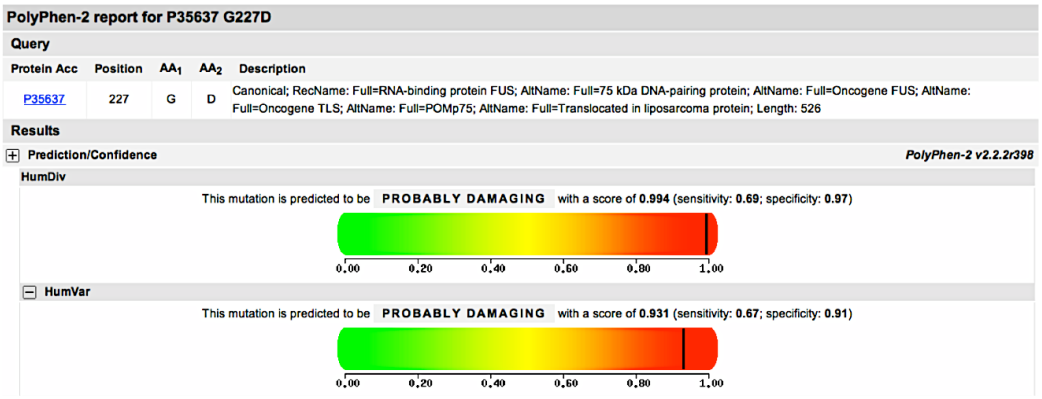


Figure 5-17. PolyPhen-2 predicted the missense change Ser135Asn in *FUS* being benign (**A**), whilst Gly227Asp (**B**) was predicted to be probably damaging. Considering that both missense changes were identified in neurologically normal controls, provided further proof, the prediction for Ser135Asn is expected, the one for Gly227Asp is surprising.

TDP-43

The screening of *TDP-43* resulted in the isolation of the Asn267Ser mutation in only one CBS patient (**Table 5-7**).

This mutation had been previously reported in an ALS [331, 374] and in a FTD patient without motor neuron disease [135]. However, PolyPhen-2 predicted this missense change to be fully benign (**Figure 5-18**). Functional studies will need to further support this prediction.

Figure 5-18. *TDP-43*; missense change Asn267Ser. PolyPhen-2 analysis

Figure 5-18. Polyphen-2 predicted the missense change Asn267Ser in *TDP-43* being benign.

5.2.2.3 Comment on variability in FUS and TDP-43

FUS

Our results clearly show a high rate of variability in the *FUS* gene, which makes it challenging to discriminate the pathogenic from the non-pathogenic variants unless there is unequivocal proof of co-segregation with disease in familial cases and/or there is support of pathogenicity through convincing functional data. Indels in *FUS* have been reported to cause fALS [142, 143], but it is difficult to argue through what mechanism the same variants would exert pathogenic effect in cases and protective effect in neurologically normal controls unless there are modifying factors yet to be determined that drive the effect of these CNVs. Thus, these variants may not be fully penetrant or

have small effect size. As a matter of fact, none of the variants we identified in *FUS* could be unequivocally labelled as pathogenic. However, the relevance of these findings lies most probably in the fact that they highlight the polymorphic nature of *FUS* driven by greater variability than what was previously thought and encompasses heterogeneous phenotypes including ALS (as originally shown), FTD, CBS as well as the normal population.

TDP-43

To date, pathogenic variants in *TDP-43* have been more common in ALS and FTD with motor neuron disease than in bvFTD or in the language variants, although TDP-43 protein has been found in the majority of FTLD-U positive brains [131-133, 135, 331, 376-378]. In this study, the screening of a robust cohort of FTLD and CBS/CBD patients resulted in the identification of only one variant in a CBS patient; the missense change Asn267Ser in exon 6 of *TDP-43* that had previously been reported in one ALS and one FTD patient [135, 374]. It needs to be said that among the FTD cases, few had overlapping symptoms with ALS. The current is the first report of this variant in a case diagnosed with CBS. The frequency rate of this variant, however, has been proven to be, overall, relatively low: 1/666 in ALS (0.15%) [374], 1/252 (0.4%) in FTD [136], whereas it was slightly higher in our CBS/CBD cohort (1/70, 1.4%) [137] but most probably because of the smaller sample size. Thus, the Asn267Ser missense change appears to be relatively rare and, although it has not been isolated in the normal

population [136], there is still no clear proof of pathogenicity because of lack of functional studies. All the more, although this is far from being conclusive, PolyPhen-2 predicted it being fully benign suggesting that, currently, the pathogenicity of Asn267Ser needs to be further verified.

5.2.2.4 Summary

In conclusion, this study supports the idea that genetic variability in *FUS* and *TDP-43* is associated with a broad range of phenotypes including neurologically normal controls (in the case of *FUS*) with a frequency that is unprecedented for candidate genes. In addition and, especially, given that these results for the most part contradict previous findings about the pathogenicity of *FUS* and, possibly, *TDP-43* variants, examination of more patients with the diagnoses of the extended FTLD spectrum as well as functional studies may be warranted to elucidate the real effects of these variants.

5.2.3 FTD and C9orf72

5.2.3.1 Introduction

There is evidence that ~15% of FTLD patients develop symptoms that are typical of motor neuron disease (MND) [379] and, conversely up to 50% of cases diagnosed with

ALS show signs typical of FTD [371]. In addition, TDP-43 inclusions represent a common and overlapping pathological hallmark in both FTLD and ALS cases [371, 380].

Several linkage and association studies of ALS, FTD and FTD-ALS cases revealed association with a locus on chromosome 9p [78, 79, 81, 82, 381]. Later, a hexanucleotide GGGGCC repeat expansion was identified in the first intron of *C9ORF72* in FTD-ALS families [76, 77] suggesting this expansion as the most probable reason underlying that association and as a genetic contributing factor to the FTLD-ALS spectrum.

In the current study the goal was to evaluate the presence, frequency and distribution of the GGGGCC hexanucleotide expansion in 53 of our FTLD patients and 174 neurologically normal controls from the Coriell Institute (plates NDPT 098 and 099). These samples are a subset of the 158 FTLD patients mentioned in section ***5.2*** and were included in the FTD-GWAS (**Chapter 4**).

These 53 FTLD cases were all recently (2012) re-evaluated by Dr Huey based on the most recent and revised diagnostic criteria [20, 21, 27, 382]. The clinical features resulted to encompass a wide range of the FTLD syndromes: 27 cases were probable bvFTD, 11 were possible bvFTD, 6 were PPA-PNFA, 2 were PPA-semantic, and 4 FTD-ALS. In addition, 2 eventually resulted being AD and 1 multiple system atrophy (MSA).

The updated diagnoses together with the *C9orf72* repeat sizes and a brief family history of each of the 53 patients used in this targeted genetic screening are summarized in **Table 5-9**.

Table 5-9. *C9orf72* expansion screening and samples characteristics

Patient ID	Diagnosis	Repeat #	Family history of neurological disorders
158	**Probable bvFTD**	**> 50**	**Father: AD; Brother: MS**
198	**Probable bvFTD**	**> 45**	**Brother: TIA; Brother: Alcoholism; Daughter: Migraines; Paternal aunt: Dementia**
211	**FTD-ALS**	**> 50**	**Mother: ALS; Brother: Alcoholism; Father: Dementia; Brother/Son: Depression; Daughter: Migraines, depression**
223	**FTD-ALS**	**> 50**	**Mother: Dementia; Father: Dementia; Sister: FTD; Brother: Alcoholism; Son: Depression**
81	Probable bvFTD	1	Father: Died of dementia at 77
83	AD	7	Paternal grandfather: Dementia
85	FTD-PPA	7	Father: Vascular dementia died at 68
88	FTD-PPA	14	Sister: PD
89	Probable bvFTD	8	Mother: LOAD; Paternal grandmother: Dementia died at 89
90	Probable bvFTD	4	Father: CVA; Nephew: OCD
95	SD	4	Maternal aunt: Dementia
96	Probable bvFTD	~ 25	Father: AD onset 58; Maternal uncle: Schizophrenia
97	FTD-PPA	3	Son: Schizophrenia
98	Possible FTLD	15	Sister: CVA
99	Probable bvFTD	10	Father: PD died at 85; Mother: Dementia, brain tumour; 3 aunts: Dementia; Maternal grandmother: Dementia
101	Probable bvFTD	3	Son: Schizophrenia, drug abuse; Daughter: Schizophrenia; Sister and Brother: Drug abuse
103	SD	5	Maternal grandmother: CVA; Maternal aunt: CVA
107	Probable bvFTD	1	Paternal uncle: Unknown mental illness
111	Possible FTLD	2	Sister: Korsakoff syndrome; Mother: OCD, depression; Sister: Alcoholism; Twin brother: Alcoholism
114	FTD-PPA	2	Father: AD, PD died at 63; 2 brothers: Schizophrenia
115	Probable bvFTD	6	Father: Dementia onset at 72; paternal grandmother: PD and memory problem died in her 70"s
118	Probable bvFTD	6	Father: AD; Mother: AD, depression both died in their early 90's
123	Probable bvFTD	8	Brother: Learning disabilities, alcoholism; Sister: Eating disorder, depression
124	Probable bvFTD	7	Father: PD; aunt: AD died at 79

Table 5-9 (continued)

Patient ID	Diagnosis	Repeat #	Family history of neurological disorders
127	MSA	1	Mother: AD; Brother: CVA; Brother: Brain cancer
128	Probable bvFTD	3	Father: OCD; Mother: Mental disorder, NOS
134	FTD-MND	1	Not available
136	Probable bvFTD	2	Mother: Dementia at 87; maternal uncle: Dementia
139	FTD-PPA	10	Mother: Depression, alcoholism
143	Probable bvFTD	17	Daughter: OCD, depression
144	Probable bvFTD	5	Not available
163	FTD-MND	9	Paternal uncle: Dementia, alcoholism
176	FTD-PPA	2	Mother: Depression; Daughter: ADD
177	AD	3	Mother: LOAD; Father: Alcoholism; Son: Alcoholism
183	Possible FTLD	2	Father: PTSD and depression; mother: DLB, mutism; mat grandfather: PD and stroke; maternal uncle: DLB; 2 paternal aunts: MND
194	Probable bvFTD	3	Sister: Dementia; Father: Stroke; Mother: Depression; Son: Depression/anxiety
195	Probable bvFTD	6	Mother: Action tremor; Paternal grandmother: dementia; maternal great grandmother: Dementia
199	Possible FTLD	2	Father: Dementia; Daughter: Seizures; Maternal grandmother: Dementia
200	Probable bvFTD	6	Daughter: ADD, anxiety; Son: Substance abuse
201	Possible FTLD	4	Father: Dementia in his 60's, stroke; Sister: Dementia; Brother: Stroke; Brother: Epilepsy, stroke; Son: Epilepsy, mental illness; grandfather dementia in his 60's
203	Probable bvFTD	3	Paternal grandmother: Late onset dementia at 91; Mother: MS; maternal grandmother Parkinson's onset at 78; maternal great aunt and uncle: LOAD
205	Possible FTLD	10	2 Sisters: Anxiety; Sister: PSP; Daughter: Eating disorder; Daughter: Anxiety, depression; Daughter: Schizophrenia; Son/2 Daughters: Depression
207	Possible FTLD	7	Siblings: Tremor, anxiety, alcoholism, substance abuse; mother died at 83 had dementia and anxiety; father died at 79 had Parkinson's and anxiety.
209	Probable bvFTD	7	Father: Stroke; Brother: Born whole in skull; Daughter: Seizure disorder
210	Possible FTLD	7	Maternal grandfather: Dementia
212	Probable bvFTD	5	Not available
215	Possible FTLD	2	Mother: AD; Father: Dementia; Brother: Depression; Son: Arthrogyrosis

Table 5-9 (continued)

Patient ID	Diagnosis	Repeat #	Family history of neurological disorders
216	Possible FTLD	9	Sister: Mild stroke, depression
217	Probable bvFTD	2	Father died at age 82 no dementia but paternal grandfather died of dementia at 81
218	Probable bvFTD	~ 25	Mother: Dementia; Son: ALS; Maternal aunt: AD
219	Probable bvFTD	2	Daughter: Anxiety, depression, neurocardiogenic syndrome; Brothers: Depression; Sister: Anxiety, depression, ADHD; Father: Dementia; Mother: Stroke
221	Probable bvFTD	10	Father: Vascular dementia at 78
222	Possible FTLD	2	Father: AD, PD, heart disease; Brother: Alcohol abuse

Table 5-9. Sample IDs with corresponding diagnosis, number of hexanucleotide expansion and a brief family history of neurological disorders for each patient. All diagnoses are updated based on [20, 27]. **Abbreviations**: AD=Alzheimer's disease; MS=Multiple sclerosis; TIA=transient ischemic attack; ALS=Amyotrophic lateral sclerosis; PD=Parkinson's disease; LOAD=late onset Alzheimer's disease; CVA=cerebrovascular accident; OCD=obsessive compulsive disorder; NOS=not otherwise specified; ADD=attention deficit disorder; PTSD=post-traumatic stress disorder; DLB=dementia with Lewy bodies; MND=motor neuron disease; PSP=progressive supranuclear palsy; ADHD=attention deficit hyperactivity disorder.

5.2.3.2 Genetic screening

Candidate genes

The DNA samples of these 53 patients (**Table 5-9**) were characterised for *MAPT*, *PGRN*, *FUS*, *TDP43* and *CHMP2B*, *PSEN1*, *PSEN2* and *APP* (**Table 5-1**). However, as the main goal here was to evaluate the presence of the hexanucleotide expansions in the *C9ORF72* gene, we performed the RP-PCR experiments as described in [77] considering the putative pathogenic repeat range as ≥40, although more recent studies have indicated that such threshold is no longer entirely accurate and predictive of a "pathogenic" expansion [90, 219, 383]. The results of the screening of *C9orf72* are summarized in **Figure 5-19**.

Figure 5-19. Sporadic FTLD: *C9orf72* expansion analysis

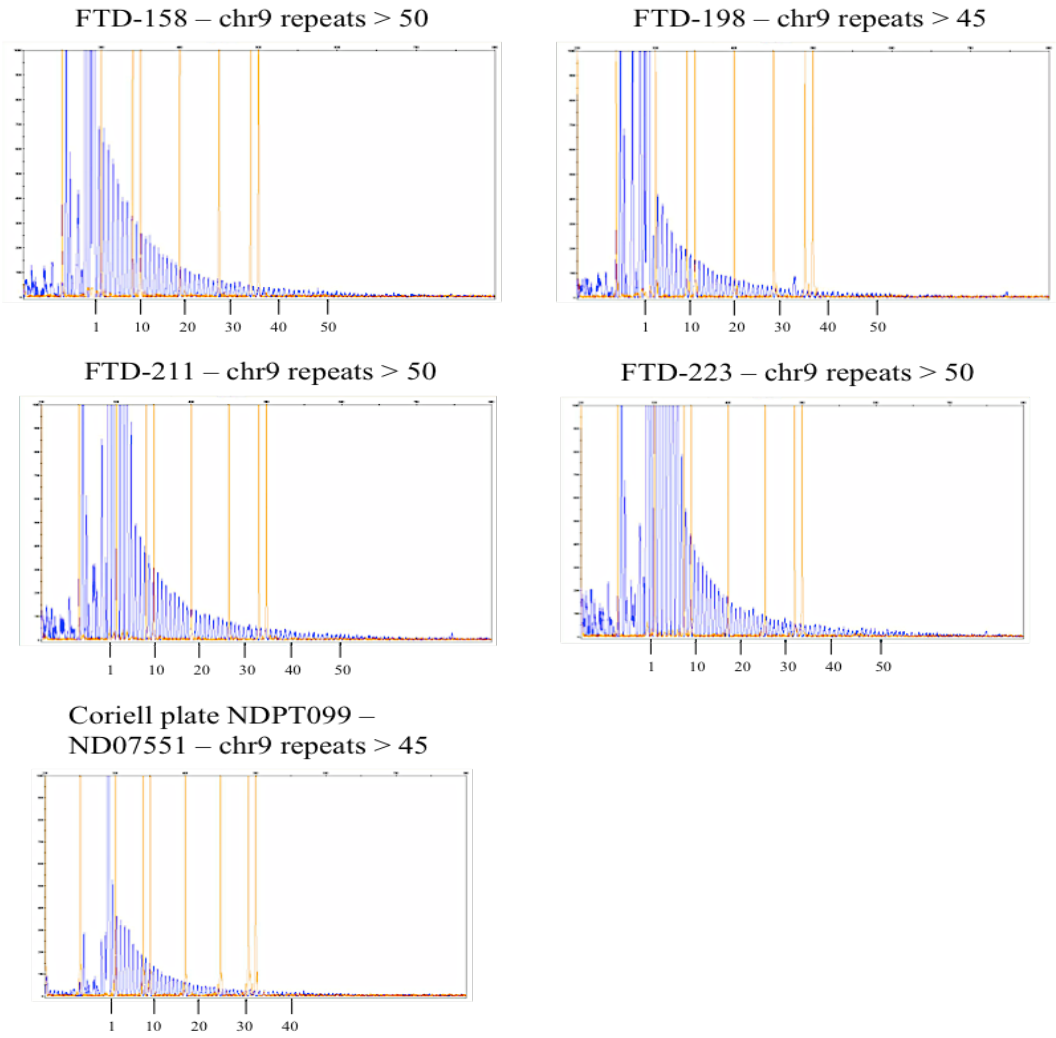


Figure_4-9a-d_Thesis_APRIL2014

Figure 5-19. The range of >30 repeats is considered as the threshold suggestive of presence of expansion even though a direct and precise relatedness between this method and the actual repeat size has not been established and/or implemented yet being the RP-PCR method, as such, simply predictive of probable repeat expansion. Counts of approximately >45 or >50 repeats are depicted for, respectively, samples FTD158 (**A**), FTD198 (**B**), FTD211 (**C**), FTD223 (**D**) and for Coriell neurologically normal control ND07551 (**E**).

Four of the 53 patients (FTD158, FTD198, FTD211 and FTD223) and one out of 174 neurologically normal controls (Coriell plate NDPT099; ID: ND07551; http://ccr.coriell.org/Sections/Search/Sample_Detail.aspx?Ref=ND07551&PgId=166) revealed presence of repeats in the putative pathogenic range (≥40) (**Figure 5-19**).

FTD158 was diagnosed with bvFTD and language signs, whilst both FTD198 and 211 were diagnosed with FTD-ALS. Finally, FTD223 was primarily diagnosed with bvFTD accompanied by signs of memory impairment.

It is of interest and noteworthy to mention that, among the carriers of the hexanucleotide expansion, one patient (FTD158) also carried a novel *GRN* mutation Tyr294Cys (**Figure 5-20A**) and a second patient (FTD223) carried a novel mutation in *PSEN2* Ile146Val (**Figure 5-20B**). None of these mutations have been previously reported and their identification represents an important piece of information as they indicate that different mutations that are all potentially pathogenic can co-occur in FTD. Conversely, this also suggests that, possibly, the penetrance of any or all of these variants may not be complete and that each may rather act as modifiers. However, prior to speculating on the pathogenicity of these variants it needs to be mentioned that the *GRN* missense change was absent from the normal population [327, 384], whilst the *PSEN2* missense change was not identified in the normal population (n=174; Coriell plates NDPT 098 and NDPT 099).

Figure 5-20. *GRN*: Chromatogram of Tyr294Cys; *PSEN2*: Chromatogram of Ile146Val

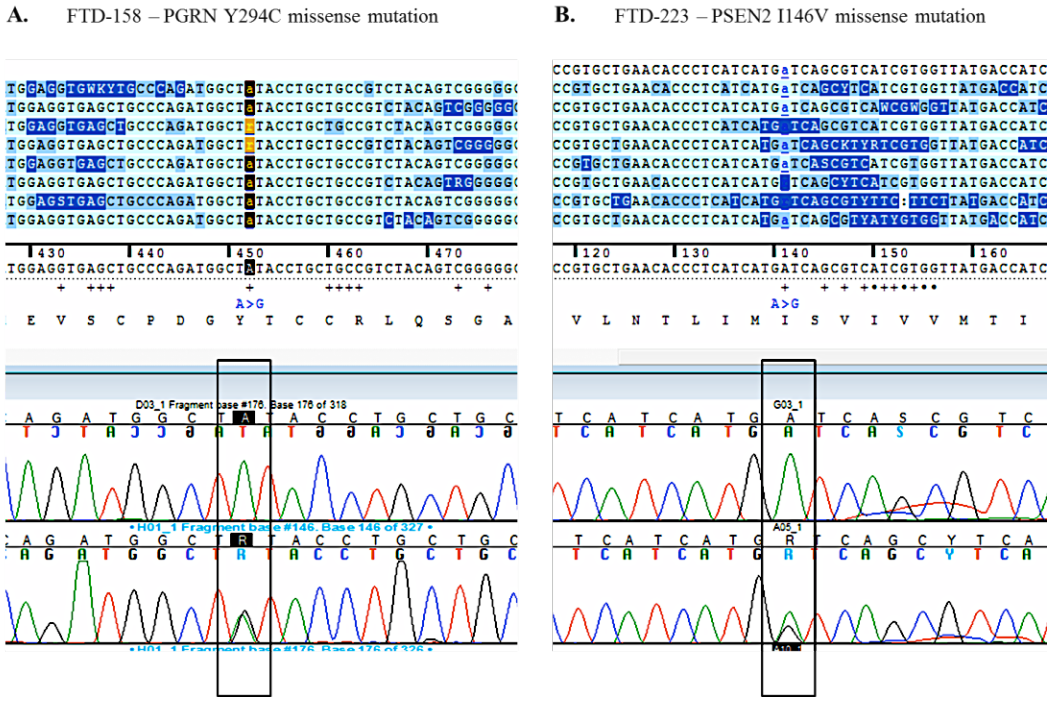
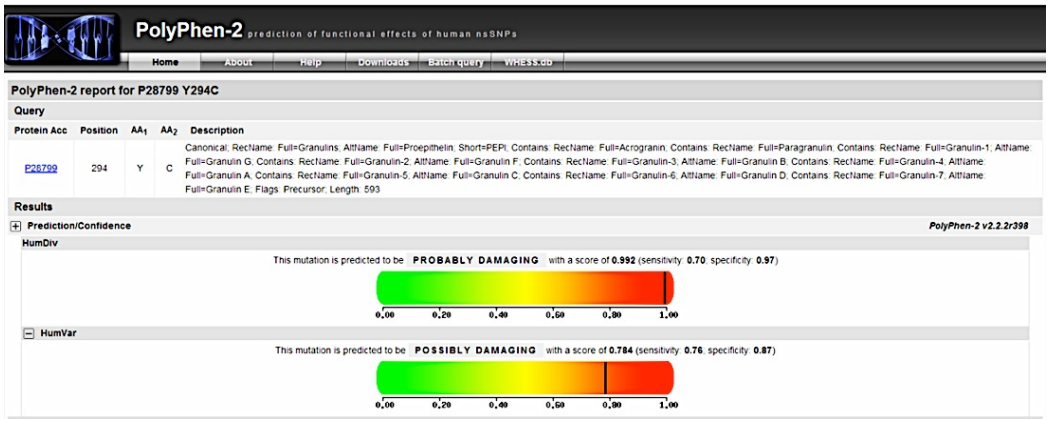


Figure 5-20. Electropherograms of the two novel missense mutations found in *GRN* (**A**) and in *PSEN2* (**B**) in two of the hexanucleotide repeat carriers, respectively individuals FTD158 (**A**) and FTD223 (**B**). Both missense mutations are novel.

In the absence of functional data, we examined, *in silico*, the prediction on their effect on the protein structure and function through the PolyPhen-2 software [225]: The *GRN* variant was predicted to be probably damaging, whilst the *PSEN2* variant was predicted to be benign (**Figure 5-21A** and **B**).

Figure 5-21. *GRN*: Missense change Tyr294Cys; *PSEN2*: Missense change Ile146Val. PolyPhen-2 analysis

A *GRN* Tyr294Cys



B *PSEN2* Ile146Val

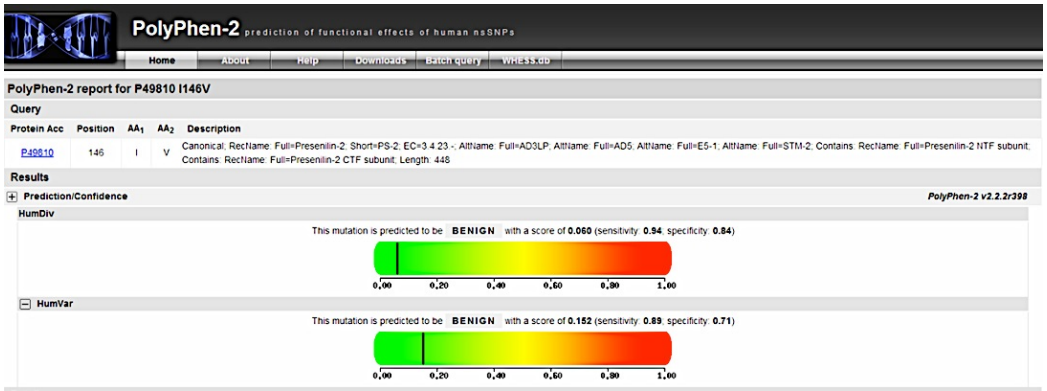


Figure 5-21. PolyPhen 2 effect prediction shows that the *GRN* missense change Tyr294Cys is probably pathogenic, whilst the *PSEN2* missense change Ile146Val is predicted being likely benign.

Finnish chromosome 9 haplotype

I genotyped the above mentioned 53 DNA samples (**Table 5-9**) as part of the FTD-GWAS (**Chapter 4**), therefore the genotypes of the 42 SNPs which identify the chromosome 9p ALS-associated Finnish haplotype [78] could be derived from the GWAS and could be used to determine whether the carriers of the expansion also harboured the same ancestral haplotype [340].

Our genotyping data revealed that the complete risk haplotype was not present in any of our patients. However, considering the SNP rs3849942 (suggested as surrogate marker for the risk haplotype [77]), we were able to verify that all the 4 patients with expansions (FTD158, FTD198, FTD211 and FTD223) did carry the risk allele A for this SNP (**Table 5-10**). Nevertheless, also 20 of the remaining 49 patients without expansion did carry the A-risk allele for rs3849942. Of note, in our patients series there were two individuals diagnosed with FTD-MND (FTD134 and FTD163): Patient FTD134 did not carry neither the expansion nor the risk allele A (rs3849942), whilst FTD163 did not harbour the expansion but the risk allele (**Table 5-10**). These data suggest that the complete chromosome 9p risk haplotype, including the expansion, is possibly a specific characteristic of the Finnish population, whereas recombination events in mixed populations (although of mainly European ancestry as in the case of the US) rearranged the SNPs series on that haplotype.

Table 5-10. Finnish *C9orf72*-risk haplotype analysis

SNPs	Risk allele	Patient					
		158	**198**	**211**	**223**	134	163
rs3849942	A	**AG**	**AG**	**AA**	**AA**	GG	**AA**
rs1330921	T	**TT**	**TT**	**TT**	**TT**	**TT**	**TC**
rs10121765	C	**AC**	AA	AA	**AC**	**AC**	**CC**
rs1110264	A	GG	GG	GG	**AG**	GG	GG
rs1110155	A	**AA**	**AA**	**AA**	**AC**	**AA**	**AA**
rs2150336	C	TT	TT	TT	**TC**	TT	TT
rs2225389	C	AA	AA	AA	AA	AA	AA
rs1161680	G	AA	AA	AA	**AG**	AA	**AG**
rs2120718	G	AA	AA	AA	**GG**	**GG**	**GG**
rs2589054	G	**GG**	**GG**	**GG**	AA	AA	AA
rs10812596	A	**AA**	**AA**	**AA**	**AG**	**AA**	**AG**
rs1058326	C	**CC**	**CC**	**CC**	**TC**	TT	**TC**
rs944404	T	**TT**	**TT**	**TT**	**TC**	**TT**	**TC**
rs765709	A	**AA**	**AA**	**AA**	**AA**	**AA**	**AC**
rs1316679	G	**GG**	**GG**	**GG**	**GG**	**GG**	**AG**
rs4406503	G	**GG**	**GG**	**GG**	**GG**	**GG**	**AG**
rs10511817	C	**CC**	**CC**	**CC**	**CC**	**CC**	**AC**
rs725804	A	**AA**	**AA**	**AA**	CC	**AC**	**AC**
rs10511816	T	**TG**	GG	**TG**	**TG**	**TG**	**TT**
rs1444533	A	**AA**	GG	**AA**	**AA**	**AA**	**AA**
rs1822723	C	**TC**	**CC**	**CC**	**CC**	**TC**	**CC**
rs4879515	T	**TT**	CC	**TT**	**TT**	CC	**TT**
rs895023	T	**TT**	**TT**	**TT**	**TT**	**TT**	**TT**
rs868856	T	**TC**	**TC**	**TT**	**TT**	CC	**TT**
rs7046653	A	**AG**	**AG**	**AA**	**AA**	GG	**AA**
rs2440622	A	**AA**	**AA**	**AA**	**AA**	**AA**	**AA**
rs1977661	C	**AC**	**CC**	**CC**	**CC**	**CC**	**CC**
rs903603	C	**TC**	**TC**	**CC**	**CC**	TT	**CC**
rs10812610	C	**AC**	**AC**	**CC**	**CC**	AA	**CC**
rs2814707	A	**AG**	**AG**	**AA**	**AA**	GG	**AA**
rs12349820	T	**TT**	**TT**	**TT**	**TT**	CC	**TT**
rs10122902	G	**GG**	**GG**	**GG**	**GG**	**GG**	**GG**
rs10757665	T	**TT**	**TT**	**TT**	**TT**	CC	**TT**
rs1565948	G	**GG**	**GG**	**GG**	**GG**	AA	**GG**
rs774359	C	**TC**	**TC**	**CC**	**CC**	TT	**CC**
rs2282241	G	**TG**	**TG**	**GG**	**GG**	**GG**	**GG**
rs1948522	C	**CC**	**CC**	**CC**	**CC**	**CC**	**CC**
rs1982915	G	**GG**	**GG**	**GG**	**GG**	AA	**GG**
rs2453556	G	**GG**	**AG**	**GG**	**GG**	AA	**GG**
rs702231	A	**AA**	**AC**	**AA**	**AA**	**AA**	CC
rs696826	G	**GG**	**GG**	**GG**	**GG**	**GG**	AA
rs2477518	T	**TT**	**TT**	**TT**	CC	**TT**	CC

Table 5-10. The 42 SNPs building the Finnish ALS risk haplotype and the distribution of genotypes among the expansion carriers and two further FTD-MND cases (134 and 163) of our cohort are shown. None of the expansion carriers (158, 198, 211, 223 in bold) held the complete series of the risk alleles (highlighted in orange). When considering only the 21 of 24 SNPs (due to chip design; highlighted in yellow) of the shorter haplotype [340], only 158, 198 and 223 held at least one of the risk alleles. None of the two FTD-MND cases (134 and 163) did fully carry the risk allele in neither the 42 nor the 21 SNPs series.

5.2.3.3 Comment on C9orf72 in this study cohort

It is noteworthy to consider that the FTLD cases analysed in this study were recruited within the United States (US) and that they were not explicitly selected on the basis of family history, pathogenic mutations, or presence of MND features. As such, these samples likely represent the average US FTLD population characterised by substantial diversity in the ethnic and genetic background.

In the current study, the expansion was identified in 4/53 (7.5%) cases, replicating, despite a rather small sample size, the original findings [76, 77]. Furthermore, our results also provided insight on a number of other aspects that can be summarised as follows:

1 – There was not a linear correlation between the presence of expansion and the previously reported chromosome 9p risk haplotype [78, 340];

2 – There was co-occurrence of the repeat expansion and missense changes in other dementia candidate genes in 2/4 cases;

3 – The expansion was isolated also in neurologically normal controls, and;

4 – One case presented heterogeneous clinical and pathological signatures indicative of co-existence of various co-morbidities.

The first point cannot be considered entirely surprising, especially, when compared to the study on the Finnish FTD kindred (section ***5.1.2***), suggesting that the haplotype block on chromosome 9p [78] is possibly population specific (Finnish/Northern European

populations), whereas it is not necessarily associated with the hexanucleotide expansion or shared among populations worldwide. For example, the chromosome 9p risk haplotype block was also absent in a large Italian ALS cohort [385].

The second point is, in fact, of interest as it highlights co-existence of several potential pathogenic mutations. In this study, *GRN* (Tyr294Cys) and *PSEN2* (Ile146Val) missense changes co-occurred with the expansion. Tyr294Cys was predicted to be pathogenic by PolyPhen-2 and the carrier had an earlier age of onset compared to the other three expansion carriers (53 vs. ~60 years of age). On the other hand, PolyPhen-2 revealed that Ile146Val may be benign. Although the Ile146Val *PSEN2* missense change was not identified in 174 neurologically normal controls functional studies are still warranted to fully confirm its pathogenicity. As much as this rather reflects a rare scenario, other mutations were reported in concomitance of the *C9orf72* expansion as in the cases of *MAPT* [98, 386], *GRN* [285, 386], and *TDP-43* [385, 387, 388] to name the most relevant. This not necessarily only underlies heterogeneous clinical and pathological features but also suggests that in some cases the contribution to the diseased phenotype of either of the aforementioned genes may not be fully penetrant.

The third point emphasises the need for more careful assessment of the penetrance of the *C9orf72* expansion. Clearly, the identification of the repeat expansion in the normal population is a rare phenomenon (<1%); however, it has been observed across various studies [77, 86, 87, 285, 383]. The repeat expansion occurs with higher frequencies in the cases than in the controls. This is indicative of its most probable contribution to the

development of disease; however, at the same time, it is suggestive of probable variable penetrance [294, 389]. All the more, although this is not likely the case with *C9orf72*, as previously shown with the *FUS* gene (section ***5.2.2***), it is possible that variants initially reported as pathogenic [142, 143] are subsequently found also in normal controls which brings their relevance/pathogenicity into question.

The fourth point suggests the need for a critical evaluation of the wide variety of phenotypic and pathological signatures that have been found associated with the expansion. Previous reports showed that the expansion is mainly detected in FTD-MND cases [76, 77, 84, 390] and that expansion carriers are likely to present memory and psychiatric symptoms [85, 391, 392]. One of our expansion carriers presented with paranoia as the predominant symptom. Two other expansion carriers showed significantly impaired memory functions and several had AD-like features [94, 393-397]. Of note, FTD211, the one patient whose autopsy report was available, showed amyloid plaques and tangles in addition to TDP-43 pathology, despite being only 65 years old at the time of death. This replicated earlier observations by Murray et al who described an expansion case with concomitant TDP-43 and AD pathology [92].

5.2.3.4 Summary

In conclusion, the findings of this study suggest that the hexanucleotide expansion is most frequently associated with the FTD-ALS spectrum, but also occasionally with a

broader and more heterogeneous group of phenotypes possibly including some clinical and pathologically confirmed AD [92, 394], Parkinsons' disease [398], CBS/CBD, PSP [85, 399] cases and neurologically normal controls.

The findings of this current study need to be interpreted with caution and more importantly, replicated in larger numbers of samples.

5.3 Final remarks

The study of two familial cases and extended screening of a robust number of sporadic cases diagnosed with FTLD and CBS/CBD contributed to the generation of valuable knowledge of the genetics of these diseases, especially in relation to the candidate genes:

1 – The Afrikaaner family (section ***5.1.1***) confirmed the complexity of the disease in that an apparent dominant mode of inheritance associated with tau pathology did not reveal any genetic aberrations in the coding regions of *MAPT* or in its dosage. This has been observed before [25], suggesting that the involvement of tau in neurodegeneration, including FTD, may be due to genetic variability within the *MAPT* gene or its associated extended haplotype, or it may be driven by genetic defect(s) to be found in other genetic factors that contribute or lead to tau pathology independently from MAPT abnormalities. Mutations in *MAPT* exon 10 or affecting exon 10 splicing are known to directly affect the function and/or the ratio of isoforms with three or four of the microtubule binding repeat domains, leading to tau pathology [29]. Nevertheless, there is evidence of tauopathies, such as PSP or CBS/CBD, in which the link to chromosome 17 and *MAPT* seems to be driven by haplotype effects rather than by pathogenic single nucleotide coding variants [179, 199]. The variations in *FUS* and *CHMP2B* that we identified in this family are of interest as they raise question about the pathogenicity and/or penetrance of these genes. The *FUS* missense change was predicted to be probably damaging by PolyPhen-2 and, normally, truncation mutations (as in the case of *CHMP2B*) are considered pathogenic

based on the potential inactivation of the protein. However, variants in both genes were identified in unaffected members of the family, and were apparently inherited from the unaffected parents who were married into the family (**Figure 5-2**);

2 – The Finnish kindred (section ***5.1.2***) highlighted specific characteristics of FTD: First, a certain heterogeneity of the clinical features at presentation and during the progression of the disease in *C9orf72* expansion carriers; Second, an almost complete LD between the expansion in *C9orf72* and the extended Finnish-risk haplotype [78] indicating this association being clearly population specific, and; Third, the pathological signatures associated with the expansion are highly replicated across different studies;

3 – The *CHMP2B* missense change (section ***5.2.1***) was non-pathogenic (PolyPhen-2 prediction, confirmed by functional study), suggesting that *CHMP2B* is highly polymorphic, with the majority of the variants being non-pathogenic (section *5.2.1.2*). Of note, based on our screening of *CHMP2B* in both a familial and a sporadic case, we concluded that none of the *CHMP2B* variants that we identified were pathogenic and/or associated with disease;

4 – The screening of *FUS* revealed that this gene is highly polymorphic (section ***5.2.2***). We provide evidence that variants previously reported as pathogenic and causing familial ALS [142, 143] are also common in the normal population (**Table 5-7** and **5-8**). In addition, we identified a novel missense change, predicted to be pathogenic (**Figure 5-17B**), in one neurologically normal control. These findings put into question the pathogenicity or penetrance of variants in *FUS*;

5 – The screening of *TDP-43* confirmed that variability in this gene is rare in FTLD (section ***5.2.2***). We, in fact, identified just one missense change, which was previously reported in ALS [331, 374] and FTLD [136], in a CBS case of our cohort. Although this suggests that *TDP-43* mutations could also cause CBS, PolyPhen-2 predicts this variant is benign. Nevertheless, as this variant is absent in 400 neurologically normal controls [136], functional studies on its effect are needed to better elucidate its pathogenicity, and;

6 – The investigation of *C9orf72* (section ***5.2.3***) revealed a highly heterogeneous picture including almost complete absence of the chromosome 9p risk haplotype in populations other than Finnish, co-occurrence of the expansion with mutations in other candidate genes and presence of diverse pathological co-morbidities (i.e. TDP-43 inclusions and amyloid plaques) in expansion carriers.

Of note, PolyPhen-2 is valuable for the prediction of potential pathogenicity of missense changes. Nevertheless, it is necessary to carry out functional studies to conclusively assess the effects of a missense change. Therefore, apart from the *CHMP2B* missense mutation Ser187Asn that PolyPhen-2 predicted to be benign, and confirmed by functional study, any other prediction that has been evaluated and reported in this chapter should be interpreted with caution in the absence of *in vitro* and/or *in vivo* functional studies.

In summary, this work has provided insight into the disease by confirming its complexity and heterogeneity in the different clinical, pathological and genetic aspects. In addition, these findings suggest that possibly some variants are not pathogenic as previously thought or have variable penetrance and emphasise the need to direct our attention to the study of splicing and expression modulation as well as to the identification of modifying factors that might contribute to disease onset and development with variable effect size.

The latter represent critical aims to be pursued in the FTLD research environment for the future.

The work presented in this chapter was performed by the following individuals (or groups):

Raffaele Ferrari: Performed genetic screening (sequencing) of the candidate genes, allelic discrimination and gene dosage experiments, and LOH, CNV and haplotype analyses; **Parastoo Momeni**: Performed genetic screening (sequencing) of the candidate genes; **Liisa Myllykangas**: linkage analysis of Finnish kindred; **Pentti Tienari** and **Mia Kero**: brain and molecular pathology analyses of Finnish kindred.

Publications derived from the work presented in this chapter:

Raffaele Ferrari, Mia Kero, Kin Mok, Anders Paetau, Tero Hiekkalinna, Pentti Tienari, Olli Tynninen, Hannu Kalimo, John Hardy, Parastoo Momeni, Auli Verkkoniemi-Ahola, Liisa Myllykangas. ***FTD kindred associated with C9orf72 repeat expansion and dysplastic gangliocytoma***. Neurobiol Aging, 2014 Feb;35(2):444.e11-4.

Raffaele Ferrari, Kin Mok, Jorge H Moreno, Stephanie Cosentino, Jill Goldman, Pierto Pietrini, Richard Mayeux, Michael C Tierney, Dimitrios Kapogiannis, Gregory A Jicha, Jill R Murrell, Bernardino Ghetti, Eric M Wassermann, Jordan Grafman, John Hardy, Edward D Huey and Parastoo Momeni. ***Screening for C9ORF72 repeat expansion in FTLD***. Neurobiol Aging, 2012 Aug;33(8):1850.e1-11.

Huey ED, **Ferrari R**, Moreno JH, Jensen C, Morris CM, Potocnik F, Kalaria RN, Tierney M, Wassermann EM, Hardy J, Grafman J, Momeni P. ***FUS and TDP43 genetic variability in FTD and CBS***. Neurobiol Aging, 2012 May;33(5):1016.e9-17.

Raffaele Ferrari, Dimitrios Kapogiannis, Edward D Huey and Parastoo Momeni. ***FTD and ALS: a tale of two diseases***. Curr Alzheimer Res. 2011 May;8(3):273-94.

CHAPTER 6 – Discussion and conclusions

6.1 Synopsis

The work presented in this thesis primarily explored the genetic underpinnings of progressive supranuclear palsy (PSP) and frontotemporal dementia (FTD) and contributed to their characterisation through the assessment of newly identified PSP loci and identification of novel disease associated loci in FTD that point to new possible disease pathways and mechanisms, opening new horizons for the study of these disorders.

The genetic analysis of PSP was designed and performed as a follow-up study based on the results of a recent PSP-GWAS [180], whereas the study of FTD (FTD-GWAS) was a pioneering endeavour aimed at identifying novel loci contributing to increased risk of developing the disease with small to moderate effect size.

In addition, this work provides an extensive study of the main candidate genes in familial FTD as well as sporadic FTLD and CBS/CBD cases. The presented results confirmed, on one hand, the extremely heterogeneous nature of this neurodegenerative disorder and, on the other, highlighted that the known candidate genes need to be cautiously assessed as they seem to contribute to disease with variable degrees of penetrance.

6.1.1 PSP

The study of a cohort of 84 pathologically confirmed PSP cases (**Chapter 3**) allowed to investigate the newly identified loci [180] in a well-defined and informative cohort. The study focused specifically on the loci surrounding the SNPs that showed highest association in the PSP-GWAS [180]. The characterisation of these loci was performed by sequencing of the genes *STX6*/*MR1*, *EIF2AK3* and *MOBP*, analysing the haplotype architecture including and surrounding the associated SNPs (within +/- 15-35kb) and by evaluating possible effects on expression exerted by the associated SNPs and by any of those SNPs in LD with the GWAS associated SNPs.

The load of information that was generated has some limitations that need to be acknowledged, in that, the sample size was relatively small (n=84) contributing to a certain decrease in the power of the study, and the expression data could only be assessed for *cis* but not for *trans* effects within a +/- 1Mb surrounding each SNP. Nevertheless, relevant information could be derived from this study.

The data indicated that it is unlikely that coding variants underlie association at any of the novel loci associated with PSP because the frequency of coding variability was too low to justify association in this manner within our study cohort. This is likely to be as well the case of the large cohort that was used to generate the GWAS data [180]. It is possible that highly pathogenic mutations in the genes within the PSP-associated loci may hold relevance in disease pathogenesis just in single or familial cases, but they do not seem to be a common cause of disease. Moreover, the minor allele frequencies of the SNPs in our

study cohort did not differ from that of the normal population (assessed through comparison with data from the 1000 Genomes project or dbSNP on the Caucasian population). This was also reflected in our haplotype analysis, in that the LD patterns observed at each locus in the PSP population were almost entirely reflecting those reported in the normal population. However, based on pairwise LD analysis of the SNPs at the *EIF2AK3* locus, there are 4 haplotypes (A, B, C, D) (**Table 3-5**), of which haplotype B has been associated with stress response mediated by the UPR [261]. Data for our cohort reflect that of a recent manuscript [262] that evaluated >1,000 pathologically confirmed PSP cases (i.e. part of those that were included in the PSP-GWAS [180]). Such outcome is important for, at least, two reasons: First, despite the difference in the cohort sizes (84 vs. >1,000), the frequency of the haplotype B at the *EIF2AK3* locus was comparable in both studies (ours and [262]) meaning that, although small, our study population is valuable and fairly representative of the PSP population at larger scale; Second, and probably more importantly, the haplotype data point to a cellular mechanism, such as that of the unfolded protein response (UPR) (**Chapter 3**, section *3.2.4.2*), being worthy of further investigation in genetic and functional studies in PSP for validation and for establishing its contribution to the pathogenesis of the disease.

Another important aspect of our study of the PSP cohort derives from the expression data, in that the minor allele (A) of the GWAS-associated SNP at the *STX6* locus (rs1411478) significantly associated in *cis* with expression levels (decreased expression) of *STX6* in white matter. White matter has been reported as among those brain areas that are affected in PSP. This finding ties the genetic association at the *STX6* locus to the

disease, pointing to a possible disease mechanism involving intracellular STX6-mediated vesicle trafficking in white matter. This hypothesis clearly needs to be further investigated and validated in functional studies to verify whether decreased levels of STX6 in white matter contribute to PSP pathogenesis.

The last point that deserves mention involves the fact that PSP is a tauopathy. It is intriguing to consider two facts: First, aberrant tau biology has been linked to the UPR-related markers in PSP and other tauopathies [263] and, second, tau pathology affects white matter in PSP. It follows that the interplay between *MAPT* H1 haplotype, tau pathology, the newly identified loci (specifically those encompassing *STX6* and *EIF2AK3*) and related cellular pathways represents an intriguing topic to be further investigated in the near future. Special attention needs to be paid to the possible interplay between tau and the UPR: Specifically, it will need to be elucidated whether tau pathology triggers EIF2AK3 activation or activated EIF2AK3 triggers tau pathology by causing tau hyperphosphorylation.

In summary, our study of PSP-GWAS loci is clearly not conclusive but indicates at least two possible and plausible disease mechanisms to be investigated in detail in the near future to better understand and characterise the molecular underpinnings that lead to the development of PSP.

6.1.2 FTD

6.1.2.1 FTD-GWAS

The study of FTD benefited from the collection and, subsequently, availability of an extremely large sample size gathered through an international consortium that was basically created through this thesis project. Two potential problems that needed to be addressed were the phenotypic characterisation and intrinsic population stratification due to samples originating from multiple countries in Europe, North America and Australia. However, the inclusion/exclusion criteria, as well as the bioinformatics and biostatistics measures used to filter samples have been extremely stringent in order to select the best possible samples for final inclusion in analysis.

Although this was not the first GWAS on FTD, given a previous study of pathologically confirmed FTLD-TDP (n=515) [41], our study included a large number of clinical FTLD cases (n>5,000 samples) comprising the subtypes bvFTD, SD, PNFA and FTD-MND (**Chapter 4**), making this an extraordinarily unique study of this genre.

Interestingly, none of the known FTD candidate loci reached genome-wide significance and this was, in fact, expected on one hand because all known *MAPT* and *GRN* mutation carriers were excluded from the study, eliminating or decreasing substantially the possible source of signal at 17q21 and, on the other hand, because, most possibly, all candidate genes and their risk alleles (including *MAPT* and *GRN*) were too rare within the

cohorts. This could also mean that those known candidate genes are rare causes of disease.

The study resulted in the identification of novel loci associated with disease: The two loci that reached genome-wide significance belonged to the bvFTD subtype (Chr11q14) and the whole cohort (Chr6p21.3), including all subtypes (bvFTD, SD, PNFA and FTD-MND). Not only were the associations significant after statistical corrections, but also quantitative trait loci for the associated SNPs revealed effects on expression and/or methylation (**Chapter 4**, section *4.2.2.3*). Although these loci will need to be further investigated they hint at possible disease mechanisms involved in FTD.

First, data point to lysosomes and phagosomes as two cellular compartments to be, from now on, considered as possibly involved in the pathogenesis of FTD and to be further investigated in both computational and functional studies; second, data suggest an involvement of the immune system in the development of disease.

Our study confirms previous hypotheses and provides new insights:

1 – Lysosomes have been suggested to play a role in FTD in association with *GRN* and *TMEM106B* variability. Our data not only support this hypothesis but clearly enhance and expand such notion including new genes within this picture, and;

2 – The immune system has been associated with several forms of neurodegenerative diseases and ours is the first report that indicates its likely involvement also in FTD.

In summary, these results are promising in two ways: On one hand, they highlight two mechanisms to be further investigated in the next years in order to understand their relation to FTD and, on the other, in turn, they provide unprecedented and critical information on possible disease mechanisms(s) to be studied and targeted for the development of therapeutic measures. At the same time, if disease mechanism(s) are elucidated not only treatment but also preventive options for FTD are likely to be established.

6.1.2.2 Candidate genes screening

The study of familial and sporadic FTD cases helped in further characterising the main FTD candidate genes and associating genetic variability with phenotype. Each study provided novel insight into the genetics of the disease (summarised in **Chapter 5**, section ***5.3***). Probably the main and most surprising outcome of this study is that it contradicts previous reports of pathogenicity of some variants. This may be due to variable penetrance and underlines the need to revisit previous genetic studies of *TDP-43*, *CHMP2B* and *FUS*, especially.

It is fundamental to continue characterising the known candidate genes in more cohorts not only to expand on genotype-phenotype correlation, but also to better characterise the incidence, prevalence and geographical distribution of the disease due to variability in these genes.

6.2 Future directions

Neurodegenerative diseases are insidious at onset and, generally, diagnosed when the patient is already years into disease.

This fact highlights probably the most critical issue associated with neurological disorders, i.e. the lack of biomarkers for detecting the disease at an early stage. In addition to that, preventive (upstream) and therapeutic (downstream) measures are still completely lacking in this field. Insights from the clinical, neuropsychological, neuroimaging and neuropathological perspectives, together with the increasing knowledge of the genetic underpinnings of disease, have undoubtedly led to ground-breaking advancements in the characterisation of neurodegenerative diseases. However, such knowledge has not been sufficient to date to address any of the aforementioned pressing issues.

It needs to be stated that, in the attempt of dissecting the aetiology of these diseases, one crucial point to keep in mind is that they are complex and multifactorial, i.e. the result of cumulative processes to which many factors contribute. The pathogenesis of complex disorders is, in fact, determined by the interplay between several elements including a gamut of genetic factors with different degrees of penetrance influenced by epigenetic and environmental factors. This not only suggests that it is recommended to critically combine and interpret the enormous and ever growing data, but also that it is warranted to expand the investigations to new fields in order to combine different fields of expertise and generate more innovative body of knowledge.

Thus, the investigation of neurodegenerative disorders will considerably benefit from the implementation of a multi-faceted approach based on the:

1 – Assessment of clinical (**a**) and neuropathological (**b**) features associated with disease to describe:

> **a** – The clinical syndromes, the physical/neuropsychological features and the associated neuroimaging profile;
>
> **b** – The atrophy patterns and associated functional loss, and the molecular pathology;

2 – Assessment of the genetics associated with disease through GWAS (**c**), WES (**d**) and WGS (**e**) approach to identify:

> **c** – Loci (specific genetic regions) associated with disease, i.e. genes as well as non-coding regions to be further investigated;
>
> **d** – Coding variants associated with disease, i.e. rare coding variants affecting the proteasome;
>
> **e** – Coding and non-coding single nucleotide and structural variations hidden in the genome, i.e. rare coding/non-coding SNVs as well as SVs affecting the transcriptome and the proteome;

3 – Assessment of transcriptome and proteome including the investigation of effects exerted by modifying factors such as regulatory elements and environmental factors,

and/or modifying mechanisms such as epigenetics and epistasis to evaluate effects on expression patterns, splicing processes and determine their impact on pathways that may underlie/lead to disease;

4 – Implementation of bioinformatics and biostatistics tools to improve and facilitate large scale computational analyses including functional genomics and pathway analyses to further examine and integrate data generated by the aforementioned different fields of expertise and identify potential disease-associated pathways/mechanisms, and;

5 – Design of pertinent functional studies aimed at validating and characterising, *in vitro* and *in vivo*, the computational findings, i.e. the pathogenic pathways and mechanisms leading to neuronal death and neurodegeneration.

This approach and these integrative steps will be fundamental in order to shed light on the disease-associated molecular underpinnings to be targeted for the development of measures to ensure effective early diagnostic/preventive and treatment options.

References

1. Sosa-Ortiz, A.L., I. Acosta-Castillo, and M.J. Prince, *Epidemiology of dementias and Alzheimer's disease.* Arch Med Res, 2012. **43**(8): p. 600-8.
2. Prince, M., et al., *The global prevalence of dementia: a systematic review and metaanalysis.* Alzheimers Dement, 2013. **9**(1): p. 63-75 e2.
3. Wimo, A., et al., *The worldwide economic impact of dementia 2010.* Alzheimers Dement, 2013. **9**(1): p. 1-11 e3.
4. Organization, W.H., *Dementia: A Public Health Priority*2012: WHO Regional Office for the Western Pacific.
5. *Diagnostic and Statistical Manual of Mental Disorders, 5th Edition: DSM-5*2013: American Psychiatric Publishing; 5 edition.
6. Berrios, G.E. and D.M. Girling, *Introduction: Pick's disease and the 'frontal lobe' dementias.* Hist Psychiatry, 1994. **5**(20 Pt 4): p. 539-47.
7. Pick, A., *Über die Beziehungen der senilen Hirnatrophie zur Aphasie.* Prag. med. Wschr., 1892. **17**: p. 165–167.
8. Pick, A., *Über primäre progressive Demenz bei Erwachsenen.* Prag. med. Wschr., 1904. **29**: p. 1-10.
9. Spatt, J., *Arnold Pick's concept of dementia.* Cortex, 2003. **39**(3): p. 525-31.
10. Alzheimer, A., *Über eigenartige Krankheitsfälle des späteren Alters.* Zeitschrift für die gesamte Neurologie und Psychiatrie, 1911. **4**: p. 356-385.
11. Murayama, S., et al., *Immunocytochemical and ultrastructural studies of Pick's disease.* Ann Neurol, 1990. **27**(4): p. 394-405.
12. *Clinical and neuropathological criteria for frontotemporal dementia. The Lund and Manchester Groups.* J Neurol Neurosurg Psychiatry, 1994. **57**(4): p. 416-8.
13. Kurz, A. and R. Perneczky, *Neurobiology of cognitive disorders.* Curr Opin Psychiatry, 2009. **22**(6): p. 546-51.
14. Hodges, J.R., et al., *Clinicopathological correlates in frontotemporal dementia.* Ann Neurol, 2004. **56**(3): p. 399-406.
15. Ratnavalli, E., et al., *The prevalence of frontotemporal dementia.* Neurology, 2002. **58**(11): p. 1615-21.
16. Rabinovici, G.D. and B.L. Miller, *Frontotemporal lobar degeneration: epidemiology, pathophysiology, diagnosis and management.* CNS Drugs, 2010. **24**(5): p. 375-98.
17. Seelaar, H., et al., *Clinical, genetic and pathological heterogeneity of frontotemporal dementia: a review.* J Neurol Neurosurg Psychiatry, 2011. **82**(5): p. 476-86.
18. Snowden, J.S., D. Neary, and D.M. Mann, *Frontotemporal dementia.* Br J Psychiatry, 2002. **180**: p. 140-3.
19. Degeneration, T.A.f.F. 2013; Available from: http://www.theaftd.org/frontotemporal-degeneration/ftd-overview.
20. Gorno-Tempini, M.L., et al., *Classification of primary progressive aphasia and its variants.* Neurology, 2011. **76**(11): p. 1006-14.
21. Neary, D., et al., *Frontotemporal lobar degeneration: a consensus on clinical diagnostic criteria.* Neurology, 1998. **51**(6): p. 1546-54.

22. Josephs, K.A., *Frontotemporal dementia and related disorders: deciphering the enigma.* Ann Neurol, 2008. **64**(1): p. 4-14.
23. Rosen, H.J., et al., *Behavioral features in semantic dementia vs other forms of progressive aphasias.* Neurology, 2006. **67**(10): p. 1752-6.
24. Gorno-Tempini, M.L., et al., *Cognition and anatomy in three variants of primary progressive aphasia.* Ann Neurol, 2004. **55**(3): p. 335-46.
25. Rohrer, J.D. and J.D. Warren, *Phenotypic signatures of genetic frontotemporal dementia.* Curr Opin Neurol, 2011. **24**(6): p. 542-9.
26. Brunden, K.R., J.Q. Trojanowski, and V.M. Lee, *Advances in tau-focused drug discovery for Alzheimer's disease and related tauopathies.* Nat Rev Drug Discov, 2009. **8**(10): p. 783-93.
27. Rascovsky, K., et al., *Sensitivity of revised diagnostic criteria for the behavioural variant of frontotemporal dementia.* Brain, 2011. **134**(Pt 9): p. 2456-77.
28. Halliday, G., et al., *Mechanisms of disease in frontotemporal lobar degeneration: gain of function versus loss of function effects.* Acta Neuropathol, 2012. **124**(3): p. 373-82.
29. Caffrey, T.M. and R. Wade-Martins, *Functional MAPT haplotypes: bridging the gap between genotype and neuropathology.* Neurobiol Dis, 2007. **27**(1): p. 1-10.
30. Liscic, R.M., et al., *ALS and FTLD: two faces of TDP-43 proteinopathy.* Eur J Neurol, 2008. **15**(8): p. 772-80.
31. Mackenzie, I.R., et al., *A harmonized classification system for FTLD-TDP pathology.* Acta Neuropathol, 2011. **122**(1): p. 111-3.
32. Mackenzie, I.R., et al., *Distinct pathological subtypes of FTLD-FUS.* Acta Neuropathol, 2011. **121**(2): p. 207-18.
33. Lashley, T., et al., *A comparative clinical, pathological, biochemical and genetic study of fused in sarcoma proteinopathies.* Brain, 2011. **134**(Pt 9): p. 2548-64.
34. Cannon, A., et al., *CHMP2B mutations are not a common cause of frontotemporal lobar degeneration.* Neurosci Lett, 2006. **398**(1-2): p. 83-4.
35. Urwin, H., et al., *The role of CHMP2B in frontotemporal dementia.* Biochem Soc Trans, 2009. **37**(Pt 1): p. 208-12.
36. Mackenzie, I.R., et al., *Nomenclature for neuropathologic subtypes of frontotemporal lobar degeneration: consensus recommendations.* Acta Neuropathol, 2009. **117**(1): p. 15-8.
37. Mackenzie, I.R., et al., *Nomenclature and nosology for neuropathologic subtypes of frontotemporal lobar degeneration: an update.* Acta Neuropathol, 2010. **119**(1): p. 1-4.
38. Susanne Froelich-Fabre, R.V.B., *Mechanisms of tauopathies.* Drug Discovery Today: Disease Mechanisms, 2004. **1**(4): p. 391–398.
39. van der Zee, J., et al., *A pan-European study of the C9orf72 repeat associated with FTLD: geographic prevalence, genomic instability, and intermediate repeats.* Hum Mutat, 2013. **34**(2): p. 363-73.
40. Rubino, E., et al., *SQSTM1 mutations in frontotemporal lobar degeneration and amyotrophic lateral sclerosis.* Neurology, 2012. **79**(15): p. 1556-62.
41. Van Deerlin, V.M., et al., *Common variants at 7p21 are associated with frontotemporal lobar degeneration with TDP-43 inclusions.* Nat Genet, 2010. **42**(3): p. 234-9.

42. Goedert, M., et al., *Multiple isoforms of human microtubule-associated protein tau: sequences and localization in neurofibrillary tangles of Alzheimer's disease.* Neuron, 1989. **3**(4): p. 519-26.
43. Niblock, M. and J.M. Gallo, *Tau alternative splicing in familial and sporadic tauopathies.* Biochem Soc Trans, 2012. **40**(4): p. 677-80.
44. van Swieten, J. and M.G. Spillantini, *Hereditary frontotemporal dementia caused by Tau gene mutations.* Brain Pathol, 2007. **17**(1): p. 63-73.
45. Hutton, M., et al., *Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17.* Nature, 1998. **393**(6686): p. 702-5.
46. Jho, Y.S., et al., *Monte carlo simulations of tau proteins: effect of phosphorylation.* Biophys J, 2010. **99**(8): p. 2387-97.
47. Cruts, M., J. Theuns, and C. Van Broeckhoven, *Locus-specific mutation databases for neurodegenerative brain diseases.* Hum Mutat, 2012. **33**(9): p. 1340-4.
48. Malkani, R., et al., *A MAPT mutation in a regulatory element upstream of exon 10 causes frontotemporal dementia.* Neurobiol Dis, 2006. **22**(2): p. 401-3.
49. Rovelet-Lecrux, A. and D. Campion, *Copy number variations involving the microtubule-associated protein tau in human diseases.* Biochem Soc Trans, 2012. **40**(4): p. 672-6.
50. Goedert, M. and R. Jakes, *Mutations causing neurodegenerative tauopathies.* Biochim Biophys Acta, 2005. **1739**(2-3): p. 240-50.
51. Ferrari, R., J. Hardy, and P. Momeni, *Frontotemporal dementia: from Mendelian genetics towards genome wide association studies.* J Mol Neurosci, 2011. **45**(3): p. 500-15.
52. Coppola, G., et al., *Evidence for a role of the rare p.A152T variant in MAPT in increasing the risk for FTD-spectrum and Alzheimer's diseases.* Hum Mol Genet, 2012. **21**(15): p. 3500-12.
53. Deramecourt, V., et al., *Clinical, neuropathological, and biochemical characterization of the novel tau mutation P332S.* J Alzheimers Dis, 2012. **31**(4): p. 741-9.
54. Rossi, G., et al., *New mutations in MAPT gene causing frontotemporal lobar degeneration: biochemical and structural characterization.* Neurobiol Aging, 2012. **33**(4): p. 834 e1-6.
55. Koolen, D.A., et al., *A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism.* Nat Genet, 2006. **38**(9): p. 999-1001.
56. Shaw-Smith, C., et al., *Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability.* Nat Genet, 2006. **38**(9): p. 1032-7.
57. Varela, M.C., et al., *A 17q21.31 microdeletion encompassing the MAPT gene in a mentally impaired patient.* Cytogenet Genome Res, 2006. **114**(1): p. 89-92.
58. Ishizuka, T., et al., *Familial semantic dementia with P301L mutation in the Tau gene.* Dement Geriatr Cogn Disord, 2011. **31**(5): p. 334-40.
59. Munoz, D.G., et al., *Progressive nonfluent aphasia associated with a new mutation V363I in tau gene.* Am J Alzheimers Dis Other Demen, 2007. **22**(4): p. 294-9.
60. Villa, C., et al., *A novel MAPT mutation associated with the clinical phenotype of progressive nonfluent aphasia.* J Alzheimers Dis, 2011. **26**(1): p. 19-26.
61. Gijselinck, I., C. Van Broeckhoven, and M. Cruts, *Granulin mutations associated with frontotemporal lobar degeneration and related disorders: an update.* Hum Mutat, 2008. **29**(12): p. 1373-86.

62. Daniel, R., et al., *Cellular localization of gene expression for progranulin.* J Histochem Cytochem, 2000. **48**(7): p. 999-1009.
63. van Swieten, J.C. and P. Heutink, *Mutations in progranulin (GRN) within the spectrum of clinical and pathological phenotypes of frontotemporal dementia.* Lancet Neurol, 2008. **7**(10): p. 965-74.
64. He, Z. and A. Bateman, *Progranulin (granulin-epithelin precursor, PC-cell-derived growth factor, acrogranin) mediates tissue repair and tumorigenesis.* J Mol Med (Berl), 2003. **81**(10): p. 600-12.
65. Mackenzie, I.R. and R. Rademakers, *The molecular genetics and neuropathology of frontotemporal lobar degeneration: recent developments.* Neurogenetics, 2007. **8**(4): p. 237-48.
66. Baker, M., et al., *Mutations in progranulin cause tau-negative frontotemporal dementia linked to chromosome 17.* Nature, 2006. **442**(7105): p. 916-9.
67. Cruts, M., et al., *Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21.* Nature, 2006. **442**(7105): p. 920-4.
68. van der Zee, J., et al., *Mutations other than null mutations producing a pathogenic loss of progranulin in frontotemporal dementia.* Hum Mutat, 2007. **28**(4): p. 416.
69. Mukherjee, O., et al., *Molecular characterization of novel progranulin (GRN) mutations in frontotemporal dementia.* Hum Mutat, 2008. **29**(4): p. 512-21.
70. Shankaran, S.S., et al., *Missense mutations in the progranulin gene linked to frontotemporal lobar degeneration with ubiquitin-immunoreactive inclusions reduce progranulin production and secretion.* J Biol Chem, 2008. **283**(3): p. 1744-53.
71. Rovelet-Lecrux, A., et al., *Deletion of the progranulin gene in patients with frontotemporal lobar degeneration or Parkinson disease.* Neurobiol Dis, 2008. **31**(1): p. 41-5.
72. Taipa, R., et al., *Clinical, neuropathological, and genetic characteristics of the novel IVS9+1delG GRN mutation in a patient with frontotemporal dementia.* J Alzheimers Dis, 2012. **30**(1): p. 83-90.
73. Bernardi, L., et al., *Epidemiology and genetics of frontotemporal dementia: a door-to-door survey in southern Italy.* Neurobiol Aging, 2012. **33**(12): p. 2948 e1-2948 e10.
74. Caso, F., et al., *The progranulin (GRN) Cys157LysfsX97 mutation is associated with nonfluent variant of primary progressive aphasia clinical phenotype.* J Alzheimers Dis, 2012. **28**(4): p. 759-63.
75. Rohrer, J.D., et al., *Distinct profiles of brain atrophy in frontotemporal lobar degeneration caused by progranulin and tau mutations.* Neuroimage, 2010. **53**(3): p. 1070-6.
76. DeJesus-Hernandez, M., et al., *Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS.* Neuron, 2011. **72**(2): p. 245-56.
77. Renton, A.E., et al., *A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD.* Neuron, 2011. **72**(2): p. 257-68.
78. Laaksovirta, H., et al., *Chromosome 9p21 in amyotrophic lateral sclerosis in Finland: a genome-wide association study.* Lancet Neurol, 2010. **9**(10): p. 978-85.
79. Morita, M., et al., *A locus on chromosome 9p confers susceptibility to ALS and frontotemporal dementia.* Neurology, 2006. **66**(6): p. 839-44.

80. Shatunov, A., et al., *Chromosome 9p21 in sporadic amyotrophic lateral sclerosis in the UK and seven other countries: a genome-wide association study.* Lancet Neurol, 2010. **9**(10): p. 986-94.
81. van Es, M.A., et al., *Genome-wide association study identifies 19p13.3 (UNC13A) and 9p21.2 as susceptibility loci for sporadic amyotrophic lateral sclerosis.* Nat Genet, 2009. **41**(10): p. 1083-7.
82. Vance, C., et al., *Familial amyotrophic lateral sclerosis with frontotemporal dementia is linked to a locus on chromosome 9p13.2-21.3.* Brain, 2006. **129**(Pt 4): p. 868-76.
83. Koppers, M., et al., *Screening for rare variants in the coding region of ALS-associated genes at 9p21.2 and 19p13.3.* Neurobiol Aging, 2013. **34**(5): p. 1518 e5-7.
84. Gijselinck, I., et al., *A C9orf72 promoter repeat expansion in a Flanders-Belgian cohort with disorders of the frontotemporal lobar degeneration-amyotrophic lateral sclerosis spectrum: a gene identification study.* Lancet Neurol, 2012. **11**(1): p. 54-65.
85. Snowden, J.S., et al., *Distinct clinical and pathological characteristics of frontotemporal dementia associated with C9ORF72 mutations.* Brain, 2012. **135**(Pt 3): p. 693-708.
86. Simon-Sanchez, J., et al., *The clinical and pathological phenotype of C9ORF72 hexanucleotide repeat expansions.* Brain, 2012. **135**(Pt 3): p. 723-35.
87. Majounie, E., et al., *Frequency of the C9orf72 hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study.* Lancet Neurol, 2012. **11**(4): p. 323-30.
88. Dols-Icardo, O., et al., *Expansion mutation in C9ORF72 does not influence plasma progranulin levels in frontotemporal dementia.* Neurobiol Aging, 2012. **33**(8): p. 1851 e17-9.
89. Dobson-Stone, C., et al., *C9ORF72 repeat expansion in clinical and neuropathologic frontotemporal dementia cohorts.* Neurology, 2012. **79**(10): p. 995-1001.
90. Xi, Z., et al., *Investigation of c9orf72 in 4 neurodegenerative disorders.* Arch Neurol, 2012. **69**(12): p. 1583-90.
91. Boeve, B.F., et al., *Characterization of frontotemporal dementia and/or amyotrophic lateral sclerosis associated with the GGGGCC repeat expansion in C9ORF72.* Brain, 2012. **135**(Pt 3): p. 765-83.
92. Murray, M.E., et al., *Clinical and neuropathologic heterogeneity of c9FTD/ALS associated with hexanucleotide repeat expansion in C9ORF72.* Acta Neuropathol, 2011. **122**(6): p. 673-90.
93. Al-Sarraj, S., et al., *p62 positive, TDP-43 negative, neuronal cytoplasmic and intranuclear inclusions in the cerebellum and hippocampus define the pathology of C9orf72-linked FTLD and MND/ALS.* Acta Neuropathol, 2011. **122**(6): p. 691-702.
94. Hsiung, G.Y., et al., *Clinical and pathological features of familial frontotemporal dementia caused by C9ORF72 mutation on chromosome 9p.* Brain, 2012. **135**(Pt 3): p. 709-22.
95. Mahoney, C.J., et al., *Longitudinal neuroimaging and neuropsychological profiles of frontotemporal dementia with C9ORF72 expansions.* Alzheimers Res Ther, 2012. **4**(5): p. 41.
96. Brettschneider, J., et al., *Pattern of ubiquilin pathology in ALS and FTLD indicates presence of C9ORF72 hexanucleotide expansion.* Acta Neuropathol, 2012. **123**(6): p. 825-39.

97. Bieniek, K.F., et al., *Tau pathology in frontotemporal lobar degeneration with C9ORF72 hexanucleotide repeat expansion.* Acta Neuropathol, 2013. **125**(2): p. 289-302.
98. King, A., et al., *Mixed tau, TDP-43 and p62 pathology in FTLD associated with a C9ORF72 repeat expansion and p.Ala239Thr MAPT (tau) variant.* Acta Neuropathol, 2013. **125**(2): p. 303-10.
99. Arighi, A., et al., *Early onset behavioral variant frontotemporal dementia due to the C9ORF72 hexanucleotide repeat expansion: psychiatric clinical presentations.* J Alzheimers Dis, 2012. **31**(2): p. 447-52.
100. Ash, P.E., et al., *Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS.* Neuron, 2013. **77**(4): p. 639-46.
101. Mori, K., et al., *The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTLD/ALS.* Science, 2013. **339**(6125): p. 1335-8.
102. Gendron, T.F., et al., *Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS.* Acta Neuropathol, 2013. **126**(6): p. 829-44.
103. Lagier-Tourenne, C., et al., *Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration.* Proc Natl Acad Sci U S A, 2013. **110**(47): p. E4530-9.
104. Mackenzie, I.R., P. Frick, and M. Neumann, *The neuropathology associated with repeat expansions in the C9ORF72 gene.* Acta Neuropathol, 2014. **127**(3): p. 347-57.
105. Mizielinska, S., et al., *C9orf72 frontotemporal lobar degeneration is characterised by frequent neuronal sense and antisense RNA foci.* Acta Neuropathol, 2013. **126**(6): p. 845-57.
106. Wojciechowska, M. and W.J. Krzyzosiak, *Cellular toxicity of expanded RNA repeats: focus on RNA foci.* Hum Mol Genet, 2011. **20**(19): p. 3811-21.
107. Mackenzie, I.R., et al., *Dipeptide repeat protein pathology in C9ORF72 mutation cases: clinico-pathological correlations.* Acta Neuropathol, 2013. **126**(6): p. 859-79.
108. Sugita, S. and T.C. Sudhof, *Specificity of Ca2+-dependent protein interactions mediated by the C2A domains of synaptotagmins.* Biochemistry, 2000. **39**(11): p. 2940-9.
109. Balch, W.E., et al., *Adapting proteostasis for disease intervention.* Science, 2008. **319**(5865): p. 916-9.
110. Forman, M.S., et al., *Novel ubiquitin neuropathology in frontotemporal dementia with valosin-containing protein gene mutations.* J Neuropathol Exp Neurol, 2006. **65**(6): p. 571-81.
111. Kondo, H., et al., *p47 is a cofactor for p97-mediated membrane fusion.* Nature, 1997. **388**(6637): p. 75-8.
112. Rabinovich, E., et al., *AAA-ATPase p97/Cdc48p, a cytosolic chaperone required for endoplasmic reticulum-associated protein degradation.* Mol Cell Biol, 2002. **22**(2): p. 626-34.
113. Rabouille, C., et al., *Syntaxin 5 is a common component of the NSF- and p97-mediated reassembly pathways of Golgi cisternae from mitotic Golgi fragments in vitro.* Cell, 1998. **92**(5): p. 603-10.
114. Wang, Q., C. Song, and C.C. Li, *Molecular perspectives on p97-VCP: progress in understanding its structure and diverse biological functions.* J Struct Biol, 2004. **146**(1-2): p. 44-57.

115. Ju, J.S. and C.C. Weihl, *Inclusion body myopathy, Paget's disease of the bone and fronto-temporal dementia: a disorder of autophagy.* Hum Mol Genet, 2010. **19**(R1): p. R38-45.
116. Weihl, C.C., A. Pestronk, and V.E. Kimonis, *Valosin-containing protein disease: inclusion body myopathy with Paget's disease of the bone and fronto-temporal dementia.* Neuromuscul Disord, 2009. **19**(5): p. 308-15.
117. Watts, G.D., et al., *Inclusion body myopathy associated with Paget disease of bone and frontotemporal dementia is caused by mutant valosin-containing protein.* Nat Genet, 2004. **36**(4): p. 377-81.
118. Kim, E.J., et al., *Inclusion body myopathy with Paget disease of bone and frontotemporal dementia linked to VCP p.Arg155Cys in a Korean family.* Arch Neurol, 2011. **68**(6): p. 787-96.
119. Johnson, J.O., et al., *Exome sequencing reveals VCP mutations as a cause of familial ALS.* Neuron, 2010. **68**(5): p. 857-64.
120. Weihl, C.C., *Valosin containing protein associated fronto-temporal lobar degeneration: clinical presentation, pathologic features and pathogenesis.* Curr Alzheimer Res, 2011. **8**(3): p. 252-60.
121. Skibinski, G., et al., *Mutations in the endosomal ESCRTIII-complex subunit CHMP2B in frontotemporal dementia.* Nat Genet, 2005. **37**(8): p. 806-8.
122. Brown, J., et al., *Familial non-specific dementia maps to chromosome 3.* Hum Mol Genet, 1995. **4**(9): p. 1625-8.
123. Ferrari, R., et al., *Novel Missense Mutation in Charged Multivesicular Body Protein 2B in a Patient With Frontotemporal Dementia.* Alzheimer Dis Assoc Disord, 2010.
124. Ghanim, M., et al., *CHMP2B mutations are rare in French families with frontotemporal lobar degeneration.* J Neurol, 2010. **257**(12): p. 2032-6.
125. Momeni, P., et al., *Genetic variability in CHMP2B and frontotemporal dementia.* Neurodegener Dis, 2006. **3**(3): p. 129-33.
126. Rizzu, P., et al., *CHMP2B mutations are not a cause of dementia in Dutch patients with familial and sporadic frontotemporal dementia.* Am J Med Genet B Neuropsychiatr Genet, 2006. **141B**(8): p. 944-6.
127. van der Zee, J., et al., *CHMP2B C-truncating mutations in frontotemporal lobar degeneration are associated with an aberrant endosomal phenotype in vitro.* Hum Mol Genet, 2008. **17**(2): p. 313-22.
128. Parkinson, N., et al., *ALS phenotypes with mutations in CHMP2B (charged multivesicular body protein 2B).* Neurology, 2006. **67**(6): p. 1074-7.
129. Ghazi-Noori, S., et al., *Progressive neuronal inclusion formation and axonal degeneration in CHMP2B mutant transgenic mice.* Brain, 2012. **135**(Pt 3): p. 819-32.
130. Geser, F., et al., *Amyotrophic lateral sclerosis, frontotemporal dementia and beyond: the TDP-43 diseases.* J Neurol, 2009. **256**(8): p. 1205-14.
131. Gijselinck, I., et al., *Neuronal inclusion protein TDP-43 has no primary genetic role in FTD and ALS.* Neurobiol Aging, 2009. **30**(8): p. 1329-31.
132. Rollinson, S., et al., *TDP-43 gene analysis in frontotemporal lobar degeneration.* Neurosci Lett, 2007. **419**(1): p. 1-4.
133. Kovacs, G.G., et al., *TARDBP variation associated with frontotemporal dementia, supranuclear gaze palsy, and chorea.* Mov Disord, 2009. **24**(12): p. 1843-7.

134. Gitcho, M.A., et al., *TARDBP 3'-UTR variant in autopsy-confirmed frontotemporal lobar degeneration with TDP-43 proteinopathy.* Acta Neuropathol, 2009. **118**(5): p. 633-45.
135. Borroni, B., et al., *Mutation within TARDBP leads to frontotemporal dementia without motor neuron disease.* Hum Mutat, 2009. **30**(11): p. E974-83.
136. Borroni, B., et al., *TARDBP mutations in frontotemporal lobar degeneration: frequency, clinical features, and disease course.* Rejuvenation Res, 2010. **13**(5): p. 509-17.
137. Huey, E.D., et al., *FUS and TDP43 genetic variability in FTD and CBS.* Neurobiol Aging, 2012. **33**(5): p. 1016 e9-17.
138. Borghero, G., et al., *A patient carrying a homozygous p.A382T TARDBP missense mutation shows a syndrome including ALS, extrapyramidal symptoms, and FTD.* Neurobiol Aging, 2011. **32**(12): p. 2327 e1-5.
139. Kabashi, E., et al., *Gain and loss of function of ALS-related mutations of TARDBP (TDP-43) cause motor deficits in vivo.* Hum Mol Genet, 2010. **19**(4): p. 671-83.
140. Yang, C., et al., *The C-terminal TDP-43 fragments have a high aggregation propensity and harm neurons by a dominant-negative mechanism.* PLoS One, 2010. **5**(12): p. e15878.
141. Budini, M., et al., *Role of selected mutations in the Q/N rich region of TDP-43 in EGFP-12xQ/N-induced aggregate formation.* Brain Res, 2012. **1462**: p. 139-50.
142. Kwiatkowski, T.J., Jr., et al., *Mutations in the FUS/TLS gene on chromosome 16 cause familial amyotrophic lateral sclerosis.* Science, 2009. **323**(5918): p. 1205-8.
143. Vance, C., et al., *Mutations in FUS, an RNA processing protein, cause familial amyotrophic lateral sclerosis type 6.* Science, 2009. **323**(5918): p. 1208-11.
144. Blair, I.P., et al., *FUS mutations in amyotrophic lateral sclerosis: clinical, pathological, neurophysiological and genetic analysis.* J Neurol Neurosurg Psychiatry, 2010. **81**(6): p. 639-45.
145. Van Langenhove, T., et al., *Genetic contribution of FUS to frontotemporal lobar degeneration.* Neurology, 2010. **74**(5): p. 366-71.
146. Rohrer, J.D., et al., *The heritability and genetics of frontotemporal lobar degeneration.* Neurology, 2009. **73**(18): p. 1451-6.
147. Cantoni, C., et al., *FUS/TLS genetic variability in sporadic frontotemporal lobar degeneration.* J Alzheimers Dis, 2010. **19**(4): p. 1317-22.
148. Lang, C.M., et al., *Membrane orientation and subcellular localization of transmembrane protein 106B (TMEM106B), a major risk factor for frontotemporal lobar degeneration.* J Biol Chem, 2012. **287**(23): p. 19355-65.
149. Chen-Plotkin, A.S., et al., *TMEM106B, the risk gene for frontotemporal dementia, is regulated by the microRNA-132/212 cluster and affects progranulin pathways.* J Neurosci, 2012. **32**(33): p. 11213-27.
150. Brady, O.A., et al., *The frontotemporal lobar degeneration risk factor, TMEM106B, regulates lysosomal morphology and function.* Hum Mol Genet, 2013. **22**(4): p. 685-95.
151. Cruchaga, C., et al., *Association of TMEM106B gene polymorphism with age at onset in granulin mutation carriers and plasma granulin protein levels.* Arch Neurol, 2011. **68**(5): p. 581-6.
152. Finch, N., et al., *TMEM106B regulates progranulin levels and the penetrance of FTLD in GRN mutation carriers.* Neurology, 2011. **76**(5): p. 467-74.

153. Vass, R., et al., *Risk genotypes at TMEM106B are associated with cognitive impairment in amyotrophic lateral sclerosis.* Acta Neuropathol, 2011. **121**(3): p. 373-80.
154. van der Zee, J. and C. Van Broeckhoven, *TMEM106B a novel risk factor for frontotemporal lobar degeneration.* J Mol Neurosci, 2011. **45**(3): p. 516-21.
155. Rutherford, N.J., et al., *TMEM106B risk variant is implicated in the pathologic presentation of Alzheimer disease.* Neurology, 2012. **79**(7): p. 717-8.
156. Steele, J.C., J.C. Richardson, and J. Olszewski, *Progressive Supranuclear Palsy. A Heterogeneous Degeneration Involving the Brain Stem, Basal Ganglia and Cerebellum with Vertical Gaze and Pseudobulbar Palsy, Nuchal Dystonia and Dementia.* Arch Neurol, 1964. **10**: p. 333-59.
157. Lubarsky, M. and J.L. Juncos, *Progressive supranuclear palsy: a current review.* Neurologist, 2008. **14**(2): p. 79-88.
158. Bower, J.H., et al., *Incidence of progressive supranuclear palsy and multiple system atrophy in Olmsted County, Minnesota, 1976 to 1990.* Neurology, 1997. **49**(5): p. 1284-8.
159. Schrag, A., Y. Ben-Shlomo, and N.P. Quinn, *Prevalence of progressive supranuclear palsy and multiple system atrophy: a cross-sectional study.* Lancet, 1999. **354**(9192): p. 1771-5.
160. Dickson, D.W., R. Rademakers, and M.L. Hutton, *Progressive supranuclear palsy: pathology and genetics.* Brain Pathol, 2007. **17**(1): p. 74-82.
161. Rojo, A., et al., *Clinical genetics of familial progressive supranuclear palsy.* Brain, 1999. **122 (Pt 7)**: p. 1233-45.
162. Burn, D.J. and A.J. Lees, *Progressive supranuclear palsy: where are we now?* Lancet Neurol, 2002. **1**(6): p. 359-69.
163. Williams, D.R. and A.J. Lees, *Progressive supranuclear palsy: clinicopathological concepts and diagnostic challenges.* Lancet Neurol, 2009. **8**(3): p. 270-9.
164. Hauw, J.J., et al., *Preliminary NINDS neuropathologic criteria for Steele-Richardson-Olszewski syndrome (progressive supranuclear palsy).* Neurology, 1994. **44**(11): p. 2015-9.
165. Williams, D.R., et al., *Pathological tau burden and distribution distinguishes progressive supranuclear palsy-parkinsonism from Richardson's syndrome.* Brain, 2007. **130**(Pt 6): p. 1566-76.
166. de Silva, R., et al., *Pathological inclusion bodies in tauopathies contain distinct complements of tau with three or four microtubule-binding repeat domains as demonstrated by new specific monoclonal antibodies.* Neuropathol Appl Neurobiol, 2003. **29**(3): p. 288-302.
167. Dickson, D.W., *Neuropathologic differentiation of progressive supranuclear palsy and corticobasal degeneration.* J Neurol, 1999. **246 Suppl 2**: p. II6-15.
168. Ahmed, Z., et al., *Clinical and neuropathologic features of progressive supranuclear palsy with severe pallido-nigro-luysial degeneration and axonal dystrophy.* Brain, 2008. **131**(Pt 2): p. 460-72.
169. Williams, D.R., et al., *Pure akinesia with gait freezing: a third clinical phenotype of progressive supranuclear palsy.* Mov Disord, 2007. **22**(15): p. 2235-41.
170. Josephs, K.A., et al., *Atypical progressive supranuclear palsy underlying progressive apraxia of speech and nonfluent aphasia.* Neurocase, 2005. **11**(4): p. 283-96.

171. Tsuboi, Y., et al., *Increased tau burden in the cortices of progressive supranuclear palsy presenting with corticobasal syndrome.* Mov Disord, 2005. **20**(8): p. 982-8.
172. Yokota, O., et al., *Phosphorylated TDP-43 pathology and hippocampal sclerosis in progressive supranuclear palsy.* Acta Neuropathol, 2010. **120**(1): p. 55-66.
173. Baker, M., et al., *Association of an extended haplotype in the tau gene with progressive supranuclear palsy.* Hum Mol Genet, 1999. **8**(4): p. 711-5.
174. Conrad, C., et al., *Genetic evidence for the involvement of tau in progressive supranuclear palsy.* Ann Neurol, 1997. **41**(2): p. 277-81.
175. Ezquerra, M., et al., *Identification of a novel polymorphism in the promoter region of the tau gene highly associated to progressive supranuclear palsy in humans.* Neurosci Lett, 1999. **275**(3): p. 183-6.
176. Vandrovcova, J., et al., *Disentangling the role of the tau gene locus in sporadic tauopathies.* Curr Alzheimer Res, 2010. **7**(8): p. 726-34.
177. Cruchaga, C., et al., *5'-Upstream variants of CRHR1 and MAPT genes associated with age at onset in progressive supranuclear palsy and cortical basal degeneration.* Neurobiol Dis, 2009. **33**(2): p. 164-70.
178. Ezquerra, M., et al., *Different MAPT haplotypes are associated with Parkinson's disease and progressive supranuclear palsy.* Neurobiol Aging, 2011. **32**(3): p. 547 e11-6.
179. Pittman, A.M., et al., *Linkage disequilibrium fine mapping and haplotype association analysis of the tau gene in progressive supranuclear palsy and corticobasal degeneration.* J Med Genet, 2005. **42**(11): p. 837-46.
180. Hoglinger, G.U., et al., *Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy.* Nat Genet, 2011. **43**(7): p. 699-705.
181. Choumert, A., et al., *G303V tau mutation presenting with progressive supranuclear palsy-like features.* Mov Disord, 2012. **27**(4): p. 581-3.
182. Delisle, M.B., et al., *A mutation at codon 279 (N279K) in exon 10 of the Tau gene causes a tauopathy with dementia and supranuclear palsy.* Acta Neuropathol, 1999. **98**(1): p. 62-77.
183. Pastor, P., et al., *Familial atypical progressive supranuclear palsy associated with homozigosity for the delN296 mutation in the tau gene.* Ann Neurol, 2001. **49**(2): p. 263-7.
184. Poorkaj, P., et al., *An R5L tau mutation in a subject with a progressive supranuclear palsy phenotype.* Ann Neurol, 2002. **52**(4): p. 511-6.
185. Rohrer, J.D., et al., *Novel L284R MAPT mutation in a family with an autosomal dominant progressive supranuclear palsy syndrome.* Neurodegener Dis, 2011. **8**(3): p. 149-52.
186. Ros, R., et al., *A new mutation of the tau gene, G303V, in early-onset familial progressive supranuclear palsy.* Arch Neurol, 2005. **62**(9): p. 1444-50.
187. Rossi, G., et al., *Progressive supranuclear palsy and Parkinson's disease in a family with a new mutation in the tau gene.* Ann Neurol, 2004. **55**(3): p. 448.
188. Spina, S., et al., *The tauopathy associated with mutation +3 in intron 10 of Tau: characterization of the MSTD family.* Brain, 2008. **131**(Pt 1): p. 72-89.
189. Morris, H.R., et al., *Tau exon 10 +16 mutation FTDP-17 presenting clinically as sporadic young onset PSP.* Neurology, 2003. **61**(1): p. 102-4.

190. Stanford, P.M., et al., *Progressive supranuclear palsy pathology caused by a novel silent mutation in exon 10 of the tau gene: expansion of the disease phenotype caused by tau gene mutations.* Brain, 2000. **123 (Pt 5)**: p. 880-93.
191. Melquist, S., et al., *Identification of a novel risk locus for progressive supranuclear palsy by a pooled genomewide scan of 500,288 single-nucleotide polymorphisms.* Am J Hum Genet, 2007. **80**(4): p. 769-78.
192. Ros, R., I. Ampuero, and J. Garcia de Yebenes, *Parkin polymorphisms in progressive supranuclear palsy.* J Neurol Sci, 2008. **268**(1-2): p. 176-8.
193. Chambers, C.B., et al., *Overexpression of four-repeat tau mRNA isoforms in progressive supranuclear palsy but not in Alzheimer's disease.* Ann Neurol, 1999. **46**(3): p. 325-32.
194. Takanashi, M., et al., *Expression patterns of tau mRNA isoforms correlate with susceptible lesions in progressive supranuclear palsy and corticobasal degeneration.* Brain Res Mol Brain Res, 2002. **104**(2): p. 210-9.
195. Luk, C., Vandrovcova, J., Malzer, E., Lees, A. and de Silva, R, *Brain tau isoform mRNA and protein correlation in PSP brain.* Translational Neuroscience, 2010. **1**: p. 7.
196. Caffrey, T.M., et al., *Haplotype-specific expression of exon 10 at the human MAPT locus.* Hum Mol Genet, 2006. **15**(24): p. 3529-37.
197. Caffrey, T.M., C. Joachim, and R. Wade-Martins, *Haplotype-specific expression of the N-terminal exons 2 and 3 at the human MAPT locus.* Neurobiol Aging, 2008. **29**(12): p. 1923-9.
198. Majounie, E., et al., *Variation in tau isoform expression in different brain regions and disease states.* Neurobiol Aging, 2013. **34**(7): p. 1922 e7-1922 e12.
199. Trabzuni, D., et al., *MAPT expression and splicing is differentially regulated by brain region: relation to genotype and implication for tauopathies.* Hum Mol Genet, 2012. **21**(18): p. 4094-103.
200. Rademakers, R., et al., *High-density SNP haplotyping suggests altered regulation of tau gene expression in progressive supranuclear palsy.* Hum Mol Genet, 2005. **14**(21): p. 3281-92.
201. Myers, A.J., et al., *The MAPT H1c risk haplotype is associated with increased expression of tau and especially of 4 repeat containing transcripts.* Neurobiol Dis, 2007. **25**(3): p. 561-70.
202. Laws, S.M., et al., *Fine mapping of the MAPT locus using quantitative trait analysis identifies possible causal variants in Alzheimer's disease.* Mol Psychiatry, 2007. **12**(5): p. 510-7.
203. Pittman, A.M., et al., *The structure of the tau haplotype in controls and in progressive supranuclear palsy.* Hum Mol Genet, 2004. **13**(12): p. 1267-74.
204. Hardy, J., et al., *Evidence suggesting that Homo neanderthalensis contributed the H2 MAPT haplotype to Homo sapiens.* Biochem Soc Trans, 2005. **33**(Pt 4): p. 582-5.
205. Stefansson, H., et al., *A common inversion under selection in Europeans.* Nat Genet, 2005. **37**(2): p. 129-37.
206. Simon-Sanchez, J., et al., *Genome-wide association study reveals genetic risk underlying Parkinson's disease.* Nat Genet, 2009. **41**(12): p. 1308-12.
207. Trabzuni, D., et al., *Quality control parameters on a large dataset of regionally dissected human control brains for whole genome expression studies.* J Neurochem, 2011. **119**(2): p. 275-82.

208. Stempler, S. and E. Ruppin, *Analyzing gene expression from whole tissue vs. different cell types reveals the central role of neurons in predicting severity of Alzheimer's disease.* PLoS One, 2012. **7**(9): p. e45879.
209. Gusella, J.F., et al., *A polymorphic DNA marker genetically linked to Huntington's disease.* Nature, 1983. **306**(5940): p. 234-8.
210. Riordan, J.R., et al., *Identification of the cystic fibrosis gene: cloning and characterization of complementary DNA.* Science, 1989. **245**(4922): p. 1066-73.
211. Rao, A.T., A.J. Degnan, and L.M. Levy, *Genetics of Alzheimer Disease.* AJNR Am J Neuroradiol, 2013.
212. Watson, J.D. and F.H. Crick, *Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid.* Nature, 1953. **171**(4356): p. 737-8.
213. Bernfield, M.R. and M.W. Nirenberg, *Rna Codewords and Protein Synthesis. The Nucleotide Sequences of Multiple Codewords for Phenylalanine, Serine, Leucine, and Proline.* Science, 1965. **147**(3657): p. 479-84.
214. Sanger, F., S. Nicklen, and A.R. Coulson, *DNA sequencing with chain-terminating inhibitors.* Proc Natl Acad Sci U S A, 1977. **74**(12): p. 5463-7.
215. Hardy, J. and A. Singleton, *Genomewide association studies and human disease.* N Engl J Med, 2009. **360**(17): p. 1759-68.
216. Litvan, I., et al., *Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDS-SPSP international workshop.* Neurology, 1996. **47**(1): p. 1-9.
217. Litvan, I., et al., *Validity and reliability of the preliminary NINDS neuropathologic criteria for progressive supranuclear palsy and related disorders.* J Neuropathol Exp Neurol, 1996. **55**(1): p. 97-105.
218. Strong, M.J., et al., *Consensus criteria for the diagnosis of frontotemporal cognitive and behavioural syndromes in amyotrophic lateral sclerosis.* Amyotroph Lateral Scler, 2009. **10**(3): p. 131-46.
219. Ferrari, R., et al., *A familial frontotemporal dementia associated with C9orf72 repeat expansion and dysplastic gangliocytoma.* Neurobiol Aging, 2013.
220. Boeve, B.F., A.E. Lang, and I. Litvan, *Corticobasal degeneration and its relationship to progressive supranuclear palsy and frontotemporal dementia.* Ann Neurol, 2003. **54 Suppl 5**: p. S15-9.
221. International Parkinson Disease Genomics, C., et al., *Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies.* Lancet, 2011. **377**(9766): p. 641-9.
222. International Parkinson's Disease Genomics, C. and C. Wellcome Trust Case Control, *A two-stage meta-analysis identifies several new loci for Parkinson's disease.* PLoS Genet, 2011. **7**(6): p. e1002142.
223. Mok, K., et al., *Homozygosity analysis in amyotrophic lateral sclerosis.* Eur J Hum Genet, 2013.
224. Genomes Project, C., et al., *An integrated map of genetic variation from 1,092 human genomes.* Nature, 2012. **491**(7422): p. 56-65.
225. Adzhubei, I.A., et al., *A method and server for predicting damaging missense mutations.* Nat Methods, 2010. **7**(4): p. 248-9.

226. *Statistical Genetics: Gene Mapping Through Linkage and Association*2008: Taylor & Francis. 608 pages.
227. Price, A.L., et al., *Principal components analysis corrects for stratification in genome-wide association studies.* Nat Genet, 2006. **38**(8): p. 904-9.
228. Grove, M.L., et al., *Best practices and joint calling of the HumanExome BeadChip: the CHARGE Consortium.* PLoS One, 2013. **8**(7): p. e68095.
229. Cheng, J., et al., *Syntaxin 6 and CAL mediate the degradation of the cystic fibrosis transmembrane conductance regulator.* Mol Biol Cell, 2010. **21**(7): p. 1178-87.
230. Jung, J.J., et al., *Regulation of intracellular membrane trafficking and cell dynamics by syntaxin-6.* Biosci Rep, 2012. **32**(4): p. 383-91.
231. Manickam, V., et al., *Regulation of vascular endothelial growth factor receptor 2 trafficking and angiogenesis by Golgi localized t-SNARE syntaxin 6.* Blood, 2011. **117**(4): p. 1425-35.
232. Kabayama, H., et al., *Syntaxin 6 regulates nerve growth factor-dependent neurite outgrowth.* Neurosci Lett, 2008. **436**(3): p. 340-4.
233. Rutkowski, D.T. and R.J. Kaufman, *A trip to the ER: coping with stress.* Trends Cell Biol, 2004. **14**(1): p. 20-8.
234. Mori, K., *Signalling pathways in the unfolded protein response: development from yeast to mammals.* J Biochem, 2009. **146**(6): p. 743-50.
235. Trinh, M.A., et al., *Brain-specific disruption of the eIF2alpha kinase PERK decreases ATF4 expression and impairs behavioral flexibility.* Cell Rep, 2012. **1**(6): p. 676-88.
236. Montague, P., et al., *Myelin-associated oligodendrocytic basic protein: a family of abundant CNS myelin proteins in search of a function.* Dev Neurosci, 2006. **28**(6): p. 479-87.
237. Yamamoto, Y., et al., *Cloning and expression of myelin-associated oligodendrocytic basic protein. A novel basic protein constituting the central nervous system myelin.* J Biol Chem, 1994. **269**(50): p. 31725-30.
238. Kaushansky, N., et al., *The myelin-associated oligodendrocytic basic protein (MOBP) as a relevant primary target autoantigen in multiple sclerosis.* Autoimmun Rev, 2010. **9**(4): p. 233-6.
239. Liu, Q.Y., et al., *An exploratory study on STX6, MOBP, MAPT, and EIF2AK3 and late-onset Alzheimer's disease.* Neurobiol Aging, 2013. **34**(5): p. 1519 e13-7.
240. Trinh, J., et al., *STX6 rs1411478 is not associated with increased risk of Parkinson's disease.* Parkinsonism Relat Disord, 2013. **19**(5): p. 563-5.
241. Barrett, J.C., et al., *Haploview: analysis and visualization of LD and haplotype maps.* Bioinformatics, 2005. **21**(2): p. 263-5.
242. Ramasamy, A., et al., *Resolving the polymorphism-in-probe problem is critical for correct interpretation of expression QTL studies.* Nucleic Acids Res, 2013. **41**(7): p. e88.
243. Millar, T., et al., *Tissue and organ donation for research in forensic pathology: the MRC Sudden Death Brain and Tissue Bank.* J Pathol, 2007. **213**(4): p. 369-75.
244. Beach, T.G., et al., *The Sun Health Research Institute Brain Donation Program: description and experience, 1987-2007.* Cell Tissue Bank, 2008. **9**(3): p. 229-45.
245. Schofield, E.C., et al., *Cortical atrophy differentiates Richardson's syndrome from the parkinsonian form of progressive supranuclear palsy.* Mov Disord, 2011. **26**(2): p. 256-63.

246. Consortium, U.K.P.s.D., et al., *Dissection of the genetics of Parkinson's disease identifies an additional association 5' of SNCA and multiple associated haplotypes at 17q21.* Hum Mol Genet, 2011. **20**(2): p. 345-53.
247. Li, Y., et al., *MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes.* Genet Epidemiol, 2010. **34**(8): p. 816-34.
248. Shabalin, A.A., *Matrix eQTL: ultra fast eQTL analysis via large matrix operations.* Bioinformatics, 2012. **28**(10): p. 1353-8.
249. Hayesmoore, J.B., et al., *The effect of age and the H1c MAPT haplotype on MAPT expression in human brain.* Neurobiol Aging, 2009. **30**(10): p. 1652-6.
250. Rosenbloom, K.R., et al., *ENCODE whole-genome data in the UCSC Genome Browser: update 2012.* Nucleic Acids Res, 2012. **40**(Database issue): p. D912-7.
251. Creyghton, M.P., et al., *Histone H3K27ac separates active from poised enhancers and predicts developmental state.* Proc Natl Acad Sci U S A, 2010. **107**(50): p. 21931-6.
252. Ong, C.T. and V.G. Corces, *Enhancer function: new insights into the regulation of tissue-specific gene expression.* Nat Rev Genet, 2011. **12**(4): p. 283-93.
253. Zhukareva, V., et al., *Unexpected abundance of pathological tau in progressive supranuclear palsy white matter.* Ann Neurol, 2006. **60**(3): p. 335-45.
254. Whitwell, J.L., et al., *Clinical correlates of white matter tract degeneration in progressive supranuclear palsy.* Arch Neurol, 2011. **68**(6): p. 753-60.
255. Armstrong, R.A., *White matter pathology in progressive supranuclear palsy (PSP): a quantitative study of 8 cases.* Clin Neuropathol, 2013.
256. Kjer-Nielsen, L., et al., *MR1 presents microbial vitamin B metabolites to MAIT cells.* Nature, 2012. **491**(7426): p. 717-23.
257. Urasawa, T., et al., *Antigenic and genetic analyses of human rotavirus with dual subgroup specificity.* J Clin Microbiol, 1990. **28**(12): p. 2837-41.
258. Pankratz, N., et al., *Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2.* Ann Neurol, 2012. **71**(3): p. 370-84.
259. International Multiple Sclerosis Genetics, C., et al., *Risk alleles for multiple sclerosis identified by a genomewide study.* N Engl J Med, 2007. **357**(9): p. 851-62.
260. Imrie, D. and K.C. Sadler, *Stress management: How the unfolded protein response impacts fatty liver disease.* J Hepatol, 2012. **57**(5): p. 1147-51.
261. Liu, J., et al., *A functional haplotype in EIF2AK3, an ER stress sensor, is associated with lower bone mineral density.* J Bone Miner Res, 2012. **27**(2): p. 331-41.
262. Stutzbach, L.D., et al., *The unfolded protein response is activated in disease-affected brain regions in progressive supranuclear palsy and Alzheimer's disease.* Acta Neuropathol Commun, 2013. **1**(1): p. 31.
263. Hoozemans, J.J., et al., *The unfolded protein response is activated in pretangle neurons in Alzheimer's disease hippocampus.* Am J Pathol, 2009. **174**(4): p. 1241-51.
264. Nijholt, D.A., et al., *The unfolded protein response is associated with early tau pathology in the hippocampus of tauopathies.* J Pathol, 2012. **226**(5): p. 693-702.
265. Ling, S.C., M. Polymenidou, and D.W. Cleveland, *Converging mechanisms in ALS and FTD: disrupted RNA and protein homeostasis.* Neuron, 2013. **79**(3): p. 416-38.
266. Singleton, A.B., et al., *Towards a complete resolution of the genetic architecture of disease.* Trends Genet, 2010. **26**(10): p. 438-42.

267. Irizarry, R.A., et al., *Exploration, normalization, and summaries of high density oligonucleotide array probe level data.* Biostatistics, 2003. **4**(2): p. 249-64.
268. Barbosa-Morais, N.L., et al., *A re-annotation pipeline for Illumina BeadArrays: improving the interpretation of gene expression data.* Nucleic Acids Res, 2010. **38**(3): p. e17.
269. Zeller, T., et al., *Genetics and beyond--the transcriptome of human monocytes and disease susceptibility.* PLoS One, 2010. **5**(5): p. e10693.
270. Skoglund, L., et al., *No evidence of PGRN or MAPT gene dosage alterations in a collection of patients with frontotemporal lobar degeneration.* Dement Geriatr Cogn Disord, 2009. **28**(5): p. 471-5.
271. Llado, A., et al., *MAPT gene duplications are not a cause of frontotemporal lobar degeneration.* Neurosci Lett, 2007. **424**(1): p. 61-5.
272. Rovelet-Lecrux, A., et al., *Partial deletion of the MAPT gene: a novel mechanism of FTDP-17.* Hum Mutat, 2009. **30**(4): p. E591-602.
273. Rovelet-Lecrux, A., et al., *Frontotemporal dementia phenotype associated with MAPT gene duplication.* J Alzheimers Dis, 2010. **21**(3): p. 897-902.
274. Varma, A.R., et al., *Evaluation of the NINCDS-ADRDA criteria in the differentiation of Alzheimer's disease and frontotemporal dementia.* J Neurol Neurosurg Psychiatry, 1999. **66**(2): p. 184-8.
275. Janssens, J. and C. Van Broeckhoven, *Pathological mechanisms underlying TDP-43 driven neurodegeneration in FTLD-ALS spectrum disorders.* Hum Mol Genet, 2013. **22**(R1): p. R77-87.
276. Spillantini, M.G. and M. Goedert, *Tau pathology and neurodegeneration.* Lancet Neurol, 2013. **12**(6): p. 609-22.
277. Wilson, R.S., et al., *TDP-43 Pathology, Cognitive Decline, and Dementia in Old Age.* JAMA Neurol, 2013.
278. Seshadri, S., et al., *Genome-wide analysis of genetic loci associated with Alzheimer disease.* JAMA, 2010. **303**(18): p. 1832-40.
279. van der Zee, J., K. Sleegers, and C. Van Broeckhoven, *Invited article: the Alzheimer disease-frontotemporal lobar degeneration spectrum.* Neurology, 2008. **71**(15): p. 1191-7.
280. Rubino, E., et al., *Apolipoprotein E polymorphisms in frontotemporal lobar degeneration: A meta-analysis.* Alzheimers Dement, 2013.
281. Seripa, D., et al., *The APOE gene locus in frontotemporal dementia and primary progressive aphasia.* Arch Neurol, 2011. **68**(5): p. 622-8.
282. Seripa, D., et al., *TOMM40, APOE, and APOC1 in primary progressive aphasia and frontotemporal dementia.* J Alzheimers Dis, 2012. **31**(4): p. 731-40.
283. Zintl, M., et al., *[Frontotemporal dementia in association with a family history of dementia and ApoE polymorphism].* Nervenarzt, 2010. **81**(1): p. 75-8.
284. Beck, J., et al., *Large C9orf72 hexanucleotide repeat expansions are seen in multiple neurodegenerative syndromes and are more frequent than expected in the UK population.* Am J Hum Genet, 2013. **92**(3): p. 345-53.
285. Ferrari, R., et al., *Screening for C9ORF72 repeat expansion in FTLD.* Neurobiol Aging, 2012. **33**(8): p. 1850 e1-11.
286. Baranzini, S.E., et al., *Genome-wide association analysis of susceptibility and clinical phenotype in multiple sclerosis.* Hum Mol Genet, 2009. **18**(4): p. 767-78.

287. Jin, Y., et al., *Variant of TYR and autoimmunity susceptibility loci in generalized vitiligo.* N Engl J Med, 2010. **362**(18): p. 1686-97.
288. Eleftherohorinou, H., et al., *Pathway-driven gene stability selection of two rheumatoid arthritis GWAS identifies and validates new susceptibility genes in receptor mediated signalling pathways.* Hum Mol Genet, 2011. **20**(17): p. 3494-506.
289. Miller, Z.A., et al., *TDP-43 frontotemporal lobar degeneration and autoimmune disease.* J Neurol Neurosurg Psychiatry, 2013. **84**(9): p. 956-62.
290. Mero, I.L., et al., *Oligoclonal band status in Scandinavian multiple sclerosis patients is associated with specific genetic risk alleles.* PLoS One, 2013. **8**(3): p. e58352.
291. Hamza, T.H., et al., *Common genetic variation in the HLA region is associated with late-onset sporadic Parkinson's disease.* Nat Genet, 2010. **42**(9): p. 781-5.
292. Rosso, S.M., et al., *Medical and environmental risk factors for sporadic frontotemporal dementia: a retrospective case-control study.* J Neurol Neurosurg Psychiatry, 2003. **74**(11): p. 1574-6.
293. Kalkonde, Y.V., et al., *Medical and environmental risk factors associated with frontotemporal dementia: a case-control study in a veteran population.* Alzheimers Dement, 2012. **8**(3): p. 204-10.
294. Galimberti, D., et al., *Incomplete Penetrance of the C9ORF72 Hexanucleotide Repeat Expansions: Frequency in a Cohort of Geriatric Non-Demented Subjects.* J Alzheimers Dis, 2013.
295. Kenna, K.P., et al., *Delineating the genetic heterogeneity of ALS using targeted high-throughput sequencing.* J Med Genet, 2013.
296. Le Ber, I., et al., *C9ORF72 repeat expansions in the frontotemporal dementias spectrum of diseases: a flow-chart for genetic testing.* J Alzheimers Dis, 2013. **34**(2): p. 485-99.
297. Jager, D., et al., *Serological cloning of a melanocyte rab guanosine 5'-triphosphate-binding protein and a chromosome condensation protein from a melanoma complementary DNA library.* Cancer Res, 2000. **60**(13): p. 3584-91.
298. Bultema, J.J., et al., *BLOC-2, AP-3, and AP-1 proteins function in concert with Rab38 and Rab32 proteins to mediate protein trafficking to lysosome-related organelles.* J Biol Chem, 2012. **287**(23): p. 19550-63.
299. Wasmeier, C., et al., *Rab38 and Rab32 control post-Golgi trafficking of melanogenic enzymes.* J Cell Biol, 2006. **175**(2): p. 271-81.
300. Seto, S., K. Tsujimura, and Y. Koide, *Rab GTPases regulating phagosome maturation are differentially recruited to mycobacterial phagosomes.* Traffic, 2011. **12**(4): p. 407-20.
301. Kierdorf, K., et al., *Microglia emerge from erythromyeloid precursors via Pu.1- and Irf8-dependent pathways.* Nat Neurosci, 2013. **16**(3): p. 273-80.
302. Hu, F., et al., *Sortilin-mediated endocytosis determines levels of the frontotemporal dementia protein, progranulin.* Neuron, 2010. **68**(4): p. 654-67.
303. Felten DL, S.A., *Netter's Atlas of Neuroscience*2010.
304. McGeer, P.L., et al., *Reactive microglia are positive for HLA-DR in the substantia nigra of Parkinson's and Alzheimer's disease brains.* Neurology, 1988. **38**(8): p. 1285-91.
305. Valentonyte, R., et al., *Sarcoidosis is associated with a truncating splice site mutation in BTNL2.* Nat Genet, 2005. **37**(4): p. 357-64.

306. Amor, S. and N. Woodroofe, *Review series on immune responses in neurodegenerative diseases: innate and adaptive immune responses in neurodegeneration and repair.* Immunology, 2013.
307. Safieh-Garabedian, B., Y. Mayasi, and N.E. Saade, *Targeting neuroinflammation for therapeutic intervention in neurodegenerative pathologies: a role for the peptide analogue of thymulin (PAT).* Expert Opin Ther Targets, 2012. **16**(11): p. 1065-73.
308. Guillot-Sestier, M.V. and T. Town, *Innate immunity in Alzheimer's disease: a complex affair.* CNS Neurol Disord Drug Targets, 2013. **12**(5): p. 593-607.
309. Cao, J.J., K.S. Li, and Y.Q. Shen, *Activated immune cells in Parkinson's disease.* J Neuroimmune Pharmacol, 2011. **6**(3): p. 323-9.
310. Evans, M.C., et al., *Inflammation and neurovascular changes in amyotrophic lateral sclerosis.* Mol Cell Neurosci, 2013. **53**: p. 34-41.
311. van Noort, J.M., D. Baker, and S. Amor, *Mechanisms in the development of multiple sclerosis lesions: reconciling autoimmune and neurodegenerative factors.* CNS Neurol Disord Drug Targets, 2012. **11**(5): p. 556-69.
312. Trager, U. and S.J. Tabrizi, *Peripheral inflammation in neurodegeneration.* J Mol Med (Berl), 2013. **91**(6): p. 673-81.
313. Cunningham, C., *Systemic inflammation and delirium: important co-factors in the progression of dementia.* Biochem Soc Trans, 2011. **39**(4): p. 945-53.
314. Rogers, J., et al., *Neuroinflammation in Alzheimer's disease and Parkinson's disease: are microglia pathogenic in either disorder?* Int Rev Neurobiol, 2007. **82**: p. 235-46.
315. Tai, Y.F., et al., *Microglial activation in presymptomatic Huntington's disease gene carriers.* Brain, 2007. **130**(Pt 7): p. 1759-66.
316. Cagnin, A., et al., *In vivo detection of microglial activation in frontotemporal dementia.* Ann Neurol, 2004. **56**(6): p. 894-7.
317. Solerte, S.B., et al., *Overproduction of IFN-gamma and TNF-alpha from natural killer (NK) cells is associated with abnormal NK reactivity and cognitive derangement in Alzheimer's disease.* Ann N Y Acad Sci, 2000. **917**: p. 331-40.
318. Sjogren, M., et al., *Increased intrathecal inflammatory activity in frontotemporal dementia: pathophysiological implications.* J Neurol Neurosurg Psychiatry, 2004. **75**(8): p. 1107-11.
319. Hauser, S.L. and J.R. Oksenberg, *The neurobiology of multiple sclerosis: genes, inflammation, and neurodegeneration.* Neuron, 2006. **52**(1): p. 61-76.
320. Kouri, N., et al., *Corticobasal degeneration: a pathologically distinct 4R tauopathy.* Nat Rev Neurol, 2011. **7**(5): p. 263-72.
321. Ferrari, R., et al., *Implication of common and disease specific variants in CLU, CR1, and PICALM.* Neurobiol Aging, 2012. **33**(8): p. 1846 e7-18.
322. Ferrari, R., et al., *Androgen receptor gene and sex-specific Alzheimer's disease.* Neurobiol Aging, 2013. **34**(8): p. 2077 e19-20.
323. Rossi, G., et al., *The G389R mutation in the MAPT gene presenting as sporadic corticobasal syndrome.* Mov Disord, 2008. **23**(6): p. 892-5.
324. Kouri, N., et al., *Novel mutation in MAPT exon 13 (p.N410H) causes corticobasal degeneration.* Acta Neuropathol, 2013.
325. Momeni, P., et al., *Clinical and pathological features of an Alzheimer's disease patient with the MAPT Delta K280 mutation.* Neurobiol Aging, 2009. **30**(3): p. 388-93.

326. Gerrish, A., et al., *The role of variation at AbetaPP, PSEN1, PSEN2, and MAPT in late onset Alzheimer's disease.* J Alzheimers Dis, 2012. **28**(2): p. 377-87.
327. Masellis, M., et al., *Novel splicing mutation in the progranulin gene causing familial corticobasal syndrome.* Brain, 2006. **129**(Pt 11): p. 3115-23.
328. Spina, S., et al., *Corticobasal syndrome associated with the A9D Progranulin mutation.* J Neuropathol Exp Neurol, 2007. **66**(10): p. 892-900.
329. Benussi, L., et al., *Progranulin Leu271LeufsX10 is one of the most common FTLD and CBS associated mutations worldwide.* Neurobiol Dis, 2009. **33**(3): p. 379-85.
330. Yu, C.E., et al., *The spectrum of mutations in progranulin: a collaborative study screening 545 cases of neurodegeneration.* Arch Neurol, 2010. **67**(2): p. 161-70.
331. Kabashi, E., et al., *TARDBP mutations in individuals with sporadic and familial amyotrophic lateral sclerosis.* Nat Genet, 2008. **40**(5): p. 572-4.
332. Saunders, A.M., et al., *Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease.* Neurology, 1993. **43**(8): p. 1467-72.
333. Goate, A., et al., *Segregation of a missense mutation in the amyloid precursor protein gene with familial Alzheimer's disease.* Nature, 1991. **349**(6311): p. 704-6.
334. Campion, D., et al., *Mutations of the presenilin I gene in families with early-onset Alzheimer's disease.* Hum Mol Genet, 1995. **4**(12): p. 2373-7.
335. Cruts, M., et al., *Estimation of the genetic contribution of presenilin-1 and -2 mutations in a population-based study of presenile Alzheimer disease.* Hum Mol Genet, 1998. **7**(1): p. 43-51.
336. Lambert, J.C., et al., *Genome-wide association study identifies variants at CLU and CR1 associated with Alzheimer's disease.* Nat Genet, 2009. **41**(10): p. 1094-9.
337. Harold, D., et al., *Genome-wide association study identifies variants at CLU and PICALM associated with Alzheimer's disease.* Nat Genet, 2009. **41**(10): p. 1088-93.
338. Bodon, G., et al., *Charged multivesicular body protein 2B (CHMP2B) of the endosomal sorting complex required for transport-III (ESCRT-III) polymerizes into helical structures deforming the plasma membrane.* J Biol Chem, 2011. **286**(46): p. 40276-86.
339. Urwin, H., et al., *Disruption of endocytic trafficking in frontotemporal dementia with CHMP2B mutations.* Hum Mol Genet, 2010. **19**(11): p. 2228-38.
340. Mok, K., et al., *Chromosome 9 ALS and FTD locus is probably derived from a single founder.* Neurobiol Aging, 2012. **33**(1): p. 209 e3-8.
341. Abel, T.W., et al., *Lhermitte-Duclos disease: a report of 31 cases with immunohistochemical analysis of the PTEN/AKT/mTOR pathway.* J Neuropathol Exp Neurol, 2005. **64**(4): p. 341-9.
342. Zhou, X.P., et al., *Germline inactivation of PTEN and dysregulation of the phosphoinositol-3-kinase/Akt pathway cause human Lhermitte-Duclos disease in adults.* Am J Hum Genet, 2003. **73**(5): p. 1191-8.
343. Eng, C., *PTEN: one gene, many syndromes.* Hum Mutat, 2003. **22**(3): p. 183-98.
344. Cordes, I., et al., *PTEN deletions are related to disease progression and unfavourable prognosis in early bladder cancer.* Histopathology, 2013.
345. Nieuwenhuis, M.H., et al., *Cancer risk and genotype-phenotype correlations in PTEN hamartoma tumor syndrome.* Fam Cancer, 2013.

346. Floris, G., et al., *Frontotemporal dementia with psychosis, parkinsonism, visuo-spatial dysfunction, upper motor neuron involvement associated to expansion of C9ORF72: a peculiar phenotype?* J Neurol, 2012. **259**(8): p. 1749-51.
347. Mahoney, C.J., et al., *Frontotemporal dementia with the C9ORF72 hexanucleotide repeat expansion: clinical, neuroanatomical and neuropathological features.* Brain, 2012. **135**(Pt 3): p. 736-50.
348. Cruts, M., et al., *Current insights into the C9orf72 repeat expansion diseases of the FTLD/ALS spectrum.* Trends Neurosci, 2013. **36**(8): p. 450-9.
349. Ning, K., et al., *PTEN depletion rescues axonal growth defect and improves survival in SMN-deficient motor neurons.* Hum Mol Genet, 2010. **19**(16): p. 3159-68.
350. Chen, C.C., et al., *Autocrine prolactin induced by the Pten-Akt pathway is required for lactation initiation and provides a direct link between the Akt and Stat5 pathways.* Genes Dev, 2012. **26**(19): p. 2154-68.
351. Domanskyi, A., et al., *Pten ablation in adult dopaminergic neurons is neuroprotective in Parkinson's disease models.* FASEB J, 2011. **25**(9): p. 2898-910.
352. Kirby, J., et al., *Phosphatase and tensin homologue/protein kinase B pathway linked to motor neuron survival in human superoxide dismutase 1-related amyotrophic lateral sclerosis.* Brain, 2011. **134**(Pt 2): p. 506-17.
353. Matenia, D., et al., *Microtubule affinity-regulating kinase 2 (MARK2) turns on phosphatase and tensin homolog (PTEN)-induced kinase 1 (PINK1) at Thr-313, a mutation site in Parkinson disease: effects on mitochondrial transport.* J Biol Chem, 2012. **287**(11): p. 8174-86.
354. Christie, K.J. and D. Zochodne, *Peripheral axon regrowth: new molecular approaches.* Neuroscience, 2013. **240**: p. 310-24.
355. Heikkinen, T., et al., *Variants on the promoter region of PTEN affect breast cancer progression and patient survival.* Breast Cancer Res, 2011. **13**(6): p. R130.
356. Mester, J. and C. Eng, *When overgrowth bumps into cancer: the PTEN-opathies.* Am J Med Genet C Semin Med Genet, 2013. **163C**(2): p. 114-21.
357. Katzmann, D.J., G. Odorizzi, and S.D. Emr, *Receptor downregulation and multivesicular-body sorting.* Nat Rev Mol Cell Biol, 2002. **3**(12): p. 893-905.
358. Williams, R.L. and S. Urbe, *The emerging shape of the ESCRT machinery.* Nat Rev Mol Cell Biol, 2007. **8**(5): p. 355-68.
359. Nickerson, D.P., M. West, and G. Odorizzi, *Did2 coordinates Vps4-mediated dissociation of ESCRT-III from endosomes.* J Cell Biol, 2006. **175**(5): p. 715-20.
360. Tsang, H.T., et al., *A systematic analysis of human CHMP protein interactions: additional MIT domain-containing proteins bind to multiple components of the human ESCRT III complex.* Genomics, 2006. **88**(3): p. 333-46.
361. Hurley, J.H. and D. Yang, *MIT domainia.* Dev Cell, 2008. **14**(1): p. 6-8.
362. Obita, T., et al., *Structural basis for selective recognition of ESCRT-III by the AAA ATPase Vps4.* Nature, 2007. **449**(7163): p. 735-9.
363. Ahmad, S.T., et al., *Genetic screen identifies serpin5 as a regulator of the toll pathway and CHMP2B toxicity associated with frontotemporal dementia.* Proc Natl Acad Sci U S A, 2009. **106**(29): p. 12168-73.
364. Lee, J.A., et al., *ESCRT-III dysfunction causes autophagosome accumulation and neurodegeneration.* Curr Biol, 2007. **17**(18): p. 1561-7.

365. Cox, L.E., et al., *Mutations in CHMP2B in lower motor neuron predominant amyotrophic lateral sclerosis (ALS).* PLoS One, 2010. **5**(3): p. e9872.
366. Hara, T., et al., *Suppression of basal autophagy in neural cells causes neurodegenerative disease in mice.* Nature, 2006. **441**(7095): p. 885-9.
367. Komatsu, M., et al., *Loss of autophagy in the central nervous system causes neurodegeneration in mice.* Nature, 2006. **441**(7095): p. 880-4.
368. Filimonenko, M., et al., *Functional multivesicular bodies are required for autophagic clearance of protein aggregates associated with neurodegenerative disease.* J Cell Biol, 2007. **179**(3): p. 485-500.
369. Nielsen, T.T., et al., *Reversal of pathology in CHMP2B-mediated frontotemporal dementia patient cells using RNA interference.* J Gene Med, 2012. **14**(8): p. 521-9.
370. Bhardwaj, A., et al., *Characterizing TDP-43 interaction with its RNA targets.* Nucleic Acids Res, 2013. **41**(9): p. 5062-74.
371. Ferrari, R., et al., *FTD and ALS: a tale of two diseases.* Curr Alzheimer Res, 2011. **8**(3): p. 273-94.
372. Neumann, M., et al., *Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis.* Science, 2006. **314**(5796): p. 130-3.
373. Sreedharan, J., et al., *TDP-43 mutations in familial and sporadic amyotrophic lateral sclerosis.* Science, 2008. **319**(5870): p. 1668-72.
374. Corrado, L., et al., *High frequency of TARDBP gene mutations in Italian patients with amyotrophic lateral sclerosis.* Hum Mutat, 2009. **30**(4): p. 688-94.
375. Belzil, V.V., et al., *Mutations in FUS cause FALS and SALS in French and French Canadian populations.* Neurology, 2009. **73**(15): p. 1176-9.
376. Benajiba, L., et al., *TARDBP mutations in motoneuron disease with frontotemporal lobar degeneration.* Ann Neurol, 2009. **65**(4): p. 470-3.
377. Rutherford, N.J., et al., *Novel mutations in TARDBP (TDP-43) in patients with familial amyotrophic lateral sclerosis.* PLoS Genet, 2008. **4**(9): p. e1000193.
378. Schumacher, A., et al., *No association of TDP-43 with sporadic frontotemporal dementia.* Neurobiol Aging, 2009. **30**(1): p. 157-9.
379. Lomen-Hoerth, C., T. Anderson, and B. Miller, *The overlap of amyotrophic lateral sclerosis and frontotemporal dementia.* Neurology, 2002. **59**(7): p. 1077-9.
380. Arai, T., et al., *TDP-43 is a component of ubiquitin-positive tau-negative inclusions in frontotemporal lobar degeneration and amyotrophic lateral sclerosis.* Biochem Biophys Res Commun, 2006. **351**(3): p. 602-11.
381. Boxer, A.L., et al., *Clinical, neuroimaging and neuropathological features of a new chromosome 9p-linked FTD-ALS family.* J Neurol Neurosurg Psychiatry, 2011. **82**(2): p. 196-203.
382. Gilman, S., et al., *Second consensus statement on the diagnosis of multiple system atrophy.* Neurology, 2008. **71**(9): p. 670-6.
383. Harms, M.B., et al., *Parkinson disease is not associated with C9ORF72 repeat expansions.* Neurobiol Aging, 2013. **34**(5): p. 1519 e1-2.
384. Schymick, J.C., et al., *Progranulin mutations and amyotrophic lateral sclerosis or amyotrophic lateral sclerosis-frontotemporal dementia phenotypes.* J Neurol Neurosurg Psychiatry, 2007. **78**(7): p. 754-6.

385. Ratti, A., et al., *C9ORF72 repeat expansion in a large Italian ALS cohort: evidence of a founder effect.* Neurobiol Aging, 2012. **33**(10): p. 2528 e7-14.
386. van Blitterswijk, M., et al., *C9ORF72 repeat expansions in cases with previously identified pathogenic mutations.* Neurology, 2013. **81**(15): p. 1332-41.
387. Chio, A., et al., *ALS/FTD phenotype in two Sardinian families carrying both C9ORF72 and TARDBP mutations.* J Neurol Neurosurg Psychiatry, 2012. **83**(7): p. 730-3.
388. Kaivorinne, A.L., et al., *Novel TARDBP Sequence Variant and C9ORF72 Repeat Expansion in a Family With Frontotemporal Dementia.* Alzheimer Dis Assoc Disord, 2012.
389. Loy, C.T., et al., *Genetics of dementia.* Lancet, 2013.
390. Troakes, C., et al., *An MND/ALS phenotype associated with C9orf72 repeat expansion: abundant p62-positive, TDP-43-negative inclusions in cerebral cortex, hippocampus and cerebellum but without associated cognitive decline.* Neuropathology, 2012. **32**(5): p. 505-14.
391. Kertesz, A., et al., *Psychosis and Hallucinations in Frontotemporal Dementia with the C9ORF72 Mutation: A Detailed Clinical Cohort.* Cogn Behav Neurol, 2013. **26**(3): p. 146-54.
392. Snowden, J.S., et al., *Psychosis, C9ORF72 and dementia with Lewy bodies.* J Neurol Neurosurg Psychiatry, 2012. **83**(10): p. 1031-2.
393. Cacace, R., et al., *C9orf72 G4C2 repeat expansions in Alzheimer's disease and mild cognitive impairment.* Neurobiol Aging, 2013. **34**(6): p. 1712 e1-7.
394. Harms, M., et al., *C9orf72 hexanucleotide repeat expansions in clinical Alzheimer disease.* JAMA Neurol, 2013. **70**(6): p. 736-41.
395. Kohli, M.A., et al., *Repeat expansions in the C9ORF72 gene contribute to Alzheimer's disease in Caucasians.* Neurobiol Aging, 2013. **34**(5): p. 1519 e5-12.
396. Murray, M.E., et al., *Progressive amnestic dementia, hippocampal sclerosis, and mutation in C9ORF72.* Acta Neuropathol, 2013. **126**(4): p. 545-54.
397. Wojtas, A., et al., *C9ORF72 repeat expansions and other FTD gene mutations in a clinical AD patient series from Mayo Clinic.* Am J Neurodegener Dis, 2012. **1**(1): p. 107-18.
398. Annan, M., et al., *Idiopathic Parkinson's disease phenotype related to C9ORF72 repeat expansions: contribution of the neuropsychological assessment.* BMC Res Notes, 2013. **6**(1): p. 343.
399. Lindquist, S.G., et al., *Corticobasal and ataxia syndromes widen the spectrum of C9ORF72 hexanucleotide expansion disease.* Clin Genet, 2013. **83**(3): p. 279-83.

Appendices

Appendix 2-1. Comparison of genotypes: Genome amplified vs. genomic DNA

General call rates barely reached 85% for the genome amplified samples. Cross-check of a sample of ~ 1000 SNPs – of which 424 overlapping between OMNI and 660K revealed average accuracy of calls of ~89-90% and miscalls of ~10-11% of which ~8-9% no calls and ~2% wrong calls.

[Header]				[Header]				
GSGT Version	1.9.4			**GSGT Version**	1.9.4			
Processing Date	7/18/2012 13:34			**Processing Date**	7/18/2012 14:53			
Content		**HumanOmniExpress-12v1_H.bpm**		**Content**		**Human660W-Quad_v1_A.bpm**		
Num SNPs	1060			**Num SNPs**	701			
Total SNPs	730525			**Total SNPs**	657366			
Num Samples	12			**Num Samples**	11			
Total Samples	12			**Total Samples**	175			
[Data]				[Data]				
Genome amplified samples				**Genomic DNA samples**				**Problem**
Sample ID	**SNP Name**	**Allele1**	**Allele2**	**Sample ID**	**SNP Name**	**Allele1**	**Allele2**	
FTD001	rs1000000	C	C	**Riem-001**	rs1000000	C	C	
FTD001	rs10000023	T	T	**Riem-001**	rs10000023	T	T	
FTD001	rs10000030	A	G	**Riem-001**	rs10000030	A	G	
FTD001	rs10000041	T	T	**Riem-001**	rs10000041	T	T	
FTD001	rs1000007	A	A	**Riem-001**	rs1000007	A	A	
FTD001	rs10000081	T	T	**Riem-001**	rs10000081	T	T	
FTD001	rs10000092	T	C	**Riem-001**	rs10000092	T	C	
FTD001	rs1000016	A	A	**Riem-001**	rs1000016	A	A	
FTD001	rs10000160	G	G	**Riem-001**	rs10000160	G	G	
FTD001	rs10000169	T	C	**Riem-001**	rs10000169	T	C	
FTD001	rs10000185	T	C	**Riem-001**	rs10000185	T	C	
FTD001	rs1000022	T	C	**Riem-001**	rs1000022	T	C	
FTD001	rs10000226	T	C	**Riem-001**	rs10000226	T	C	
FTD001	rs10000255	T	C	**Riem-001**	rs10000255	C	C	WRONG CALL
FTD001	rs10000266	T	C	**Riem-001**	rs10000266	T	C	
FTD001	rs10000272	T	C	**Riem-001**	rs10000272	T	C	
FTD001	rs10000282	C	C	**Riem-001**	rs10000282	C	C	
FTD001	rs10000300	T	G	**Riem-001**	rs10000300	T	G	
FTD001	rs1000031	C	C	**Riem-001**	rs1000031	C	C	
FTD001	rs1000032	C	C	**Riem-001**	rs1000032	C	C	
FTD001	rs10000388	T	C	**Riem-001**	rs10000388	T	C	
FTD001	rs1000040	A	A	**Riem-001**	rs1000040	A	A	
FTD001	rs1000041	G	G	**Riem-001**	rs1000041	G	G	
FTD001	rs10000432	T	C	**Riem-001**	rs10000432	T	C	
FTD001	rs10000435	T	C	**Riem-001**	rs10000435	T	C	
FTD001	rs10000438	T	C	**Riem-001**	rs10000438	T	C	
FTD001	rs10000456	T	T	**Riem-001**	rs10000456	T	T	

FTD001	rs10000471	T	C	**Riem-001**	rs10000471	T	C	
FTD001	rs10000487	A	A	**Riem-001**	rs10000487	A	A	
FTD001	rs1000050	T	C	**Riem-001**	rs1000050	T	C	
FTD001	rs10000538	T	C	**Riem-001**	rs10000538	T	C	
FTD001	rs10000543	-	-	**Riem-001**	rs10000543	C	C	NO CALL
FTD001	rs1000055	C	C	**Riem-001**	rs1000055	C	C	
FTD001	rs10000595	C	C	**Riem-001**	rs10000595	C	C	
FTD001	rs1000061	A	G	**Riem-001**	rs1000061	A	G	
FTD001	rs1000068	A	G	**Riem-001**	rs1000068	A	G	
FTD001	rs10000708	G	G	**Riem-001**	rs10000708	G	G	
FTD001	rs1000071	G	G	**Riem-001**	rs1000071	G	G	
FTD001	rs10000719	-	-	**Riem-001**	rs10000719	T	T	NO CALL
FTD001	rs1000073	T	T	**Riem-001**	rs1000073	T	T	
FTD001	rs10000762	C	C	**Riem-001**	rs10000762	C	C	
FTD001	rs1000079	T	C	**Riem-001**	rs1000079	T	C	
FTD001	rs1000083	C	C	**Riem-001**	rs1000083	C	C	
FTD001	rs10000918	A	A	**Riem-001**	rs10000918	A	A	
FTD001	rs1000094	T	T	**Riem-001**	rs1000094	T	T	
FTD001	rs10000959	T	T	**Riem-001**	rs10000959	T	T	
FTD001	rs10000976	T	C	**Riem-001**	rs10000976	T	C	
FTD001	rs1000104	A	C	**Riem-001**	rs1000104	A	C	
FTD001	rs10001148	A	G	**Riem-001**	rs10001148	A	G	
FTD001	rs1000115	T	C	**Riem-001**	rs1000115	T	C	
FTD001	rs10001154	T	C	**Riem-001**	rs10001154	T	C	
FTD001	rs10001188	G	G	**Riem-001**	rs10001188	G	G	
FTD001	rs10001198	-	-	**Riem-001**	rs10001198	G	G	NO CALL
FTD001	rs1000121	-	-	**Riem-001**	rs1000121	T	T	NO CALL
FTD001	rs10001214	T	C	**Riem-001**	rs10001214	T	C	
FTD001	rs10001236	-	-	**Riem-001**	rs10001236	G	G	NO CALL
FTD001	rs10001280	G	G	**Riem-001**	rs10001280	G	G	
FTD001	rs10001348	A	A	**Riem-001**	rs10001348	A	A	
FTD001	rs1000137	T	C	**Riem-001**	rs1000137	T	C	
FTD001	rs10001385	T	T	**Riem-001**	rs10001385	T	T	
FTD001	rs1000140	T	C	**Riem-001**	rs1000140	T	C	
FTD001	rs1000141	A	G	**Riem-001**	rs1000141	A	G	
FTD001	rs10001415	T	T	**Riem-001**	rs10001415	T	T	
FTD001	rs10001462	G	G	**Riem-001**	rs10001462	G	G	
FTD001	rs10001478	G	G	**Riem-001**	rs10001478	G	G	
FTD001	rs10001480	C	C	**Riem-001**	rs10001480	C	C	
FTD001	rs10001483	A	A	**Riem-001**	rs10001483	A	A	
FTD001	rs10001495	A	G	**Riem-001**	rs10001495	A	G	

FTD001	rs1000152	T	T	**Riem-001**	rs1000152	T	T	
FTD001	rs10001565	C	C	**Riem-001**	rs10001565	C	C	
FTD001	rs10001577	T	C	**Riem-001**	rs10001577	T	C	
FTD001	rs10001580	A	A	**Riem-001**	rs10001580	A	A	
FTD001	rs100016	A	G	**Riem-001**	rs100016	A	G	
FTD001	rs10001608	T	C	**Riem-001**	rs10001608	T	C	
FTD001	rs10001613	A	A	**Riem-001**	rs10001613	A	A	
FTD001	rs10001638	T	T	**Riem-001**	rs10001638	T	T	
FTD001	rs10001657	A	A	**Riem-001**	rs10001657	A	A	
FTD001	rs10001661	A	G	**Riem-001**	rs10001661	A	G	
FTD001	rs10001689	-	-	**Riem-001**	rs10001689	T	G	NO CALL
FTD001	rs10001694	-	-	**Riem-001**	rs10001694	A	A	NO CALL
FTD001	rs10001729	G	G	**Riem-001**	rs10001729	G	G	
FTD001	rs1000174	G	G	**Riem-001**	rs1000174	G	G	
FTD001	rs10001776	A	G	**Riem-001**	rs10001776	A	G	
FTD001	rs10001818	C	C	**Riem-001**	rs10001818	C	C	
FTD001	rs10001834	-	-	**Riem-001**	rs10001834	T	G	NO CALL
FTD001	rs10001851	C	C	**Riem-001**	rs10001851	C	C	
FTD001	rs10001877	C	C	**Riem-001**	rs10001877	C	C	
FTD001	rs1000198	T	G	**Riem-001**	rs1000198	T	G	
FTD001	rs10001989	G	G	**Riem-001**	rs10001989	G	G	
FTD001	rs10002028	T	T	**Riem-001**	rs10002028	T	T	
FTD001	rs10002068	G	G	**Riem-001**	rs10002068	G	G	
FTD001	rs10002121	A	G	**Riem-001**	rs10002121	A	G	
FTD001	rs10002142	A	C	**Riem-001**	rs10002142	A	C	
FTD001	rs10002171	C	C	**Riem-001**	rs10002171	C	C	
FTD001	rs10002189	C	C	**Riem-001**	rs10002189	C	C	
FTD001	rs1000219	T	G	**Riem-001**	rs1000219	T	G	
FTD001	rs1000221	G	G	**Riem-001**	rs1000221	G	G	
FTD001	rs10002224	C	C	**Riem-001**	rs10002224	C	C	
FTD001	rs10002254	A	G	**Riem-001**	rs10002254	A	G	
FTD001	rs1000226	T	C	**Riem-001**	rs1000226	T	C	
FTD001	rs10002325	A	A	**Riem-001**	rs10002325	A	A	
FTD001	rs1000236	A	G	**Riem-001**	rs1000236	A	G	
FTD001	rs10002388	A	A	**Riem-001**	rs10002388	A	A	
FTD001	rs10002393	T	T	**Riem-001**	rs10002393	T	T	
FTD001	rs10002472	A	A	**Riem-001**	rs10002472	A	A	
FTD001	rs1000251	G	G	**Riem-001**	rs1000251	G	G	
FTD001	rs1000255	T	T	**Riem-001**	rs1000255	T	T	
FTD001	rs1000256	T	T	**Riem-001**	rs1000256	T	T	
FTD001	rs10002573	T	T	**Riem-001**	rs10002573	T	T	

FTD001	rs10002598	C	C	**Riem-001**	rs10002598	C	C	
FTD001	rs1000264	A	G	**Riem-001**	rs1000264	A	G	
FTD001	rs10002688	A	G	**Riem-001**	rs10002688	A	G	
FTD001	rs10002695	C	C	**Riem-001**	rs10002695	C	C	
FTD001	rs1000270	-	-	**Riem-001**	rs1000270	A	C	NO CALL
FTD001	rs10002703	G	G	**Riem-001**	rs10002703	G	G	
FTD001	rs1000272	C	C	**Riem-001**	rs1000272	C	C	
FTD001	rs10002743	A	A	**Riem-001**	rs10002743	A	A	
FTD001	rs10002760	T	C	**Riem-001**	rs10002760	T	C	
FTD001	rs1000280	A	G	**Riem-001**	rs1000280	A	G	
FTD001	rs10002811	-	-	**Riem-001**	rs10002811	T	T	NO CALL
FTD001	rs10002842	C	C	**Riem-001**	rs10002842	C	C	
FTD001	rs1000286	C	C	**Riem-001**	rs1000286	C	C	
FTD001	rs1000290	A	A	**Riem-001**	rs1000290	A	A	
FTD001	rs10002917	T	C	**Riem-001**	rs10002917	T	C	
FTD001	rs10002965	C	C	**Riem-001**	rs10002965	C	C	
FTD001	rs10002981	A	A	**Riem-001**	rs10002981	A	A	
FTD001	rs1000299	G	G	**Riem-001**	rs1000299	G	G	
FTD001	rs10003025	T	C	**Riem-001**	rs10003025	T	C	
FTD001	rs10003040	T	C	**Riem-001**	rs10003040	C	C	WRONG CALL
FTD001	rs1000308	G	G	**Riem-001**	rs1000308	G	G	
FTD001	rs10003106	G	G	**Riem-001**	rs10003106	G	G	
FTD001	rs1000313	G	G	**Riem-001**	rs1000313	G	G	
FTD001	rs10003143	A	C	**Riem-001**	rs10003143	A	C	
FTD001	rs10003194	A	A	**Riem-001**	rs10003194	A	A	
FTD001	rs1000320	T	T	**Riem-001**	rs1000320	T	T	
FTD001	rs10003214	C	C	**Riem-001**	rs10003214	C	C	
FTD001	rs10003271	G	G	**Riem-001**	rs10003271	G	G	
FTD001	rs10003327	A	A	**Riem-001**	rs10003327	A	A	
FTD001	rs10003330	A	A	**Riem-001**	rs10003330	A	A	
FTD001	rs10003349	A	A	**Riem-001**	rs10003349	A	A	
FTD001	rs10003420	A	A	**Riem-001**	rs10003420	A	A	
FTD001	rs10003467	T	C	**Riem-001**	rs10003467	T	C	
FTD001	rs10003471	C	C	**Riem-001**	rs10003471	C	C	
FTD001	rs10003496	A	A	**Riem-001**	rs10003496	A	A	
FTD001	rs10003497	A	G	**Riem-001**	rs10003497	A	G	
FTD001	rs10003585	T	T	**Riem-001**	rs10003585	T	T	
FTD001	rs10003606	G	G	**Riem-001**	rs10003606	G	G	
FTD001	rs10003638	A	A	**Riem-001**	rs10003638	A	A	
FTD001	rs10003688	-	-	**Riem-001**	rs10003688	A	A	NO CALL
FTD001	rs10003689	-	-	**Riem-001**	rs10003689	T	T	NO CALL

FTD001	rs1000374	G	G	**Riem-001**	rs1000374	G	G	
FTD001	rs10003746	T	C	**Riem-001**	rs10003746	T	C	
FTD001	rs10003762	A	A	**Riem-001**	rs10003762	A	A	
FTD001	rs10003777	A	G	**Riem-001**	rs10003777	G	G	WRONG CALL
FTD001	rs1000380	T	T	**Riem-001**	rs1000380	T	T	
FTD001	rs10003861	C	C	**Riem-001**	rs10003861	C	C	
FTD001	rs10003864	T	T	**Riem-001**	rs10003864	T	T	
FTD001	rs10003889	A	A	**Riem-001**	rs10003889	A	A	
FTD001	rs10003904	A	A	**Riem-001**	rs10003904	A	A	
FTD001	rs1000394	G	G	**Riem-001**	rs1000394	G	G	
FTD001	rs10003973	C	C	**Riem-001**	rs10003973	C	C	
FTD001	rs10003999	C	C	**Riem-001**	rs10003999	C	C	
FTD001	rs10004010	A	A	**Riem-001**	rs10004010	A	A	
FTD001	rs10004030	A	A	**Riem-001**	rs10004030	A	A	
FTD001	rs10004038	A	A	**Riem-001**	rs10004038	A	A	
FTD001	rs10004063	T	T	**Riem-001**	rs10004063	T	T	
FTD001	rs1000408	C	C	**Riem-001**	rs1000408	C	C	
FTD001	rs1000411	T	T	**Riem-001**	rs1000411	T	T	
FTD001	rs10004128	T	C	**Riem-001**	rs10004128	T	C	
FTD001	rs10004130	G	G	**Riem-001**	rs10004130	G	G	
FTD001	rs1000416	C	C	**Riem-001**	rs1000416	C	C	
FTD001	rs10004181	T	C	**Riem-001**	rs10004181	T	C	
FTD001	rs1000421	T	T	**Riem-001**	rs1000421	T	T	
FTD001	rs10004266	T	C	**Riem-001**	rs10004266	T	C	
FTD001	rs1000427	G	G	**Riem-001**	rs1000427	G	G	
FTD001	rs10004274	A	A	**Riem-001**	rs10004274	A	A	
FTD001	rs10004285	-	-	**Riem-001**	rs10004285	G	G	NO CALL
FTD001	rs1000433	G	G	**Riem-001**	rs1000433	G	G	
FTD001	rs10004397	A	C	**Riem-001**	rs10004397	A	C	
FTD001	rs10004437	A	A	**Riem-001**	rs10004437	A	A	
FTD001	rs10004440	G	G	**Riem-001**	rs10004440	G	G	
FTD001	rs10004459	G	G	**Riem-001**	rs10004459	G	G	
FTD001	rs10004475	T	T	**Riem-001**	rs10004475	T	T	
FTD001	rs1000449	A	G	**Riem-001**	rs1000449	A	G	
FTD001	rs10004500	C	C	**Riem-001**	rs10004500	C	C	
FTD001	rs10004503	C	C	**Riem-001**	rs10004503	C	C	
FTD001	rs10004516	G	G	**Riem-001**	rs10004516	G	G	
FTD001	rs10004523	A	C	**Riem-001**	rs10004523	C	C	WRONG CALL
FTD001	rs10004557	C	C	**Riem-001**	rs10004557	C	C	
FTD001	rs10004565	T	T	**Riem-001**	rs10004565	T	T	
FTD001	rs1000459	C	C	**Riem-001**	rs1000459	C	C	

FTD001	rs10004643	A	A	**Riem-001**	rs10004643	A	A	
FTD001	rs1000465	C	C	**Riem-001**	rs1000465	C	C	
FTD001	rs1000466	T	T	**Riem-001**	rs1000466	T	T	
FTD001	rs10004753	T	G	**Riem-001**	rs10004753	T	G	
FTD001	rs10004765	C	C	**Riem-001**	rs10004765	C	C	
FTD001	rs10004776	-	-	**Riem-001**	rs10004776	T	T	NO CALL
FTD001	rs10004779	T	G	**Riem-001**	rs10004779	T	G	
FTD001	rs10004837	A	G	**Riem-001**	rs10004837	A	G	
FTD001	rs10004910	A	G	**Riem-001**	rs10004910	A	G	
FTD001	rs10004929	C	C	**Riem-001**	rs10004929	C	C	
FTD001	rs10004976	-	-	**Riem-001**	rs10004976	A	A	NO CALL
FTD001	rs10004988	C	C	**Riem-001**	rs10004988	C	C	
FTD001	rs1000499	G	G	**Riem-001**	rs1000499	G	G	
FTD001	rs1000505	A	A	**Riem-001**	rs1000505	A	A	
FTD001	rs10005058	A	A	**Riem-001**	rs10005058	A	A	
FTD001	rs1000507	C	C	**Riem-001**	rs1000507	C	C	
FTD001	rs10005074	T	C	**Riem-001**	rs10005074	T	C	
FTD001	rs1000510	-	-	**Riem-001**	rs1000510	A	C	NO CALL
FTD001	rs10005140	G	G	**Riem-001**	rs10005140	G	G	
FTD001	rs10005156	C	C	**Riem-001**	rs10005156	C	C	
FTD001	rs1000516	A	G	**Riem-001**	rs1000516	A	G	
FTD001	rs10005160	G	G	**Riem-001**	rs10005160	G	G	
FTD001	rs10005175	-	-	**Riem-001**	rs10005175	A	A	NO CALL
FTD001	rs1000521	A	A	**Riem-001**	rs1000521	A	A	
FTD001	rs10005242	G	G	**Riem-001**	rs10005242	G	G	
FTD001	rs10005281	A	A	**Riem-001**	rs10005281	A	A	
FTD001	rs1000530	-	-	**Riem-001**	rs1000530	C	C	NO CALL
FTD001	rs1000531	C	C	**Riem-001**	rs1000531	C	C	
FTD001	rs10005317	C	C	**Riem-001**	rs10005317	C	C	
FTD001	rs1000533	T	C	**Riem-001**	rs1000533	T	C	
FTD001	rs10005380	G	G	**Riem-001**	rs10005380	G	G	
FTD001	rs1000543	C	C	**Riem-001**	rs1000543	C	C	
FTD001	rs10005483	A	G	**Riem-001**	rs10005483	A	G	
FTD001	rs10005500	C	C	**Riem-001**	rs10005500	C	C	
FTD001	rs1000551	A	G	**Riem-001**	rs1000551	A	G	
FTD001	rs10005531	-	-	**Riem-001**	rs10005531	T	G	NO CALL
FTD001	rs10005550	A	G	**Riem-001**	rs10005550	G	G	WRONG CALL
FTD001	rs10005569	C	C	**Riem-001**	rs10005569	C	C	
FTD001	rs1000559	C	C	**Riem-001**	rs1000559	C	C	
FTD001	rs10005640	A	A	**Riem-001**	rs10005640	A	A	
FTD001	rs10005702	T	T	**Riem-001**	rs10005702	T	T	

FTD001	rs10005718	-	-	**Riem-001**	rs10005718	T	T	NO CALL
FTD001	rs1000574	-	-	**Riem-001**	rs1000574	C	C	NO CALL
FTD001	rs1000578	C	C	**Riem-001**	rs1000578	C	C	
FTD001	rs10005789	A	G	**Riem-001**	rs10005789	A	G	
FTD001	rs1000579	A	A	**Riem-001**	rs1000579	A	A	
FTD001	rs10005809	T	T	**Riem-001**	rs10005809	T	T	
FTD001	rs10005811	A	A	**Riem-001**	rs10005811	A	A	
FTD001	rs1000585	A	G	**Riem-001**	rs1000585	A	G	
FTD001	rs1000588	A	A	**Riem-001**	rs1000588	A	A	
FTD001	rs10005886	T	T	**Riem-001**	rs10005886	T	T	
FTD001	rs10005912	T	T	**Riem-001**	rs10005912	T	T	
FTD001	rs10005935	-	-	**Riem-001**	rs10005935	T	G	NO CALL
FTD001	rs10005962	C	C	**Riem-001**	rs10005962	C	C	
FTD001	rs10006044	-	-	**Riem-001**	rs10006044	A	G	NO CALL
FTD001	rs10006144	C	C	**Riem-001**	rs10006144	C	C	
FTD001	rs10006195	C	C	**Riem-001**	rs10006195	C	C	
FTD001	rs10006196	T	C	**Riem-001**	rs10006196	T	C	
FTD001	rs1000620	T	C	**Riem-001**	rs1000620	T	C	
FTD001	rs10006225	-	-	**Riem-001**	rs10006225	G	G	NO CALL
FTD001	rs10006242	A	G	**Riem-001**	rs10006242	A	G	
FTD001	rs10006245	A	A	**Riem-001**	rs10006245	A	A	
FTD001	rs10006257	C	C	**Riem-001**	rs10006257	C	C	
FTD001	rs10006308	-	-	**Riem-001**	rs10006308	A	A	NO CALL
FTD001	rs10006314	T	T	**Riem-001**	rs10006314	T	T	
FTD001	rs10006329	T	T	**Riem-001**	rs10006329	T	T	
FTD001	rs10006363	G	G	**Riem-001**	rs10006363	G	G	
FTD001	rs10006395	C	C	**Riem-001**	rs10006395	C	C	
FTD001	rs10006423	T	T	**Riem-001**	rs10006423	T	T	
FTD001	rs10006439	C	C	**Riem-001**	rs10006439	C	C	
FTD001	rs10006443	G	G	**Riem-001**	rs10006443	G	G	
FTD001	rs10006458	T	T	**Riem-001**	rs10006458	T	T	
FTD001	rs1000658	T	T	**Riem-001**	rs1000658	T	T	
FTD001	rs10006622	C	C	**Riem-001**	rs10006622	C	C	
FTD001	rs1000664	-	-	**Riem-001**	rs1000664	G	G	NO CALL
FTD001	rs10006644	G	G	**Riem-001**	rs10006644	G	G	
FTD001	rs1000672	-	-	**Riem-001**	rs1000672	G	G	NO CALL
FTD001	rs10006741	T	C	**Riem-001**	rs10006741	T	C	
FTD001	rs10006747	A	G	**Riem-001**	rs10006747	A	G	
FTD001	rs1000676	G	G	**Riem-001**	rs1000676	G	G	
FTD001	rs10006766	A	A	**Riem-001**	rs10006766	A	A	
FTD001	rs10006864	T	T	**Riem-001**	rs10006864	T	T	

FTD001	rs10006892	A	A	**Riem-001**	rs10006892	A	A	
FTD001	rs1000691	G	G	**Riem-001**	rs1000691	G	G	
FTD001	rs10006955	T	C	**Riem-001**	rs10006955	T	C	
FTD001	rs10006997	T	T	**Riem-001**	rs10006997	T	T	
FTD001	rs1000703	A	G	**Riem-001**	rs1000703	A	G	
FTD001	rs10007066	G	G	**Riem-001**	rs10007066	G	G	
FTD001	rs10007083	A	A	**Riem-001**	rs10007083	A	A	
FTD001	rs10007197	C	C	**Riem-001**	rs10007197	C	C	
FTD001	rs10007205	T	C	**Riem-001**	rs10007205	T	C	
FTD001	rs10007245	T	C	**Riem-001**	rs10007245	T	C	
FTD001	rs1000726	A	A	**Riem-001**	rs1000726	A	A	
FTD001	rs1000727	C	C	**Riem-001**	rs1000727	C	C	
FTD001	rs1000730	-	-	**Riem-001**	rs1000730	G	G	NO CALL
FTD001	rs10007367	T	T	**Riem-001**	rs10007367	T	T	
FTD001	rs1000737	T	C	**Riem-001**	rs1000737	T	C	
FTD001	rs1000738	G	G	**Riem-001**	rs1000738	G	G	
FTD001	rs10007396	A	A	**Riem-001**	rs10007396	A	A	
FTD001	rs10007440	C	C	**Riem-001**	rs10007440	C	C	
FTD001	rs10007472	T	T	**Riem-001**	rs10007472	T	T	
FTD001	rs1000752	A	A	**Riem-001**	rs1000752	A	A	
FTD001	rs10007543	A	A	**Riem-001**	rs10007543	A	A	
FTD001	rs10007576	T	C	**Riem-001**	rs10007576	T	C	
FTD001	rs10007590	A	A	**Riem-001**	rs10007590	A	A	
FTD001	rs10007601	A	G	**Riem-001**	rs10007601	G	G	WRONG CALL
FTD001	rs10007635	T	T	**Riem-001**	rs10007635	T	T	
FTD001	rs10007704	C	C	**Riem-001**	rs10007704	C	C	
FTD001	rs10007707	A	G	**Riem-001**	rs10007707	A	G	
FTD001	rs1000772	C	C	**Riem-001**	rs1000772	C	C	
FTD001	rs10007743	A	A	**Riem-001**	rs10007743	A	A	
FTD001	rs1000775	-	-	**Riem-001**	rs1000775	A	A	NO CALL
FTD001	rs1000778	G	G	**Riem-001**	rs1000778	G	G	
FTD001	rs10007784	T	T	**Riem-001**	rs10007784	T	T	
FTD001	rs10007790	G	G	**Riem-001**	rs10007790	G	G	
FTD001	rs10007810	A	G	**Riem-001**	rs10007810	A	G	
FTD001	rs10007812	G	G	**Riem-001**	rs10007812	G	G	
FTD001	rs10007887	C	C	**Riem-001**	rs10007887	C	C	
FTD001	rs10007934	-	-	**Riem-001**	rs10007934	G	G	NO CALL
FTD001	rs10007938	A	G	**Riem-001**	rs10007938	A	G	
FTD001	rs1000797	C	C	**Riem-001**	rs1000797	C	C	
FTD001	rs10007998	C	C	**Riem-001**	rs10007998	C	C	
FTD001	rs10008006	A	A	**Riem-001**	rs10008006	A	A	

FTD001	rs1000810	A	A	**Riem-001**	rs1000810	A	A	
FTD001	rs10008174	T	G	**Riem-001**	rs10008174	T	G	
FTD001	rs1000821	T	C	**Riem-001**	rs1000821	T	C	
FTD001	rs10008238	C	C	**Riem-001**	rs10008238	C	C	
FTD001	rs10008278	T	C	**Riem-001**	rs10008278	T	C	
FTD001	rs10008281	-	-	**Riem-001**	rs10008281	A	C	NO CALL
FTD001	rs10008418	-	-	**Riem-001**	rs10008418	A	G	NO CALL
FTD001	rs10008442	G	G	**Riem-001**	rs10008442	G	G	
FTD001	rs10008462	G	G	**Riem-001**	rs10008462	G	G	
FTD001	rs10008492	C	C	**Riem-001**	rs10008492	C	C	
FTD001	rs1000850	G	G	**Riem-001**	rs1000850	G	G	
FTD001	rs1000858	C	C	**Riem-001**	rs1000858	C	C	
FTD001	rs10008587	T	T	**Riem-001**	rs10008587	T	T	
FTD001	rs10008621	C	C	**Riem-001**	rs10008621	C	C	
FTD001	rs10008623	A	A	**Riem-001**	rs10008623	A	A	
FTD001	rs10008636	T	T	**Riem-001**	rs10008636	T	T	
FTD001	rs10008679	T	C	**Riem-001**	rs10008679	T	C	
FTD001	rs10008710	T	C	**Riem-001**	rs10008710	-	-	NO CALL
FTD001	rs10008744	A	A	**Riem-001**	rs10008744	A	A	
FTD001	rs10008779	A	C	**Riem-001**	rs10008779	A	C	
FTD001	rs10008808	A	G	**Riem-001**	rs10008808	A	G	
FTD001	rs10008834	T	T	**Riem-001**	rs10008834	T	T	
FTD001	rs10008858	T	C	**Riem-001**	rs10008858	T	C	
FTD001	rs10008892	C	C	**Riem-001**	rs10008892	C	C	
FTD001	rs10008894	A	G	**Riem-001**	rs10008894	A	G	
FTD001	rs10008923	A	G	**Riem-001**	rs10008923	A	G	
FTD001	rs10008952	G	G	**Riem-001**	rs10008952	G	G	
FTD001	rs10008978	C	C	**Riem-001**	rs10008978	C	C	
FTD001	rs10008981	C	C	**Riem-001**	rs10008981	C	C	
FTD001	rs10009030	C	C	**Riem-001**	rs10009030	C	C	
FTD001	rs10009033	G	G	**Riem-001**	rs10009033	G	G	
FTD001	rs10009093	G	G	**Riem-001**	rs10009093	G	G	
FTD001	rs1000911	G	G	**Riem-001**	rs1000911	G	G	
FTD001	rs10009111	A	G	**Riem-001**	rs10009111	A	G	
FTD001	rs10009152	T	T	**Riem-001**	rs10009152	T	T	
FTD001	rs10009186	C	C	**Riem-001**	rs10009186	C	C	
FTD001	rs1000919	A	G	**Riem-001**	rs1000919	G	G	WRONG CALL
FTD001	rs10009221	A	G	**Riem-001**	rs10009221	A	G	
FTD001	rs10009228	A	G	**Riem-001**	rs10009228	A	G	
FTD001	rs10009253	T	C	**Riem-001**	rs10009253	T	C	
FTD001	rs10009269	C	C	**Riem-001**	rs10009269	C	C	

FTD001	rs1000927	A	A	**Riem-001**	rs1000927	A	A	
FTD001	rs10009272	T	C	**Riem-001**	rs10009272	T	C	
FTD001	rs1000933	T	C	**Riem-001**	rs1000933	T	C	
FTD001	rs10009342	T	T	**Riem-001**	rs10009342	T	T	
FTD001	rs10009368	T	C	**Riem-001**	rs10009368	T	C	
FTD001	rs1000939	G	G	**Riem-001**	rs1000939	G	G	
FTD001	rs10009448	A	A	**Riem-001**	rs10009448	A	A	
FTD001	rs10009456	T	G	**Riem-001**	rs10009456	T	G	
FTD001	rs10009464	C	C	**Riem-001**	rs10009464	C	C	
FTD001	rs1000950	A	C	**Riem-001**	rs1000950	A	C	
FTD001	rs10009522	T	C	**Riem-001**	rs10009522	T	C	
FTD001	rs10009533	G	G	**Riem-001**	rs10009533	G	G	
FTD001	rs10009566	T	T	**Riem-001**	rs10009566	T	T	
FTD001	rs10009607	T	T	**Riem-001**	rs10009607	T	T	
FTD001	rs1000962	-	-	**Riem-001**	rs1000962	A	A	NO CALL
FTD001	rs10009693	C	C	**Riem-001**	rs10009693	C	C	
FTD001	rs10009701	A	G	**Riem-001**	rs10009701	A	G	
FTD001	rs10009750	T	T	**Riem-001**	rs10009750	T	T	
FTD001	rs10009807	G	G	**Riem-001**	rs10009807	G	G	
FTD001	rs10009816	C	C	**Riem-001**	rs10009816	C	C	
FTD001	rs10009834	T	T	**Riem-001**	rs10009834	T	T	
FTD001	rs10009892	C	C	**Riem-001**	rs10009892	C	C	
FTD001	rs10009898	T	C	**Riem-001**	rs10009898	T	C	
FTD001	rs10009967	A	G	**Riem-001**	rs10009967	A	G	
FTD001	rs10009998	G	G	**Riem-001**	rs10009998	T	G	WRONG CALL
FTD001	rs10010069	T	C	**Riem-001**	rs10010069	T	C	
FTD001	rs10010106	A	G	**Riem-001**	rs10010106	A	G	
FTD001	rs10010115	T	T	**Riem-001**	rs10010115	T	T	
FTD001	rs10010188	-	-	**Riem-001**	rs10010188	C	C	NO CALL
FTD001	rs10010217	T	C	**Riem-001**	rs10010217	T	C	
FTD001	rs10010236	G	G	**Riem-001**	rs10010236	G	G	
FTD001	rs10010240	G	G	**Riem-001**	rs10010240	G	G	
FTD001	rs10010241	-	-	**Riem-001**	rs10010241	T	T	NO CALL
FTD001	rs10010285	T	T	**Riem-001**	rs10010285	T	T	
FTD001	rs1001034	C	C	**Riem-001**	rs1001034	C	C	
FTD001	rs10010346	T	T	**Riem-001**	rs10010346	T	T	
FTD001	rs10010358	-	-	**Riem-001**	rs10010358	A	C	NO CALL
FTD001	rs10010385	T	T	**Riem-001**	rs10010385	T	T	
FTD001	rs10010445	A	G	**Riem-001**	rs10010445	A	G	
FTD001	rs10010466	T	T	**Riem-001**	rs10010466	T	T	
FTD001	rs10010472	A	G	**Riem-001**	rs10010472	A	G	

FTD001	rs1001049	A	C	**Riem-001**	rs1001049	A	C	
FTD001	rs10010505	G	G	**Riem-001**	rs10010505	G	G	
FTD001	rs10010565	C	C	**Riem-001**	rs10010565	C	C	
FTD001	rs10010623	G	G	**Riem-001**	rs10010623	G	G	
FTD001	rs10010638	C	C	**Riem-001**	rs10010638	C	C	
FTD001	rs1001073	G	G	**Riem-001**	rs1001073	G	G	
FTD001	rs10010758	T	T	**Riem-001**	rs10010758	T	T	
FTD001	rs1001098	T	C	**Riem-001**	rs1001098	T	C	
FTD001	rs1001099	T	T	**Riem-001**	rs1001099	T	T	
FTD001	rs10010994	A	C	**Riem-001**	rs10010994	A	C	
FTD001	rs10011007	A	A	**Riem-001**	rs10011007	A	A	
FTD001	rs1001110	T	C	**Riem-001**	rs1001110	C	C	WRONG CALL
FTD001	rs10011107	A	A	**Riem-001**	rs10011107	A	A	
FTD001	rs1001111	T	T	**Riem-001**	rs1001111	T	T	
FTD001	rs10011134	T	C	**Riem-001**	rs10011134	T	C	
FTD001	rs10011142	T	C	**Riem-001**	rs10011142	T	C	
FTD001	rs10011174	G	G	**Riem-001**	rs10011174	G	G	
FTD001	rs1001118	T	C	**Riem-001**	rs1001118	T	C	
FTD001	rs10011190	-	-	**Riem-001**	rs10011190	G	G	NO CALL
FTD001	rs1001120	C	C	**Riem-001**	rs1001120	C	C	
FTD001	rs10011202	A	A	**Riem-001**	rs10011202	A	A	
FTD001	rs1001131	A	A	**Riem-001**	rs1001131	A	A	
FTD001	rs10011351	A	A	**Riem-001**	rs10011351	A	A	
FTD001	rs10011402	T	G	**Riem-001**	rs10011402	T	G	
FTD001	rs1001145	A	G	**Riem-001**	rs1001145	A	G	
FTD001	rs10011452	C	C	**Riem-001**	rs10011452	C	C	
FTD001	rs1001148	T	C	**Riem-001**	rs1001148	T	C	
FTD001	rs1001149	T	C	**Riem-001**	rs1001149	T	C	
FTD002	rs1000000	C	C	**Riem-002**	rs1000000	C	C	
FTD002	rs10000023	T	T	**Riem-002**	rs10000023	T	T	
FTD002	rs10000030	-	-	**Riem-002**	rs10000030	A	A	NO CALL
FTD002	rs10000041	T	T	**Riem-002**	rs10000041	T	T	
FTD002	rs1000007	A	A	**Riem-002**	rs1000007	A	A	
FTD002	rs10000081	T	T	**Riem-002**	rs10000081	T	T	
FTD002	rs10000092	T	T	**Riem-002**	rs10000092	T	T	
FTD002	rs1000016	A	A	**Riem-002**	rs1000016	A	A	
FTD002	rs10000160	G	G	**Riem-002**	rs10000160	G	G	
FTD002	rs10000169	C	C	**Riem-002**	rs10000169	C	C	
FTD002	rs10000185	T	C	**Riem-002**	rs10000185	T	C	
FTD002	rs1000022	T	C	**Riem-002**	rs1000022	T	C	

Appendix 2-2. Discovery phase: Phenotype data collection

Sample ID	Age at Onset	Age at Death	Gender	Diagnosis					Family History	Neuropsych data available (y/n)	Imaging data available (y/n)	Path data available (y/n)	Proband ID	MAPT	GRN	TDP43	FUS	CHMP2B	APOE
				bvFTD	SD	PNFA (excluding all LPA cases)	FTD-MND	FTLD unspecified	(in first degree relative)	(and supportive of FTLD)	(and supportive of FTLD)	(or from family member)	(if other family member	exon 1, 9-13 Hap					

Appendix 2-3. Replication phase: Phenotype data collection

Sample ID	Age at Onset	Age at Death	Gender	Diagnosis					Family History	Neuropsych data available (y/n)	Imaging data available (y/n)	Path data available (y/n)	Proband ID	MAPT			MAPT hap	PGRN	TDP43	FUS	CHMP2B	APOE	chr9 repeats
				bvFTD	SD	PNFA (excluding all LPA cases)	FTD-MND	FTLD unspecified	(in first degree relative)	(and supportive of FTLD)	(and supportive of FTLD)	(or from family member)	(if other family member included)	1, 9	10	11, 12, 13							

Appendix 2-4. NDPT098 Coriell plate

NDPT098: Neurologically Normal Caucasian Control Plate

(Replaces NDPT022)

Control Individuals

Age Range at Time of Collection: 56-91

46 Males/46 Females

Ref	Well Number	blank well for user control
ND08998	A01	
ND10875	A02	
ND06787	A03	
ND11515	A04	
ND08394	A05	
ND03937	A06	
ND09880	A07	
ND02206	A08	
ND05808	A09	
ND10728	A10	
ND05414	A11	
ND02747	A12	
ND07950	B01	
ND05771	B02	
ND10393	B03	
ND09830	B04	
ND05008	B05	
ND08492	B06	
ND06434	B07	
ND05684	B08	
ND05682	B09	
ND05899	B10	
ND09791	B11	
ND07536	B12	
ND01680	C01	
ND10977	C02	
ND06350	C03	
ND06432	C04	
ND10275	C05	
ND10864	C06	
ND10769	C07	
ND10827	C08	
ND08240	C09	
ND10496	C10	
ND05653	C11	
ND10297	C12	
ND05900	D02	
ND10541	D03	
ND06362	D04	
ND06234	D05	
ND01789	D06	
ND09589	D07	
ND08978	D08	
ND10876	D09	
ND10743	D1	

ND07112	D10
ND05295	D11
ND03927	D12
ND04237	E01
ND09957	E02
ND09939	E03
ND07615	E04
ND07427	E05
ND09579	E06
ND05868	E07
ND07423	E08
ND10897	E09
ND06235	E10
ND10529	E11
BLANK	E12
ND03363	F01
ND12254	F02
ND07980	F03
ND10906	F04
ND10896	F05
ND06010	F06
ND06232	F07
ND07300	F08
ND05981	F09
ND06228	F10
ND07535	F11
BLANK	F12
ND07105	G01
ND08927	G02
ND07048	G03
ND06351	G04
ND08955	G05
ND06026	G06
ND07428	G07
ND08079	G08
ND10469	G09
ND06855	G10
ND09671	G11
BLANK	G12
ND05501	H01
ND06756	H02
ND05997	H03
ND09777	H04
ND08059	H05
ND10189	H06
ND10655	H07
ND05809	H08
ND09789	H09
ND10668	H10
ND10577	H11
BLANK	H12

Appendix 2-5. NABEC and UKBEC

NABEC and **UKBEC** members:

Group	First Name	Middle	Surname	Affiliation 1	Affiliation 2
NABEC	Andrew	B	Singleton	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA	
NABEC	Mark	R	Cookson	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA	
NABEC	J. Raphael		Gibbs	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA	Reta Lila Weston Institute and Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK
NABEC	Dena	G	Hernandez	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA	Reta Lila Weston Institute and Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK
NABEC	Allissa		Dillman	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA	Department of Neuroscience, Karolinska Institutet, 171 77 Stockholm, Sweden
NABEC	Michael	A	Nalls	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA	
NABEC	Alan	B	Zonderman	Research Resources Branch, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA	
NABEC	Sampath		Arepalli	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA	
NABEC	Luigi		Ferrucci	Clinical Research Branch, National Institute on Aging, Baltimore, MD, USA	
NABEC	Robert		Johnson	NICHD Brain and Tissue Bank for Developmental Disorders, University of Maryland Medical School, Baltimore, Maryland 21201, USA	
NABEC	Dan	L	Longo	Lymphocyte Cell Biology Unit, Laboratory of Immunology, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA	
NABEC	Richard		O'Brien	Brain Resource Center, Johns Hopkins University, Baltimore, MD, USA	
NABEC	Bryan		Traynor	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA	

NABEC Juan Troncoso Brain Resource Center, Johns Hopkins University, Baltimore, MD, USA

NABEC Marcel van der Brug Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA ITGR Biomarker Discovery Group, Genentech, South San Francisco, CA, USA

NABEC Ronald H Zielke NICHD Brain and Tissue Bank for Developmental Disorders, University of Maryland Medical School, Baltimore, Maryland 21201, USA

UKBEC John Hardy Reta Lila Weston Institute and Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK

UKBEC Michael E Weale Department of Medical and Molecular Genetics, King s College London, 8th Floor, Tower Wing, Guy s Hospital, London SE1 9RT, UK

UKBEC Mina Ryten Reta Lila Weston Institute and Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK

UKBEC Adaikalavan Ramasamy Department of Medical and Molecular Genetics, King s College London, 8th Floor, Tower Wing, Guy s Hospital, London SE1 9RT, UK Reta Lila Weston Institute and Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK

UKBEC Daniah Trabzuni Reta Lila Weston Institute and Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London WC1N 3BG, UK Department of Genetics, King Faisal Specialist Hospital and Research Centre, PO Box 3354, Riyadh 11211, Saudi Arabia

UKBEC Colin Smith Department of Neuropathology, MRC Sudden Death Brain Bank Project, University of Edinburgh, Wilkie Building, Teviot Place, Edinburgh EH8 9AG

UKBEC Robert Walker Department of Neuropathology, MRC Sudden Death Brain Bank Project, University of Edinburgh, Wilkie Building, Teviot Place, Edinburgh EH8 9AG

Appendix 4-1. FTD-GWAS: Initial questionnaire

General questions for the collaborators for the WGA about their samples. We will need to check if the quality of their information is enough detailed.

Also prepare an excel file for specific questions.

1. Information about the patient: do they know where exactly the patient comes from (to see if the patients could be distantly related)? What is their geography?
2. Information about experiments performed on the samples:
   - MAPT: sequenced for exons 1, 9, 10, 11, (12), 13?

     Gene dosage?
   - PGRN: sequenced for exons 1,-, 12?

     Gene dosage?
   - CHMP2B: sequenced for exons 1,-, 6 (John suggests that from what is known from the literature it is not worth it anymore to check on exons 1, 2, 3, 4).
   - Haplotype genotyping
3. Information on the diagnosis for sporadic cases:
   - Behavioral variant
   - Semantic dementia
   - Progressive non-fluent aphasia
   - FTD/MND/ALS
   - FTD not further described
4. Information on the pathology for sporadic cases:
   - Tangle? (cases not sent to UPENN)
   - TDP43? (cases not sent to UPENN)
   - Ubiquitin-positive inclusions? (cases not sent to UPENN)
5. Information on the familial cases; do they show;
   - MAPT mutations?

- PGRN mutations
- CHMP2B mutations?
- Other familial cases?

After having received the information by phone we will send out an excel file to cover all the details.

We will send bar coded 96 wells plates to the collaborative groups with 4 blanks (so that we know what the direction of the plate is).

Appendix 4-2. FTD-GWAS meeting 2009: Minutes

Minutes/slide

Slide1

Names and contacts of the people leading the WGA-FTD project are shown to acknowledge the attendants

Slide2

The 1) list of the collaborative groups taking part of the project so far and 2) the number of samples that are going to be collected for initial screening are displayed for everyone to have an idea about who is involved. John acknowledges the attendants that there is a verbal confirmation for funding of the study and genotyping of ≥ 3000 samples. Also he says that there is availability of a large amount of control data for samples coming from UK, USA and ITALY and there is less data available from other European countries like BELGIUM, GERMANY, NETHERLANDS, DENMARK, SWEDEN and SPAIN.

Slide3

Summarizes the main issues to be discussed:

- **a**. Sample quality and sample shipping
- **b**. Diagnostic criteria and extra clinical information
- **c**. The genotyping and analysis of samples with mutations
- **d**. The genotyping and analysis of more than one sample from the same pedigree.
- **e**. Data availability

Slide4

Point **a**. is discussed: John says that if we are going to do chipping only 2ug of DNA are going to be enough. The amount of DNA should be increased up to at least 5ug if also sequencing is going to be performed. Andy says that first he will need to prepare the IRB for permission to collect the samples, process that would take, more or less, 3-4 months.
John says that the time frame to collect samples should be between now and January 2010.
The mailing contact of Cynthia Crews is shown.

PM&RF
As of the meeting we think it is clear that we will need 2ug DNA for Genotyping or 5ug if we decide to perform sequencing. We should decide if we want to do sequencing for the samples involved in the study before the genotyping (fill the gaps for those samples that have not been screened yet) and if yes what genes (TAU, PGRN, FUS, TAU haplotype and ApoE genotype?).
In that case people should then prepare 2 aliquotes: 1) with 2ug for genotyping to be sent to Andy's lab and 2) with 3ug for sequencing to be shipped to our lab in Texas. So we can do the sequencing before and after genotyping.

Slide5

Point **b**. is discussed: John summarizes that we should consider the Neary criteria for diagnosis considering enrollable patients those with Behavioural variant, language variant (semantic dementia,

progressive non fluent aphasia), and maybe FTD-MND. John says that also imaging (MRI, SPEC [single-photon emission computed tomography], FDG-PET) is considered as a good diagnostic tool.
Diagnosis: we had some controversial views on how to manage this issue, in fact some attendants were keen on reviewing the cases with a novel agreed upon set of criteria some were not. Neary criteria plus imaging (if available) should be followed/used to identify clinical characteristics of the samples/patients. Rohrer said that trying to collect all these information is time consuming (time is money) and that some groups might have more information than others and there would be discrepancies in the type and amount of available clinical data. Vivianna says that MRI/PET SCAN would raise the quality of the clinical information. Mackenzie was wondering how much an accurate diagnosis would influence the study since FTD will mainly be BV and PNFA. Neil said that in any way it would be good if every group would add most possible clinical data for their samples and Andy agreed that it would be good to know a priori the characteristics of the samples to refine genotyping and statistical analysis.
As a conclusion for this section it was agreed to keep CBS/CBD samples out of the study. Also, autopsy cases are accepted for enrollment. It was suggested to have a meeting between physicians to further discuss and define clinical criteria. Right now we would go by Neary criteria (should we include FTD-MND?). For clinical data it would be good, if possible, to include neuropsychology, imaging and pathology. But, in the end, as a minimum, for enrollment, the Neary criteria should be met.

Slide6

Point **c**. and **d**. are discussed:
c. John emphasizes that the first important question regards whether to include or not those samples with mutations. Further what mutations do we need to care about? *TAU*, *PGRN*, *CHMP2B*, *VCP*, *TDP-43*, *FUS*? John said that it would be helpful that every group would share the sequencing analysis results for the samples they want to share for the project. Vivianna says that probably not all the groups will have sequencing data available for all the samples they want to share and that it would be good to have an idea of what has been sequenced per group.
Rosa said that in case we care about the sequencing data we should only worry about *TAU* and *PGRN*. Based on Van Broeckhoven suggestions, only samples with no variants in *TAU* and *PGRN* should be genotyped.
At the end of the discussion it was proposed to have at least sequencing data available for *TAU* and *PGRN*.

PM&RF
We think that it would be important to include mutation carriers in the study and divide the samples for genotyping into 2 groups: 1) mutation carriers and 2) non mutation carriers. It would be an internal control and a tool to compare the results at the end of genotyping.

If John agrees Raf can start screening of the samples which have not been sequenced for TAU, PGRN, FUS, TAU haplotype and ApoE genotype as soon as the samples are available. If John wants to do so, Parastoo will submit the IRB protocol soon.
We think that before starting genotyping we should have a clear sequencing characterization. It is understandable that doing the sequencing is time consuming and that some labs might even not be able to perform those experiments. The DNA can be shipped to us in Texas for sequencing purposes. Raf can do all the necessary sequencing: our lab is well equipped, we have PCR machines and 3730XL and every other equipment necessary for sequencing. Our sequencing protocol is highly standardized and efficient. So genotyping will be performed at NIH, while sequencing could all be performed at Texas Tech. We have developed protocols for gene dosage for TAU and PGRN that we could perform if necessary (we do it for all of our samples).

d. Big question to be answered here is whether we want to include or not family members of the proband. John says **NO**, we should not, unless we want to produce linkage data and also Amalia Bruni said that linkage could be useful. A final decision on that is pending.

Slide7-8

Data analysis and data availability. The last two slides have been presented but not deeply discussed. For now the analysis will regard association, maybe long haplotype analysis (for founders), sequencing after loci will be (hopefully) identified and experiments for deletions and gene duplications. It will also be interesting to consider the results of the van Deerlin study (TDP-43 WGA)/MND and PSP(Schellenberg WGA)/AD misdiagnosis studies (?).
Data could be made available through dbGAP that, Andy said,will need 6-9 months to be prepared. In this case everyone of the groups could be able to check for their updated data online at anytime. Important is to communicate when results are close to be delivered and discuss results by conference calls and eventually meetings prior to data analysis.

GENERAL COMMENTS.

Rosa said that we could delay the start date of the project in our request for money to have time to do the IRB and collect the samples.

John said that he wants to create a Google doc to communicate and share information/files between groups.

Van Broeckhoven said:
1. Is money there or not yet?
2. Speed up the process to get genotyping data as soon as possible.
3. Focus on FTD and try to be sure to exclude AD
4. Include samples showing no variants and no multiple samples from the same family
5. Yes to most complete clinical data (MRI, SPEC, FDG-PET) if available
6. What about money for follow up?

Mackenzie gave his point of view for diagnosis and clinical information. He said that there should be a minimum of clinical information for the samples. He suggested having a meeting for physicians to identify the best minimum criteria to categorize the samples clinically and he suggested that neuro-imaging and basic initial criteria should be used before beginning the study.

The group of Italians who were present (A. Bruni, G. Rossi and G. Binetti) asked to develop general and universal inclusion/exclusion criteria for samples enrollment and for a copy of the Grant application so that they can do their IRB. Specifically A. Bruni said that she has problems in doing sequencing because she does not have an actual lab. So I told her for *TAU* and *PGRN* we can do sequencing for her in Texas (and she said that it is fine but she wants to be sure legally that we won't use data obtained from her samples to write our own papers [!!!!!]). She also said that it is not clear to her how to consider sporadic or familial cases for the study…. Finally she said that she has well characterized clinical data (neuropsychology, neuroimaging, etc) and that she thinks it would be good to write a legal agreement on how to manage data on the dbGAP (!!).

TIME TABLE.

1) Now to 3-4 months IRB approval at NIH and at Texas Tech University.
2) Start collecting the samples at the two facilities.

Appendix 4-3. FTD-GWAS protocol

WGA

Whole genome association studies (WGAs) represent a proven successful method for genotyping and for mapping novel common variants underlying common diseases. The requirements for successful GWAs can be summarized as follows: large cohort sizes, well-characterized phenotypes and, robust association and follow-up platforms.

The Principal Investigator of the proposed study, Dr. John Hardy, organized an open meeting at the Rotterdam FTD meeting in September 2008 to promote the participation in the project of all centers that have been and are leaders in the investigation of Frontotemporal Dementia (FTD).

FTD

Frontotemporal Dementia (FTD) belongs to a wide spectrum of neurological disorders and is the focus of this study: FTD is the second or third most common cause of dementia in the developed world (depending on disease definitions). It is clinically, pathologically and genetically heterogeneous and presents with a frequency that increases with age. The clinical diagnosis of FTD relies on the presence of a progressive dementia together with variable initial features that can include behavioral changes, semantic problems or aphasia. FTD has been defined, clinicopathologically, by the presence of either TDP-43 or tau/tangle inclusions (1). Mutations in the 3 known genes (MAPT, PGRN and CHMP2B) explain many, but not all of the autosomal dominant cases and truncating mutations in PGRN (which is incompletely penetrant) explains a high proportion, though not all, of the familial cases. The clinical phenotype, as in most neurologic disorders, is variable due to the location and density of the lesions in the brain. Four clinical subgroups of the syndrome have been characterized:

1. Behavioural variant (50% of cases with evenly split tau and TDP-43 pathology);
2. Semantic Dementia variant (20% of cases with usually TDP-43 pathology);
3. Progressive nonfluent aphasia (20% of cases with largely tau pathology) and
4. Cases with MND complications (10% of cases all with TDP-43 pathology).

As of today, even if there has been tremendous progress in FTD genetics the role of genetic factors for common forms of FTD is still poorly understood.

AIM

The proposed study is the largest and most comprehensive GWA study for FTD to date.

The major aim is to discover common genetic variants that influence disease risk, progression and variation.

STUDY DESIGN

Samples collection

A large series of well characterized patients is expected to be collected (~3500 cases). This study will be based on a series of ongoing detailed clinical studies from nearly all the centers who have published on FTD in the last 5 years.

Samples will be collected either in tubes (sent by the owner) or in 96 wells plates sent by the group organizing the study to be filled by the owner of samples and sent back by courier. Samples will be sent to and managed by Cynthia Crews, Building 35 Room 1A 1000 35 Lincoln Drive Bethesda MD20892 at NIH and by John Hardy and Raffaele Ferrari at UCL.

Samples will be stored with a specific ID and a description of their main clinical, pathological and genetic known characteristics will be integrated.

The samples are all of European-North American extraction. Whole genome amplified samples will not be used for the WGA.

The amount of DNA requested will be varying between 2-5 µg. DNA will be used for genotyping in the initial phase of the study. In a second phase DNA will be used to perform sequencing, gene dosage and replicates. Unused DNA will be returned to the source.

Inclusion/exclusion criteria

Clinically: samples will be collected based on the Neary Criteria. Clinical features as behavioral variant, language variant (semantic dementia, progressive non fluent aphasia) and FTD-MND will be considered as relevant for sample enrollment.

Genetically: samples carrying mutations in MAPT and PGRN will be initially excluded from the study. Sporadic cases of FTD will be included in the study and, for familial cases, just one sample (index patient) will be included in the study. Samples should be characterized for MAPT haplotype, ApoE and MAPT/PGRN sequencing in pre-genotyping phase.

Genetic screening: genotyping

Illumina 610-660 Quad chips will be used for genotyping. Dr. Hardy's group has extensive experience in generating and handling the data from this robust and reliable platform. Control data are already available for this platform: the WTCCC data will be used for UK; own data on >2000 US controls (now publicly available at dbGAP) will be used for the US; own data generated on ~600 samples on the inCHIANTI study will be used for Italy; 50 control samples are available for Spain and more are going to be generated as part of our NIH intramurally funded whole genome study in ischaemic stroke. Additionally, Swedish and German Illumina control data from other centers are available as well as the CEPH-Human Genome Diversity Panel.

Analysis

Polymorphisms with real effects will not necessarily have the lowest p values in the study, but may be hidden amongst many other polymorphisms that have similar or greater levels of association. The issue of false positives will be best addressed with replication and by increasing numbers of samples.

Data cleaning: after gender and call quality checks, Eigenstrat will be used to assess whether cases and controls are matched. Cases or controls which are outliers won't be used. SNPs which have a minor allele frequency of <5% will be discarded.

Primary analysis will be looking for additive associations using PLINK. The top 1,000 SNPs in the UPenn FTD, the MND and the PSP/CBD databases and those which replicate in any of the 3 other studies* (at $p<0.01$ for this stage alone) will be initially assessed.

At this stage, also the SNPs which show putative association for their affects on gene expression, (particularly of the MAPT and TDP43 genes) as well as for *cis* effects on their own expression in the expression databases will be assessed.

Because much of the control data is historical and because samples are of different ethnicities, association analysis will be arduous process and will probably show many apparent strong positives (due to chip differences). Several have shown up in previous analyses and thus easily identifiable; undoubtedly, there will be new ones which will need to be checked visually and bioinformatically. Unreliable associations will be discarded.

Additional analyses: in datasets as rich as whole genome sets, there are also additional ways in which we may find risk genes for FTD. 1. homozygosity mapping since, even if in ostensibly outbred populations, a significant amount of autozygosity occurs (2): this approach could point to causative loci (3), 2. identification of insertions and deletions (an ataxia locus was recently found by this approach [4]) and 3. identification of long common haplotypes in affected individuals (5). There is a history of useful common founders in FTD syndromes: for example the occurrence of many families with a tau mutation and a common Welsh founder (6) and a common Belgian founder for many of the FTD kindreds with (inpenetrant) PGRN mutations (7). Data obtained by this study will be parsed to search for such common haplotypes.

Database entry and management

A central database (dbGAP) will be established. The db will contain age at onset, sex and centre codes together with the whole genome data set. Data collection and uploading will take approximately 6-9 months.

EXPECTED OUTCOMES

One of the major difficulties underlying complex disease is the presumed and likely heterogeneous nature of the disorder. One of the strengths of this consortium is the previous work by the applicants in their patient series to permit further analysis on subsets of the samples.

We expect to identify novel loci relevant to disease onset and progression and, possibly, we will be able to begin understanding the genetic basis of the different subtypes of the disease.

Finally, we expect this study to be also useful in enlightening on the mechanisms of cell death in Alzheimer's disease (AD), motor neuron disease (MND) and progressive supranuclear palsy (PSP) because FTD shares pathologic features with these diseases.

REFERENCES

1. Mackenzie IR, Neumann M, Bigio EH, Cairns NJ, Alafuzoff I, Kril J, Kovacs GG, Ghetti B, Halliday G, Holm IE, Ince PG, Kamphorst W, Revesz T, Rozemuller AJ, Kumar-Singh S, Akiyama H, Baborie A, Spina S, Dickson DW, Trojanowski JQ, Mann DM. Nomenclature for neuropathologic subtypes of frontotemporal lobar degeneration: consensus recommendations. Acta Neuropathol. 2009 Jan;117(1):15-8.

2. Simon-Sanchez, J., S. Scholz, H.C. Fung, M. Matarin, D. Hernandez, J.R. Gibbs, A. Britton, F. Wavrant de Vrieze, E. Peckham, K. Gwinn-Hardy, A. Crawley, J.C. Keen, J. Nash, D. Borgaonkar, J. Hardy, and A. Singleton. 2006. Genome-wide SNP assay reveals structural genomic variation, extended homozygosity and cell-line induced alterations in normal individuals. Hum Mol Genet 16:1-14.

3. Simon-Sanchez J, Scholz S, Matarin M del M, Fung HC, Hernandez D, Gibbs JR, Britton A, Hardy J, Singleton A. Genomewide SNP assay reveals mutations underlying Parkinson disease. Hum Mutat. 2008 Feb;29(2):315-22

4. van de Leemput J, Chandran J, Knight MA, Holtzclaw LA, Scholz S, Cookson MR, Houlden H, Gwinn-Hardy K, Fung HC, Lin X, Hernandez D, Simon-Sanchez J, Wood NW, Giunti P, Rafferty I, Hardy J, Storey E, Gardner RJ, Forrest SM, Fisher EM, Russell JT, Cai H, Singleton AB. Deletion at ITPR1 Underlies Ataxia in Mice and Spinocerebellar Ataxia 15 in Humans. PLoS Genet. 2007 Jun 22;3(6):e108

5. Kong A, Masson G, Frigge ML, Gylfason A, Zusmanovich P, Thorleifsson G, Olason PI, Ingason A, Steinberg S, Rafnar T, Sulem P, Mouy M, Jonsson F, Thorsteinsdottir U, Gudbjartsson DF, Stefansson H, Stefansson K. Detection of sharing by descent, long-range phasing and haplotype imputation. Nat Genet. 2008 Aug 17

6. Pickering-Brown SM, Richardson AM, Snowden JS, McDonagh AM, Burns A, Braude W, Baker M, Liu WK, Yen SH, Hardy J, Hutton M, Davies Y, Allsop D, Craufurd D, Neary D, Mann DM. Inherited frontotemporal dementia in nine British families associated with intronic mutations in the tau gene. Brain. 2002 Apr;125(Pt 4):732-51.

7. van der Zee J, Rademakers R, Engelborghs S, Gijselinck I, Bogaerts V, Vandenberghe R, Santens P, Caekebeke J, De Pooter T, Peeters K, Lübke U, Van den Broeck M, Martin JJ, Cruts M, De Deyn PP, Van Broeckhoven C, Dermaut B. A Belgian ancestral haplotype harbours a highly prevalent mutation for 17q21-linked tau-negative FTLD. Brain. 2006 Apr;129(Pt 4):841-52.

*

The small majority of cases of FTD (TDP-43 pathology) has two replication series:

(**a**) the WGA at UPenn on ~600 pathologically defined FTD cases which is also being run on an Illumina platform (PI: Dr. Vivianna van Deerlin)
(**b**) the collaborative WGA on motor neuron disease which we are part of: we have run ~3000 MND samples from the US, Ireland, Italy and Finland on the Illumina platform (analysis is ongoing) and a collaborative group is being put together to perform a meta-analysis on ~7,000 MND samples including ours by the (US) ALS Association which is due to be complete ~July 2009.

The large minority of cases of FTD has MAPT pathology has an immediately relevant replication series:

(**c**) the collaborative whole genome analysis of progressive supranuclear palsy (1,200 samples) and corticobasal degeneration (140 samples) which is being run at UPenn (PI: Dr. Gerry Schellenberg).

Appendix 4-4. FTD-GWAS meeting October 2011: Minutes

FTD-GWAS – Montreal October 12, 2011

Minutes

Meeting started at **11:00AM** EST-US time. Attendees were as follows:

Attendees	Centers
Stuart Pickering-Brown	University of Manchester, Clinical Neuroscience, Manchester, UK
Rosa Rademakers	Mayo Clinic Jacksonville, Florida, USA
Julie van der Zee, Ilse Gijselinck, Marc Cruts	Department of Molecular Genetics, University of Antwerp, BELGIUM
John van Swieten	Department of Neurology, Erasmus Medical Center, Rotterdam, NETHERLANDS
Pau Pastor, Oswaldo Lorenzo	Center for Applied Medical Research, Division of Neurology, University of Navarra, Pamplona, SPAIN
Parastoo Momeni, Raffaele Ferrari	Texas Tech University Health Sciences Center, Lubbock, Texas, USA
Huei-hsin Chiang	Karolinska Institute, Department of NVS, Stokholm, SWEDEN
Alexandra Dürr	French Consortium, FRANCE
Riemenschneider Matthias	Universität des Saarlandes, Klinikum für Psychiatrie & Psychotherapie, Homburg/Saar, GERMANY
Andy Singleton, Mike Nalls, Bryan Traynor	Laboratory of Neurogenetics, NIA, NIH, Bethesda, MA, USA
Carlos Cruchaga	Dept. of Psychiatry, Washington University School of Medicine, St Louis, MO, USA
John Hardy, Rita Guerreiro, Jose Bras	Department of Molecular Neuroscience, Institute of Neurology, UCL, Queen's Square, London, UK

John Hardy introduced the FTD-GWAS project mentioning, at first, the initial funding issues. In 2007 the project was submitted to the Wellcome Trust UK, but, even if appreciated by the reviewing committee, it was not funded at the time. Afterwards, the project was presented at NINDS and was funded by Intramural funding through intercession of **Story Landis** and **Andy Singleton**. This allowed providing Illumina 660K chips to perform genotyping. Samples collection and genotyping was then split between the two facilities, University College of London (UCL) and National Institute of Health (NIH). Samples collection and genotyping happened between January 2010 and May 2011. **John Hardy** acknowledged **Jonathan Rohrer** for his help in defining clinical criteria for samples' enrollment. **John Hardy** also mentioned that in July 2011 genotyping data were sent back to every individual group for them to start looking into their own cohorts and that, recently, 1. **Raf Ferrari** has checked for genotyping quality as well as for duplications/deletions in chromosome 17 (especially at the *MAPT*/*PGRN* loci) and 2. **Mike Nalls** has performed preliminary checks and analysis on the whole cohort. **John Hardy** finally mentioned that he would think of performing the analysis as a whole group, the FTD-Consortium, and that the FTD-GWAS meeting in Montreal would be crucial for strategizing and planning the next steps of the project.

Then **John Hardy** introduced the first talk held by **Raf Ferrari** at ~ **11:30AM**.

INSERT RAF'S SLIDES HERE

During and after **Raf**'s talk some discussion points came up. **John van Swieten** asked what happened to 1 plate that was sent from Amsterdam and that seemingly was not genotyped and **Rosa Rademakers** asked what happened to **Ian Mackenzie**'s samples as they were not shown in the slides. **John Hardy**'s answer was that we will need to double check internally. Another point that was briefly discussed was concerning the criteria for performing primary analysis: **Julie van der Zee** asked if we are going to include mutation carriers at all and **Rosa Rademakers** asked how will Pick's – Tau – FUS pathology be considered in terms of analysis. **John Hardy** said that we are probably going to exclude mutation carriers even though we don't know about all samples as many haven't been screened and, yes, samples with pathology as above (corresponding to a clinical diagnosis falling between bvFTD, SD, PNFA and FTD-MND) will be included. **Mike Nalls** said that all samples with diagnosis other than bvFTD, SD, PNFA and FTD-MND will need to be taken out from the study. **John Hardy** said that samples with the diagnosis of (or path confirmed) AD, PSP, CBS-CBD and, of course, normal controls will be fished out and excluded before actual primary analysis. **Andy Singleton** added, concerning replication phase, that it will depend on hits identified after primary analysis: 1. if hits will be found we will replicate those hits and a cohort of ≥ 1,000 samples would suffice or 2. if we don't get hits we should consider a sample size of at least ≥ 2,000 samples. Afterwards, **Mike Nalls**' talk followed at ~ **12:00PM**.

INSERT MIKE'S SLIDES HERE

From **Mike Nalls**' talk it seemed clear that there were some issues due to low overlap among chips (decreasing the number of SNPs to ≤ 300K to be used for analysis) and to population stratification. **Mike Nalls** said that even if we had ≤ 300K other SNPs could be imputed through 1000 Genomes with a high chance of calling SNPs correctly (at the same level or even better than if those SNPs would have actually been genotyped). Though, the issue representing major concern was the population stratification and substructure problem. **Mike Nalls** showed that in the study population there are 3 substructures northern Europeans, southern and "other Europeans. **Mike Nalls** suggested to find a way to better match controls: this would mean to collect control samples and genotype them on a similar platform. This could help regarding the stratification problem but would mean putting more time and funding into this project, fact that, as of now, is not affordable (without forgetting that we'd probably lose further ~ 300 samples). **Mike Nalls** suggested that we could match and stratify by ancestry. Finally, **Mike Nalls** said that, prior primary

analysis, we should define the phenotypes we want to included in the analysis, delete all the samples that have different diagnosis/pathology and, ideally, we would need to precisely match cases with controls in a ratio 1:3. Nonetheless, we should decide for a pooling vs. meta-analysis strategy.

12:40PM: lunch. After lunch discussion on FTD-GWAS followed.

Main points that were discussed are the following:

- **Primary Analysis – samples inclusion criteria**: primary analysis will depend on the precise definition of the phenotype of the samples. Samples with clinical diagnosis of **bvFTD** – **SD** – **PNFA** – **FTD-MND** will be included in the analysis (if there are path data that confirm the diagnosis this is even better). Of course, if path data reverse the diagnosis to some syndromes that differ from the aforementioned phenotypes those will be excluded. As a matter of fact, some normal controls seem to be present within the whole cohort, therefore, once identified, those will be excluded. As well, based on updated diagnosis or path info, cases designated as AD, PSP, CBS-CBD and logopenic aphasia will be excluded from the study and primary analysis. **Raf Ferrari** proposed to identify samples eligible for primary analysis in no longer than 2 weeks (by October 28th). Primary analysis could then start by October 31st and 3-4 weeks later results (hits), if any, should be identified. Importantly, **Mike Nalls** made clear that we will need to decide for a pooling vs. a meta-analysis type of approach.

- **Use of Van deerlin dataset**: these data are freely available. It was discussed whether these data could be relevant during primary analysis or during replication analysis or at all. The debate didn't bring to any conclusive decision as this dataset contains, in vast majority, *PGRN* mutation carriers and it is not clear what the pros of including them in analysis would be. This is still an open question. Nevertheless here two points of view: **John Hardy** keen on using them during primary phase, whilst **Andy Singleton** leaned towards using them during replication. Nobody else expressed a clear position.

- **Replication arrays**: for this purpose **Andy Singleton** proposed the use of Illumina's exome arrays convenient from, both, SNPs coverage and costs ($15/sample) points of view. **Mike Nalls** added that not only this platform is reliable and suitable for replication purposes but also that it could provide more genetic information to be evaluated and interpreted. Another option proposed by **Andy Singleton** was designing an "FTD-chip" including also SNPs for other diseases (this

would make it cheaper for users and more reasonable for the providing company to invest on such a chip). This option, though, would have some critical issues, specifically, it would decrease the number of SNPs relevant to FTD and would have anyway a bigger cost ($70/sample); therefore, it seemed not a suitable option. Finally, **Parastoo Momeni** proposed the use of genotyping assays to be run on Illumina Veracode; this option was not convincing in part in terms of SNPs coverage, but mainly because of the costs ($18,000 for 384SNPs/well for 480 samples).

- **Replication costs**: to cover for the replication phase, which was evaluated being ~$150,000, it was proposed that all participating groups (the whole FTD-Consortium) could allocate some funds into a "common account" to cover for replication expenses. This possibility needs to be further discussed. **John Hardy** suggested that some funds for this could be available through the Alzheimer's Association. **Andy Singleton** also said to keep in mind that costs could be discussed and negotiated with Illumina.

- **Chr9 screening**: **John Hardy** said that knowing which samples hold repeats in intron 1 of *C9ORF72* would, possibly, be useful prior to and during analysis as it would help subdividing samples as well as having an idea about to what degree we should/could expect eventual hits on chr9 considering it as an internal positive control. Therefore, **John Hardy** suggested that everybody should start looking into their samples and that, in case specific sites would not be able to perform such experiments, those could be carried out at the following facilities: UCL, NIH, John van Swieten's, Stuart Pickerin-Brown's or van Broeckhoven's labs.

- **Meeting ~December 16th in London**: it was proposed by **John Hardy** that the whole consortium should participate in the analysis phase after results from primary analysis and that data after primary analysis should be free to circulate among collaborators. Moreover, **John Hardy** suggested that in this direction the FTD-Consortium should meet to discuss the next steps of the project, after the results from primary analysis, around December 16th in London.

- **Very next steps**: **Raf Ferrari** will circulate an email to all collaborators to gather one more time the most updated and final info on the samples (specifically in relation to clinical, path and genetic data) to identify the samples that will go into primary analysis. This should happen in the next two weeks and be done by October 28th. In the following 3-4 weeks results for primary analysis will be generated by **Mike Nalls** at NIH. By nearly the end of November 2011we should

have results of primary analysis. Meanwhile, we should all keep in mind that sample collection for replication needs to be planned during the coming month starting from now.

At **1:30PM** the Chr 9 session started.

Data were presented in the following order: **Bryan Traynor**, **Rosa Rademakers**, **John van Swieten**, **Julie van der Zee**, **Ilse Gijselinck**, **Stuart Pickerin-Brown**. Power points of the talks follow.

INSERT ALL SLIDES

During and after the presentations of chr9 repeats some issues were raised among participants: specifically, it is still not clear how reliable and indicative the methods that have been used are, especially the primed PCR. It seems that cases show a higher frequency of repeats than controls. But also controls have shown repeats (with lower overall frequency though). Furthermore, it is not clear how the repeats in an intronic region would affect the phenotype to the extent that they would explain the disease. No functional studies could give a hint of explanation. Does frequency alone explain the difference between cases and controls? None of the presenters was able to answer/address these questions. **Parastoo Momeni**, **Raf Ferrari** and **John van Swieten** were pointing out that the repeats were also found in some controls which were showing, graphically, the same repeats pattern as in the cases. **Parastoo Momeni**, **Raf Ferrari** and **Rosa Rademakers** highlighted that primed PCR method does not give an exact estimate of the repeats. **Rosa Rademakers** also pointed out that the southern blot experiments could be biased due to cloning of the cell lines. **Andy Singleton** remarked the possibility of cloning being responsible for seeing 700-1,500 bands as well as he suggested that it seems hard to associate the repeats with the diseased phenotype. **Stuart Pickerin-Brown** suggested that the differences in the frequencies between cases and controls are self explanatory.

At **3:30PM Rita Guerreiro** introduced the relevance and possible link between Nasu Hakola disease and FTD highlighting the possibility of evaluating homozygosity regions at the *TREM2* (chr6) and *TYROBP* (chr19) loci in FTD cases. She asked for permission to look for this purpose into the FTD-GWAS samples. **John van Swieten** agreed.

At almost **4:00PM** the meeting ended with the understanding that **John Hardy** and **Raf Ferrari** would send minutes and presentations of the meeting to the whole FTD-Consortium in a week time.

Appendix 4-5. FTD-GWAS meeting December 2011: Minutes

December 11, 2011

FTD-GWAS primary association analysis – results meeting

Attendees:

John Hardy, Matthias Rimenschneider, Jon Rohrer, Rita Guerreiro, Jose Bras, Stuart Pickering Brown, Peter Hyslop, Raph Gibbs, Christine Van Broeckhoven, Andrew Singleton, Julie Van der Zee, Raf Ferrari, Dena Hernandez, Mike Nalls, Caroline Graff.

Main points:

- Data was presented from the ~2500 samples with FTD and ~7500 controls (for data cleaning see slides)

- Highlights of analysis:

    - ApoE was highest. Could be interpreted as either a real effect or diagnostic bleed thru of ~15% from AD
    - HLA was high: this could be population stratification or real.
    - Ch9 data had just sub significant score
    - TMEM data was unclear

Resolved to do:

- Site specific analyses to see if HLA disappeared
- Diagnosis subtype specific analyses to try and understand ApoE signal
- How to replicate and what samples to replicate in.

Raf has emailed around and has ascertained that the sites who have contributed have in total about 1000 new cases. Julie Van der Zee noted she had 500 additional samples from non-contributors which could contribute to these analyses with their permissions.

There was a discussion of the need for replication in these 1000-2000 new samples. Two possibilities were discussed:

1) The new exome chip + custom content, which would cost about $100 a samples, would take about 3 months to design and require the raising of about $300K (controls would be needed to be genotyped). We will send the content of the exome chip round when we have it, but essentially it has on it every coding SNP with a frequency >0.5%

2) **Christine Van Broeckhoven** suggested the use of Sequenome platform for a few top SNPs with ancestry informative markers. This would allow the assessment of ~30 top SNPs with ancestry and gender markers if more than one population was assessed.

Further resolved:

- **Andy Singleton** to try and get NIH money for exome plus chip
- **John Hardy** and **Stuart Pickering-Brown** to try and get Alzheimer's Research UK (ARUK) money for exome plus chip
- **Matthias Riemenschneider** and **John Hardy** to try and get money for exome plus chip
- **Julie Van Der Zee** and **Christine Van Broeckhoven** to see if they have budget or capability within Belgium for exome plus chip
- $300K will be needed in total.

Decision point for this will be in March while samples are gathered together.

Conclusions:

- The first analysis for publication will probably be a subtype analysis with a side by side with the Van Deerlin data as this will be very informative as to effects which are subtype and pathology specific.
- As in other such consortia, there is a no-surprises policy and if any of these are written up we will circulate well before the final draft stage
- It would be great if groups wished to do their own analyses (e.g. haplotype sharing). It would be good to hear of these
- Next meeting will be 8th September in Manchester (the day after the International FTD meeting).

Following the slides presented at the meeting by **Raf Ferrari** and **Mike Nalls**.

Appendix 4-6. FTD-GWAS meeting September 2012: Minutes

FTD-GWAS – Manchester September 6, 2012

Minutes

Meeting started at **18:00** local time. Attendees were as follows:

Attendees	Centers
Caroline Graff; Huei-hsin Chiang	Karolinska Institute, Department of NVS, Stockholm, SWEDEN
Matthias Riemenschneider	Universität des Saarlandes, Klinikum für Psychiatrie & Psychotherapie, Homburg/Saar, GERMANY
Javier Simon-Sanchez; Sasja Heetveld; Peter Heutink	Department of Clinical Genetics, VU University Medical Centre, Amsterdam, Netherlands
Innocenzo Rainero	Department of Neuroscience, University of Torino, Italy
Amalia C Bruni; Chiara Cupidi	Regional Center of Neurogenetic, Lamezia Terme, Italy
Agustin Ruiz; Isabel Hernandez	ACE Foundation. Catalan Institute of Applied Neuroscience, Barcelona, Spain
Agnes Camuzat; Isabelle Leber; Alexis Brice; Latouche	GICM, Paris, France
Marc Suarez-Calvet; Jordi Clarimon	Genetics of Neurodegenerative Diseases Unit | IIB Hospital Sant Pau, Barcelona, Spain
Giancarlo Logroscino	University of Bari, Italy
Markus Otto	Uni Ulm, Germany
Daniela Galimberti; Elio Scarpini; Giorgio Fumagalli; Andrea Arighi	Department of Neurological Sciences, Dino Ferrari Institute, University of Milan, Italy
Jorgen Nielsen; Lena Hjermino	Section of Neurogenetics, University of Copenhagen, Denmark
Peter Schofield; Carol Dobson-Stone	Neuroscience Research Australia, Sidney, Australia
Adrian Danek	Department of Neurology, Ludwig-Maximilian University of Munich, Germany
John van Swieten	Department of Neurology, Erasmus Medical Center, Rotterdam, Netherlands
Vivianna Van deerlin	University of Pennsylvania Health System, Department of Pathology, Philadelphia, PA, USA
Simon Mead; Jon Beck	Institute of Neurology, Prion Unit, UCL, London, UK
Giuliano Binetti; Roberta Ghidoni; Luisa Benussi	NeuroBioGen Lab-Memory Clinic, Fatebenefratelli, Brescia, Italy
Benedetta Nacmias	Department of Neurological and Psychiatric Sciences, University of Florence, Italy
Barbara Borroni; Alessandro Padovani	Department of Medical Sciences, Neurological Clinic, University of Brescia, Italy
Ekatrina Rogaeva	Center for research in Neurodegenerative diseases, Toronto, Canada
Elisabet Englund; Christen Nilsson; Maria Landqvist	University of Lund, Sweden
James Rowe	University of Cambridge, Cambridge, UK
Alfredo Ramirez	University Bonn, Germany
John Hardy; Jonathan Rohrer	Institute of Neurology, UCL, London, UK
Stuart Pickering-Brown (UK)	University of Manchester, Clinical Neuroscience, Manchester, UK
Bryan Traynor	Laboratory of Neurogenetics, NIA, NIH, Bethesda, MD, USA
Parastoo Momeni, Raffaele Ferrari	Texas Tech University Health Sciences Center, Lubbock, Texas, USA

Raf Ferrari introduces the main topics to be discussed in the meeting:

1. **FTD-GWAS**: results – replication timeline – future directions
2. **FTD-GWAS**: presentation & discussion of side projects
3. **FTD-ALS *C9orf72***: presentation of analysis spreadsheet and comments
4. **Conclusions**

1. FTD-GWAS: results – replication timeline – future directions

Raf showed the slides that summarize Manhattan plots along with associated SNPs with p-values with genome wide significance (p<E-8) and close to genome wide significance (p<E-7) for each subtype and for the meta-analysis. Then he discusses the final summary of primary analysis results together with all attendees. These slides are attached again.
Main points are as follows:

- We need to replicate our results (primary focus)
- Association with **chr 11**, *RAB38*, in bvFTD subtype, especially if replicated, is going to be interesting and worthy of further investigations
- Association with **chr 9**, *C9orf72* locus, in FTD-MND subtype overlaps with previous ALS-GWAS
- Strong association with **chr 6**, *HLA* locus, in meta-analysis:
    - might be true – overlaps with some previous PD-GWAS
    - might be due to population stratification – redo analysis within a homogeneous cohort from the whole cohort
- Association with **chr 19**, *APOE* locus, in bvFTD subtype and meta-analysis:
    - might represent presence of samples other than FTLD in our cohort
    - could it be true or does it represent minor (~10% contamination) with AD?
    - no association with *CLU*, *PICALM*, *CR1* loci
- Lack of association with **chr 17**, *MAPT* & *GRN* loci, might be due to exclusion of mutation carriers
- Lack of association with **chr 9**, *C9orf72* locus, in all but FTD-MND subtype is worth consideration

- Lack of statistical association with **chr 7**, *TMEM106B* locus, in all subtypes and meta-analysis. See the attached file for this. Clearly, the OR cannot be close to the reported 1.6, but a weak association is possible (see data). The replication chip will be important in this regard
- All SNPs with p-values E-05/E-06/E-07 are included in replication chip
- **Raf** has additionally hand checked all the files for *MAPT* and *PGRN* duplications and deletions and not found any

At this point, **John Hardy** briefly introduced the NeuroX chip that will be used for replication. Characteristics of the chip, that will be available in October 2012, are as follows:

Standard Human Exome v1.1 microarray:

- **200K SNPs including**:
    - SNPs from the exome sequencing project (ESP) dataset
    - ancestry informative makers
    - X/Y markers to verify gender
    - published GWAS hits
- **30K SNPs for a number of neurological disorders**:
    - known rare/coding variants
    - mendelian genes
    - follow-up for exome sequencing studies
    - candidate SNPs for current and future GWAS
    - GWAS replication SNPs (~4K hits from phase I for FTD-GWAS replication)

Following, **Raf** proceeds in explaining the plan and timeline for replication: replication phase (or phase II) has already started and several hundreds of samples and controls have been and are being collected at UCL. **Raf** emphasizes that, in fact, one of the aims of phase II is to collect both cases as well as controls to be run together using the same genotyping platform as well as the same type of chip to ease and empower analysis.
Based on the timeline, the intention is, as of September 2012, to collect samples, to perform Quality Control (QC) on them, to genotype them, to strategize and perform analysis – all together – by ideally March-April 2013. A manuscript could be submitted by May 2013. It is, in fact, in the interest of the whole international FTD-consortium to proceed towards the end of this project to identify novel loci associated with FTD and publish them.

To help in this direction, **Raf** suggests to finalize samples collection asap: a strategy to ease this step could be promoting samples collection at UCL, NIH and also at the Texas Tech University Health Sciences Center (TTUHSC), where **Raf** works for most of the time. Sending samples to TTUHSC could be an idea to facilitate samples collection, QC and preparation for chips. Once **Raf** QCs samples and prepares them at the right dilution, he can send the plates directly either to UCL or NIH ready to be put on chip and to be run. Once the genotyping phase is concluded, the samples collected at TTUHSC will be sent back to the original source.
Along, **Raf** explains the inclusion/exclusion criteria for phase II:

- path confirmed FTD cases (if available)
- bvFTD – SD – PNFA – FTD/MND diagnosed cases (Neary criteria – Neary et al., 1998)
- firmly exclude any possible LPA
- exclude mutation carriers in the candidate genes, but, if including, identify mutations
- include results of screening of *C9orf72* hexanucleotide repeats (if available)
- amount of DNA (samples): 2 – 2.5 ug
- include neurologically normal controls (to be genotyped as well): 2 – 2.5 ug (if available)

It was gratifying that several other groups at the FTD meet were made known to **Raf** at the meeting and he will contact them for the GWAS analysis.
Following, **Raf** mentioned that some side studies could already be started and performed during replication phase. Such studies are suggested and open for discussion as follows:

- Redo analysis with and without *C9orf72* positive cases
- Tackle HLA association by looking just into a homogeneous cohort
- Redo analysis with and without *ApoE* allele 4
- Meta-analysis with FTD-TDP dataset (van Deerlin *et al*, 2010) using SD, PNFA and FTD-MND (TDP-43 path)
- Meta-analysis with ALS-GWAS: our cohort with and without *C9orf72* positive cases vs. Finnish cohort with and without *C9orf72* positive cases (van Es *et al*, 2009; Laaksovirta *et al*, 2010)
- Age of onset sub-analysis

Then **Raf** mentions what could/should be done after conclusion of replication:

- Targeted re-sequencing

- Pathways analysis
- Functional studies

2. FTD-GWAS: presentation & discussion of side projects

At this point **Raf** opens the discussion for comments on results and on suggestions for side studies to be performed using the data generated during phase I.

Dr. Riemenschneider mentions that it should be told that genomiphied samples are not suitable for genotyping unless, probably, samples are enriched and stabilized through certain kits. **John** and **Raf** replied that they will look into this possibility. Further, **Dr. Riemenschneider** mentions that his group would like to and could perform CNV analysis on the whole cohort.

Dr. Mead mentions that it would be worth looking into our data and extract the exact p-values for the SNPs associated with *TMEM106B* providing as a reason that there are a lot of functional studies being performed on that gene and its interaction with *PGRN*. These data are now attached.

Also, **Dr. Mead** mentioned that an interesting side study could be to look for relatedness within homogeneous cohorts to possibly identify founder effects. **Dr. Hardy** said this type of study was encouraged and we will follow up to offer help to anyone in this regard, but we would be delighted if people made individual contacts to do this as long as they kept the larger group informed.

Drs. Mead and **Pickering-Brown** suggested that they could perform this type of analysis on the UK cohort.

Dr. van Deerlin mentions that she would like to perform a new analysis on her ~500 samples TDP-43 path confirmed cohort using a different set of normal controls and that she is looking for the availability of genotyping data of US neurologically normal controls. **Dr. Singleton** has such a data set.

Dr. Ramirez mentions that the University of Bonn could participate in the study providing DNA of German neurologically normal controls.

Dr. van Swieten asks if there is the possibility to further assess and refine the clinical data; on that point **Dr. Rohrer** answers that him and **Raf** have been spending hours to decide which samples to include or exclude into/from primary analysis. Moreover, **Dr. Rohrer** mentions that **Raf** has spent hours on the phone with each single participating colleague and has been in almost constant email contact with them to progressively discuss any update on the clinical information for each single sample prior primary analysis. **John Hardy** confirms and says that we really did all our best to thoroughly characterize all samples.

Finally, **Dr. Dobson-Stone** asks about the possibility to contribute with samples from Australia and if also Australian normal controls would be needed.
After the discussion **Raf** introduces **Dr. Bryan Traynor**'s presentation of the FTD-ALS C9orf72 project.

3. FTD-ALS *C9orf72*: presentation of analysis spreadsheet and comments

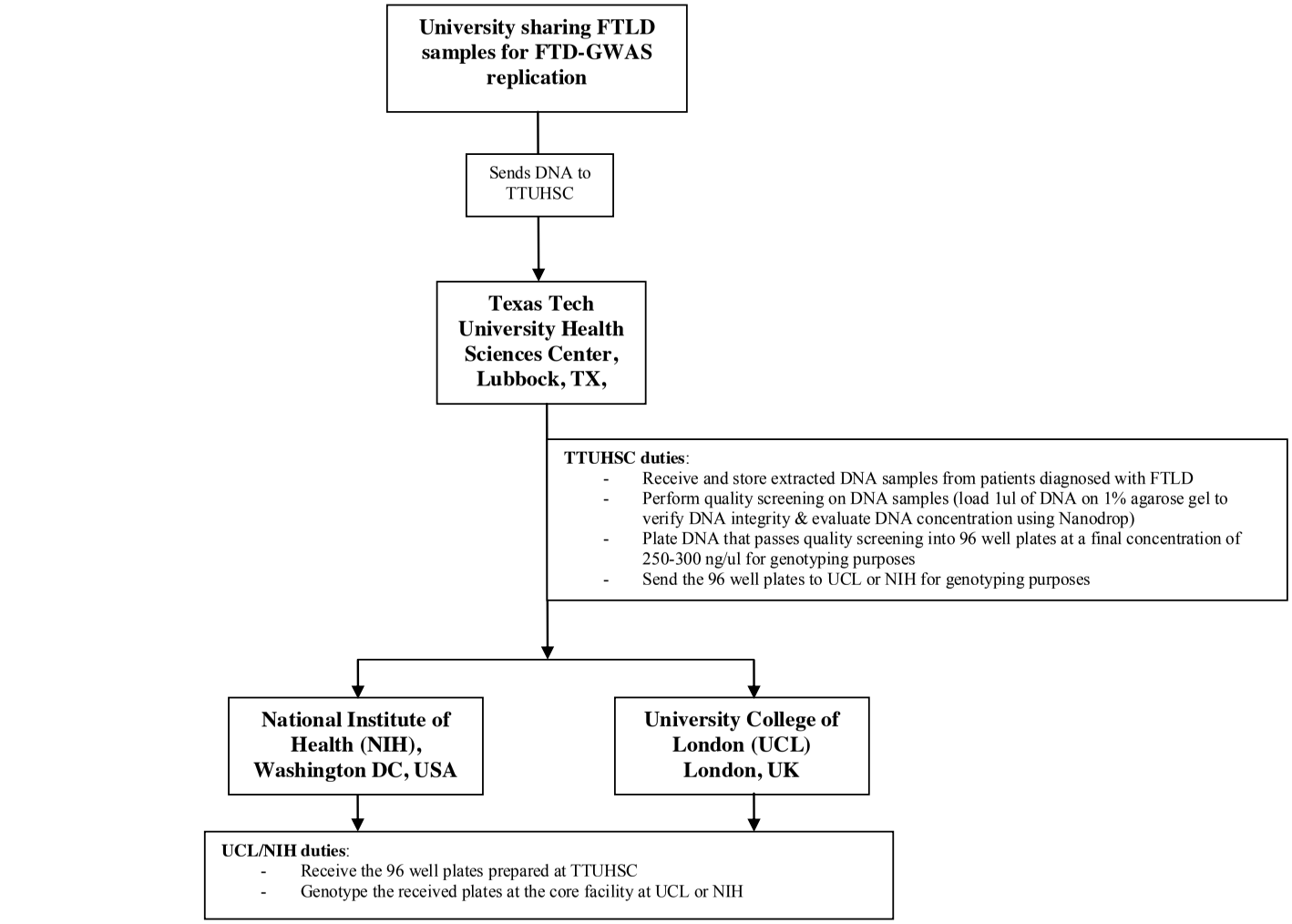
Dr. Traynor discusses the excellent progress that has been made. That FTD/ALS group is invited to circulate a progress report.

4. Conclusion

At this point **John** and **Raf** conclude the meeting.
In summary, main points of the meeting are as follows:

- Replication (phase II) is ongoing and samples collection is open
- Contact Raffaele Ferrari (raffaele.ferrari@ttuhsc.edu) to organize samples shipment/collection
- NeuroX replication chip will be available in October 2012
- Minutes of the meeting will be sent out in ~ one week time
- Several possible side projects will soon be listed and discussed to be approved by the whole consortium prior start. The approval will depend upon agreement of all sites that genotyping data can circulate among the consortium members (especially to those who will be directly responsible and in charge of a side-project)
- Side projects will be performed by groups within the consortium under the understanding that generated results from side studies might be incorporated in the main manuscript or be published in satellite papers after the first comprehensive publication of the complete FTD-GWAS study
- Next meeting for the members of the international FTD-consortium to discuss updates on the FTD-GWAS project will be at AD/PD (Florence) March 6th-10th or the AAN (San Diego) March 16th-23rd

Appendix 4-7. FTD-GWAS: Replication phase workflow



NOTES:

- **Genotyping data generated at UCL/NIH are stored at UCL, NIH and TTUHSC and are sent back to each of the above listed Institutions individually through secure ftp websites**
- **Sender of DNA covers material shipment expenses to and from TTUHSC**
- **UCL/NIH cover for shipment of the 96 well plates from TTUHSC to UCL/NIH**