The fluid biomarker and genetic determinants of phenotype and survival in

progressive supranuclear palsy

Dr Edwin Jabbari

UCL Queen Square Institute of Neurology

For the award of Doctor of Philosophy

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

#### <u>Acknowledgments</u>

I would like to begin by acknowledging the hundreds of PSP patients and their families that I've had the immense privilege of looking after over the past 4 years. I feel honoured to have been let into your lives and to go through this brutal illness with you. Your willingness to sign up to research projects/trials to 'help others' when you were at your most vulnerable was so inspiring. I truly believe that, like all neurodegenerative diseases, effective treatments for PSP patients are on the horizon and I hope that the findings from this PhD will contribute to that process. I'd like to thank the Medical Research Council for awarding me a clinical research training fellowship which allowed me to do this work and to travel to the National Institutes of Health, Washington DC, in the summer of 2019. This was an invaluable experience which enabled me to learn new skills and build collaborations. I'd also like to thank the PSP Association for initially funding me and for always being an incredible source of support for our patients.

None of this would have been possible without the support and guidance of Prof Huw Morris. More than just "the boss" – Huw has been a supervisor, mentor, teacher and friend, and gave me the platform and confidence to grow as a researcher. I really am grateful for the positive impact that he has had on my career and I am sure that this is not the end of the road for us working together.

Alongside Huw, I have also been very lucky to work with and learn from other powerhouses in the field at UCL – Prof John Hardy, Prof Andrew Lees, Prof Henrik Zetterberg and Dr Mina Ryten. There is also a plethora of super-intelligent junior scientist colleagues who at some point have taken the time out of their busy schedules to teach me something or run analyses, especially Dr Maryam Shoai and soon-to-be Drs Manuela Tan and Regina Reynolds – I've no doubt that you will all be future leaders in the field! Also, a special shout out to Dr John Woodside and Dr Ruth Lamb whose efforts in the initiation of the PROSPECT study allowed me to hit the ground running when I joined the team in August 2016.

ii

As with any major accomplishment, there is also a wider team of champions to acknowledge and thank for everything they have done for me, not only during this PhD, but for many years before that too.

To my mum, Alice – I will always be grateful to you for giving me the foundation of lifelong love and support. There is no doubt that I have inherited my work ethic from you and I hope that I continue to make you proud for many more years to come.

To my partner-in-life-and-crime, Kate – You have the patience of a saint! Thanks for putting up with the late nights glued to my laptop and supporting me throughout this journey. You're the best teammate I could ever ask for and I've no doubt that we'll continue to achieve our goals together.

Thanks also to other friends and family who have been a source of sanity whenever I needed a break from the demands of the PhD. A special mention to James Murphy who, aside from being a lifelong friend and partner in exercise/music/writing, has allowed me to hide my lack of artistic flair by creating and/or editing figures that appear in chapters 2, 3 and 4.

If anything, this PhD has taught me that hard graft, teamwork and collaboration is the recipe for success!

iii

#### <u>Abstract</u>

Progressive supranuclear palsy (PSP) is a neurodegenerative disorder and common cause of atypical parkinsonism (APS), with an estimated prevalence of 5-7 per 100,000. The pathology of PSP is centred on the microtubule associated protein tau, encoded by *MAPT*. Richardson's syndrome (PSP-RS) is the classical clinical phenotype related to PSP pathology. Recently, variant clinical phenotypes related to PSP pathology have been operationalised in the 2017 Movement Disorder Society diagnostic criteria.

Early and accurate clinical diagnosis of PSP is challenging as there is heterogeneity within PSP and significant clinical overlap with other neurodegenerative conditions including Parkinson's disease (PD) and frontotemporal dementia. Fluid biomarker studies have identified that levels of CSF and plasma neurofilament light chain (NF-L) are able to differentiate PSP-RS from PD and controls but are unable to differentiate PSP from other causes of APS. Case-control genome-wide association studies using pathologically confirmed PSP cases have identified risk loci at *MAPT*, *MOBP*, *STX6* and *EIF2AK3*. In addition, studying the progression of neurodegenerative disorders is central to discovering determinants of disease progression and assessing the effects of therapeutic interventions.

During this PhD I have:

- Used baseline data from the PROSPECT study to highlight distinct clinical trajectory, cognitive, neuroimaging and fluid biomarker profiles of different PSP subtypes.
- 2) Used baseline CSF and plasma NF-L to predict survival in PSP, and used proximity extension assay to discover novel diagnostic biomarkers of APS.
- 3) Defined early onset PSP and characterised its genetic and clinico-pathological profile.
- Discovered the *TRIM11/17* locus as a genetic determinant of PSP phenotype, and the *LRRK2* locus as a genetic determinant of survival in PSP.

iv

#### Impact statement

The impact and potential benefits of the research presented in this thesis are far-reaching, both inside and outside academia.

Within academia, the genetic findings from this PhD have already generated subsequent funding for myself (MRC Clinical Research Training Fellowship) and research projects in other research groups at UCL and the Dementia Research Institute, such as the functional characterisation of TRIM11/17 (PI – Prof Karen Duff). Both the PSP phenotype and disease progression/survival GWAS have fostered ongoing active collaborations between UCL and other institutions across the world, including the National Institutes of Health, the Mayo Clinic, University of California San Francisco Memory and Aging Centre and the German Centre for Neurodegenerative Diseases. Additionally, the methodology used to discover genetic determinants of disease progression/survival in PSP are currently being used by UK-based and international research groups to study other neurodegenerative disease groups, including Parkinson's disease, Alzheimer's disease and Frontotemporal dementia. The data and samples obtained from patients in the PROSPECT study will undoubtedly stimulate future research in the field of atypical parkinsonian syndromes. All of the research presented in this thesis has been published in high impact peer-reviewed journals, with the exception of the PSP survival GWAS which is currently under review at *Lancet Neurology*, and has also been presented at national and international conferences. Mostly recently, this led to me winning the Lightning Prize at the international Tau2020 conference in Washington DC.

Commercially, our discovery of the *LRRK2* locus as a genetic determinant of survival in PSP (and associated UCL press release) has generated great interest from pharmaceutical companies regarding potentially trialling LRRK2-inhibitors and/or anti-sense oligonucleotides as disease modifying therapies in tauopathy patients.

v

Outside academia, the main impact of this research has been in engaging with patient support groups. I have presented the findings of our research to local PSP Association (PSPA) support groups and have also written blog pieces for the PSPA, including one as part of a funding appeal to the general public. The PSPA have also done their own press releases for each of the publications associated with this thesis. All of these activities have resulted in generating further funding for the PSPA and the recruitment of PSP patients to both the PROSPECT study and current tau-antibody trials.

#### **Publications**

Jabbari E, Tan MMX, Reynolds RH, et al. Common variation at the *LRRK2* locus is associated with survival in the primary tauopathy progressive supranuclear palsy. *bioRxiv* 2020; doi: <u>https://doi.org/10.1101/2020.02.04.932335</u> - under review at *Lancet Neurology*.

Jabbari E, Holland N, Chelban V, et al. Diagnosis across the spectrum of progressive supranuclear palsy and corticobasal syndrome. *JAMA Neurology* 2020; 77:377-387.

Jabbari E, Woodside J, Tan MMX, et al. The genetic and clinico-pathological profile of early onset progressive supranuclear palsy. *Mov Disorders* 2019; 34:1307-1314.

Jabbari E, Woodside J, Guo T, et al. Proximity extension assay testing reveals novel diagnostic biomarkers of atypical parkinsonian syndromes. *JNNP* 2019; 90:1-7.

Jabbari E, Woodside J, Tan MMX, et al. Variation at the *TRIM11* locus modifies Progressive Supranuclear Palsy phenotype. *Ann Neurol.* 2018; 84:485-496.

Jabbari E, Zetterberg H, Morris HR. Predicting and tracking disease progression in Progressive Supranuclear Palsy: CSF and blood biomarkers. *JNNP* 2017; 88:883-888.

# Table of contents

Acknowledgments	ii
Abstract	iv
Impact statement	v
Publications	vii
List of tables	x
List of figures	xii
List of abbreviations	xiv
Chapter 1 – Introduction and background	1
1.1 – Epidemiology, pathology, clinical/radiological features and natural	
history of progressive supranuclear palsy	1
1.2 – Clinical trials in progressive supranuclear palsy	4
1.3 – Genetic studies in progressive supranuclear palsy	4
1.4 – Fluid biomarker studies in progressive supranuclear palsy	7
1.5 – The challenges that have been addressed by this PhD	9
Chapter 2 – Phenotypic variation of progressive supranuclear palsy in the	
PROSPECT study cohort	10
2.1 – Abstract	10
2.2 – Introduction	12
2.3 – Methods	13
2.4 – Results	20
2.5 – Discussion	33
Chapter 3 – Blood + CSF neurofilament light chain and novel proximity	
extension assay markers as predictors of PSP pathology and survival	37
3.1 – Abstract	37
3.2 – Introduction	38
3.3 – Methods	40

3.4 – Results	44
3.5 – Discussion	54
Chapter 4 – The genetic and clinico-pathological profile of early-onset	
progressive supranuclear palsy	57
4.1 – Abstract	57
4.2 – Introduction	58
4.3 – Methods	59
4.4 – Results	63
4.5 – Discussion	71
Chapter 5 – Progressive supranuclear palsy phenotype genome-wide	
association study	75
5.1 – Abstract	75
5.2 – Introduction	76
5.3 – Methods	77
5.4 – Results	85
5.5 – Discussion	95
Chapter 6 – Progressive supranuclear palsy survival genome-wide association	
study	100
6.1 – Abstract	100
6.2 – Introduction	101
6.3 – Methods	102
6.4 – Results	108
6.5 – Discussion	123
Chapter 7 – Conclusions and future directions	126
References	131

# List of tables

Table 2.1 – Overview of assessments in the PROSPECT study	16
Table 2.2 – Clinical rating scales, functional rating scales and cognitive screening	
tests used in the PROSPECT study	16
Table 2.3 – Baseline clinical features of the PROSPECT study cohort	23
Table 2.4 – Cognitive screening measures in the PROSPECT study cohort	25
Table 2.5 – Genetic analyses in the PROSPECT study cohort	28
Table 2.6 – Volumetric MRI measures in the PROSPECT study cohort	30
Table 2.7 – Cognitive, fluid biomarker and imaging AUC values from ROC curve	
analyses	32
Table 3.1 – Summary of baseline clinical characteristics, CSF NF-L and survival data.	44
Table 3.2 – Hazard ratios for survival in PSP and APS over 5 year follow up,	
stratified by CSF and plasma NFL levels	46
Table 3.3 – Disease-specific summary statistics for significant PEA CSF biomarkers	52
Table 3.4 – Biological function (UniProt database) and tissue expression (GTEx	
database) of significant Olink panel markers	53
Table 4.1 – Neuropathological and genetic mutation status of clinically diagnosed	
PSP and PD cases	64
Table 4.2 – Clinical profile of EOPSP, LOPSP and PD study cohorts	67
Table 4.3 – Genetic profile of EOPSP, LOPSP and PD study cohorts	69
Table 4.4 – Impact of alternative age at onset cut-off points to define EOPSP	71
Table 5.1 – Clinical features of subjects included in genotype-phenotype analyses	86
Table 5.2 – Comparison of PSP risk variant MAFs between PSP case-control GWAS	
and PSP phenotype GWAS	87
Table 5.3 – PSP-RS vs non-PSP-RS association statistics for rs564309, and the	
most significant flanking SNPs located at neighbouring genes	90

Table 5.4 – PSP-RS vs non-PSP-RS association statistics for rs1188473, in	
pathological, clinical and 2011 case-control GWAS cohorts respectively	91
Table 6.1 – PSP sample sources for survival GWAS	109
Table 6.2 – Clinical characteristics and rs2242367 association statistics from PSP	
survival GWAS, stratified by cohort and PSP phenotype	110
Table 6.3 – Association statistics of genome-wide significant SNPs in the PSP	
survival GWAS	112
Table 6.4 – rs2242367 association statistics in PSP case-control and phenotype	
GWAS'	115
Table 6.5 – Impact of rs2242367 genotype on disease duration in deceased PSP	
cases	117
Table 6.6 – PSP survival GWAS association statistics for LRRK2 locus lead SNPs	
from Parkinson's disease, Crohn's disease and leprosy case-control GWAS'	119

# List of figures

<b>Figure 1.1 –</b> A schematic diagram showing <i>MAPT</i> exons and locations of mutations	5
Figure 1.2 – Tau antibody binding sites and capture antibodies used for standard	
(INNOTEST) ELISA	7
Figure 2.1 – Recruitment of patients to the PROSPECT study	21
Figure 2.2 – Clinical disease trajectory profiles of groups in the PROSPECT study	24
Figure 2.3 – Fluid biomarker profiles of groups in the PROSPECT study	27
Figure 2.4 – ROC curve analysis of serum NF-L differentiating between all PSP	
cases vs. controls	31
Figure 3.1 – APS group Kaplan Meier survival curve according to high and low CSF	
NF-L groups	47
Figure 3.2 – Volcano plot of CSF proximity extension assay biomarkers, highlighting	
no significant markers in the Tau vs Synuclein group analysis	48
Figure 3.3 – Volcano plot of CSF proximity extension assay biomarkers, highlighting	
significant markers that differentiated the APS group from controls	49
Figure 3.4 – Volcano plot of CSF proximity extension assay biomarkers, highlighting	
significant markers that differentiated the APS group from the PD group	50
Figure 3.5 – ROC curve analyses of: (A) APS vs controls; (B) APS vs PD	54
Figure 4.1 – Initial(A) and final/current(B) clinical profile of EOPSP, LOPSP and PD	66
Figure 4.2 – ROC curve analysis of the PSP GRS differentiating between PD vs. all	
PSP cases (EOPSP and LOPSP combined)	69
Figure 5.1 – Genetic PCA analysis using HapMap reference dataset	80
Figure 5.2 – Manhattan plot of whole cohort PSP-RS versus non-PSP-RS GWAS	88
Figure 5.3 – Regional association plot of PSP-RS versus non-PSP-RS association	
analysis	89
Figure 5.4 – MAGMA analyses revealing significance at TRIM11, TRIM17, and	
HIST3H3 genes	92

Figure 5.5 – TRIM11, TRIM17, and HIST3H3 brain expression in 3 databases	93
Figure 5.6 – TRIM11 expression in human frontal cortex shown by semi-quantitative	
RT-PCR	95
Figure 6.1 – Kaplan-Meier survival curves for covariates used in the PSP survival	
GWAS	111
Figure 6.2 – Manhattan and regional association plots, with supporting eQTL data,	
highlighting LRRK2 association with PSP survival	113
Figure 6.3 – Kaplan-Meier survival curve for rs2242367 genotypes	114
Figure 6.4 – Meta-analysis of rs2242367 association statistics in UCL, 2011 GWAS	
and Mayo cohorts	116
Figure 6.5 – LD structure around rs2242367	120
Figure 6.6 – Colocalisation of blood-derived eQTLs regulating LINC02555	
expression and overlapping PSP survival loci	122

### List of abbreviations

- ACE-III Addenbrooke's Cognitive Examination 3
- AD Alzheimer's disease
- AGD Argyrophilic grain disease
- ALS Amyotrophic lateral sclerosis
- APS Atypical parkinsonian syndromes
- AUC Area under the curve
- CBD Corticobasal degeneration
- CBS Corticobasal syndrome
- CBS-4RT CBS with underlying 4-repeat tau pathology
- CBS-AD CBS with underlying AD pathology
- CBS-NK CBS with unknown underlying pathology
- CD Crohn's disease
- CSF Cerebrospinal fluid
- eQTL Expression quantitative trait loci
- ECAS Edinburgh Cognitive and Behavioural ALS Screen
- ELISA Enzyme-linked immunosorbent assay
- EOPSP Early-onset PSP
- ET Essential tremor
- FDR False discovery rate
- FTD Frontotemporal dementia

- FTLD Frontotemporal lobar degeneration
- GRS Genetic risk score
- GWAS Genome-wide association study
- HC Healthy controls
- HD Huntington's disease
- IDT Indeterminate
- lincRNA Long intergenic non-coding RNA
- LBD Lewy body dementia
- LD Linkage disequilibrium
- LOPSP Late-onset PSP
- LP Lumbar puncture
- MAF Minor allele frequency
- MDS Movement Disorder Society
- MRI Magnetic resonance imaging
- MSA Multiple system atrophy
- MSA-C MSA-Cerebellar
- MSA-P MSA-Parkinsonism
- MoCA Montreal cognitive assessment
- NF-L Neurofilament light chain
- NPH Normal pressure hydrocephalus
- NPV Negative predictive value

- NPX Normalised protein expression
- P-tau Phosphorylated tau
- PCA Principal components analysis
- PCR Polymerase chain reaction
- PD Parkinson's disease
- PEA Proximity extension assay
- PPV Positive predictive value
- PSP Progressive supranuclear palsy
- PSP-F PSP-frontal
- PSP-P PSP-Parkinsonism
- PSP-PGF PSP-Progressive gait freezing
- PSP-RS PSP-Richardson syndrome
- PSP/CBS Progressive supranuclear palsy-corticobasal syndrome overlap
- PSPRS PSP rating scale
- QSBB Queen Square brain bank
- **RESTBI Retrospective Screening of Traumatic Brain Injury**
- ROC Receiver operating characteristic
- RT-PCR Reverse-transcription PCR
- SD Standard deviation
- SE Standard error
- SEADL Schwab and England Activities of Daily Living

SNP - Single nucleotide polymorphism

- SWEDD Scans without evidence of dopaminergic deficit
- T-tau Total tau
- TDP-43 TAR DNA-binding protein 43
- TIV Total intracranial volume
- UCL University College London
- UMSARS Unified Multiple System Atrophy Rating Scale
- UPDRS Unified Parkinson's disease rating scale
- UPS Ubiquitin proteasome system
- VascP Vascular parkinsonism
- WGS Whole genome sequencing

#### Chapter 1 – Introduction and background

# 1.1 Epidemiology, pathology, clinical/radiological features and natural history of progressive supranuclear palsy

Progressive supranuclear palsy (PSP) is a progressive neurodegenerative condition and a common cause of atypical parkinsonism, with an estimated prevalence of 5–7 per 100,000 **(1) (2)**. The pathology of PSP is centred on the structural microtubule associated protein tau, encoded by *MAPT*. In PSP there is neuronal and glial accumulation of hyperphosphorylated tau aggregates in subcortical and cortical brain regions. Resulting neuronal dysfunction and disruption of normal microtubule function eventually leads to neuronal cell death (neurodegeneration) **(3)**. Disease progression involves the topographic spread of tau pathology in a stereotypical fashion (which is the basis for Braak and Braak staging of Alzheimer's disease (AD)) or the worsening of pathology in specific brain regions. There are several possible mechanisms that drive progression including the formation of toxic oligomers and prion-like cell-to-cell spread of pathogenic tau **(4)**.

Tau pathology in PSP also involves an alteration in tau isoform homoeostasis. Tauopathies can be classified according to the predominant isoform of tau that accumulates through alternative splicing *MAPT* exon 10, leading to tau protein with 3 (3R) or 4 (4R) repeats of ~32 amino acids in the carboxy-terminus microtubule binding domain. 4R-tau is the dominant tau isoform in both PSP and corticobasal degeneration (CBD), in contrast to mixed 3R/4R-tau in AD. Furthermore, differences in the clinico-pathological phenotypes of tauopathies (including AD) may relate to differences in the strain properties of toxic tau species **(4)**.

Epidemiological studies estimating the prevalence of PSP **(1) (2)** have been largely based on studying cases with the classical clinical phenotype, Richardson syndrome (PSP-RS). PSP-RS is characterised by a levodopa–unresponsive akinetic–rigid syndrome with falls, a

vertical supranuclear gaze palsy, and dementia (5), and the diagnosis of this form of PSP was operationalized in the NINDS-Society for Progressive Supranuclear Palsy criteria of 1996 (6). Previous studies on the natural history of PSP-RS have shown that the mean age of disease onset is 65 to 67 years, and the median disease duration is 6 to 7 years (7). In the last two decades, variant clinical phenotypes relating to underlying PSP pathology have been identified in relatively small case series. These include PSP-parkinsonism (PSP-P), characterised by an onset of asymmetrical parkinsonism which is partially responsive to levodopa (8), and PSP with progressive gait freezing (PSP-PGF), characterised by gradual onset freezing of gait or speech, absent limb rigidity and tremor, no sustained response to levodopa, and no dementia or ophthalmoplegia in the first 5 years of disease (9). PSP-P and PSP-PGF both have a similar age of disease onset to PSP-RS and clinically resemble PSP-RS in the latter stages of disease, but have a significantly longer mean disease duration (PSP-P = 9 years, PSP-PGF = 13 years) (8) (9). The basis for this clinical variation related to a core pathology is unknown. PSP clinical subtypes have been related to the regional distribution and density of pathogenic tau accumulation and neuronal loss (10). PSP-P and PSP-PGF have been formally operationalised in the 2017 diagnostic criteria from the Movement Disorder Society (MDS) PSP study group (11), along with other variant presentations relating to underlying PSP pathology such as PSP-corticobasal syndrome (PSP/CBS) and PSP-frontal (PSP-F) phenotypes.

The gold standard of diagnosing PSP is observing the characteristic pathological changes described above at post-mortem. However, diagnosing PSP (and its various subtypes) in life is challenging, owing to a lack of objective diagnostic biomarkers and the high degree of clinical overlap with other, more common, neurodegenerative diseases such as Parkinson's disease (PD), and other causes of atypical parkinsonism such as CBD and multiple system atrophy (MSA), especially in the early stages of disease. Indeed, the positive predictive value of a clinical diagnosis of PD has been shown to be considerably higher than that of

PSP (98.6% vs 80%) in a prospective cohort of patients seen in specialist movement disorder service (12).

Although outside the scope of this thesis, there are certain radiological features of PSP that aid in differential diagnosis. Using volumetric magnetic resonance imaging (MRI), midbrain atrophy and a low midbrain:pons ratio has been shown to differentiate PSP-RS from controls, MSA-P and PD **(13)**. Other imaging techniques, including tau positron emission tomography to quantify the burden of tau pathology for diagnostic and prognostic purposes, are currently being evaluated in PSP and other tauopathies **(13)**.

Measuring progression in neurodegenerative disorders is central to defining the effects of therapeutic interventions and disease biology. In clinical and therapeutic studies, clinical disease progression is usually measured using a combination of motor scales such as the PSP rating scale (PSPRS), cognitive scales such as the MoCA and functional scales such as the Schwab and England Activities of Daily Living (SEADL) scale. In assessing the PSPRS score, which includes behaviour, bulbar, ocular motor, limb motor and gait/midline domains, Golbe and colleagues followed up 162 patients with PSP over 11 years. They showed that the mean rate of progression was 11.3 points per year and the median actuarially corrected survival was 7.3 years from symptom onset to death (14). In addition, the PSPRS score was a good independent predictor of survival (14). However, such scales may be affected by intra-rater and inter-rater variability. In addition, their use in clinical trials may be hindered by differences in the time interval between pathological disease progression/response to therapeutics and change in clinical state. Therefore, the need for reliable disease progression biomarkers to complement clinical rating scales is clear. Predicting disease progression provides early prognostic information and may enable better powered clinical trials with more homogenous groups, and give insights into the determinants of disease progression. From a therapeutic trials perspective, measuring nonclinical biomarkers of disease progression may enable early demonstration of target

engagement. However, no studies have shown that individual biomarkers are better than clinical scales at tracking disease progression and so, for now, clinical scales remain the only reliable option available.

#### 1.2 Clinical trials in progressive supranuclear palsy

To date, there have been no effective disease-modifying therapies for PSP identified from clinical trials, which have all used 1-year change in the PSPRS as the primary outcome measure **(15)**. Initially, small-scale trials (n<100) of Lithium and coenzyme Q10 were found to be negative. This was followed by the first large, multi-national, multi-centre randomised controlled trial to include PSP patients, with the NNIPPS trial of Riluzole. In the last decade, the focus has shifted to target tau dysfunction. Microtubule dysfunction and tau hyperphosphorylation were targeted in the Davunetide and Tideglusib trials, although neither showed an effect on disease modification in PSP **(16) (17)**. Current clinical trials in AD and PSP aim to prevent cell-to-cell spread of tau by using neutralising antibodies against tau in the extracellular space **(18)**, with full results expected by the end of this year. Alternative approaches that are expected to be trialled in the near future include *MAPT* anti-sense oligonucleotide therapy for *MAPT* mutation carriers, and targeting post-translational modifications of tau, including ubiquitination and OGlcNacylation **(15)**.

#### 1.3 Genetic studies in progressive supranuclear palsy

Traditionally, PSP has been considered a sporadic disorder, with less than 1% of cases being due to an underlying monogenic cause. However, prior to the GWAS-era, the genetic architecture of PSP was explored by studying familial PSP. These studies identified rare *MAPT* and *LRRK2* mutations (19), and also revealed a higher rate of a family history of dementia or parkinsonism in PSP cases vs. controls (see Chapter 4). Of note, *MAPT* mutations have been shown to cause both PSP and Frontotemporal dementia (FTD)-like (FTDP-17) presentations. A majority of *MAPT* mutations are in exon 10 and its splicing region (Figure 1.1) so may enhance production of the more fibrillogenic 4R isoform (20).

Whilst common and rare variants at *LRRK2* have typically been associated with PD, rare variants (G2019S, R1441C and A1413T) have been identified in patients with a PSP phenotype and PSP-type tau pathology at post-mortem **(21)**. Although a majority of *MAPT* mutations lead to early-onset presentations of PSP, it should be noted that there are monogenic early-onset PSP mimics including Perry syndrome (*DCTN1* mutations) and Niemann-Pick type C (*NPC1/NPC2*) mutations **(22)**.

Additionally, there have no twin studies in PSP, likely owing to the fact that this is an exceptionally rare occurrence. Of note, only one PSP locus has been identified from a linkage analysis of a large autosomal dominant Spanish PSP family. This solitary study identified a 3.4cM candidate disease locus between markers D1S238 and D1S2823 on chromosome 1q31.1 (23). However, a causative gene has not been identified and there has been no replication of this finding.



**<u>Pigure 1.1** (nom met al 2013 (19))</u> – A schematic diagram showing exons of *MAP1* and locations of mutations. Mutations discovered in patients with PSP or PSP-like phenotypes were marked in the upper half of the diagram and those with frontotemporal lobar degeneration presentation were marked in the lower half.

The first studies on PSP risk variants focused on the MAPT locus and identified a common haplotype (H1 haplotype – determined by an ancestral inversion of 900kb on chromosome 17q21) that was significantly over-represented in PSP patients vs. controls (24). There have been two large-scale case-control genome-wide association studies (GWAS) to identify common risk variants of PSP. Stage 1 of the original GWAS of 2011 was based on neuropathologically defined 1,114 cases of European ancestry and 3,287 controls. This identified genome-wide significant risk variants at MAPT (H1 haplotype and H1c subhaplotype), STX6, EIF2AK3 and MOBP. These implicate biological pathways involved in microtubule function, vesicle trafficking, the endoplasmic reticulum unfolded protein response, and myelin structure (25). A subsequent GWAS showed that risk variants at MAPT (H1 haplotype and H1c sub-haplotype) and MOBP) were shared between PSP and CBD (26). The most recent (2018) PSP case-control GWAS, which was comprised of cases from the 2011 GWAS and an additional 600 neuropathologically defined cases), replicated signals at MAPT, STX6 and MOBP, and identified novel genome-wide significant signals at RUNX2 and SLCO1A2, thereby implicating biological pathways involved in osteoblastic differentiation and transporters present (among other places) at the bloodbrain barrier where it regulates solute trafficking (27). Like many other neurodegenerative diseases, follow-up functional studies of risk variants identified from GWAS are lacking in PSP. In particular, there have been inconsistent results for the functional impact of EIF2AK3 which encodes PERK, a part of the endoplasmic reticulum's unfolded protein response, with evidence of both PERK activation (28) and inhibition (29) leading to a reduction in tau-mediated neurodegeneration in P301L and P301S mouse models of frontotemporal lobar degeneration. Outside this PhD, there have been no studies on the genetic determinants of PSP phenotype or the rate of disease progression/survival.

#### 1.4 Fluid biomarker studies in progressive supranuclear palsy

Many cross-sectional CSF studies have assessed whether protein biomarkers can reliably differentiate PSP from controls and patients with relevant differential diagnoses such AD, PD and other forms of APS.

Based on the reproducible evidence that CSF tau levels are raised in AD versus controls, thought to reflect both the active secretion of tau by cells and the release of tau by dying cells into the extracellular space (3), one would expect that increased levels of CSF tau could be detected in PSP. Importantly, most of tau in CSF appears to be in fragments containing N-terminal and/or mid-domain epitopes; virtually no full-length tau containing C-terminal sequences can be detected (30). The majority of CSF total tau (t-tau) and phosphorylated tau (p-tau) data reported in the literature are based on commercially available ELISAs, such as the INNOTEST assay (Fujirebio). In these assays, t-tau and p-tau measurements are dependent on anti-tau capture antibodies, such as AT120 and AT270 (pT181), specific for the mid-domain region of the protein, encoded by exons 4–8 (31) (Figure 1.2).



antibodies used for standard (INNOTEST) ELISA.

However, in established PSP disease there has been no observed consistent elevation of CSF t-tau or p-tau compared with healthy controls **(32) (33)**. In contrast, raised CSF tau

levels form a robust part of the diagnostic criteria for AD (34). Recently, raised levels of the p-tau181 fragment in plasma have been shown to both correlate with CSF p-tau 181 and differentiate AD from controls and non-AD neurodegenerative diseases including PSP. However, it is uncertain as to why there is a difference between the CSF/plasma tau profile in PSP and AD considering that both conditions involve extensive neurofibrillary tangle pathology related to hyperphosphorylated tau protein. A potential explanation for the divergent tau profiles described above is that standard tau ELISA may not detect specific tau strains that are elevated in primary tauopathies such as PSP. In addition, the tau fragment profile may differ between AD and PSP, which will determine epitopes and protein levels identified by ELISA (35). Of note, future studies may be able to achieve ultrasensitive measurement of fluid biomarkers by using novel technologies such as single molecule array (36), single molecular counting (37) and proximity extension assay (38).

Non-tau related fluid biomarker candidates have also been studied extensively. The most promising of these has been neurofilament light chain (NF-L). Similar to tau, neurofilaments are found in axons of neurons and are important components of the cytoskeleton. However, it appears that neurofilaments are highly expressed in large-calibre myelinated axons in white matter, while tau is predominantly expressed in thin unmyelinated axons of the cortex (39). Correlated CSF and plasma levels of NF-L have been shown to be significantly higher in patients with APS, including PSP, compared with controls and patients with PD (40) (41). However, there are no significant differences between the individual atypical parkinsonian conditions. These results are thought to reflect the more extensive and rapid neuronal loss that occurs in APS compared with PD. In addition, compared to controls, NF-L is raised in other neurodegenerative disorders with neuronal loss, including FTD (42), Huntington's disease (HD) (43), and multiple sclerosis where it has also been shown to be a marker of treatment response (44). With regards to tracking disease progression, the levels of CSF NF-L have been shown to increase over 1 year and be associated with worsening clinical rating scale scores including the PSPRS (16) (45). Similarly, baseline levels of plasma NF-

L have been shown to predict both clinical and radiological disease progression in PSP using gender and age-adjusted mixed linear models and Pearson's correlation **(46)**. Of note, all of the diagnostic and prognostic biomarker studies discussed above included PSP patients with the PSP-RS phenotype such that outside this PhD, there have been no studies that have measured the levels of relevant biomarkers (including tau and NF-L) in variant presentations of PSP.

#### 1.5 The challenges that have been addressed by this PhD

This PhD has addressed the following challenges which have also been defined and discussed in this introductory chapter:

1) Enhancing the diagnosis of PSP across its phenotypic spectrum by characterising the profile of specific PSP subtypes (Chapter 2).

2) Enhancing the early diagnosis of PSP with the use of fluid and genetic markers, which has the potential to refine the design of future clinical trials (Chapters 3 and 4).

3) Using genetic studies to understand the biological determinants of variation in PSP phenotype and the rate of survival, which has the potential to discover novel therapeutic targets (Chapters 5 and 6).

# <u>Chapter 2 – Phenotypic variation of progressive supranuclear palsy in the</u> <u>PROSPECT study cohort</u>

#### 2.1 Abstract

Atypical parkinsonian syndromes (APS), including progressive supranuclear palsy (PSP), corticobasal syndrome (CBS), and multiple system atrophy (MSA), may be difficult to distinguish in early stages and are often misdiagnosed as Parkinson disease (PD). The diagnostic criteria for PSP have been updated to encompass a range of clinical subtypes but have not been prospectively studied.

This cohort study recruited patients with APS and PD from movement disorder clinics across the United Kingdom from September 1, 2015, through December 1, 2018. Patients with APS were stratified into the following groups: those with Richardson syndrome (PSP-RS), PSP-subcortical (including PSP-parkinsonism and progressive gait freezing subtypes), PSP-cortical (including PSP-frontal and PSP-CBS overlap subtypes), MSA-parkinsonism, MSA-cerebellar, CBS–Alzheimer disease (CBS-AD), and CBS–non-AD. Data were analysed from February 1, through May 1, 2019. Baseline group comparisons used (1) clinical trajectory; (2) cognitive screening scales; (3) serum neurofilament light chain (NF-L) levels; (4) *TRIM11*, *ApoE*, and *MAPT* genotypes; and (5) volumetric magnetic resonance imaging measures.

A total of 222 patients with APS (101 with PSP, 55 with MSA, 40 with CBS, and 26 indeterminate) were recruited (129 [58.1%] male; mean [SD] age at recruitment, 68.3 [8.7] years). Age-matched control participants (n = 76) and patients with PD (n = 1967) were included for comparison. Concordance between the ante-mortem clinical and pathologic diagnoses was achieved in 12 of 13 patients with PSP and CBS (92.3%) undergoing post-mortem evaluation. Applying the Movement Disorder Society PSP diagnostic criteria almost doubled the number of patients diagnosed with PSP from 58 to 101. Forty-nine of 101

patients with reclassified PSP (48.5%) did not have the classic PSP-RS subtype. Patients in the PSP-subcortical group had a longer diagnostic latency and a more benign clinical trajectory than those in PSP-RS and PSP-cortical groups. The PSP-subcortical group was distinguished from PSP-cortical and PSP-RS groups by cortical volumetric magnetic resonance imaging measures (area under the curve [AUC], 0.84-0.89), cognitive profile (AUC, 0.80-0.83), serum NF-L level (AUC, 0.75-0.83), and *TRIM11* rs564309 genotype. Midbrain atrophy was a common feature of all PSP groups. Eight of 17 patients with CBS (47.1%) undergoing cerebrospinal fluid analysis were identified as having the CBS-AD subtype. Patients in the CBS-AD group had a longer diagnostic latency, relatively benign clinical trajectory, greater cognitive impairment, and higher *APOE*- $\epsilon$ 4 allele frequency than those in the CBS–non-AD group (AUC, 0.80-0.87; *P* < .05). Serum NF-L levels distinguished PD from all PSP and CBS cases combined (AUC, 0.80; *P* < .05).

These findings suggest that studies focusing on the PSP-RS subtype are likely to miss a large number of patients with underlying PSP tau pathology. Analysis of cerebrospinal fluid defined a distinct CBS-AD subtype. The PSP and CBS subtypes have distinct characteristics that may enhance their early diagnosis.

#### 2.2 Introduction

Atypical parkinsonian syndromes (APS) – including PSP, corticobasal syndrome (CBS) and multiple system atrophy (MSA) – are characterized by a more rapid deterioration and poorer levodopa response than is usually seen in PD (47). In addition, APS are rarer than PD, with an estimated combined prevalence of 10 to 18 per 100,000 population (1) (2). Within APS, there is a high degree of clinical overlap, particularly in early disease, leading to greater misdiagnosis than occurs in PD (12). The lack of proven disease-specific diagnostic markers means that post-mortem neuropathologic analysis is the gold standard for confirming the clinical diagnosis.

The recent therapeutic trials of davunetide and tideglusib in PSP-RS did not result in improved outcomes **(16) (17)**. The power of clinical trials is limited by individual variability in disease progression and misclassification. Moreover, trials that focus on classic presentations may not be applicable to the full disease spectrum. Accurate diagnosis and prognosis based on clinical and biomarker data may increase statistical power and reduce the required sample size for trials **(46)**. The new era of potential disease-modifying therapies for APS has made the need for early and accurate biomarker-supported clinical diagnosis even greater. Since publication of the 2017 MDS PSP clinical diagnostic criteria **(11)**, which recognise variant presentations of PSP, there has been independent retrospective neuropathologic validation showing improved sensitivity compared with the NINDS-SPSP criteria **(6)**. Until now, prospective evaluation of these variant PSP presentations has been lacking.

CBS is a clinical syndrome characterized by progressive asymmetrical limb apraxia, parkinsonism, dystonia, and particular cognitive impairments **(48)**. The underlying neuropathologic features of CBS are heterogeneous, with CBD, PSP, AD, and TAR DNA-binding protein 43 (TDP-43) pathologic changes seen at post-mortem, even when using the

clinical consensus criteria (49) (50). MSA is an  $\alpha$ -synuclein–linked oligodendrogliopathy manifesting with variable combinations of progressive autonomic failure, parkinsonism with poor levodopa response, and cerebellar ataxia (51).

Here, I describe the UK-wide Progressive Supranuclear Palsy–Corticobasal Syndrome– Multiple System Atrophy (PROSPECT) study and compare our baseline data with that of patients with PD in the UK-wide Tracking Parkinson's study to provide a comprehensive prospective picture of the diagnosis and clinical features of PSP and CBS. A strength of the PROSPECT study was the breadth of clinical subtypes that were studied systematically with multiple candidate biomarkers, including indeterminate cases that lay outside of diagnostic criteria when the study was started but came to lie within the current classifications of APS after publication of the MDS PSP criteria. I examined clinical, cognitive, fluid, genetic, and imaging biomarkers and performed group comparisons, including receiver operating characteristic curves for patient classification.

#### 2.3 Methods

#### Study Design

The PROSPECT study natural history cohort consists of 7 UK study sites (University College London [UCL], Oxford, Cambridge, Newcastle, Brighton, Newport, and Manchester), funded by the PSP Association and MSA Trust. Prof Morris and Dr Woodside obtained study-wide ethical approval from the UCL Queen Square Institute of Neurology research ethics committee. I was part of a UK-wide research network that recruited participants, and obtained written informed consent from September 1, 2015, through December 1, 2018. We invited all participants to register for post-mortem brain donation at 1 of 4 UK brain banks (Queen Square [London], Cambridge, Oxford, and Manchester). Tracking Parkinson's is a UK-wide longitudinal study of PD led by Prof Donald Grosset (Glasgow). Participants with a baseline clinical diagnosis of PD at 72 sites in the United

Kingdom, with multi-centre ethics committee and local research and development department approvals, were recruited and provided written informed consent from January 1, 2012, through December 31, 2014 **(52)**. Post-mortem data from patients with PD in the Tracking Parkinson's study were not available for analysis.

#### **Participants**

I defined patients entering the study as having PSP, following the NINDS-SPSP criteria (6); CBS, following the Armstrong criteria (48); or MSA, following the revised Gilman criteria (51). I also included patients with progressive movement or cognitive disorders, thought likely to have APS (based on having atypical clinical features for PD) but not meeting any of the above diagnostic criteria, as indeterminate (IDT) cases. Recruited control participants included a spouse or a friend of the case or were recruited through the Join Dementia Research volunteer registry. Cases with PD from the Tracking Parkinson's study were diagnosed using the Queen Square Brain Bank clinical diagnostic criteria (53).

#### Phenotyping

I then reclassified PROSPECT study cases with a diagnosis of PSP, CBS, or IDT according to current MDS PSP criteria at the end of baseline recruitment using consensus guidelines **(54)**. All reclassified PSP cases fulfilled at least "possible" diagnostic criteria. I stratified PSP cases into PSP-RS, PSP-subcortical, and PSP-cortical groups. The PSP-subcortical group includes cases with PSP-P, PSP-PGF, and PSP-oculomotor; the PSP-cortical group includes cases with PSP/CBS overlap and PSP-frontal. Baseline CBS cases with cerebrospinal fluid (CSF) evidence of underlying AD pathologic features (described in the "Fluid Biomarkers" sub-section below) were defined as CBS-AD, and those with normal CSF analysis were defined as CBS-4RT because they are likely to have underlying CBD or PSP pathologic features. Cases of CBS without CSF or post-mortem examination were defined as CBS-unknown. Baseline MSA cases were divided into MSA-parkinsonism and MSA-cerebellar groups according to the revised Gilman criteria. Cases with PD who have

had a change in clinical diagnosis since their baseline Tracking Parkinson's clinical assessment were excluded from this study.

#### **Clinical Assessments**

Core and optional study assessments were completed at baseline. These assessments were repeated after 6, 12, 24, and 36 months of follow-up, with brief assessments at the 48- and 60-month study visits (Tables 2.1 and 2.2). At each study visit, a neurological history was obtained, and an examination was performed. The PSPRS (scores range from 0-100, with higher scores indicating greater impairment) (14) for PSP, CBS, and IDT cases or the Unified Multiple System Atrophy Rating Scale (scores range from 0-104, with higher scores indicating greater impairment) (55) for MSA cases were administered. In addition, all cases were assessed using the MDS Unified Parkinson's Disease Rating Scale parts II (MDS-UPDRS II; scores range from 0-52, with higher scores indicating greater impairment) and III (MDS-UPDRS III; scores range from 0-132, with higher scores indicating greater impairment) (56) and the Schwab and England Activities of Daily Living Scale (SEADL; scores range from 0-100, with lower scores indicating greater impairment) (57). Myself, Dr Ruth Lamb and Dr Viorica Chelban administered these clinical assessments at the UCL site. At other sites, the assessments were administered by a mixture of clinical research fellows and principal investigators. Cases and controls were screened for cognitive dysfunction using the Montreal Cognitive Assessment (MoCA; scores range from 0-30, with higher scores indicating greater impairment) (58). Addenbrooke's Cognitive Examination 3 (ACE-III) (59) and Edinburgh Cognitive and Behavioural ALS Screen (60) were administered as additional, optional cognitive screening assessments. Cases with PD from the Tracking Parkinson's study underwent baseline testing with the MDS Unified Parkinson's Disease Rating Scale parts II and III, SEADL, and MoCA. Dr John Woodside and Miss Charlotte Rawlinson administered these cognitive assessments at the UCL site. At other sites, the assessments were administered by a mixture of research nurses, clinical research fellows and principal investigators.

## Table 2.1: Overview of assessments in the PROSPECT study natural history cohort

	Study visit							
	<b>0</b> m	6m	12m	24m	36m	48m	60m	
Core study assessments								
Clinical rating scales	$\checkmark$							
Functional rating scales	$\checkmark$							
Cognitive screening tests	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$			
Neuropsychometric tests	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$			
Blood sample for biomarkers (and genetics at 0m only)	~	~	✓	~	✓			
Smell test	$\checkmark$							
Optional biomarker assessments								
Skin biopsy	$\checkmark$		$\checkmark$					
Magnetic resonance imaging	$\checkmark$		$\checkmark$					
Cerebrospinal fluid collection	$\checkmark$	$\checkmark$	$\checkmark$					

## Table 2.2: Clinical rating scales, functional rating scales and cognitive screening

	Study visit							
	0m	6m	12m	24m	36m	48m	60m	
Clinician assessed								
PSPRS/UMSARS	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	>	>	$\checkmark$	
MDS UPDRS-III	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	>	~	$\checkmark$	
SEADL	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	>	~	$\checkmark$	
MoCA	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	>			
ECAS	$\checkmark$	$\checkmark$	~	$\checkmark$	~			
ACE-III	$\checkmark$	$\checkmark$	$\checkmark$	$\checkmark$	>			
Patient self-assessed								
MDS UPDRS-II	$\checkmark$							

tests used in the PROSPECT study natural history cohort

Addenbrooke's cognitive examination 3: ACE-III, Edinburgh cognitive and behavioural ALS screen:

ECAS, Montreal cognitive assessment: MoCA, Schwab and England activities of daily living scale:

SEADL, PSP rating scale: PSPRS, Unified Multiple System Atrophy rating scale: UMSARS,

Movement Disorder Society Unified Parkinson's Disease Rating Scale part II: MDS UPDRS-II,

Movement Disorder Society Unified Parkinson's Disease Rating Scale part III: MDS UPDRS-III.

## Fluid Biomarkers

Serum NF-L levels were measured in a subset of PROSPECT and Tracking Parkinson's cases, and PROSPECT controls. CSF t-tau and  $\beta$ -amyloid 1-42 (A $\beta$ 1-42) levels were measured in a subset of PROSPECT cases. A subset of PD cases from the Tracking

Parkinson's study underwent baseline serum NF-L testing. At each site, non-fasting blood and CSF samples were collected using standardised protocols and stored 0.5ml aliquots at -80°C within 60 minutes of sample collection. Serum NF-L was measured using an ultrasensitive single molecule array (Simoa) assay **(61)**. CSF samples from cases were tested for t-tau and A $\beta$ 1-42 levels using INNOTEST ELISA – Fujirebio Europe N.V., Gent, Belgium **(62)**. These biomarker assays were run by Dr Tong Guo and Dr Amanda Heslegrave at the UK Dementia Research Institute Fluid Biomarker Laboratory at UCL. I stratified CBS cases into groups with likely underlying AD pathologic features (CBS-AD), defined as cases with a CSF t-tau:A $\beta$ 1-42 ratio of greater than 1 **(63) (64)**; likely 4R tau pathologic features (CBS-4RT), defined as cases with a CSF t-tau:A $\beta$ 1-42 ratio of less than 1; and unknown pathologic features (CBS-unknown), defined as cases with no CSF analysis.

#### Genetics

A subset of PROSPECT and Tracking Parkinson cases had DNA extracted from blood samples. For PROSPECT study cases, I was responsible for carrying out DNA sample quality control and preparation for genotyping by doing gel electrophoresis and Qubit fluorometry at the UCL Institute of Neurology. I subsequent delivered plates of DNA samples to the UCL Institute of Child Health where they underwent genotyping using the Illumina NeuroChip (65). Tracking Parkinson's cases were genotyped at Cardiff University under the supervision of Prof Nigel Williams using the Illumina Human Core Exome array (66). Once both datasets were merged, I carried out standard steps for data quality control including sex checking, heterozygosity, identity by descent to exclude related individuals and a principal components analysis (PCA) merged with the HapMap reference dataset to exclude all cases of non-European ancestry and screening for *MAPT*, *LRRK2* and *DCTN1* mutations covered by both genotyping platforms (67). I then carried out SNP imputation using the Haplotype Reference Consortium v1.1 panel (https://imputation.sanger.ac.uk/). I

used the following post-imputation data quality control steps to filter out SNPs with: INFO score <0.70, posterior probability <0.90, genetic missingness >0.05, Hardy-Weinberg equilibrium p-value <1.0x10<sup>-7</sup> and minor allele frequency <0.01. I extracted *MAPT* H1/H1, *APOE*  $\varepsilon$ 4 allele, and *TRIM11* rs564309 minor allele group frequencies from the imputed dataset using the --recodeA command in Plink, version 1.9. Caucasian control allele/haplotype frequencies were derived from dbSNP (www.ncbi.nlm.nih.gov/snp).

#### Neuroimaging

A subset of PROSPECT participants attended 3 scanning centres (UCL, Cambridge, and Oxford) and underwent baseline volumetric T1-weighted magnetic resonance imaging (MRI) on 3T scanners (Siemens, Prisma, or TRIO). All of the following methods were designed and carried out by Prof James Rowe and his team at the University of Cambridge. Scan protocols were designed at the outset of the study to closely match across centres, based on the international Genetic Frontotemporal Dementia Initiative protocols (MP-RAGE, TR 2s, TE 2.93ms, Flip angle 8deg, 1.1mm isotropic) (68). T1weighted images were processed using the recon-all pipeline with brainstem structures of FreeSurfer v6.0.0 (surfer.nmr.mgh.harvard.edu) into subcortical segments and cortical surface parcellations. Regional analyses were performed using volume measures from 68 Desikan-Killiany atlas cortical regions and 38 subcortical volume measures from the segmentation (69) (70). We combined volume measures from the parcellation to calculate cortical grey matter volumes of the frontal, temporal, parietal, occipital lobes. Hippocampi and amygdalae volumes are included in the temporal lobe data. The remaining segmentation regions were combined into central regions (basal ganglia, thalamus, accumbens), cerebellar grey matter, brainstem, and ventricles. The images were additionally segmented into grey, white and CSF modulated probability maps using CAT12 (neuro.uni-jena.de/cat) in order to obtain total intracranial volume (TIV) measures. We also calculated the ratio of the pons to midbrain volume. Imaging data from Tracking Parkinson's participants were not available.

#### **Statistical Analysis**

Data were analysed from February 1 through May 1, 2019, using Plink, version 1.9 (Harvard University), GraphPad, version 8 (Prism), and Stata, version 15 (StataCorp LLC). I conducted all of the statistical analyses included in this chapter aside from the analysis of imaging data which was conducted by Prof Rowe's team. For missing data in clinical scales, an adjusted mean score was used if at least 80% of the assessment was complete. Group comparisons of clinical, cognitive, and biomarker measures were made using logistic regression analyses with sex, age at symptom onset, and disease duration at testing as covariates. I calculated the clinical disease trajectory by dividing PSPRS and MDS-UPDRS II and III scores at baseline by the number of years since reported motor symptom onset, assuming a score of 0 immediately before symptom onset. For the SEADL, the clinical disease trajectory was calculated as (100 – baseline score) divided by the number of years since reported motor symptom of years since reported motor symptom onset, assuming a score of 100 immediately before symptom onset. Statistical significance for the clinical, cognitive, and biomarker group comparisons described above was defined as a false discovery rate–corrected, 2-sided P < .05.

Group comparisons of genetic data were made using Fisher exact tests, and statistical significance was defined as a Bonferroni-corrected 2-sided P < .05. We performed an analysis of covariance on imaging volumetric measures from each brain region, with diagnosis and sex as factors and age at scan and total intracranial volume as covariates. Regional marginal mean values were compared post hoc using unpaired *t* tests. The significance of mean differences was adjusted using false discovery rate correction with 2-sided P < .05 considered significant.

In addition, receiver operating characteristic curve analyses were performed on cognitive scale, serum NF-L level, and regional imaging volumetric values from group pairs with the area under the curve (AUC) used as a measure of separation between the groups.
#### 2.4 Results

#### **Recruitment and Phenotyping**

I analysed 222 cases with APS (93 female [41.9%] and 129 male [58.1%]; mean [SD] age at recruitment, 68.3 [8.7] years), 76 controls, and 1967 cases with PD. At study entry, application of clinical diagnostic criteria, including the NINDS-SPSP PSP criteria, identified 58 cases with PSP, 55 cases with MSA, 55 cases with CBS, and 54 IDT cases (**Figure 2.1**). Reclassification of PROSPECT PSP, CBS, and IDT cases was possible after the publication of the 2017 MDS PSP criteria, resulting in 101 cases with PSP, 55 cases with MSA, 40 cases with CBS, and 26 IDT cases (**Figure 2.1**). Of note, 15 cases with CBS were reclassified as PSP/CBS overlap under the MDS PSP diagnostic criteria because they had the presence of slowed vertical saccades and/or a vertical supranuclear gaze palsy, both of which are associated with underlying PSP pathologic findings. In total, 17 of 40 reclassified CBS cases (42.5%) had CSF collection, of whom8 (47.1%) had an AD-like CSF profile. The following disease groups were defined (**Figure 2.1**): PSP-RS, PSP-subcortical (consisting of PSP-P, PSP-PGF, and PSP–oculomotor predominant subtypes), PSP-cortical (consisting of PSP/CBS overlap and PSP-frontal subtypes), MSA-parkinsonism, MSA-cerebellar, CBS-unknown, CBS-4RT, CBS-AD, and IDT.

#### Pathologic Confirmation of Diagnosis

Forty-four of 222 cases in the PROSPECT cohort (19.8%) had died at the point of censoring, with a mean (SD) disease duration of 5.9 (2.3) years. Seventeen of 44 cases (38.6%) had pathologic confirmation of diagnosis at the Queen Square, Cambridge, Oxford, and Manchester brain banks with concordance between ante-mortem clinical and pathologic diagnoses achieved in 12 of 13 cases with PSP and CBS (92.3%) **(Table 2.3)**. Of note, the one clinically diagnosed case with PSP-RS who had CBD pathologic findings at post-mortem had typical features of PSP-RS with no evidence of apraxia throughout the

disease course and was therefore classified as a case of PSP-RS under the NINDS-SPSP and MDS diagnostic criteria.



**Figure 2.1** – Recruitment of patients to the natural history cohort of the PROSPECT study from September 1, 2015, through December 1, 2018.

#### **Clinical Features**

All PROSPECT cases underwent baseline clinical testing, whereas baseline clinical data were obtained from 1763 of 1967 recruited patients with PD (89.6%) in the Tracking Parkinson's study. Thirty two of 204 patients with PD (15.7%) were excluded from this analysis owing to missing data or a change in diagnosis at the point of data analysis. The baseline clinical features of reclassified cases and controls are summarized in **Table 2.3**. There was a long diagnostic delay for CBS-AD (mean [SD], 4.6 [3.2] years) and PSP-subcortical (mean [SD], 4.2 [3.2] years) groups compared with the CBS-4RT (mean [SD], 3.1 [4.7] years) and PSP-RS (mean [SD], 2.3 [1.8] years) groups.

#### Disease progression analyses using clinical trajectory measures

There was also variation in the burden of disease at study enrolment as measured by the baseline PSPRS and SEADL scores, with the highest degree of impairment seen in the PSP-RS (mean [SD] scores, 35.7 [15.1] and 53.3 [18.1], respectively), PSP-cortical (mean [SD] scores, 39.1 [13.5] and 48.8 [15.0], respectively), and CBS-4RT (mean [SD] scores, 41.3 [15.6] and 50.1 [19.8], respectively) groups. Clinical trajectory analyses (**Figure 2.2**), in particular the SEADL, showed that the PSP-subcortical (mean [SD] decline in score, -8.5 [8.6] points per year), CBS-AD (mean [SD] decline in score, -12.0 [7.0] points per year), and PD (mean [SD] decline in score, -3.9 [5.1] points per year) groups had more benign disease trajectories than all other groups.

	Controls	PSP- All	PSP- RS	PSP- Subcortical	PSP- Cortical	MSA- Ali	MSA- P	MSA- C	CBS- All	CBS- NK	CBS- 4RT	CBS- AD	IDT	PD**
N	76	101	52	24	25	55	32	23	40	23	9	8	26	1763
Gender - % male	37	62	63	75	48	65	53	83	35	32	44	38	58	65
Age at enrolment (years) - mean, (SD)	66.7 (8.4)	70.6 (7.3)	69.2 (7.8)	73.1 (6.0)	71.1 (7.0)	63.6 (10.1)	63.7 (10.4)	63.5 (9.9)	68.4 (7.4)	68.3 (6.8)	67.5 (8.6)	69.7 (8.3)	69.2 (9.3)	67.3 (9.5)
Age at motor symptom onset (years) - mean, (SD)	-	66.3 (7.4)	65.9 (7.3)	66.5 (8.2)	66.9 (7.2)	58.1 <b>†</b> (10.8)	58.5 (11.3)	57.5 (10.1)	63.6 (7.2)	63.9 (7.6)	61.8 (9.6)	64.6 (7.6)	65.2 (9.1)	64.4 (9.7)
Disease duration at enrolment (years) - mean, (SD)	-	4.4 (2.7)	3.4 (2.2)	6.6* (3.7)	4.3 (2.2)	5.4 (2.8)	5.2 (2.8)	5.9 (2.7)	4.9 (3.2)	4.3 (3.0)	5.8 (3.7)	5.2 (3.3)	4.1 (2.9)	3.0 <b>Ω</b> (3.1)
Diagnostic latency (years) - mean, (SD)	-	2.8 (2.2)	2.3 (1.8)	4.2* (3.2)	2.7 (1.3)	3.2 (2.5)	2.8 (2.2)	3.6 (2.8)	3.2 (3.0)	2.8 (2.0)	3.1 (4.7)	4.6 (3.2)	-	1.8 <b>Ω</b> (2.8)
Deceased - n	-	23	9	7	7	10	5	5	8	3	4	1	3	-
Post-mortem - n, diagnosis	-	10 PSP-9 CBD-1	5 PSP-4 CBD-1	3 PSP–3	2 PSP–2	4 MSA–3 PD–1	2 MSA–1 PD–1	2 MSA–2	3 CBD–3	1 CBD–1	2 CBD–2	0	0	-

Table 2.3: Baseline clinical features of the PROSPECT study natural history cohort

Baseline clinical features of the PROSPECT study natural history cohort, defined by patients' reclassified baseline diagnoses. Indeterminate: IDT, PSP-Richardson syndrome: PSP-RS, MSA-parkinsonism: MSA-P, MSA-cerebellar: MSA-C, CBS-Alzheimer's disease: CBS-AD, CBS-4 repeat tau: CBS-4RT, CBS-unknown: CBS-NK, Parkinson's disease: PD, standard deviation: SD, PSP rating scale: PSPRS, Schwab and England activities of daily living scale: SEADL. Group and sub-group comparisons made using t-testing -  $\dagger$  = FDR adjusted P<0.05 vs. PSP-All, \* = FDR adjusted P<0.05 vs. PSP-RS and PSP-Cortical,  $\Omega$  = FDR adjusted P<0.05 vs. PSP-All, CBS-All and MSA-All, \*\* = data from the Tracking Parkinson's study.



**Figure 2.2** – Clinical disease trajectory profiles. A) Schwab and England activities of daily living scale: SEADL; B) PSP rating scale: PSPRS; C) Movement Disorder Society Unified Parkinson's Disease Rating Scale part II: MDS UPDRS-II; D) Movement Disorder Society Unified Parkinson's Disease Rating Scale part III: MDS UPDRS-III. Indeterminate: IDT, Parkinson's disease: PD, PSP Richardson's syndrome: PSP-RS, CBS-Alzheimer's Disease: CBS-AD, CBS-4 repeat tau: CBS-4RT. Mean group values with standard deviation error bars shown. Group comparisons adjusted for sex and age at symptom onset. a= FDR adjusted P<0.05 PSP-Subcortical vs PSP-RS and PSP-Cortical, b= FDR adjusted P<0.01 PD vs PSP-All and CBS-All, c= FDR adjusted P<0.05 PD vs PSP-All and CBS-All.

#### **Cognitive Profiles**

I evaluated cognitive function using the MoCA, Edinburgh Cognitive and Behavioural ALS Screen, and ACE-III. Among the PROSPECT cases and controls, the MoCA was completed in 235 of 243 participants (96.7%); the Edinburgh Cognitive and Behavioural ALS screen, in 211 of 243 (86.8%); and the ACE-III, in 223 of 243 (91.8%) **(Table 2.4)**. The 3 assessments

were strongly correlated (all comparisons, r > 0.80). Among the 1967 patients with PD in the Tracking Parkinson's study, 1833 (93.2%) had baseline MoCA testing. With regard to total scores, the PD group had better cognition (mean [SD] score, 24.9 [3.6]) compared with the PSP-all (mean [SD] score, 21.9 [4.7]) and CBS-all (20.4 [7.4]) groups. The PSP-cortical group was more impaired across all 3 scales compared with the PSP-RS and PSP-subcortical groups (false discovery rate corrected, P < .05). The CBS-AD group had worse cognition in all scales compared with CBS-4RT, but the statistical comparisons were likely limited by small group sizes, with significance reached only in MoCA total score (mean [SD] score, 22.9 [5.3] for CBS-4RT and 12.4 [9.0] for CBS-AD) and ACE-III attention, memory, and language subscale measures **(Table 2.4)**.

Table 2.4: Cognitive screening measures in the PROSPECT study natural his	story
cohort	

	Controls	PSP- All	PSP- RS	PSP- SC	PSP- Cort	CBS- All	CBS- NK	CBS- 4RT	CBS- AD	IDT	PD**
MoCA - n, mean, (SD)	76 26.8 <b>†</b> (2.2)	96 21.9 (4.7)	51 22.5 <b>x</b> (3.6)	22 23.1 <b>x</b> (5.0)	23 19.7 (6.0)	39 20.4 (7.4)	23 21.8 <b>ф</b> (6.1)	9 22.9 <b>ф</b> (5.3)	7 12.4 (9.0)	24 21.9 (6.5)	1833 24.9* (3.6)
ECAS - n, mean, (SD)	76 116.3 <del>†</del> (10.6)	86 88.7 (22.1)	46 89.4 <b>x</b> (17.8)	21 97.0 <b>x</b> (22.0)	19 78.3 (27.8)	27 84.3 (33.0)	14 96.7 <b>ф</b> (24.9)	7 88.1 (26.6)	6 55.0 (39.0)	22 95.3 (18.8)	-
ACE-III - n, mean, (SD)	76 95.1 <b>†</b> (5.3)	91 78.8 <b>Ω</b> (13.1)	50 80.5 <b>x</b> (9.5)	20 84.0 <b>x</b> (9.8)	21 69.8 (18.6)	31 71.9 (23.5)	17 76.3 <b>ф</b> (20.1)	7 78.1 (17.0)	7 54.3 (31.2)	25 80.7 (12.1)	-
Attention -	17.3*	15.8 <b>Ω</b>	15.8	16.7 <b>x</b>	14.9	14.3	14.8 <b>φ</b>	16.1 <b>ф</b>	10.8	16.8	-
mean, (SD)	(1.2)	(2.3)	(2.1)	(1.9)	(2.3)	(4.1)	(3.2)	(2.5)	(5.7)	(2.1)	
Memory -	24.4 <b>†</b>	21.0 <b>Ω</b>	21.6 <b>x</b>	21.8	18.5	18.4	19.6	20.0 <b>ф</b>	13.7	19.5	-
mean, (SD)	(2.6)	(4.6)	(3.5)	(4.4)	(6.4)	(7.3)	(7.5)	(4.9)	(5.8)	(4.1)	
Fluency -	12.3 <b>†</b>	6.4	6.3	7.9 <b>x</b>	5.3	7.3	8.4	6.5	5.7	7.6	-
mean, (SD)	(1.6)	(3.4)	(3.0)	(3.2)	(3.9)	(4.1)	(3.4)	(4.4)	(5.2)	(4.1)	
Language -	25.7 <b>†</b>	23.4 <b>Ω</b>	23.9 <b>x</b>	24.8 <b>x</b>	20.5	21.5	22.5 <b>ф</b>	23.6 <b>ф</b>	16.6	23.6	-
mean, (SD)	(0.6)	(3.8)	(2.3)	(1.5)	(6.6)	(6.0)	(4.7)	(2.3)	(8.1)	(3.4)	
Visuospatial	15.4 <b>†</b>	12.2 <b>Ω</b>	12.6 <b>x</b>	12.9 <b>x</b>	10.4	10.4	11.0	11.9	7.4	13.5	-
- mean, (SD)	(1.0)	(3.1)	(2.5)	(2.9)	(3.8)	(4.9)	(4.5)	(5.4)	(5.7)	(2.3)	

Cognitive screening measures in PROSPECT study natural history cohort. Addenbrooke's cognitive

examination 3: ACE-III, Edinburgh cognitive and behavioural ALS screen: ECAS, Indeterminate: IDT,

PSP-Richardson syndrome: PSP-RS, PSP-subcortical: PSP-SC, PSP-cortical: PSP-Cort, Montreal cognitive assessment: MoCA, CBS-Alzheimer's disease: CBS-AD, CBS-4 repeat tau: CBS-4RT, CBS-unknown: CBS-NK, Parkinson's disease: PD, standard deviation: SD. Group and sub-group comparisons made with logistic regression analyses that used gender, age at symptom onset and disease duration at testing as covariates -  $\mathbf{t}$  = FDR adjusted P<0.05 vs. all disease groups, \* = FDR adjusted P<0.05 vs. CBS All,  $\mathbf{\phi}$  = FDR adjusted P<0.05 vs. CBS-AD,  $\mathbf{x}$  = FDR adjusted P<0.05 vs. PSP-Cort, \*\* = data from the Tracking Parkinson's study.

#### **Fluid Biomarkers**

Testing of serum samples for NF-L levels was performed in 186 of 243 PROSPECT cases and controls (76.5%) and 140 of 1967 PD cases (7.1%) in the Tracking Parkinson's study. Forty-four of 167 cases (26.3%) in the PROSPECT study had CSF testing for T-tau and Aβ1-42 levels. At the group level, serum NF-L levels in patients with PD (26.5pg/L) were significantly higher than in controls (16.4pg/L) and the PSP-all (47.4 pg/L) and CBS-all (53.1 pg/L) groups. Serum NF-L levels did not distinguish between the PSP-all and CBS-all groups (Figure 2.3). With respect to disease subgroups, there was a trend toward higher mean serum NF-L levels in PSP-cortical (58.6 pg/L) vs PSP-RS (45.3 pg/L) and PSP-subcortical (41.6 pg/L) and in CBS-4RT (52.4 pg/L) vs CBS-AD (36.5 pg/L) (Figure 2.3).



**Figure 2.3** – Fluid biomarker profiles. Cerebrospinal fluid: CSF, Neurofilament light chain: NF-L, total tau: T-tau, amyloid beta 1-42: A $\beta$ 1-42, Indeterminate: IDT, Parkinson's disease: PD, PSP Richardson's syndrome: PSP-RS, CBS-Alzheimer's Disease: CBS-AD, CBS-4 repeat tau: CBS-4RT. Mean group values with standard deviation error bars shown. Group comparisons adjusted for sex, age at symptom onset and disease duration at testing. a= False discovery rate (FDR)–adjusted *P* < .01, controls vs all disease groups; b= FDRadjusted *P* < .05, PD vs PSP-all; c= FDR-adjusted *P* < .05, PD vs CBS-all; d= FDR-adjusted *P* < .01, CBS-AD vs all other disease groups.

#### Genetics

Genotype data were obtained from 134 of 167 PROSPECT cases (80.2%) and 1566 of 1967 PD cases (79.6%) in the Tracking Parkinson's study **(Table 2.5)**. In the analysis of cases of European ancestry, I found significantly higher *MAPT* H1/H1 frequencies in the PSP-all (88.9%) and CBS-all (78.8%) groups compared with the PD group (67.2%) and reference controls (67.1%) (Bonferroni-corrected P < .05). At the subgroup level, I found significantly

higher *APOE*- $\varepsilon$ 4 allele frequencies in CBS-AD (35.7%) compared with CBS-4RT (Bonferroni corrected *P* < .05). Although analyses were underpowered to reach significance, as reported previously **(71)** I found higher *TRIM11* rs564309 minor allele frequencies in PSP-subcortical (15.0%) compared with PSP-RS (7.1%).

	Controls*	PSP- All	PSP- RS	PSP- SC	PSP- Cort	CBS- All	CBS- NK	CBS- 4RT	CBS- AD	IDT	PD**
No. of cases genotyped	-	81	42	20	19	33	20	6	7	20	1566
MAPT H1/H1 frequency (%)	67.1	88.9	90.5	85.0	89.5	78.8	80.0	83.3	71.4	75.0	67.2 <b>‡</b>
<i>APOE</i> -ε4 allele frequency (%)	15.1	6.8	9.5	2.5	5.3	12.1	7.5	0 <b>ф</b>	35.7	10.0	12.8 <b>Ω</b>
<i>TRIM11</i> (rs564309) MAF (%)	9.6	9.3	7.1	15.0	7.9	12.1	10.0	33.3 <b>ф</b>	0	25.0	14.7

Table 2.5: Genetic ana	yses in the PROSPECT stu	udy natural history cohort
------------------------	--------------------------	----------------------------

Genetic analyses in the PROSPECT study natural history cohort. Apolipoprotein E: APOE,

Indeterminate: IDT, PSP Richardson syndrome: PSP-RS, PSP-subcortical: PSP-SC, PSP-cortical: PSP-Cort, minor allele frequency: MAF, microtubule associated protein tau: *MAPT*, CBS-Alzheimer's disease: CBS-AD, CBS-4 repeat tau: CBS-4RT, CBS-unknown: CBS-NK, Parkinson's disease: PD, tripartite motif-containing protein 11: *TRIM11*. Group and sub-group comparisons of genetic data made with Fisher's exact test -  $\ddagger$  = Bonferroni corrected P<0.05 vs. PSP-All and CBS-All,  $\Omega$  = Bonferroni corrected P<0.05 vs. PSP-All,  $\phi$  = Bonferroni corrected P<0.05 vs. CBS-AD. \* = Caucasian control frequencies derived from dbSNP, \*\* = data from the Tracking Parkinson's study.

#### Magnetic Resonance Imaging

Volumetric measures from T1-weighted MRI scans were derived for 108 of 243 PROSPECT cases and controls (44.4%). Table 2.6 outlines the differences in cortical and subcortical volumetric measures across all groups, with post hoc pairwise group comparisons of each patient group vs controls and selected comparisons between patient groups (Table 2.7). Midbrain atrophy was a consistent neuroimaging feature in all PSP groups (marginal mean [SD] volume: 5.99 [0.53] mL in controls; 5.01 [0.54] mL in PSPRS; 5.23 [0.54] mL in PSPsubcortical; and 5.16 [0.55] mL in PSP-cortical). However, there was a dissociation between subcortical and cortical variants of PSP: the PSP-subcortical group showed less atrophy in the midbrain, medulla, and central structures, with relatively preserved cortical volumes; the PSP-cortical group showed additional severe frontotemporal atrophy. Corticobasal syndromes were associated with relative preservation of the pons and midbrain (marginal mean [SD] volume: 14.72 [1.70] mL and 5.99 [0.53] mL, respectively, in controls; 13.67 [1.68] mL and 5.54 [0.52] mL, respectively, in CBS-all) but severe atrophy of the central structures and cerebral cortex. Atrophy varied according to whether the CSF AD biomarkers were positive or not, with especially prominent ventriculomegaly in cases with CBS-AD (marginal mean [SD] volume: 35.80 [19.25] mL in controls; 60.81 [18.83] mL in CBS-4RT; 75.75 [18.81] mL in CBS-AD). The IDT cases were notable for their preserved posterior fossa structures, with atrophy of central structures and cerebral cortex.

#### **Receiver Operating Characteristic Curve Analyses**

As shown in previous studies, I found that serum NF-L levels were very effective at differentiating all PSP cases from controls (AUC, 0.88) (Figure 2.4). Additionally, I showed that PD was distinguished from an APS group, which consisted of all PSP and CBS cases, using serum NF-L levels (AUC, 0.80) and the MoCA score (AUC, 0.78). I also highlight measures that had high diagnostic accuracy (defined by an AUC  $\geq$  0.80) in differentiating between subgroups (Table 2.7). All cognitive measures (MoCA, Edinburgh Cognitive and Behavioural ALS Screen, and ACE III) differentiated CBS-AD from CBS-4RT (AUC, 0.80-

0.87) and PSP-subcortical from PSP-RS and PSP-cortical (AUC, 0.80-0.83). In addition, PSP-subcortical was distinguished from PSP-RS using serum NF-L levels (AUC, 0.83) and from PSP-cortical using cortical volumetric MRI measures (AUC, 0.80-0.89).

	Controls	PSP- All	PSP- RS	PSP- SC	PSP- Cort	CBS- All	CBS- NK	CBS- 4RT	CBS- AD	IDT	p-
n	35	44	25	11	8	17	7	5	5	12	value
Pons- Midbrain ratio	2.45 (0.22)	2.60* (0.24)	2.61* (0.23)	2.61 (0.22)	2.55 (0.23)	2.48 (0.23)	2.57 (0.22)	2.40 (0.22)	2.44 (0.22)	2.52 (0.22)	2.46e <sup>-9</sup>
Pons	14.72 (1.70)	13.25** (1.76)	13.06** (1.71)	13.70 (1.70)	13.17* (1.75)	13.67* (1.68)	14.09 (1.67)	13.73 (1.66)	13.57 (1.66)	14.55 (1.66)	2.73e-⁰
Midbrain	5.99 (0.53)	5.09** (0.54)	5.01** (0.54)	5.23** (0.54)	5.16** (0.55)	5.54* (0.52)	5.54 (0.52)	5.73 (0.52)	5.55 (0.52)	5.77 (0.52)	5.99e <sup>-7</sup>
Medulla	4.47 (0.43)	3.99** (0.44)	3.94** (0.43)	4.09* (0.43)	4.03* (0.44)	4.27 (0.42)	4.38 (0.42)	4.16 (0.42)	4.41 (0.42)	4.29 (0.42)	0.003
Cerebellum	103.10 (9.84)	99.95 (10.35)	102.24 (9.86)	102.24 (9.86)	94.88* (10.18)	99.70 (9.91)	101.47 (9.66)	98.33 (9.62)	101.77 (9.61)	103.72 (9.63)	1.96e <sup>-₄</sup>
Frontal Lobe	155.76 (11.12)	144.06** (11.78)	150.79 (11.14)	150.79 (11.14)	133.94** (11.50)	137.95** (11.28)	141.90* (10.92)	138.12** (10.88)	135.24** (10.87)	144.94* (10.89)	1.25e⁻⁵
Parietal Lobe	103.75 (9.22)	99.14* (10.38)	102.26 (9.23)	102.26 (9.23)	90.10** (9.53)	86.73** (9.94)	95.29 (9.05)	76.99** (9.02)	88.34** (9.01)	97.47* (9.02)	3.89e-⁵
Temporal Lobe	115.03 (10.12)	106.83** (10.67)	109.23 (10.13)	109.23 (10.13)	99.87** (10.46)	100.79** (10.22)	109.60 (9.93)	95.18** (9.89)	99.89* (9.88)	106.91* (9.90)	1.98e <sup>-₄</sup>
Occipital Lobe	47.30 (4.87)	45.58 (5.19)	46.49 (4.88)	46.49 (4.88)	41.97* (5.03)	43.34* (4.97)	48.29 (4.78)	40.28** (4.76)	42.09* (4.76)	42.63* (4.76)	0.003
Central Structures	41.56 (3.56)	36.33** (3.56)	36.35** (3.57)	36.35** (3.57)	35.01** (3.69)	36.54** (3.41)	37.07* (3.50)	37.42* (3.49)	36.96* (3.48)	38.81* (3.49)	1.42e⁵
Ventricles	35.80 (19.25)	48.76* (20.42)	39.56 (19.28)	39.56 (19.28)	57.01* (19.91)	59.83** (19.56)	44.93 (18.90)	60.81* (18.83)	75.75** (18.81)	48.53* (18.84)	2.25e⁻⁴

Table 2.6: Volumetric MRI measures in the PROSPECT study natural history cohort

Volumetric MRI measures in the PROSPECT study natural history cohort. Indeterminate: IDT, PSP-Richardson syndrome: PSP-RS, PSP-subcortical: PSP-SC, PSP-cortical: PSP-Cort, CBS-Alzheimer's disease: CBS-AD, CBS-4 repeat tau: CBS-4RT, CBS-unknown: CBS-NK. Marginal mean volumes in ml (and standard deviation). Final column p-value of the overall F-test from between subjects design including total intracranial volume, age at scan and sex. Other within-cell p-values are FDR adjusted. \*\* = P<0.001, \* = P<0.05 from t-test comparing each group to controls.



**Figure 2.4** – ROC curve analysis of serum NF-L differentiating between all PSP cases vs. controls.

	Controls vs PSP-All	Controls vs CBS-All	PSP-AII vs CBS-AII	PSP-RS vs PSP-Cort	PSP-RS vs PSP-SC	PSP-Cort vs PSP-SC	CBS-AD vs CBS-4RT
MoCA	0.84*	0.79	0.66	0.75	0.83*	0.80*	0.87*
ECAS	0.91*	0.83*	0.61	0.73	0.82*	0.80*	0.80*
ACE-III	0.94*	0.88*	0.64	0.76	0.81*	0.81*	0.80*
Serum NF-L	0.88*	0.91*	0.72	0.74	0.83*	0.75	0.76
Pons-Midbrain ratio	0.69	0.73	0.84*	0.71	0.58	0.60	0.66
Pons	0.76	0.92*	0.78	0.59	0.58	0.70	0.57
Midbrain	0.89*	0.91*	0.56	0.63	0.65	0.60	0.60
Medulla	0.78	0.82*	0.51	0.54	0.59	0.52	0.69
Cerebellum	0.59	0.79	0.55	0.73	0.54	0.80*	0.63
Frontal lobe	0.80*	0.71	0.75	0.76	0.70	0.89*	0.57
Parietal lobe	0.69	0.60	0.73	0.73	0.60	0.84*	0.69
Temporal lobe	0.79	0.67	0.77	0.65	0.60	0.75	0.51
Occipital lobe	0.59	0.63	0.61	0.72	0.52	0.78	0.54
Central structures	0.88*	0.81*	0.74	0.62	0.50	0.64	0.57
Ventricles	0.71	0.51	0.81*	0.60	0.68	0.79	0.66

Table 2.7: Cognitive, fluid biomarker and imaging AUC values from ROC curveanalyses

Cognitive, fluid biomarker and imaging receiver operating characteristic (ROC) curve analyses. Addenbrooke's cognitive examination 3: ACE-III, Edinburgh cognitive and behavioural ALS screen: ECAS, Montreal Cognitive Assessment: MoCA, PSP-Richardson syndrome: PSP-RS, PSP-subcortical: PSP-SC, PSP-cortical: PSP-Cort, CBS-Alzheimer's disease: CBS-AD, CBS-4 repeat tau: CBS-4RT, atypical parkinsonian syndromes: APS, cerebrospinal fluid: CSF, neurofilament light chain: NF-L. Cognitive scales and serum NF-L area under the curve (AUC) values based on logistic regression analyses that used sex, age at symptom onset and disease duration at testing as covariates. Imaging group comparisons adjusted for sex, age at scan and total intracranial volume. \* = AUC value≥0.80.

#### 2.5 Discussion

In this cohort study, I assessed a large number of patients with PSP and CBS recruited to the natural history arm of the PROSPECT study. Although this is not a community-based epidemiologic study, patients with APS were recruited across the United Kingdom. I characterised the different clinical presentations of PSP using the new MDS PSP diagnostic criteria. I identified disease- and subtype-specific markers that are likely to improve the early and accurate differentiation of PD from PSP and CBS and increase the power of future clinical trials with more homogeneous disease groups. I believe these findings should have a direct effect on the new era of anti-tau clinical trials that aim to recruit patients with earlystage PSP (18). The PROSPECT study's PSP-RS group data are consistent with data from patients with PSP-RS in the davunetide and tideglusib trials (16) (17) with regard to clinical (age at symptom onset and recruitment), genetic, and imaging profiles and the degree of motor and/or functional impairment at study recruitment. In addition, the baseline PSPRS and SEADL scores for patients with PSP-RS were consistent with those seen in patients with PSP-RS recruited to the 4RT neuroimaging initiative longitudinal cohort (72). Use of the 2017 MDS PSP criteria increased the number of clinical PSP cases from 58 to 101, implying that non-Richardson syndrome presentations (49 of 101 [48.5%]) are common. The subcortical presentations of PSP, consisting of PSP-P and PSP-PGF phenotypes, have a long delay to diagnosis that can at least in part be attributed to frequent initial misdiagnoses as PD, because they share similar clinical trajectories and initial clinical features. Although the present study was not adequately powered to detect significant differences between PSP and CBS subgroups and PD, I was able to detect trends of greater cognitive impairment and higher levels of serum NF-L in the PSP-subcortical group compared with PD. The PSPsubcortical group had a more benign clinical trajectory, less cognitive impairment, lower serum NF-L levels, higher TRIM11 rs564309 minor allele frequency, and more restricted midbrain and cortical atrophy than the PSP-RS and PSP-cortical groups. However, I identified midbrain atrophy to be a core neuroimaging feature of PSP across the different

subtypes, which may enable early separation from Lewy body PD. The finding of higher serum NF-L levels in the PSP-RS and PSP-cortical groups may indicate higher disease intensity or be a consequence of greater cortical atrophy seen in these phenotypes.

Pathologically proven CBS-AD and CBS-CBD may be difficult to distinguish in clinical practice **(49) (50)**. The advent of AD biomarkers is likely to improve this differentiation. I found that a biomarker-defined CBS-AD group had a milder clinical trajectory, greater ventriculomegaly, higher *APOE*-ɛ4 allele frequency, and greater cognitive impairment compared with the CBS-4RT group. In particular, the ACE-III revealed significant differences in attention, memory, and language sub-scores between CBS-AD and CBS-4RT. My data shows that despite the clinical overlap, CBS-AD can be distinguished from CBS-4RT in life. This finding is further supported by the fact that both of the CSF biomarker-defined CBS-4RT cases with post-mortem evaluation had CBD pathologic findings. Although all the major syndromes studied herein are bilateral brain diseases, CBS is typically asymmetrical, in contrast to PSP-RS and MSA. Such asymmetry can be quantified by a laterality index of motor features, but less so in terms of cognitive asymmetry. Prof Rowe therefore opted for a simple general linear model for MRI analysis without laterality. Further increases in the diagnostic accuracy of MRI may be gained in future studies by incorporating phenotypic data, including laterality effects, in the model.

To compare the discriminant usefulness of multimodal biomarkers, accommodating widely different scales and variances, I have also presented their performance as AUC values for cases vs controls and comparisons of disease groups. Using an AUC cut-off of at least 0.80 to represent high diagnostic accuracy, I confirmed the role of cognitive screening scales in differentiating CBS-AD from CBS-4RT and the role of cognitive screening scales, serum NF-L levels, and cortical volumetric MRI measures in differentiating PSP-subcortical from PSP-RS and PSP-cortical. In addition, serum NF-L level (AUC, 0.80) was able to accurately distinguish PD from a combined APS group consisting of all PSP and CBS cases. The

comparison of tables 2.4, 2.6 and 2.7 highlights that the utility of a biomarker to discriminate patient groups (such as the AUC) cannot simply be inferred from the significance of an unpaired *t* test between groups, especially where group sizes vary.

The 2017 MDS PSP diagnostic criteria were published during study recruitment, and so I am able to report the prospective characterization of variant PSP phenotypes using clinical, cognitive, fluid biomarker, genetic, and imaging measures with neuropathologic confirmation of diagnosis. In addition, I present a biomarker-defined CBS-AD group that has distinct clinical, cognitive, and genetic features that allow it to be distinguished from CBS-4RT. I found that as many as 50% of cases with PSP presented with non-classic variant PSP phenotypes, and in retrospective case series, this frequency has been shown to be as high as 76% (73). Until now, these PSP variants have been missed by clinical, therapeutic, and epidemiologic studies that have largely focused on the classic PSP-RS presentation. Similarly, as many as 50% of CBS cases with CSF analysis had a biomarker profile consistent with underlying AD pathologic features. Of note, my estimates are higher than those seen in a similar-sized retrospective case series with pathologic confirmation in which 5 of 21 CBS cases (23.8%) had primary pathologic AD findings at post-mortem (49). Alongside these phenotype-specific markers, my inclusion of data from a large PD cohort allowed me to confirm that the use of serum NF-L levels and cognitive screening scales may aid the early differentiation of PD from APS (41) (74).

I acknowledge limitations to the present study. First, most of the PROSPECT study cases were diagnosed using clinical criteria without neuropathologic verification. Cases with CBS-4RT and CBS-AD were defined using CSF biomarker criteria, and I acknowledge that in late life, AD biomarker positivity may be coincidental alongside CBD and does not prove that AD pathologic features are the primary cause of the clinical symptoms. I anticipate that follow-up of this natural history cohort, with further cases undergoing post-mortem assessment, will allow me to validate the sensitivity and specificity of the clinical and biomarker criteria used

to stratify patients. Although not currently available in this cohort, in-depth pathologic characterization of APS subtypes and associations with their ante-mortem biomarker profiles are informative. Previously, pathologic variants of PSP have been described (75). Of interest, that study found a higher density of cortical tau pathology in variants of PSP presenting with focal cortical syndromes compared with PSP-P and PSP-PGF, a finding that is in line with differences in cortical atrophy seen in PSP-cortical vs PSP-subcortical groups in my study. My clinical disease trajectory analyses were based on baseline clinical and functional rating scale scores. I believe longitudinal data from this cohort will be essential to accurately characterize the clinical progression of PSP and CBS and identify markers that predict and track progression. Although my AUC results are promising, as we gather more longitudinal data, I expect the diagnostic accuracy of PSP and CBS to further improve with a well powered multivariate approach, including cross-validated machine learning algorithms. Although a proportion of the IDT cases will eventually have non-APS diagnoses such as PD and vascular gait disorders, I expect that some cases will eventually fulfil diagnostic criteria for defined APS, representing cases that have been recruited at the very earliest disease stages.

My multimodal assessment of clinical, cognitive, fluid, genetic, and imaging data from the PROSPECT study has identified markers that enable the differentiation of PD from PSP and CBS. In addition, I present confirmatory data on the changes across modalities in classical phenotypes of PSP and CBS and evaluate biomarkers of variant PSP syndromes included in the most recent diagnostic criteria, and in a distinct biomarker-defined CBS-AD syndrome. These findings may enhance the early diagnosis of PSP and CBS for accurate prognostication and stratification of patients for clinical trials.

# <u>Chapter 3 – Blood + CSF neurofilament light chain and novel proximity extension</u> <u>assay markers as predictors of PSP pathology and survival</u>

#### 3.1 Abstract

The high degree of clinical overlap between atypical parkinsonian syndromes (APS) and Parkinson's disease (PD) makes diagnosis challenging. We aimed to identify novel diagnostic protein biomarkers of APS using multiplex proximity extension assay (PEA) testing.

Cerebrospinal fluid (CSF) samples from two independent cohorts, each consisting of APS and PD cases, and controls, were analysed for neurofilament light chain (NF-L) and Olink Neurology and Inflammation PEA biomarker panels. Whole-cohort comparisons of biomarker concentrations were made between APS (n=114), PD (n=37) and control (n=34) groups using logistic regression analyses that included gender, age and disease duration as covariates. We used baseline CSF and plasma NF-L values to predict survival in one of our cohorts that was followed up longitudinally over 5 years using a cox-proportional hazards survival model.

In our PEA marker APS versus controls analyses, 11 CSF markers had significantly different levels in cases and controls (p<0.002). Four of these markers also reached significance (p<0.05) in APS versus PD analyses. Disease-specific analyses revealed lower group levels of FGF-5, FGF-19 and SPOCK1 in multiple system atrophy compared with progressive supranuclear palsy and corticobasal syndrome. Receiver operating characteristic curve analyses suggested that the diagnostic accuracy of NF-L was superior to the significant PEA biomarkers in distinguishing APS, PD and controls. The biological processes regulated by the significant proteins include cell differentiation and immune cell migration. Delta and notch-like epidermal growth factor-related receptor (DNER) had the strongest effect size in APS versus controls and APS versus PD analyses. DNER is highly expressed in substantia

nigra and is an activator of the NOTCH1 pathway which has been implicated in the aetiology of other neurodegenerative disorders including Alzheimer's disease. Baseline levels of NF-L (highest tertile vs lowest tertile) from APS cases in the UCL biomarker cohort were able to predict survival over a 5 year period in both CSF (hazard ratio = 4.47, p = 0.001) and plasma (hazard ratio = 2.38, p = 0.03) measurements.

PEA testing has identified potential novel diagnostic biomarkers of APS. Additionally, baseline CSF and plasma NF-L levels were able to predict survival in APS, which has implications for the design of future clinical trials.

#### **3.2 Introduction**

Many cross-sectional CSF studies have assessed whether protein biomarkers can reliably differentiate disease groups from controls. In the case of Alzheimer's disease (AD), such studies have led to the discovery of biomarkers (t-tau, p-tau and A $\beta$ 1-42) which can be used as reliable and disease-specific diagnostic biomarkers, and provide insights into the biology of disease, i.e. low CSF levels of A $\beta$ 1-42 thought to reflect incorporation of A $\beta$ 1-42 in pathogenic amyloid plaques (**76**).

Similar to tau, neurofilaments are found in axons of neurons and are important components of the cytoskeleton. However, it appears that neurofilaments are highly expressed in large-calibre myelinated axons in white matter, while tau is predominantly expressed in thin unmyelinated axons of the cortex (39). NF-L has been shown to be a reliable differentiator of APS, PD and controls in CSF and plasma (40) (41). In neurodegenerative diseases, axonal degeneration releasing NF-L into the extracellular space is likely to be the reason behind raised CSF NF-L levels compared to controls (41). Although the exact mechanism remains unclear, it is thought that accompanying rises in plasma NF-L mostly originate in the CNS and subsequently cross the blood-brain barrier, although it should be noted that high levels of plasma NF-L have been shown in peripheral nerve disorders, suggesting peripheral as

well as central origins (41). Additionally, studies with small numbers of cases have shown that baseline measurements of plasma NF-L can predict the severity of subsequent clinical and radiological disease progression in PSP. However, there are limitations in the use of NF-L as a diagnostic and/or prognostic fluid biomarker: 1) Levels of NF-L do not differentiate between the individual atypical parkinsonian syndromes, i.e. PSP, CBS and MSA (40) (41); 2) Levels of NF-L are also raised in other neurodegenerative disorders such as FTD and HD (42) (43), suggesting that it is a non-specific marker of axonal degeneration, and does not offer insights into the dysfunctional biological pathways that lead to axonal degeneration in tauopathies specifically; 3) It is unknown as to whether baseline levels of NF-L predict survival in PSP and other atypical parkinsonian syndromes.

A majority of recent studies on NF-L have used ultrasensitive single molecule array (**36**). Proximity extension assay (PEA) technology, another ultrasensitive method of biomarker quantification, is a 96-plex immunoassay for high-throughput fluid protein biomarker detection, using unique antibody–oligonucleotide protein binding for quantitative real-time polymerase chain reaction (PCR)-based measurement (**38**). PEA biomarker measurement has previously been explored in neurological conditions such as traumatic brain injury (**77**). In addition, PEA, like single molecule array, appears to be less affected by the technical issues of multiplex enzyme-linked immunosorbent assay (ELISA) such as antibody crossreactivity, inter-assay variability and the need for larger sample volumes (**78**).

Here, my specific aims were: 1) to use PEA biomarker measurements to identify diagnostic markers of APS that differentiate them from PD and controls, and offer novel biological insights into these disorders using two independent cohorts. In particular, this study aims to explores the role of inflammation which has been identified as a key component in the pathogenesis of AD and FTD via microglial activation pathways in biomarker **(79)** and genetic **(80)** studies; 2) to discover PEA markers that may differentiate between tau (PSP and CBD) and synuclein (MSA and PD) disorders, which may aid in differential diagnosis

and offer novel biological insights into these disorders; 3) to assess whether baseline CSF and plasma levels of NF-L predict survival in APS in one of our cohorts which has been followed up for over five years.

#### 3.3 Methods

#### Standard protocol approvals, registrations and patient consents

Two prospective cohorts, each consisting of subjects with PSP, CBS, MSA and PD, and controls, were recruited and followed up longitudinally. Cohort 1 was recruited by Dr Nadia Magdalinou and Prof Andrew Lees at The National Hospital for Neurology and Neurosurgery, London, between 2012 and 2015 (Research Ethics Committee reference— 12/LO/0640) for a Queen Square biomarker study of patients with APS **(45)**. Cohort 2 was recruited by myself, Prof Huw Morris and other investigators between 2015 and 2017 as part of their involvement in the PROSPECT study (Research Ethics Committee reference 14/LO/1575), a longitudinal observational study of patients with APS in the UK **(81)**. Cohort 2 PD subject samples and clinical data were obtained from PD patients in the placebo arm of the Exenatide trial (trial approval by Brent NHS Research Ethics Committee, London) by Dr Dilan Athauda and Prof Tom Foltynie **(82)**. Patient consent for the trial covered the use of samples and clinical data in related studies such as this one. CSF sampling, quality control and storage protocols implemented in the trial were identical to the protocol outlined below.

Patients were assigned diagnoses according to current clinical diagnostic criteria. Of note, my application of the MDS PSP diagnostic criteria **(11)** identified probable PSP cases with clinical syndromes other than classical PSP-RS, such as PSP-P and PSP-PGF. Baseline PSPRS scores were obtained from patients with PSP and CBS on the same day of, but prior to, lumbar puncture (LP) testing. For each patient, I recorded the following clinical data: gender, age at motor symptom onset, date of motor symptom onset, age at the point of LP, disease duration at the point of LP, alive/deceased status of subjects at the point of

censoring (7 December 2018) and the total disease duration in subjects (defined as date of motor symptom onset to the date of death/date of censoring). In addition, I conducted a thorough review of current clinical notes for all patients to ensure that their clinical diagnosis had not changed. A subset of deceased patients underwent post-mortem examination at the Queen Square Brain Bank, London, for neuropathological confirmation of diagnosis.

#### LP/blood testing, sample handling and initial biomarker testing

Cases and controls underwent baseline LP testing. Cases and controls from cohort 1 also underwent venepuncture for blood testing at the same time as LP. CSF and blood samples were frozen and stored at  $-80^{\circ}$ C within 1 hour of sampling. Biomarker testing was performed on 0.5 mL aliquots and no aliquots had undergone interim freeze–thaw cycles. Blood contaminated CSF samples (>500 red blood cells/µL) were excluded. Prior to PEA testing, CSF and plasma levels of the following markers were obtained by Dr Nadia Magdalinou, Dr Tong Guo and Dr Amanda Heslegrave at the UCL Institute of Neurology using a separate aliquot of CSF/plasma: t-tau/p-tau/A $\beta$ 1–42 (INNOTEST ELISA—Fujirebio Europe N.V., Gent, Belgium) and NF-L (Simoa platform; Quanterix, Lexington, Massachusetts, USA). Samples from cohorts 1 and 2 were analysed on separate runs for the above biomarkers.

#### **PEA testing**

Biomarker panel testing was performed by the Olink biomarker team in Uppsala, Sweden, using multiplex PEA technology as previously described by Olink **(38)**. Samples were simultaneously run on two panels, (1) Neurology and (2) Inflammation, each consisting of 92 biomarkers, with 96 samples tested simultaneously on each run. The full list of Olink Neurology and Inflammation panel markers and Olink panel validation data are freely available online (https://www.olink.com/data-you-can-trust/validation/). I obtained the biological function of PEA markers of interest from the UniProt database (www.uniprot.org). Tissue RNA expression of PEA markers of interest was assessed using publicly available data on the GTEx database (www.gtexportal.org). The GTEx database consists of 8555

samples from 53 tissues (including 13 brain regions) of 544 donors for which RNAseq was conducted. The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by NCI, NHGRI, NHLBI, NIDA, NIMH and NINDS. I obtained the data used for the analyses described in this study from the GTEx Portal on 31 July 2018. Samples from cohorts 1 (CSF and plasma) and 2 (CSF alone) were analysed on separate runs. The resulting data for each biomarker were generated as a normalised protein expression (NPX) value. NPX is an arbitrary unit on a Log2 scale with data being normalised to minimise both intra-assay and inter-assay variation. I subsequently performed intensity normalisation (detailed description at

www.olink.com/content/uploads/2018/05/Data- normalization-and-standardization\_v1.0.pdf) of CSF NPX values across cohorts 1 and 2 to allow the combination of data sets to carry out whole-cohort analyses.

#### Statistical analyses

I carried out all of the following statistical analyses using Stata v.14.0. I generated figures using Stata v.14.0 and R v.3.3.2. Group comparisons of baseline clinical data (continuous variables) were carried out using t-tests, with statistical significance defined as p value <0.05.

My survival analyses were performed on APS cases in cohort 1 using both CSF and plasma NF-L values separately. By dividing patients into high, middle and low NF-L groups based on the tertiles of NF-L concentration, a Cox proportional hazards survival analysis (with associated Kaplan Meier survival curves) was used to assess the relationship between baseline NF-L concentration and the rate of survival, adjusting for sex, age at motor symptom onset and disease duration at the point of blood/LP testing. This was performed separately on: 1) PSP group alone; 2) PSP, CBS and MSA groups combined as one APS group.

All PEA biomarkers that had a >5% subject failure rate and all subjects that had a >5% marker failure rate were excluded from analyses. Group comparisons of intensity-normalised PEA marker data were carried out using two separate approaches. First, I combined subjects with PSP and CBS to form a 'tau' group, and subjects with MSA and PD combined to form a 'synuclein' group in order to carry out whole-cohort tau versus synuclein group comparisons with gender, age at the point of testing and disease duration at the point of testing as covariates. Second, subjects with PSP, CBS and MSA were combined to form one 'APS' group in order to carry out whole-cohort APS versus controls and APS versus PD group comparisons. Initially, whole-cohort APS versus controls analyses for all PEA markers and NF-L were carried out using logistic regression analyses, with gender and age at the point of testing as covariates. I then carried out whole-cohort APS versus PD group analyses for NF-L and all PEA markers that had reached significance in the whole-cohort APS versus controls analyses. For APS versus PD comparisons, I used logistic regression analyses, with gender, age at the point of testing and disease duration at the point of testing as covariates. I set the threshold for significant group differences in both analyses using the Benjamini-Hochberg correction method for multiple testing (83) with a false discovery rate of 5%. I calculated disease-specific association statistics for all PEA markers that reached significance in whole-cohort APS versus controls and APS versus PD analyses.

I carried out separate ROC curve analyses for whole-cohort APS versus controls, and APS versus PD group comparisons. In each analysis, ROC curves were generated for the following variables: (1) covariates; (2) CSF NF-L; (3) combined significant CSF PEA markers; (4) covariates + CSF NF-L + combined significant CSF PEA markers. In subjects with APS, the relationship between the levels of significant PEA markers and (1) NF-L levels (log-transformed) and (2) PSPRS scores was assessed using linear regression analyses, with gender, age at the point of testing and disease duration at the point of testing as covariates.

#### 3.4 Results

A total of 151 cases and 34 controls were recruited to the study across two independent cohorts (**Table 3.1**). All subjects had baseline CSF sampling for NF-L and PEA marker testing. All subjects from cohort 1 underwent paired blood sampling for plasma NF-L and PEA marker testing. Of note, I excluded eight patients with CBS who fulfilled criteria for a CSF profile that was indicative of underlying AD pathology—defined as a CSF tau:A $\beta$ 1–42 > 1 (63) (64). In addition, all PSP cases from cohort 1 fulfilled probable PSP-RS criteria while the breakdown of cohort 2 PSP cases fulfilling probable criteria were as follows: 15 PSP-RS; 4 PSP-P; 2 PSP-PGF.

			Cohort 1			Cohort 2					
	PSP	CBS	MSA	PD	HC	PSP	CBS	MSA	PD	HC	
	(n=33)	(n=11)	(n=29)	(n=25)	(n=30)	(n=21)	(n=11)	(n=9)	(n=12)	(n=4)	
Gender	48/52	57/43	55/45	64/36	50/50	81/19	18/82	89/11	67/33	25/75	
Male/Female											
(%)											
Age at motor	65.1*	64.5*	60.2*	55.8¶		63.6*	59.4*	61.0*	50.7¶		
symptom onset											
(years)-mean, SD	6.2	8.0	5.9	8.1		6.2	7.8	9.2	6.0		
Age at LP	69.6 <b>†</b>	68.4†	64 4 +	67 4 <b>†</b> ¶	59.8*	67.2*	64 1*	65.6*	59.9 <b>†</b> ¶	65.0*	
(years)-mean,	00.01	00.11	01	0	00.0	01.2	0	00.0	00.011	00.0	
ŚD	5.9	8.4	5.8	9.1	9.9	5.7	7.5	8.7	6.1	2.9	
Disease	4.5*	3.9*	4.2*	11.6		3.6*	4.7*	4.6*	9.2		
duration at LP											
(years)-mean, SD	2.5	1.2	2.2	6.1		2.0	1.9	2.6	3.1		
PSPRS score-	42.0¶	38.9				30.6¶	30.4				
mean, SD	44.5	15.0				<u> </u>	10.0				
	11.5	15.2	0004.0	000.0	000.0	9.1	10.0	4074.0	000.0	000.5	
CSF NF-L	2225.2	2268.6	2991.0	963.0	630.6	3228.5	2547.3	4671.6	980.3	868.5	
concentration	†^¶\$	T^	T^¶	Ť	Ŷ	T^¶	<u>†</u> ^\$	T^¶			
(IIg/L)=IIIeaII,									040 7	226 4	
30	913.4	1291.4	1462.2	565.8	278.9	1611.3	1255.7	2768.9	218.7	230.4	
% of subjects	88	64	86	48		24	0	0	0		
deceased at											
point of											
censoring											
Disease	7.2*	7.0*	7.1*	13.7		6.0	N/A	N/A	N/A		
duration in											
deceased	2.5	2.5	2.4	6.8		2.3					
group (years)-											
Number of	0	2	6	0		2	NI/A	NI/A	ΝΙ/Δ		
nathologically	9 PSP-0	CBD-2	MSA-5	U		PSP-2	IN/A	IN/A	IN/A		
confirmed	. 0. – 3	566-2	PSP=1			. 02					
cases											

Table 3.1: Summary of baseline clinical characteristics, CSF NF-L and survival data

CBS = corticobasal syndrome, HC = healthy controls, MSA = multiple system atrophy, PD = Parkinson's disease, PSP = progressive supranuclear palsy, PSPRS = PSP rating scale, SD = standard deviation. Only statistically significant (p-value <0.05) differences between continuous variables are noted from t-tests: \* = vs PD in same cohort;  $\dagger$  = vs HC in same cohort; \$ = vs MSA in same cohort;  $\P$  = vs same measure in other cohort.

Pearson correlation revealed strong correlation between CSF and plasma NF-L levels in the whole-cohort PSP group alone (r = 0.82, p < 0.0001) and the whole-cohort combined APS group (r = 0.70, p < 0.0001).

Survival analyses were restricted to cohort 1 which had a mean follow-up period of 5.1 years. By dividing patients in the PSP group into high, medium and low baseline NF-L groups for CSF and plasma separately, I found that the high NF-L group was more likely to die during the monitored time period vs the low NF-L group in CSF (hazard ratio = 11.04, p < 0.001) and plasma (hazard ratio = 7.39, p < 0.04), when adjusting for gender, age and disease duration at testing and baseline PSPRS score **(Table 3.2)**. By dividing patients in the APS group (PSP, CBS and MSA groups combined) into high, medium and low baseline NF-L groups for CSF and plasma separately, I found that patients in the high NF-L group were more likely to die during the monitored time period vs the low NF-L group in CSF (hazard ratio = 4.47, p = 0.001) and plasma (hazard ratio = 2.38, p = 0.03), when adjusting for gender, age and disease duration at testing **(Table 3.2)**. A similar result was obtained when the APS group was divided into high and low CSF NF-L groups based on the median CSF NF-L concentration **(Figure 3.1)**.

Table 3.2: Hazard ratios for survival in PSP and APS over 5 year follow up, stratified

	PSP grou	<u>ıp</u>	Combined APS	<u>group</u>	
Baseline CSF NF-L level*	Hazard ratio	<u>p-value</u>	Hazard ratio	<u>p-value</u>	
	<u>(95% CI)**</u>		<u>(95% CI)**</u>		
Lowest tertile	1 (ref)	-	1 (ref)	-	
Middle tertile	2.80 (0.48-16.29)	0.25	2.56 (1.10-5.94)	0.03	
Highest tertile	11.04 (1.96-62.02)	0.006	4.47 (1.90-10.51)	0.001	
Decelling places o ME L level*					
Baseline plasma NF-L level*	PSP grou	<u>ip</u>	Combined APS	<u>s group</u>	
Baseline plasma NF-L level*	Hazard ratio	ip p-value	Hazard ratio	<u>p-value</u>	
Baseline plasma NF-L level*	Hazard ratio (95% CI)**	<u>p-value</u>	<u>Combined APS</u> <u>Hazard ratio</u> (95% CI)**	<u>p-value</u>	
Lowest tertile	<u>Hazard ratio</u> (95% CI)** 1 (ref)	<u>p-value</u> -	<u>Combined APS</u> <u>Hazard ratio</u> (95% CI)** 1 (ref)	<u>p-value</u> -	
Lowest tertile	PSP grou           Hazard ratio           (95% CI)**           1 (ref)           11.18 (1.51-82.66)	<u>p-value</u> - 0.02	Combined APS           Hazard ratio           (95% CI)**           1 (ref)           1.95 (0.86-4.45)	<u>p-value</u> - 0.11	

#### by CSF and plasma NFL levels

\* NF-L group cut-off values for tertiles are cohort specific: PSP group CSF concentration ranges (ng/ml): Lowest third 531-1843; Middle third 1902-2405; Highest third 2540-4023. Combined APS group CSF concentration ranges (ng/ml): Lowest third 513-1902; Middle third 1914-2870; Highest third 3020-6223. PSP group plasma concentration ranges (pg/ml): Lowest third 11.7-26.9; Middle third 28.7-41.5; Highest third 42.1-77.4. Combined APS group plasma concentration ranges (pg/ml): Lowest third 11-7-31.6; Middle third 31.8-42.3; Highest third 43.8-77.6.

\*\* Hazard ratios in PSP group analysis are adjusted for gender, age at testing, disease duration at testing and disease severity according to baseline PSPRS. Hazard ratios in combined APS group analysis are corrected for gender, age at testing and disease duration at testing.



**Figure 3.1** – APS group Kaplan Meier survival curve according to high and low CSF NF-L groups based on the median CSF concentration (2320ng/ml), adjusted for gender, age and disease duration at testing.

CSF samples from all subjects underwent PEA marker testing for 184 markers (92 from the Neurology panel and 92 from the Inflammation panel). In total, 119/184 biomarkers (65 from the Neurology panel and 54 from the Inflammation panel) were detectable in the CSF of >95% of all subjects in both cohorts and were therefore used for further analyses. All subjects had a detectable result for >95% of the remaining 119 biomarkers. I carried out intensity normalisation across both cohorts to enable the combination of data sets to conduct whole-cohort analyses.

My initial analyses did not reveal any markers that reached statistical significance in differentiating between tau and synuclein groups (Figure 3.2).



Volcano plot of Tau vs Synuclein (n = 130 CSF biomarkers)

**Figure 3.2** – Volcano plot of CSF proximity extension assay biomarkers, highlighting no significant markers in the Tau vs Synuclein group analysis. Markers to the right of 0 on the x-axis were higher in the Tau group and markers to the left of 0 on the x-axis were higher in the Synuclein group. The threshold for p value significance (<0.002) was set using the Benjamini-Hochberg correction method for multiple testing with a false discovery rate of 5%.

I then combined PSP, CBS and MSA groups into one APS group. The ability of each biomarker to differentiate between the APS group and controls was assessed by carrying out individual logistic regression analyses that used gender and age at the point of testing as covariates. The resulting coefficient values, used as markers of group fold change, and p values were used to construct a volcano plot to highlight significant markers **(Figure 3.3)**. I identified 11 markers that reached statistical significance.



**Figure 3.3** – Volcano plot of CSF proximity extension assay biomarkers, highlighting significant markers that differentiated the APS group from controls. Markers to the right of 0 on the x-axis were higher in the APS group and markers to the left of 0 on the x-axis were higher in controls. The threshold for p value significance (<0.002) was set using the Benjamini-Hochberg correction method for multiple testing with a false discovery rate of 5%.

I took forward the 11 significant markers from the whole-cohort APS versus controls analysis and assessed their ability to differentiate the APS group from the PD group by carrying out logistic regression analyses that used gender, age at the point of testing and disease duration at the point of testing as covariates **(Figure 3.4)**. I identified four markers with reduced CSF concentrations in APS cases as compared with PD cases, which were also reduced in APS versus controls. Although the remaining seven markers followed the same trends as in the APS versus controls analysis, these group differences did not reach statistical significance. I did not detect any statistically significant (p<0.05) APS versus controls or APS versus PD differences using cohort 1 plasma data when carrying out logistic regression analyses using the same covariates as in the CSF analyses above.





significant markers that differentiated the APS group from the PD group. Markers to the right

of 0 on the x-axis were higher in the APS group and markers to the left of 0 on the x-axis were higher in the PD group. The threshold for p value significance (<0.05) was set using the Benjamini-Hochberg correction method for multiple testing with a false discovery rate of 5%.

I assessed for disease-specific differences between PSP, CBS, MSA, PD and control groups in the significant PEA markers by carrying out whole-cohort logistic regression analyses using gender, age at testing and disease duration at testing as covariates. Although I found significantly different levels of our identified markers between controls and each of the PSP, CBS and MSA groups, I found that the APS versus PD signals for FGF-5, FGF-19 and DNER were primarily being driven by lower levels in the MSA group. In addition, I found significantly lower group levels of FGF-5, FGF-19 and SPOCK1 in MSA compared with PSP and CBS (Table 3.3). The biological function and tissue expression of the significant PEA markers are summarised below (Table 3.4).

# Table 3.3: Disease-specific summary statistics for our significant PEA CSF biomarkers

CSF Biomarker	PSP group (n=54) mean [NPX]	CBS group (n=22) mean [NPX]	MSA group (n=38) mean [NPX]	PD group (n=37) mean [NPX]	HC group (n=34) mean [NPX]
FGE-5	3 26+¥	3 28+¥	2 93*+\$¶	3.60¥	3.52\$¶¥
1013	0.54	0.52	0.58	0.53	0.46
MSR1	2.11†	2.08†	2.02†	1.95†	1.58*\$¶¥
	0.40	0.39	0.48	0.50	0.48
VWC2	4.78†	4.93	4.55*†	5.13¥	5.06\$¥
	0.49	0.55	0.54	0.51	0.51
VEGF-A	9.24†	9.34	8.98*†	9.61¥	9.48\$¥
	0.46	0.54	0.53	0.46	0.55
ADAM22	7.64*†	7.67	7.56*†	7.84\$¥	7.78\$¥
	0.29	0.24	0.26	0.25	0.26
DNER	10.02†	10.00†	9.96*†	10.11¥	10.08\$¶¥
	0.15	0.17	0.16	0.14	0.13
UNC5C	2.14†	2.26	2.09*†	2.43¥	2.46\$¥
	0.51	0.51	0.51	0.43	0.47
ADAM23	3.00†	3.04	2.89*†	3.21¥	3.17\$¥
	0.30	0.34	0.33	0.26	0.28
SPOCK1	7.15*†¥	7.12¥	6.94*†\$¶	7.36\$¥	7.23\$¥
	0.29	0.37	0.32	0.29	0.29
N2DL-2	4.13†	3.92*†	3.90*†	4.60¶¥	4.52\$¶¥
	0.72	0.71	0.85	0.74	0.86
FGF-19	4.53†¥	4.70¥	4.23*†\$¶	4.90¥	4.79\$¥
	0.55	0.53	0.59	0.51	0.54

CBS = corticobasal syndrome, HC = healthy controls, MSA = multiple system atrophy, PD =

Parkinson's disease, PSP = progressive supranuclear palsy; NPX = normalised protein expression value; SD = standard deviation. Only statistically significant (p-value <0.05) differences are noted from logistic regression analyses: \* = vs PD;  $\dagger$  = vs HC; \$ = vs PSP;  $\P$  = vs CBS; ¥ = vs MSA.

I carried out separate whole-cohort APS versus controls and APS versus PD ROC curve analyses to assess the diagnostic strength of the significant PEA markers in comparison with NF-L (Figure 3.5). Using whole-cohort APS group data, I did not detect any statistically significant relationships between the CSF levels of the significant PEA markers and both logtransformed CSF NF-L levels and PSP rating scale scores by conducting linear regression analyses using gender, age at testing and disease duration at testing as covariates.

## Table 3.4: Biological function (UniProt database) and tissue expression (GTEx

## database) of significant Olink panel markers

CSF Biomarker	Olink panel	Associated	Biological function	Tissue with	Brain region
		Gene (Chromosomal location)		highest expression (median TPM)	with highest expression (median TPM)
FGF-5 (Fibroblast growth factor 5)	Inflammation	FGF5 (4q21.21)	Regulation of cell proliferation and cell differentiation	Fibroblasts (18.9)	Cerebellum (6.0)
MSR1 (Macrophage scavenger receptor 1)	Neurology	MSR1 (8p22.1)	A membrane glycoprotein implicated in the pathologic deposition of cholesterol in arterial walls during atherogenesis	Lung (33.2)	Hypothalamus (1.0)
VWC2 (von Willebrand factor C domain- containing protein 2)	Neurology	VWC2 (7p12.2)	Bone morphogenetic protein antagonist which may play a role in neural development	Cerebellum (24.4)	Cerebellum (24.4)
VEGF-A (Vascular endothelial growth factor A)	Inflammation	VEGFA (6p21.1)	Growth factor active in angiogenesis, vasculogenesis and endothelial cell growth	Thyroid (613.1)	Cerebellum (39.7)
ADAM22 (Disintegrin and metalloproteinase domain-containing protein 22)	Neurology	ADAM22 (7p21.12)	Regulation of cell adhesion and inhibition of cell proliferation	Cerebellum (87.4)	Cerebellum (87.4)
DNER (Delta and Notch- like epidermal growth factor- related receptor)	Inflammation	DNER (2q36.3)	Activator of the NOTCH1 pathway. May mediate neuron- glia interaction during astrocytogenesis	Substantia nigra (124.9)	Substantia nigra (124.9)
UNC5C (Netrin receptor)	Neurology	UNC5C (4q22.3)	Mediates axon repulsion of neuronal growth cones in the developing nervous system	Thyroid (11.7)	Cervical cord (8.2)
ADAM23 (Disintegrin and metalloproteinase domain-containing protein 23)	Neurology	ADAM23 (2q33.3)	May play a role in cell- cell and cell-matrix interactions	Frontal cortex (45.5)	Frontal cortex (45.5)
SPOCK1 (Testican 1)	Neurology	SPOCK1 (5q31.2)	May play a role in cell- cell and cell-matrix interactions	Cerebellum (314.3)	Cerebellum (314.3)
N2DL-2 (UL16-binding protein 2)	Neurology	ULBP2 (6q25.1)	Binds and activates the KLRK1/NKG2D receptor, mediating natural killer cytotoxicity	Fibroblasts (9.6)	Cerebellum (5.0)
FGF-19 (Fibroblast growth factor 19)	Inflammation	FGF19 (11q13.3)	Involved in the suppression of bile acid biosynthesis through downregulation of	Testis (0.5)	Cerebellum (0.4)

TPM = Transcripts per Kilobase Million.



**Figure 3.5** – ROC curve analyses of: (A) APS vs controls; (B) APS vs PD.

Combined plot = significant CSF Olink proximity extension assay markers + CSF NF-L + covariates. Covariates in APS vs controls analysis = age at testing and gender. Covariates in APS vs PD analysis = age at testing, disease duration at testing and gender.

#### 3.5 Discussion

This study highlights the application of high-throughput multiplex PEA testing to reveal novel biological insights into atypical parkinsonian disorders. The statistical methods used in the APS versus controls and APS versus PD analyses provides robust evidence for the diagnostic markers that I have identified.

The baseline clinical characteristics, CSF NF-L data and pathological diagnosis data suggest a high level of diagnostic accuracy in the two cohorts which enabled a combined wholecohort study. My thorough current clinical notes review for each case ensured that patients had not subsequently developed clinical features suggestive of alternative diagnoses. In particular, none of the cohort 1 PD group cases had evidence of oculomotor dysfunction suggesting inadvertent inclusion of PSP-P cases. The mean disease duration at the point of diagnosis in this group was 11.6 years, and at this stage I would certainly expect to see oculomotor dysfunction in patients with PSP-P (73). Similarly, patients with PSP-P included in the cohort 2 PSP group fulfilled 'probable' MDS PSP diagnostic criteria (11). This means that all cases would have had the presence of slowed vertical saccades and/or a vertical supranuclear gaze palsy, both of which are highly suggestive of underlying PSP pathology, such that I do not believe this group inadvertently contained PD cases. In addition, I intensity-normalised my PEA biomarker data across both cohorts to reduce inter-run variability, allowing me to combine datasets for better powered whole-cohort analyses.

Although my initial approach of creating tau and synuclein groups led to more pathologically homogeneous group comparisons, I did not identify any markers that were significantly different between the two groups. However, my subsequent whole-cohort analyses identified 11 CSF biomarkers that were reduced in APS as compared to controls. Four of these markers (FGF-5, FGF-19, DNER and N2DL-2) also differentiated APS from PD, with levels of the remaining seven markers not reaching statistical significance but following the same trend as in the APS versus controls analyses. Despite having heterogeneous pathology within the combined APS group, it is possible that the markers reaching significance in the above analyses are, like NF-L, non-specific markers of more rapid rates of neurodegeneration seen in PSP, CBS and MSA in comparison with PD. However, it is more likely that we were underpowered to detect pathology-specific differences in biomarker concentrations when carrying out tau (PSP and CBS) versus synuclein (MSA and PD) analyses. This is suggested by the fact that I found significantly lower group levels of FGF-5, FGF-19 and SPOCK1 in MSA compared with PSP and CBS groups. It remains premature to suggest that these are disease-specific markers until further replication data in larger cohorts is obtained.

All but one of the significant PEA markers had lower levels in the APS group compared with both PD and controls. Although this trend draws parallels with the observation of lower CSF levels of A $\beta$ 1–42 in patients with AD compared with controls, thought to reflect the incorporation of A $\beta$ 1–42 into amyloid plaques **(76)**, it is unclear as to why this is the case in the identified markers from this study. The biological processes regulated by the significant
PEA markers include cell proliferation/differentiation, cell apoptosis, immune cell migration and neural development. Of particular interest, DNER had the strongest effect size (coefficient) of all of the markers in both APS versus controls and APS versus PD analyses. DNER is highly expressed in substantia nigra, is an activator of the NOTCH1 pathway which has a role in neuronal and glial cell differentiation, and has previously been implicated in the aetiology of AD **(84)**.

APS versus controls and APS versus PD ROC curve analyses revealed that the individual area under the curve values for CSF NF-L alone were superior to the combination of the significant PEA markers.

There is strong evidence that levels of NF-L can track and predict the rate of disease progression in PSP (31). The lack of a linear relationship between the levels of the significant PEA markers and both the level of NF-L and PSPRS scores suggest that these markers are unlikely to have prognostic value. However, this needs to be explored further with longitudinal biomarker measurements to assess the temporal pattern of these markers in relation to changing clinical rating scale scores.

In summary, I present promising findings using PEA biomarker technology to discover novel diagnostic markers of APS. Although outside of the scope of this study, follow-up work includes replication of my findings in larger cohorts of subjects with APS which, in turn, may lead to the discovery of disease-specific and pathology-specific markers. A similar approach using phenotype group comparisons such as PSP-RS versus PSP-P/PSP-PGF would also be of interest. In addition, I would aim to validate PEA as a reliable multiplex technique by comparing the levels of markers measured by PEA, single-molecule array and ELISA.

# <u>Chapter 4 – The genetic and clinico-pathological profile of early-onset progressive</u> <u>supranuclear palsy</u>

# 4.1 Abstract

Studies on early-onset presentations of progressive supranuclear palsy (PSP) have been limited to those where a rare monogenic cause has been identified. Here, I have defined early-onset PSP (EOPSP) and investigated its genetic and clinico-pathological profile in comparison with late-onset PSP (LOPSP) and Parkinson's disease (PD).

I included subjects from the Queen Square Brain Bank, PROSPECT study, and Tracking Parkinson's study. Group comparisons of data were made using Welch's t-test and Kruskal-Wallis analysis of variance. EOPSP was defined as the youngest decile of motor age at onset (≤55 years) in the Queen Square Brain Bank PSP case series.

I studied 33 EOPSP, 328 LOPSP, and 2000 PD subjects. The early clinical features of EOPSP usually involve limb parkinsonism and gait freezing, with 50% of cases initially misdiagnosed as having PD. I found that an initial clinical diagnosis of EOPSP had lower diagnostic sensitivity (33%) and positive predictive value (38%) in comparison with LOPSP (80% and 76%) using a post-mortem diagnosis of PSP as the gold standard. Frequently pathologically confirmed EOPSP was initially misdiagnosed as PD, and conversely other pathologically diagnosed conditions were misdiagnosed as EOPSP. 3/33 (9%) of the EOPSP group had an underlying monogenic cause. Using a PSP genetic risk score (GRS), I showed that the genetic risk burden in the EOPSP (mean z-score, 0.59) and LOPSP (mean z-score, -0.08).

The initial clinical profile of EOPSP is often PD-like. At the group level, a PSP GRS was able to differentiate EOPSP from PD, and this may be helpful in future diagnostic algorithms.

#### 4.2 Introduction

Numerous cohort studies of PSP patients have shown that the mean age of symptom onset is approximately 65-67 years across the phenotypic spectrum **(7) (73) (81)**. Studies involving younger patients with a PSP syndrome have focused on familial cases with an identified single gene mutation despite the fact that PSP is considered to be a sporadic disease. Studies of rare autosomal-dominant familial PSP phenotypes (which represent <1% of all PSP cases) with variable disease durations and PSP-type tau pathology at post-mortem examination have identified *MAPT* and *LRRK2* mutations **(19)**. Although MAPT mutations are a common Mendelian cause of PSP phenotypes, it should be noted that there are other monogenic early-onset PSP mimics including Perry syndrome (*DCTN1* mutations) and Niemann-Pick type C (*NPC1/NPC2*) mutations **(22)**.

However, in my experience of seeing PSP patients, there are a portion of patients with a significantly younger age at onset in whom a known monogenic cause is not identified. One possibility is that these patients have a novel monogenic cause of PSP that has yet to be identified. The detection of such novel genes, although outside the scope of this PhD, is possible through the study of large families with PSP and/or parkinsonism. Another possibility is that a younger age of onset is driven primarily by a higher burden of genetic risk variants that have been identified in previous case-control GWAS' **(25) (27)**.

The aims of this study were: 1) to define early-onset PSP (EOPSP); 2) to characterise the clinical profile, diagnostic accuracy and clinical progression of EOPSP in comparison to late-onset PSP (LOPSP) and Parkinson's disease (PD); 3) to identify potential determinants of early-onset disease, including infection, head injury and the burden of genetic risk variants measured via a genetic risk score.

#### 4.3 Methods

#### Patient consent

All patients gave written informed consent for the use of their medical records and brain tissue/blood samples for research purposes, including the analysis of DNA. I identified all subjects, regardless of their pathological diagnosis, who had an ante-mortem clinical diagnosis of PSP at any point in their disease course from the Queen Square Brain Bank (QSBB), with the year of death ranging from 2000 to 2018. The brain donor program was approved by a London multicentre research ethics committee, and tissue is stored for research under a license from the Human Tissue Authority, No. 12198. To enable the calculation of diagnostic sensitivity and positive predictive value (PPV), I later included patients with a post-mortem diagnosis of PSP where PSP was neither the initial or final clinical diagnosis.

Progressive Supranuclear Palsy-Corticobasal Syndrome-Multiple System Atrophy (PROSPECT) is a U.K.-wide longitudinal study of patients with atypical parkinsonian syndromes, including PSP (Queen Square Research Ethics Committee 14/LO/1575), led by Prof Huw Morris (University College London). I identified subjects with a baseline clinical diagnosis of PSP from the PROSPECT study, with the year of recruitment ranging from 2015 to 2018 **(81)**.

Tracking Parkinson's is a U.K.-wide longitudinal study of Parkinson's disease (PD) across 72 sites, with multicentre ethics committee and local research and development department approvals, led by Prof Donald Grosset (University of Glasgow). I identified subjects with a baseline clinical diagnosis of PD, with the year of recruitment ranging from 2012 to 2014 **(52)**. PD cases were diagnosed consistent with QSBB clinical diagnostic criteria **(53)**.

#### Defining EOPSP

I defined EOPSP in patients with a clinical diagnosis of PSP, consistent with MDS clinical diagnostic criteria (11), and a motor symptom onset ≤55 years of age. This threshold was used as it represented the youngest decile of age at motor symptom onset in the QSBB series of pathologically diagnosed PSP cases. Late-onset PSP (LOPSP) was defined as cases with a clinical diagnosis of PSP and a motor symptom onset >55 years of age.

#### **Clinical Data Collection and Phenotyping**

I recorded the following clinical features for each case: sex, ethnicity, family history of dementia and/or parkinsonism in first-degree relatives, age at motor symptom onset, initial clinical diagnosis and PSP phenotype, final/current clinical diagnosis and PSP phenotype, diagnostic latency (from motor symptom onset to correct diagnosis), and disease duration (from motor symptom onset to death) in deceased cases. Of note, the initial clinical diagnosis/phenotype was defined as the clinical diagnosis/MDS criteria PSP phenotype given to patients in the first 3 years after their motor symptom onset. Final/current clinical diagnosis/phenotype was defined as the clinical diagnosis/MDS criteria PSP phenotype given to patients at least 2 years after the date of their initial clinical diagnosis/phenotype. Although I emphasise the age at motor symptom onset to define EOPSP/LOPSP, I screened each case for the onset and burden of cognitive symptoms relative to motor symptom onset to identify patients with frontal presentations of PSP.

In conjunction with the phenotyping methods described above, I used the presence or absence of MDS PSP criteria clinical features (11) in the initial and final/current disease stages to produce radar charts to further highlight the phenotypic differences between EOPSP, LOPSP, and PD. PROSPECT subjects had serial PSP rating scale (PSPRS) scores recorded, and both PROSPECT and Tracking Parkinson's subjects had serial MDS-UPDRS part III scores recorded, to assess the rates of clinical disease progression. To compare the rates of encephalitis and head injury between the EOPSP and LOPSP cases, I

collected data on the presence or absence of a documented past medical history of encephalitis in case notes (QSBB cases only) and the presence or absence of any mode of head injury resulting in loss of consciousness prior to the onset of PSP motor symptoms, identified from the Retrospective Screening of Traumatic Brain Injury (RESTBI) questionnaire (PROSPECT cases only).

#### Neuropathological Diagnosis

The neuropathological examinations of EOPSP and LOPSP cases in this study were carried out at QSBB by Prof Janice Holton and Prof Tamas Revesz. The pathological diagnoses of these cases were used to calculate the diagnostic sensitivity and PPV of EOPSP and LOPSP using a pathological diagnosis of PSP as the gold standard. Neuropathological data from Tracking Parkinson's PD cases were not available for analysis.

# Genotyping

PSP and PD cases had DNA extracted from either brain tissue or blood. For PROSPECT study cases, I was responsible for carrying out DNA sample quality control and preparation for genotyping by doing gel electrophoresis and Qubit fluorometry at the UCL Institute of Neurology. I subsequent delivered plates of DNA samples to the UCL Institute of Child Health where they underwent genotyping using the Illumina NeuroChip **(65)**. Tracking Parkinson's cases were genotyped at Cardiff University under the supervision of Prof Nigel Williams using the Illumina Human Core Exome array **(66)**. Once both datasets were merged, I carried out standard steps for data quality control including sex checking, heterozygosity, identity by descent to exclude related individuals and a principal components analysis (PCA) merged with the HapMap reference dataset to exclude all cases of non-European ancestry. I then carried out SNP imputation using the Haplotype Reference Consortium v1.1 panel (https://imputation.sanger.ac.uk/). I used the following post-imputation data quality control steps to filter out SNPs with: INFO score <0.70, posterior probability <0.90, genetic missingness >0.05, Hardy-Weinberg equilibrium p-value <1.0x10<sup>-7</sup>

and minor allele frequency <0.01. I screened all cases for known pathogenic *MAPT*, *LRRK*2, *PRKN*, *PINK1*, *SNCA*, *GRN*, and *DCTN1* mutations, which are directly genotyped on both the Illumina NeuroChip and Illumina Human Core Exome Array **(65) (66)**. I extracted *MAPT* H1/H1 frequency (determined by rs1800547 genotype), *TRIM11* minor allele frequency (determined by rs564309 genotype), and *APOE*  $\epsilon$ 4 allele frequency (determined by rs429358 and rs7412 genotypes) for all cases using the --recodeA command in Plink, version 1.9. In addition, one biochemically proven case of Niemann-Pick type C disease had targeted sequencing to identify pathogenic mutations in the *NPC1* and *NPC2* genes by the neurogenetics lab team at the National Hospital for Neurology and Neurosurgery.

#### **Statistical Analyses**

I carried out all of the following statistical analyses using Stata version 15 (StataCorp. 2017. Stata Statistical Software: Release 15. College Station, TX: StataCorp LLC) and Plink version 1.9 (Harvard University, Cambridge, MA). I generated figures using R version 3.3.2 (R Foundation for Statistical Computing, Vienna, Austria). Statistical significance was defined as P < 0.05.

I studied the pathological diagnoses and/or genetic mutation status of the EOPSP, LOPSP, and PD cases. Cases with alternative pathological and/or genetic diagnoses were excluded from subsequent analyses of clinical and genetic data. Group comparisons of clinical features were made using Welch's t-test.

Clinically diagnosed PSP and PD cases with an alternative current clinical diagnosis were excluded from genetic analyses. I calculated a PSP genetic risk score (GRS), based on weighted odds ratios for all risk loci (*MAPT* H1 haplotype and H1c subhaplotype, *MOBP*, *STX6*, and *EIF2AK3*) identified in the original PSP case-control genome-wide association study **(25)**, for all PSP and PD cases of European ancestry (confirmed by principal component analysis) using the --score function. Group comparisons of PSP GRS z-scores

were made using Kruskal-Wallis analysis of variance. I used t-tests to look for clinical and genetic differences between QSBB and PROSPECT cases within our EOPSP and LOPSP groups.

I collected the following data from QSBB cases to compare the diagnostic sensitivity and PPV of initial and final clinical diagnoses of EOPSP and LOPSP in the cohort: (1) the primary pathological diagnosis of all cases with initial and/or final clinical diagnoses of EOPSP/LOPSP; (2) the initial and final clinical diagnoses of all cases with a primary pathological diagnosis of PSP, which includes cases that never had an ante-mortem diagnosis of PSP.

### Variable Age at Onset to Define EOPSP

Alongside my arbitrary age at onset cut-off point (youngest 10% of QSBB PSP series =  $\leq$ 55 years) to define EOPSP, I assessed the impact of changing the age at onset cut-off point on the clinical profile, PSP GRS z-score, and diagnostic sensitivity/PPV of EOPSP in comparison with LOPSP. Specifically, the alternative age at onset cut-off points studied were as follows: (1) youngest 5% of QSBB PSP series =  $\leq$ 52 years; (2) youngest 15% of QSBB PSP series =  $\leq$ 62 years; (4) youngest 25% of QSBB PSP series =  $\leq$ 64 years. The definition of LOPSP varied with each alternative age at onset cut-off point accordingly.

# 4.4 Results

I identified 361 subjects with a clinical diagnosis of PSP at any point in their disease course from the QSBB and PROSPECT study with detailed clinical data available throughout the entire disease course. Of the PSP cases, 33/361 (9%) fulfilled criteria for EOPSP. In addition, 2000 PD cases from the Tracking Parkinson's study were included. The neuropathological and genetic mutation status of these groups are summarised in **Table 4.1**.

All clinically diagnosed PSP subjects with alternative pathological diagnoses were excluded from subsequent analyses. 30% of subjects with clinically diagnosed EOPSP who came to post-mortem had an alternative pathological diagnosis (Table 4.1). In addition, in the EOPSP group, two cases had *MAPT* mutations that have previously been reported in subjects with and clinical and/or pathological diagnosis of PSP (85) (86). One EOPSP case had previously described pathogenic *NPC1* mutations (87) and was therefore excluded from subsequent analyses. A total of 21 PD cases had *LRRK2*, *PRKN*, and *SNCA* mutations previously reported in subjects with PD pathology (88).

# Table 4.1: Neuropathological and genetic mutation status of clinically diagnosed PSP and PD cases

	Clinically diagnosed EOPSP	Clinically Diagnosed LOPSP	Clinically diagnosed PD
No. of	33	328	2000
subjects			
% of post-mortem			
cohort with a	14/20 (70%)	129/158 (82%)	
of PSP			
% of post-mortem	AD – 2/20 (10%)	CBD – 10/158 (6%)	
cohort with alternative	PD – 2/20 (10%)	PD – 5/158 (3%)	
pathological	CBD – 1/20 (5%)	AD – 4/158 (3%)	
diagnoses	MSA – 1/20 (5%)	MSA – 4/158 (3%)	
		ALS – 2/158 (1%)	
		FTLD – 2/158 (1%)	
		AGD – 2/158 (1%)	
% of cohort with a	3/33 (9%)	0/328 (0%)	21/1566 (1%)
pathogenic genetic mutation			
Pathogenic genetic mutations	MAPT IVS10+16 (1)	None	<i>LRRK</i> 2 G2019S (17)
(number of cases)	<i>MAPT</i> L284R (1)		LRRK2 R1441C
	NPC1 heterozygous c.1844G>T p.(Arg615Leu) and NPC1 heterozygous c.3182T>C p.(Ile1061Thr) (1)		PRKN p.P113Xfs/PARK2 exon 5 hemizygous deletion (compound heterozygous) (1) SNCA heterozygous duplication of exons 1- 6 (1)

EOPSP = early onset PSP, LOPSP = late onset PSP, PD = Parkinson's disease, QSBB = Queen Square brain bank. AD = Alzheimer's disease. CBD = Corticobasal degeneration. MSA = Multiple system atrophy. ALS = Amyotrophic lateral sclerosis. FTLD = Frontotemporal lobar degeneration. AGD = Argyrophilic grain disease. *MAPT* = Microtubule associated protein tau. *NPC1* = Niemann-Pick type C1. *LRRK2* = Leucine-rich repeat kinase 2. *PRKN/PARK2* = Parkin. *SNCA* = Alpha-synuclein.

The clinical profiles of the groups are summarised in **Table 4.2**. The initial clinical profile of EOPSP was more PD-like when compared with LOPSP while the final/current clinical profiles of EOPSP and LOPSP both resembled PSP-RS (**Figure 4.1**). Deceased EOPSP cases had a longer disease duration than deceased LOPSP and PD cases (**Table 4.2**). However, only 5% of the Tracking Parkinson's study cohort were deceased at the point of censoring, so a majority of these deceased PD cases are likely to be atypical fast progressing cases, with no neuropathological data available. In comparison with an agematched (motor symptom onset  $\leq$ 55 years) cohort of PD cases from the Tracking Parkinson's study (n = 328, mean age at onset = 48.5 years), the EOPSP group had a significantly longer mean diagnostic latency (3.2 vs. 2.5 years; P < 0.05).



**Figure 4.1** – Initial (A) and final/current (B) clinical profile of EOPSP, LOPSP and PD. Radar charts comparing the percentage (%) of EOPSP, LOPSP and PD cases with MDS PSP diagnostic criteria clinical features in early (A) and late (B) stages of disease. EOPSP = early onset PSP, LOPSP = late onset PSP, PD = Parkinson's disease.

# Table 4.2: Clinical profile of EOPSP, LOPSP and PD study cohorts

	EOPSP	LOPSP	PD
No. of subjects	26	299	2000
% male	69%	62%	65%
Ethnicity (% of cases)	CEU (92%) Non-CEU (8%)	CEU (92%) Non-CEU (8%)	CEU (97%) Non-CEU (3%)
Family history of dementia and / or parkinsonism (% of cases)	27%	15%	22%
Age at motor symptom	51.0* <b>†\$</b>	68.1 <b>\$</b>	64.4
onset (years) – mean, (SD), range	(4.8) 40 – 55	(6.3) 56 – 88	(9.7) 23 – 90
Initial clinical diagnoses (% of cases)	PD (50%) PSP (31%) VascP (7%) Dementia (4%) CBS (4%) ET (4%)	PSP (80%) PD (10%) CBS (7%) FTD (2%) NPH (1%)	PD (100%)
Initial PSP clinical subtype (% of cases)	s.o. PSP-P (31%) prob. PSP-RS (27%) poss. PSP-PGF (19%) prob. PSP-P (15%) s.o. PSP-F (4%) s.o. PSP-CBS (4%)	prob. PSP-RS (64%) prob. PSP-P (12%) poss. PSP-PGF (8%) s.o. PSP-CBS (7%) s.o. PSP-P (6%) s.o. PSP-F (3%)	
Final / current clinical diagnoses (% of cases)	PSP (96%) CBS (4%)	PSP (91%) CBS (5%) MSA (1%) FTD (1%) LBD (1%) APS (1%)	PD (98%) SWEDD (0.4%) MSA (0.3%) PSP (0.2%) Other <b>‡</b> (1.1%)
Final / current PSP clinical subtype (% of cases)	prob. PSP-RS (84%) prob. PSP-P (8%) prob. PSP-PGF (4%) s.o. PSP-CBS (4%)	prob. PSP-RS (84%) s.o. PSP-CBS (5%) prob. PSP-P (4%) poss. PSP-CBS (3%) prob. PSP-PGF (2%) prob. PSP-F (2%)	
Diagnostic latency (years) – mean, SD	3.2* <b>†\$</b>	2.2\$	1.8
	(1.5)	(1.3)	(2.8)
% of cohort deceased	69%	56%	5%
Disease duration in deceased subjects (years) – mean, (SD), range	10.5* <b>†\$</b> (3.9) 4.4 – 15.2	6.2 <b>\$</b> (2.6) 2.4 – 15.9	6.0 (5.0) 2.3 – 45.2

The study cohort were pathologically confirmed, where available. Group comparisons made using Welch's t-test – \* = p < 0.05 vs PD,  $\dagger$  = p < 0.05 vs LOPSP, \$ = No significant intra-group difference between QSBB and PROSPECT cases,  $\ddagger$  = other diagnoses consist of essential tremor, corticobasal

syndrome, dystonic tremor, functional neurological disorder, multiple sclerosis, vascular parkinsonism. APS = atypical parkinsonian syndrome, CBS = corticobasal syndrome, CEU = Caucasian residents of European ancestry, EOPSP = early onset PSP, ET = essential tremor, FTD = frontotemporal dementia, LBD = Lewy body dementia, LOPSP = late onset PSP, MSA = multiple system atrophy, NPH = normal pressure hydrocephalus, PD = Parkinson's disease, PSP = progressive supranuclear palsy, SD = standard deviation, SWEDD = scans without evidence of dopaminergic deficit, VascP = vascular parkinsonism, prob. = probable, poss. = possible, s.o. = suggestive of, PSP-P = PSPparkinsonism, PSP-RS = PSP-Richardson syndrome, PSP-PGF = PSP-progressive gait freezing, PSP-F = PSP-frontal, PSP-CBS = PSP-corticobasal syndrome overlap.

None of the QSBB EOPSP or LOPSP cases had a past medical history of encephalitis. Of the PROSPECT EOPSP cases, 1/12 (8%) reported experiencing a head injury that resulted in a loss of consciousness prior to the onset of PSP motor symptoms compared with 7/183 (4%) of the PROSPECT LOPSP cases. The mean latency of time from concussive head injury to onset of PSP motor symptoms was 44.5 years in the EOPSP group and 38.9 years in the LOPSP group.

A total of 1878 cases had genotyping data available that passed quality control thresholds to be included in genetic analyses. Of note, I excluded the two EOPSP *MAPT* mutation cases from these analyses. The minor allele frequency of the *TRIM11* rs564309 single nucleotide polymorphism was higher in the EOPSP group in comparison with the LOPSP group, coinciding with a higher rate of non-PSP-RS phenotypes in EOPSP. I found significantly higher PSP GRS z-scores in the EOPSP and LOPSP groups when compared with the PD group, with no significant differences between the EOPSP and LOPSP groups (Table 4.3). The PSP GRS was effective at differentiating between PD vs. all PSP cases (EOPSP and LOPSP combined), with an AUC of 0.68 (Figure 4.2).

### Table 4.3: Genetic profile of EOPSP, LOPSP and PD study cohorts

	EOPSP	LOPSP	PD				
No. of subjects	24	288	1566				
MAPT H1/H1 haplotype	92%	90%	67%				
frequency (%)							
ApoE E4 allele	18%	10%	13%				
frequency (%)							
<i>TRIM11</i> rs564309 MAF	21%	11%	15%				
(%)							
PSP GRS z-score –	0.59*\$	0.48* <b>\$</b>	-0.08				
mean, (SE)	(0.13)	(0.05)	(0.03)				
Group comparisons of PSP GRS made using Kruskal-Wallis ANOVA - * = p < 0.05 vs PD, \$ = No							

significant intra-group difference between QSBB and PROSPECT cases. *ApoE* = apolipoprotein E, EOPSP = early onset PSP, GRS = genetic risk score, LOPSP = late onset PSP, *MAPT* = microtubule associated protein tau, PD = Parkinson's disease, PSP = progressive supranuclear palsy, SE = standard error, *TRIM11* = tripartite motif-containing protein 11, MAF = minor allele frequency.



**Figure 4.2** – ROC curve analysis of the PSP GRS differentiating between PD vs. all PSP cases (EOPSP and LOPSP combined).

A subset of PROSPECT EOPSP (n = 5) and LOPSP (n = 42) cases, and Tracking Parkinson's PD (n = 570) cases, had baseline and 2-year or 3-year PSPRS and/or MDS-UPDRS part III scores recorded. There were similar mean baseline clinical rating scale scores between EOPSP (PSPRS = 34.2, MDS-UPDRS part III = 33.2) and LOPSP (PSPRS = 33.4, MDS-UPDRS part III = 37.5) groups, although the mean disease duration at baseline testing in the EOPSP group was higher in comparison with the LOPSP group (4.1 years vs. 2.6 years). I found no significant differences in the subscale (history, mentation, bulbar, ocular, limb, and gait) scores of the baseline PSPRS between the EOPSP and LOPSP groups. The mean baseline MDS-UPDRS part III score was significantly higher in the LOPSP group when compared with the PD group (37.5 vs. 20.9; P < 0.05). I found that there was a non-significant trend toward the mean annualised change in PSPRS being lower in the EOPSP group when compared with the LOPSP group (+9.9 vs. +12.8). In contrast, the mean annualised change in MDS-UPDRS part III scores were significantly lower in the PD group when compared with both EOPSP and LOPSP groups (+3.3 vs. +19.6 and +14.9; P < 0.05), with no significant difference between the EOPSP and LOPSP groups.

Using a pathological diagnosis of PSP as the gold standard, I reviewed 160 QSBB cases to calculate sensitivity (true positive / true positive + false negative) and PPV (true positive / true positive + false positive). True positive was defined as the number of cases with an ante-mortem clinical diagnosis of EOPSP/LOPSP and subsequent pathological diagnosis of PSP. False negative was defined as the number of cases with an ante-mortem clinical diagnosis of anything other than EOPSP/LOPSP and subsequent pathological diagnosis of PSP. False positive was defined as the number of cases with an ante-mortem clinical diagnosis of anything other than EOPSP/LOPSP and subsequent pathological diagnosis of PSP. False positive was defined as the number of cases with an ante-mortem clinical diagnosis of EOPSP/LOPSP and subsequent pathological diagnosis of PSP. False positive was defined as the number of cases with an ante-mortem clinical diagnosis of EOPSP/LOPSP and subsequent pathological diagnosis of anything other than PSP. Using this approach, I found that an initial clinical diagnosis of EOPSP had a lower diagnostic sensitivity (33%) and PPV (38%) in comparison with an initial clinical diagnosis of LOPSP (80% and 76%). In contrast, the diagnostic sensitivity and PPV of a final clinical

diagnosis of EOPSP (93% and 76%) were higher and similar to that of LOPSP (89% and 84%).

I carried out analyses to investigate the impact of using alternative age at onset cut-off points to define EOPSP **(Table 4.4)**. This revealed that as the age at onset cut-off point was increased, the EOPSP and LOPSP groups became more homogeneous in their clinical phenotype and PSP GRS z-scores.

	EOPSP					LOPSP				
Age at onset cut-off (years) *	≤52	≤55	≤59	≤62	≤64	>52	>55	>59	>62	>64
No. of subjects	12	26	49	94	117	313	299	276	231	208
% of cases with PSP- P/PSP-PGF initial clinical phenotypes	67%	65%	39%	29%	26%	22%	26%	21%	21%	22%
Mean PSP GRS z-	0.57	0.59	0.37	0.40	0.44	0.50	0.48	0.52	0.53	0.52
score										
Sensitivity **	20%	33%	50%	63%	71%	81%	80%	83%	84%	82%
PPV **	33%	38%	53%	56%	61%	75%	76%	77%	81%	82%

Table 4.4: Impact of alternative age at onset cut-off points to define EOPSP

Analysis on the impact of alternative age at onset cut-off points to define EOPSP. \* age at onset cutoff points represent the 5%, 10%, 15%, 20% and 25% age at onset percentile cut-off points from the Queen Square brain bank PSP case series, \*\* = Analyses restricted to only cases with a pathological diagnosis from Queen Square brain bank. EOPSP = early onset PSP, GRS = genetic risk score, LOPSP = late onset PSP, PPV = positive predictive value, PSP = progressive supranuclear palsy, PSP-P = PSP-parkinsonism, PSP-PGF = PSP-pure akinesia with gait freezing.

# 4.5 Discussion

This is the first study to define EOPSP and describe its genetic and clinico-pathological profile. Using my definition of an age at motor symptom onset ≤55 years of age, I found that up to 10% of PSP cases were early-onset in nature. A similar frequency of cases has been observed in young-onset PD (defined as age at onset <50 years of age) **(89)** and young-onset multiple system atrophy (defined as age at onset <40 years of age) **(90)**.

My study highlights a number of important points that are relevant to both clinical and research settings. First, I show the value of screening patients presenting with an EOPSP syndrome for PSP mimics and rare genetic mutations known to cause familial PSP. Of note, higher rates of a family history of dementia and/or parkinsonism were noted in the EOPSP group when compared with the LOPSP group, even when I discount the two identified *MAPT* mutation EOPSP cases. This observation has been noted in PSP previously **(91)** and suggests that there may be novel genetic causes of familial PSP that have yet to be identified.

The overall diagnostic sensitivity and PPV of an initial clinical diagnosis of EOPSP was considerably lower than that of LOPSP while the diagnostic sensitivity and PPV of a final clinical diagnosis of EOPSP and LOPSP were predictably higher and similar to values obtained by Hughes and colleagues **(12)**. This highlights the need for robust and objective biomarkers which complement clinical assessment to allow accurate differentiation of EOPSP from PD.

In comparison with LOPSP and PD, a clinical diagnosis of EOPSP had a longer diagnostic latency; 50% of the EOPSP cases were initially misdiagnosed as PD, and this coincided with the initial clinical profile of EOPSP being dominated by limb parkinsonism and gait freezing. The most common initial MDS PSP criteria phenotypes in association with these presentations were "suggestive of" PSP-P and "possible" PSP-PGF. In these cases, although abnormal eye movements had yet to occur to permit "probable" PSP-P/PSP-PGF diagnoses, the presence of early postural instability and/or progressive gait freezing (in the context of parkinsonism) were key clinical features that enabled the differentiation of EOPSP from PD.

The final clinical profile of EOPSP closely resembled that of LOPSP and mirrored previous studies which have shown that initial non-PSP-RS phenotypes come to resemble PSP-RS in the latter stages of disease (73). As we move into a new era of potential anti-tau therapies (18), early and accurate distinction between PSP and PD will become increasingly important. Therefore, I explored the value of a PSP GRS and found that, at the group level, EOPSP and LOPSP z-scores were significantly higher in comparison with PD.

In the absence of an identified genetic mutation, the biological drivers of early-onset presentations of neurodegenerative diseases such as PSP are likely to be multifactorial. I explored the potential aetiological roles of encephalitis and head injury and found no significant differences in rates between EOPSP and LOPSP.

My findings of a longer disease duration in deceased EOPSP cases in comparison with LOPSP is consistent with a previous clinico-pathological study that compared early-onset and late-onset PD subjects (92). One potential explanation for this is the fact that our EOPSP group mostly consisted of cases with PSP-P and PSP-PGF phenotypes, which have been associated with slower rates of disease progression (8) (9) (73) (81). However, another possibility, outside the scope of this study, is the likelihood of lower rates of co-pathologies in EOPSP cases, with a similar study demonstrating an increased age at death in neurodegenerative diseases with minimal co-pathology (93).

One of the major strengths of this study was the in-depth clinical data that was available to me from both the QSBB case notes and serial clinical assessments of patients in the PROSPECT and Tracking Parkinson's studies. Within the EOPSP and LOPSP groups, the similarity in clinico-genetic profiles between the QSBB and PROSPECT cases suggest that my findings are robust and that the diagnostic accuracy of the PROSPECT EOPSP and LOPSP cases is high. Similar to previous early-onset studies of movement disorders **(89) (90)**, I used an arbitrary age at onset cut-off point to define EOPSP. However, in my study

I've gone further by investigating the impact of changing the EOPSP age at onset cut-off point. This approach was particularly useful as it further justified my initial approach of using an age at onset cut-off point of  $\leq$ 55 years of age to define EOPSP. When compared with the other cut-off points studied, a cut-off point of  $\leq$ 55 years of age highlighted the greatest difference in clinical phenotype and PSP GRS z-scores between the EOPSP and LOPSP groups.

The main limitation of this study was the relatively small number of EOPSP cases that were available for analysis. In addition, the genetic mutation rates in EOPSP are based on pathogenic mutations that are directly genotyped on the Illumina NeuroChip, and targeted *NPC1* and *NPC2* sequencing was limited to cases that had biochemical evidence of Niemann-Pick type C disease. A more accurate estimate of pathogenic genetic mutation rates will be achieved by carrying out whole-genome sequencing of these cases. I also acknowledge that the absence of pathological confirmation in the large PD cohort may lead to the inclusion of non-PD patients. However, this is likely to be applicable to a very small number of patients as previous studies have shown that the PPV of a clinical diagnosis of PD going on to have pathological confirmation of PD is as high as 98.6% **(12)**.

In conclusion, EOPSP was defined as cases with a clinical diagnosis of PSP and a motor symptom onset ≤55 years of age. Genetic testing for familial *MAPT* mutations and PSP mimics is recommended in this patient group. The diagnostic accuracy of EOPSP is lower than that of LOPSP in the early stages of disease, and this coincides with the initial clinical profile of EOPSP being similar to PD. At the group level, a PSP GRS was able to differentiate EOPSP from PD, and this may be helpful in future diagnostic algorithms.

# <u>Chapter 5 – Progressive supranuclear palsy phenotype genome-wide association</u> <u>study</u>

### 5.1 Abstract

The basis for clinical variation related to underlying progressive supranuclear palsy (PSP) pathology is unknown. I performed a genome-wide association study (GWAS) to identify genetic determinants of PSP phenotype.

Two independent pathological and clinically diagnosed PSP cohorts were genotyped and phenotyped to create Richardson syndrome (PSP-RS) and non-PSP-RS (consisting of PSP-P and PSP-PGF cases) groups. I carried out a logistic regression GWAS to compare PSP-RS and non-PSP-RS groups in each independent cohort and then combined datasets to carry out a whole cohort analysis (PSP-RS = 367, non-PSP-RS = 130). I validated my findings in a third cohort by referring to data from 100 deeply phenotyped cases from the original PSP case-control GWAS. I assessed the expression/coexpression patterns of the identified genes from this phenotype GWAS and used my data to carry out gene-based association testing.

In the whole-cohort analysis, I found a genome-wide significant signal in chromosome 1. The lead single nucleotide polymorphism (SNP), rs564309, showed an association signal in both cohorts, reaching genome-wide significance in the whole cohort analysis (odds ratio = 5.5, 95% confidence interval = 3.2-10.0, p =  $1.7 \times 10^{-9}$ ). rs564309 is an intronic variant of the tripartite motif-containing protein 11 (*TRIM11*) gene, a component of the ubiquitin proteasome system (UPS). In the third cohort, minor allele frequencies of surrogate SNPs in high linkage disequilibrium with rs564309 replicated my findings. Gene-based association testing confirmed an association signal at *TRIM11* and *TRIM17*. I found that TRIM11/17 is predominantly expressed neuronally, in the cerebellum and basal ganglia. However, at the moment the functional effects of rs564309 are unknown.

This study suggests that the *TRIM11/17* locus may be a genetic modifier of PSP phenotype and potentially adds further evidence for the UPS having a key role in tau pathology, therefore representing a target for disease-modifying therapies.

# 5.2 Introduction

To date, GWAS in PSP has largely focused on identifying common risk variants using a case vs control approach. As such, the two previous large-scale GWAS, based on neuropathologically defined cases (total n = 1,646), have identified risk variants at *MAPT* (H1 haplotype and H1c sub-haplotype), *STX6*, *EIF2AK3*, *MOBP*, *RUNX2* and *SLCO1A2*. These implicate biological pathways involved in microtubule function, vesicle trafficking, the endoplasmic reticulum unfolded protein response, myelin structure and transporters present (among other places) at the blood-brain barrier where they regulate solute trafficking (25) (27).

The classical clinical phenotype of PSP, Richardson syndrome (PSP-RS), is characterised by a levodopa–unresponsive akinetic–rigid syndrome with falls, a vertical supranuclear gaze palsy, and dementia **(5)**. A diagnosis of this form of PSP was operationalized in the NINDS– Society for Progressive Supranuclear Palsy criteria of 1996 **(6)**. In the last two decades, variant clinical phenotypes relating to underlying PSP pathology have been identified in relatively small case series. These include PSP-parkinsonism (PSP-P), characterised by an onset of asymmetrical parkinsonism which is partially responsive to levodopa **(8)**, and PSP with progressive gait freezing (PSP-PGF), characterised by gradual onset freezing of gait or speech, absent limb rigidity and tremor, no sustained response to levodopa, and no dementia or ophthalmoplegia in the first 5 years of disease **(9)**. PSP-P and PSP-PGF both have a similar age of disease onset to PSP-RS and clinically resemble PSP-RS in the latter stages of disease, but have a significantly longer mean disease duration (PSP-P = 9 years,

PSP-PGF = 13 years) (8) (9). Although post-mortem remains the gold standard for diagnosing PSP, PSP-P and PSP-PGF have been formally operationalised in the 2017 MDS PSP diagnostic criteria, along with other variant presentations relating to underlying PSP pathology such as PSP/CBS overlap and PSP-F, enabling diagnosis in life (11).

The basis for this clinical variation related to a core pathology is unknown. PSP clinical subtypes have been related to the regional distribution and density of pathogenic tau accumulation and neuronal loss **(10)**. Here, I have used a large clinico-pathological PSP cohort based on the latest MDS diagnostic criteria to show that the clinical phenotype of PSP relates in part to genetic variants which may determine regional susceptibility.

#### 5.3 Methods

#### **Study Design and Participants**

All patients gave written informed consent for the use of their medical records and blood/brain tissue for research purposes, including the analysis of DNA. I identified patients with a neuropathological diagnosis of PSP from the following UK brain banks: MRC London Neurodegenerative Diseases Brain Bank (Research Ethics Committee reference 08/MRE09/38 + 5), Multiple Sclerosis and Parkinson's UK Brain Bank, London (Research Ethics Committee reference 08/MRE09/31 + 5), and Queen Square Brain Bank (the brain donor program was approved by a London Multi-Centre Research Ethics Committee, and tissue is stored for research under a license from the Human Tissue Authority, No. 12198). The year of death for cases ranged from 1998 to 2017.

I identified patients with a clinical diagnosis of a PSP syndrome from the PROSPECT study, a longitudinal study of patients with atypical parkinsonian syndromes undergoing serial clinical, imaging, and biomarker measures (Queen Square Research Ethics Committee

14/LO/1575) **(81)**. Cases were recruited between 2015 and 2017. A subset of these patients also underwent post-mortem neuropathological diagnosis at the Queen Square Brain Bank.

#### Phenotyping of Cases

I performed a retrospective clinical notes review of all neuropathological PSP cases to extract the following demographic and clinical information: gender, age at motor symptom onset, date of motor symptom onset, and date of death. I used this information to calculate the total disease duration (defined as date of motor symptom onset to date of death). Cases that did not have the above clinical information available were excluded from the study. Exclusion criteria used in the MDS diagnostic criteria were not considered, as the presence of alternative diseases would have been identified at post-mortem.

Using the MDS diagnostic criteria **(11)**, I assigned each case an initial and final clinical phenotype. This was based on the clinical features documented in clinical letters in the first 3 years from motor symptom onset and the clinical features documented in clinical letters in the last 2 years of life. I focused on 3 clinical phenotypes of interest: PSP-RS, PSP-P, and PSP-PGF; and only assigned these phenotypes if their corresponding "possible" or "probable" criteria were fulfilled. Cases were assigned a diagnosis of "unclassified" if there was insufficient evidence from the clinical notes to assign one of the phenotypes of interest. In cases where there was an overlap of clinical phenotype features, a consensus decision (between myself and Prof Morris) was made to assign the most appropriate clinical phenotype. I collected the same clinical assessments. To ensure accuracy in assigning a phenotype, living subjects were only included if their latest clinical assessment was carried out at least 3 years after motor symptom onset. In addition, cases were excluded from analyses if they had the presence of any MDS diagnostic exclusion criteria or if they fulfilled both MDS criteria for one of the PSP phenotypes of interest as well as Armstrong criteria for

probable PSP-CBS, as these subjects may have underlying corticobasal degeneration pathology (48).

#### **Genotyping and Quality Control**

All pathologically diagnosed cases had DNA extracted from frozen brain tissue (cerebellum or frontal cortex). Clinically diagnosed PROSPECT study cases had DNA extracted from whole blood. I was responsible for carrying out DNA sample quality control and preparation for genotyping by doing gel electrophoresis and Qubit fluorometry at the UCL Institute of Neurology. I subsequent delivered plates of DNA samples to the UCL Institute of Child Health where they underwent genotyping using the Illumina NeuroChip **(65)**. I carried out standard steps for data quality control including sex checking, heterozygosity, identity by descent to exclude related individuals and a principal components analysis (PCA) merged with the HapMap reference dataset to exclude all cases of non-European ancestry **(Figure 5.1)**. I then carried out SNP imputation using the Haplotype Reference Consortium v1.1 panel (https://imputation.sanger.ac.uk/). I used the following post-imputation data quality control steps to filter out SNPs with: INFO score <0.70, posterior probability <0.90, genetic missingness >0.05, Hardy-Weinberg equilibrium p-value <1.0x10<sup>-7</sup> and minor allele frequency <0.01. I screened all cases for known pathogenic *MAPT*, *LRRK2*, and *DCTN1* mutations, which are directly genotyped on the Illumina NeuroChip **(65)**.

To confirm the validity of the NeuroChip genotyping and imputation, a subset of both directly genotyped and imputed SNPs underwent re-genotyping using the LGC KASP genotyping service for coverage of significant regions in association.



**Figure 5.1** – Genetic PCA analysis using HapMap reference dataset. Grey = Reference cases of European (CEU) ancestry); Yellow = Reference cases of African ancestry; Blue = Reference cases of Asian ancestry; Orange = Cases considered for inclusion in PSP phenotype and survival GWAS'. All cases with a PC1 value >2SD away from the mean reference CEU group PC1 value were excluded.

# **Statistical Analyses**

I carried out all statistical analyses using Plink v1.9 and I generated images using R v3.3.2 and LocusZoom.

By dividing the whole cohort into PSP-RS and non-PSP-RS (combined PSP-P and PSP-PGF) groups based on their initial clinical phenotype, group comparisons of clinical features were carried out using t tests. In addition, the PSP-RS and non-PSP-RS group minor allele frequencies (MAFs) of all PSP case–control GWAS risk variants were extracted from the imputed data using the --recodeA command in Plink, version 1.9.

I performed a logistic regression GWAS using the imputed dataset to compare PSP-RS and non-PSP-RS groups. Based on their assigned initial clinical phenotypes, non-PSP-RS subjects were defined as "cases" and PSP-RS subjects were defined as "controls." The regression model used gender, age at motor symptom onset, study site of subject recruitment, and the first 2 principal components as covariates. This analysis was first carried out on the pathological cohort and then subsequently on the clinical cohort before combining datasets to carry out a whole cohort analysis. The Bonferroni correction for multiple SNP testing was used to set the genome-wide significance p value threshold at 9  $\times$  10<sup>-9</sup>. The whole cohort GWAS analysis was used to generate Manhattan and regional association plots.

All significant SNPs from the association analysis were assessed for their MAFs in European controls. I acquired this data from the Genome Aggregation Database (gnomad.broadinstitute.org), which is based on data from ~120,000 exome sequences and ~15,500 whole genome sequences from unrelated individuals. All significant SNPs from the association analysis were assessed for their level of significance in phase 1 of the original PSP case–control GWAS **(25)** using publicly available data at the National Institute on Aging Genetics of Alzheimer's Disease Data Storage Site (www.niagads.org).

# **PSP Case–Control GWAS Validation Cohort**

A separate subset of 100 pathologically confirmed PSP cases from phase 1 of the original PSP case–control GWAS **(25)** had in-depth phenotype data available (from Prof Günter Höglinger, Prof Ulrich Müller, Dr Gesine Respondek and Dr Max-Joseph Grimm) to assign an initial clinical phenotype according to the MDS criteria, as per my study methods. These cases had undergone genotyping using the Illumina Human 660W-Quad Infinium Beadchip with standard data quality control steps taken by Prof Höglinger, including a PCA to exclude cases that were not of European ancestry. PSP-RS and non-PSP-RS group

MAFs for directly genotyped SNPs that were significant in the phenotype GWAS were extracted to further validate my findings.

#### Whole Exome Sequencing

Sixty-nine cases from the pathological cohort had previously undergone whole exome sequencing (WES) by Dr Kin Mok and Dr Alan Pittman at the UCL Institute of Neurology using the Illumina Truseq Capture in Illumina HiSeq platform. Dr Pittman and I used this data to look for the presence of rare coding variants in genes of interest to arise from my GWAS. Read data were aligned to hg19 by use of novoalign (v3.02.04) and indexed bam files were de-duplicated of polymerase chain reaction artefacts by use of Picard Tools MarkDuplicates. The Genome Analysis Toolkit was then used to perform all subsequent steps according to their good practice guidelines; local realignments around possible indels and variant calling were conducted with HaplotypeCaller. Variants were filtered by use of variant quality score recalibration (truth tranche = 99.0%). In addition, hard-filtering based on low-depth and low-genotype quality was performed. Annotation was performed by use of Annovar software.

#### **Gene-Based Association Testing**

Gene-level p values were calculated by Miss Regina Reynolds and Dr Mina Ryten (UCL Institute of Neurology) using MAGMA v1.06 as outlined in de Leeuw et al (94). MAGMA tests the joint association of all SNPs in a gene with the phenotype while accounting for linkage disequilibrium (LD) between SNPs. This presents a powerful alternative to SNP-based analyses, as it reduces the multiple testing burden and thus increases the possibility of detecting effects consisting of multiple weaker associations (94). SNPs were mapped to genes using National Center for Biotechnology Information (NCBI) definitions (GRCh37/hg19, annotation release 105); only genes in which at least one SNP mapped were included in downstream analyses. These were run both with and without a window of 35kb upstream and 10kb downstream of each gene, as most transcriptional regulatory elements fall within this interval (95). Furthermore, the major histocompatibility complex

region was excluded. The gene p value was computed based on the mean association statistic of SNPs within a gene, with genome-wide significance set to  $p < 2.74 \times 10^{-6}$ , and LD was estimated from the European subset of 1000 Genomes Phase 3.

#### **Colocalization Analyses**

To evaluate the probability that the same causal SNP was responsible for modifying the phenotype of PSP and modulating gene expression, the Coloc method, described by Giambartolomei et al (96), was used by Mr David Zhang and Dr Mina Ryten (UCL Institute of Neurology), using my GWAS summary statistics coupled with expression quantitative trait loci (eQTLs) from BRAINEAC and GTEx. GTEx eQTLs included those originating from all brain regions. Analyses were restricted to genes within 1Mb of the significant region of interest ( $p < 5 \times 10^{-8}$ ) and coloc.abf was run with default priors. Tests with a posterior probability of hypothesis 4 (PPH4)  $\geq$  0.75 were considered to have strong evidence for colocalization.

### Assessment of Gene Expression

I assessed gene expression profiles using publicly available BRAINEAC (<u>www.braineac.org</u>) (97), GTEx (<u>www.gtexportal.org</u>), and Allen Brain Atlas (<u>www.brain-map.org</u>) (98) webbased resources.

The BRAINEAC database contains brain tissues from 134 healthy controls from the following brain regions: frontal cortex, temporal cortex, parietal cortex, occipital cortex, hippocampus, thalamus, putamen, substantia nigra, medulla, cerebellum, and white matter. RNA isolation and processing of brain samples were performed and analyzed using Affymetrix (Santa Clara, CA) Exon 1.0 ST arrays. The GTEx database consists of 8,555 samples from 53 tissues (including 13 brain regions) of 544 donors for which RNAseq was conducted. The GTEx Project was supported by the Common Fund of the Office of the Director of the National Institutes of Health, and by the National Cancer Institute, National Human Genome

Research Institute, National Heart, Lung, and Blood Institute, National Institute on Drug Abuse, National Institute of Mental Health, and NINDS. The data used for the analyses described in this article were obtained from the GTEx Portal on April 31, 2018. The Allen Human Brain Atlas database contains microarray data from 8 neuropathologically normal individuals of varying ethnicity. Microarray data were generated using the Agilent Technologies (Santa Clara, CA) 4 × 44 Whole Human Genome array and covers ~62,000 gene probes per profile and ~150 brain regions.

I analysed gene expression at the cellular level in the brain using RNAseq data from the Brain RNA-Seq database (www.brainrnaseq.org/) as per Zhang et al (99). Of note, these data were generated from healthy temporal lobe samples that were resected from 14 patients to gain access to deeper epileptic hippocampi. The number of different cell types obtained from these samples were as follows: mature astrocyte, n = 12; microglia, n = 3; neuron, n = 1; oligodendrocyte, n = 3. I also referred to single cell RNA-Seq data provided by DropViz (www.dropviz.org), which provides gene expression on 690,000 individual cells derived from 9 different regions of the adult mouse brain (100).

### Semi-quantitative RT-PCR measurement of gene expression

To determine the brain expression of splice variants of the candidate protein, TRIM11, semiquantitative RT-PCR was performed by myself, Dr Tammaryn Lashley, Dr Christina Toomey and Dr Nuria Seto-Salvia, as described previously **(101)**, using frontal cortex tissue from 6 neuropathologically diagnosed PSP cases and 6 age-matched neurologically normal controls from Queen Square brain bank. Of note, the 6 PSP cases consisted of 2 PSP-RS, 2 PSP-P and 2 PSP-PGF cases, with each group having 1 case homozygous for the major allele (CC) and 1 case heterozygous (AC) in relation to the rs564309 variant. We used validated primers from Thermo Fisher located in exons 1-2, 2-4, 4-6 and 6 of TRIM11. GAPDH was used as a housekeeping gene for comparison.

For each sample, total RNA was extracted from tissue using Trizol (Life Technologies) according to the manufacturer's protocol resulting in 0.5–2 µg of RNA in a total volume of 20 µl, measured by the NanoDrop which was also used for detection of RNA quality control measures including the 260/280nm ratio (~2 for "pure RNA"). Reverse transcription was performed with the SuperScript III first strand kit (Invitrogen) with oligo(dT) or an equimolar ratio of oligo(dT) and random hexamers. Product amplification was obtained using Taq DNA polymerase with standard Taq buffer (BioLabs). PCR products were separated on 1% agarose gels and stained with GelRed (Biotium).

### 5.4 Results

A total of 497 subjects were included for analyses. Their clinical features are summarised by cohort and disease group in **Table 5.1**. Forty-four subjects were deemed unclassifiable and therefore not included in subsequent analyses. An initial screen of the genotype data revealed similar MAFs for risk variants identified in the PSP case-control GWAS **(Table 5.2)**. In addition, none of the cases carried known pathogenic variants covered by the NeuroChip for the following genes: *MAPT* (40 variants), *LRRK2* G2019S (1 variant), and *DCTN1* (12 variants). To highlight the fact that this study was underpowered, I used the University of Michigan's GWAS power calculator

(http://csg.sph.umich.edu/abecasis/gas\_power\_calculator/) to calculate the minimum achievable genotype relative risk at 80% statistical power. Using my parameters (sample size = 497; significance level =  $9 \times 10^{-9}$ ; additive model; disease prevalence = 0.0001; disease allele frequency = 0.05), I found that at 80% statistical power the minimum achievable genotype relative risk was 4.71.

Table 5.1: Clinical features of subjects included in genotype-phenotype analyses	

	Pathological cohort		<u>Clinica</u>	l cohort	Whole cohort		
	PSP-RS	non-PSP-RS	PSP-RS	non-PSP-RS	PSP-RS	non-PSP-RS	
		76		54		130	
No. of subjects	230	PSP-P = 60	137	PSP-P = 42	367	PSP-P = 102	
		PSP-PGF = 16		PSP-PGF = 12		PSP-PGF = 28	
% of subjects male	60.0%	53.9%	57.7%	51.9%	59.1%	53.1%	
Age at motor							
symptom onset	68.9	65.9	66.5	65.3	68.1 <sup>a b</sup>	65.6 <sup>a b</sup>	
(years)–Mean,	49-89	46-86	51-87	54-82	49-89	46-86	
range, SD	7.4	8.8	6.8	6.9	7.3	8.0	
Final/current		PSP-RS = 71		PSP-RS = 28		PSP-RS = 99	
clinical phenotype	PSP-RS = 230	PSP-P = 4	PSP-RS = 137	PSP-P = 20	PSP-RS = 367	PSP-P = 24	
		PSP-PGF = 1		PSP-PGF = 6		PSP-PGF = 7	
Mean disease							
duration in	5.9	10.7	5.6	9.2	5.8 b <b>†</b>	10.6 <sup>b</sup> <b>†</b>	
deceased subjects	1.9-15.9	2.2-16.3	2.4-13.5	8.1-10.8	1.9-15.9	2.4-13.5	
(years)–Mean,	1.9	2.9	2.2	1.2	1.9	2.8	
range, SD							
No. of subjects							
undergoing post-							
mortem (% with a	230	76	10	1	240	77	
pathological	(100%)	(100%)	(100%)	(100%)	(100%)	(100%)	
diagnosis of PSP)							

<sup>a</sup> No statistically significant difference between PSP-RS and non-PSP-RS groups. <sup>b</sup> No statistically

significant difference between pathological and clinical cohorts. **†** Statistically significant (p < 0.05) difference between PSP-RS and non-PSP-RS groups using Welch's t-test.

# Table 5.2: Comparison of PSP risk variant MAFs between PSP case-control GWAS and PSP phenotype GWAS

Chr.	SNP	Gene	PSP case-con	trol GWAS	PSP phenotype GWAS		
Band	GRCh37 position (BP)		MAF in healthy controls	MAF in PSP	MAF in PSP-RS	MAF in non-PSP-RS	
1q25.3	rs1411478	STX6	0.42	0.50	0.44 <b>†</b>	0.43	
	180,962,282						
2p11.2	rs7571971	EIF2AK3	0.26	0.31	0.34 <b>†</b>	0.30	
	88,895,351						
3p22.1	rs1768208	MOBP	0.29	0.36	0.32 <b>†</b>	0.32	
	39,523,003						
	rs8070723	MAPT	0.23	0.05	0.05 <b>†</b>	0.05	
17q21.31	44,081,064	(H1 haplotype)					
	rs242557	MAPT	0.35	0.53	0.44 <b>†</b>	0.49	
	44,019,712	(H1c sub-					
		haplotype)					

PSP case-control GWAS data taken from Höglinger et al **(25)**. **†** No statistically significant difference between PSP-RS and non-PSP-RS groups using Fisher's exact test.

After SNP imputation and data quality control, 6,215,948 common variants were included in my analysis. Assigning non-PSP-RS subjects as "cases" and PSP-RS subjects as "controls," I applied a logistic regression association analysis using gender, age at motor symptom onset, study site of subject recruitment, and the first 2 principal components as covariates. I first carried out this analysis using data from the pathological cohort and then validated my findings using data from the independent clinical cohort before combining both datasets for a whole cohort analysis. The whole cohort analysis revealed 27 SNPs, all located on chromosome 1, which passed the threshold for genome-wide significance ( $p < 9 \times 10^{-9}$ ). These results are summarized in **Figure 5.2**. Following the exclusion of non-European subjects as part of my genotype data quality control, population stratification was not evident in the cohort, as evidenced by obtaining a genomic inflation factor (lambda) value of 1.05. A further locus on chromosome 12 approached genome-wide significance with the lead SNP (rs621042) p = 7.8 × 10<sup>-7</sup>.

An in-depth analysis of the significant SNPs reveal that they are all in high LD with each other (as defined by r2 > 0.80) and are all located at the chromosome 1q42.13 locus. A regional association plot **(Figure 5.3)** reveals that the lead SNP, rs564309, is an intronic variant located between exons 3 and 4 of the tripartite motif-containing protein 11 (*TRIM11*) gene. Alongside the directly genotyped lead SNP, the imputation quality (INFO) score for imputed significant SNPs ranged from 0.96 to 1. Ninety-six cases from the pathological cohort underwent re-genotyping for 8 SNPs (rs564309, rs35670307, rs12065815, rs10158354, rs3795811, rs6426503, rs138782220, and rs7555298) that span the significant chromosome 1q42.13 locus. Three of the 8 SNPs, including the lead SNP, were originally directly genotyped via the NeuroChip, whereas the remaining 5 SNPs were originally imputed in the dataset. The p value of these SNPs in the whole cohort GWAS ranged from 1.7 x 10<sup>-9</sup> to 7.3 x 10<sup>-5</sup>. The results of this re-genotyping run showed 100% concordance with my original NeuroChip and imputation data.



**PSP Phenotype GWAS** 

<u>Figure 5.2</u> – Manhattan plot of whole cohort PSP-RS versus non-PSP-RS association analysis, highlighting genome-wide significance at chromosome 1. The red line indicates the threshold for genome-wide significance ( $p < 9 \times 10^{-9}$ ).



**Figure 5.3** – Regional association plot of PSP-RS versus non-PSP-RS association analysis using imputed SNP data, implicating the chromosome 1q42.13 locus and identifying rs564309, an intronic variant of *TRIM11*, as our lead SNP. SNP positions, recombination rates, and gene boundaries are based on GRCh37/hg19.

Furthermore, 3 of the significant SNPs in LD with rs564309 are non-synonymous (missense) coding variants of the Obscurin (*OBSCN*) gene. *OBSCN* is mainly expressed in skeletal muscle and may have a role in the organisation of myofibrils during assembly as well as mediating interactions between the sarcoplasmic reticulum and myofibrils (**102**). Related diseases include fibromuscular dysplasia and hypertrophic obstructive cardiomyopathy.

The association statistics for rs564309, and the most significant flanking SNPs located at neighbouring genes within the chromosome 1q42.12 locus, are summarised below **(Table 5.3)**. The MAF of these SNPs was shown to be 0.10 in healthy European controls in the gnomAD database.

Table 5.3: PSP-RS vs non-PSP-RS association statistics for rs564309, and the most significant flanking SNPs located at neighbouring genes, in pathological, clinical and whole cohorts respectively

SNP	Gene	Pathological cohort				Clinical col	Whole cohort		
Position (BP)		MAF in	MAF in	OR	MAF in	MAF in	OR	OR	Р
Genotyped /		PSP-	non-	(95% CI)	PSP-	non-	(95% CI)	(95% CI)	
Imputed		RS	PSP-RS		RS	PSP-RS			
rs564309	TRIM11	0.04	0.19	6.25 <b>†</b>	0.04	0.16	4.76 <b>†</b>	5.55	1.7x10⁻⁰
228,585,562				(3.12-12.5)			(1.96-12.5)	(3.22-10.0)	
Genotyped									
rs61827276	TRIM17	0.04	0.18	5.88 <b>†</b>	0.04	0.16	5.55 <b>†</b>	5.55	6.2x10⁻⁰
228,597,130				(2.78-12.5)			(2.17-14.3)	(3.12-10.0)	
Imputed									
rs61825312	OBSCN	0.04	0.19	5.88 <b>†</b>	0.04	0.16	4.35 <b>†</b>	5.26	7.1x10⁻⁰
228,530,748				(2.86-12.5)			(1.78-11.1)	(2.94-9.09)	
Imputed									
rs2230656	HIST3H3	0.06	0.23	4.35 <b>†</b>	0.06	0.20	3.70 <b>†</b>	4.00	1.3x10⁻ଃ
228,612,838				(2.32-8.33)			(1.75-8.33)	(2.50-6.67)	
Imputed									

SNP positions based on GRCh37/hg19. MAF = Minor Allele Frequency. OR = Odds Ratio. P = p-value in whole cohort analysis.  $\dagger$  = p-value > 9x10<sup>-9</sup>.

To explore the impact of inadvertently including PD cases in the clinically diagnosed non-PSP-RS group, I referred to genotyping data from 484 European clinically diagnosed PD cases which were genotyped by Miss Manuela Tan (UCL Institute of Neurology) alongside my PSP cases and had undergone the same quality control steps outlined above. I found that the MAF of rs564309 in PD cases was 7%, similar to the MAF in healthy controls and considerably lower than the MAF in the non-PSP-RS group which contains PSP-P cases. When referring to publicly available p value data from phase 1 of the 2011 PSP case–control GWAS, I found that none of the significant SNPs reached even nominal significance (p < 0.05). One hundred European pathologically confirmed PSP cases from the PSP case-control GWAS underwent retrospective phenotyping according to the MDS diagnostic criteria using available clinical notes. Of those, 83 cases fulfilled probable criteria for initial clinical phenotypes of relevance to this study (PSP-RS, n = 45; PSP-P, n = 38). rs1188473, a SNP that was directly genotyped in the case–control GWAS, in high LD (r<sup>2</sup> 0.87, D' 1.0) with rs564309 and found to be significant in our whole cohort phenotype GWAS (p =  $2.6 \times 10^{-9}$ ), was shown to have similar MAFs when comparing the GWAS datasets in both PSP-RS and non-PSP-RS groups, therefore further validating our findings **(Table 5.4)**.

# Table 5.4: PSP-RS vs non-PSP-RS association statistics for rs1188473, in

pathological, clinical and 2	2011 case-control GWAS	cohorts respectively
------------------------------	------------------------	----------------------

	Pathological cohort			Clinical cohort			2011 case-control GWAS cohort		
SNP	MAF in	MAF in	OR	MAF in	MAF in	OR	MAF in	MAF in	OR
Position (BP)	PSP-	non-	(95% CI)	PSP-	non-	(95% CI)	PSP-	non-	(95% CI)
	RS	PSP-RS	p-value	RS	PSP-RS	p-value	RS	PSP-RS	p-value
rs1188473			5.35			3.75			2.90
228,612,444	0.05	0.23	(2.79-10.26)	0.06	0.19	(1.70-8.28)	0.06	0.16	(1.9-9.1)
			4.5x10⁻ <sup>7</sup>			1.0x10⁻³			9.3x10⁻³

SNP position based on GRCh37/hg19. MAF = Minor Allele Frequency. OR = Odds Ratio. Odds ratios based on logistic regression (PSP-RS vs non-PSP-RS) analysis, adjusting for gender and age at motor symptom onset.

Analysis of WES data from 65 subjects (49 RS, 16 non-RS) within the pathological cohort did not identify any nonsynonymous coding variants in *TRIM11* or *TRIM17* genes.

MAGMA analyses (Figure 5.4) revealed that 4 genes passed genome-wide significance in analyses run with and without 35kb upstream and 10kb downstream of each gene (*TRIM11*,  $p = 5.64 \times 10^{-9}$ ; *TRIM17*,  $p = 8.99 \times 10^{-9}$ ; *HIST3H3*,  $p = 1.29 \times 10^{-8}$ ; *LOC101927401*,
$p = 5.72 \times 10^{-8}$ ). *LOC101927401* appeared only in NCBI annotation and was absent in the queried gene expression databases; thus, it was excluded in downstream analyses.



**Figure 5.4** – MAGMA analyses revealing significance at *TRIM11*, *TRIM17*, and *HIST3H3* genes. Gene boundaries are based on GRCh37/hg19. Red lines indicate the threshold for genome-wide significance ( $p < 2.74 \times 10^{-6}$ ).

I then used human brain gene expression and co-expression data from the BRAINEAC, GTEx, and Allen Atlas databases to assess the expression profiles of genes identified in the MAGMA analysis above: *TRIM11*, *TRIM17*, and *HIST3H3*. All 3 datasets revealed high levels of TRIM11 and TRIM17 expression in the brain, particularly cerebellum and putamen, whereas HIST3H3 expression appeared to be at the lower limit of detection in human brain (Figure 5.5).



**Figure 5.5** – TRIM11, TRIM17, and HIST3H3 brain expression in 3 databases. (A) BRAINEAC database. CRBL = cerebellum; FCTX = frontal cortex; HIPP = hippocampus; MEDU = medulla; OCTX = occipital cortex; PUTM = putamen; SNIG = substantia nigra; TCTX = temporal cortex; THAL = thalamus; WHMT = white matter. (B) GTEx database. (C) Allen Atlas database (Caucasian subjects). Amg = amygdala; BF = basal forebrain; BG = basal ganglia; CAU = caudate; ET = epithalamus; HiF = hippocampal formation; Hy = hypothalamus; MES = mesencephalon; MET = metencephalon; MY = myelencephalon; PUT

= putamen; TH = thalamus; TPM = Transcripts Per Kilobase Million. Image credit: Allen
 Institute.

I explored the cellular specificity of TRIM11, TRIM17, and HIST3H3 expression in human brain using data provided by the Brain RNA-Seq database. This demonstrated higher neuronal expression (n = 1) of TRIM11 (0.38FPKM) compared to both TRIM17 (0.12FPKM) and HIST3H3 (0.10FPKM), which was expressed at the lower limit of detection. In comparison to its neuronal expression, TRIM11 expression in glial cell types was lower (mature astrocytes,  $0.14 \pm 0.02FPKM$ , n = 12; microglia,  $0.12 \pm 0.02FPKM$ , n = 3; oligodendrocytes,  $0.11 \pm 0.01FPKM$ , n = 3). I also explored cell-specific expression of the TRIM11 and TRIM17 mouse orthologues across the brain using single cell RNA-Seq data (http://dropviz.org/) generated from mouse brain tissue. These data suggested that expression of TRIM11 was highest in the spiny projection neurons of the striatum, with high expression in SPNs of both the "direct" and "indirect" pathways. In contrast, TRIM17 expression was generally lower, with the highest expression detected in neurons of the substantia nigra.

The colocalization analysis did not reveal any significant colocalization between the GWAS signals and eQTL data in the BRAINEAC and GTEx databases in all brain regions. However, of note, expression analyses using GTEx data revealed that several SNPs from the chromosome 1q42.13 locus reaching genome-wide significance in our GWAS were significant eQTLs for TRIM11 and TRIM17 in skin and thyroid tissues when analysed individually.

I extracted RNA from frontal cortex tissue of 6 neuropathologically diagnosed PSP cases (2 PSP-RS, 2 PSP-P and 2 PSP-PGF) and 6 age-matched controls with no neuropathology, with a mixture of rs564309 genotypes (CC and AC) in each of the PSP groups. RNA samples passed quality control steps outlined in the methods and were at a concentration of

0.5–2 µg in a total volume of 20µl. Semi-quantitative RT-PCR was carried out using validated primers located in exons 1-2, 2-4, 4-6 and 6 of TRIM11 with GAPDH used as a housekeeping gene for comparison. This confirmed the expression of TRIM11 in human brain with no apparent difference in expression between PSP vs controls, PSP-RS vs non-PSP-RS, and rs564309 CC genotype vs AC genotype. Of note, the exon 1-2 primer failed in both PSP and control RNA samples (Figure 5.6).



**Figure 5.6** – TRIM11 expression in human frontal cortex shown by semi-quantitative RT-PCR. A) Exon 1-2 primer, B) Exon 2-4 primer, C) Exon 4-6 primer, D) Exon 6 primer, E) GAPDH (housekeeping gene) expression. Lanes in each image (from left to right): 1 – Ladder; 2 – PSP-RS (rs564309 CC genotype); 3 – PSP-RS (rs564309 AC genotype); 4 – PSP-P (rs564309 CC genotype); 5 – PSP-P (rs564309 AC genotype); 6 – PSP-PGF (rs564309 CC genotype); 7 – PSP-PGF (rs564309 AC genotype); 8 – Blank; 9-14 – Agematched controls (unknown rs564309 genotype).

### 5.5 Discussion

To my knowledge, this is the first GWAS of clinical phenotype in PSP. I have shown that variation at the chromosome 1q42.13 locus determines clinical phenotype in PSP with a very strong effect size (odds ratio = 5.5). The validity of my GWAS results is increased by

similarly sized association signals and minor allele frequencies being observed in 2 independent cohorts, with a genome-wide significant association achieved when the 2 cohorts were combined. Furthermore, when considering the subset of MDS criteria– phenotyped cases from the original PSP case–control GWAS, PSP-RS and non-PSP-RS group MAFs of directly genotyped SNPs in high LD with the lead SNP are supportive of my findings.

I am also reassured by the PD genotyping data, which revealed that the GWAS signal would have been attenuated if PD cases had been inadvertently included in the clinically diagnosed non-PSP-RS group. However, this is unlikely, as I only included PSP-P and PSP-PGF cases that fulfilled "possible" and "probable" (not "suggestive of") PSP MDS criteria for these phenotypes. I suspect that none of the significant SNPs reached genome-wide significance in the original PSP case–control GWAS because pathologically diagnosed PSP cases would have contained a mixture of PSP-RS and non-PSP-RS cases. When considering my data, I can infer that the combined MAFs of PSP-RS and non-PSP-RS cases would have resulted in overall PSP group MAFs that were similar to those of healthy controls.

The validity of my NeuroChip genotyping and imputation data were highlighted by the confirmatory genotyping I carried out to span the chromosome 1q42.13 locus. The homogeneity of the independent cohorts is suggested by the following: (1) in both cohorts, a majority of cases with an initial non-PSP-RS phenotype had a final clinical diagnosis of PSP-RS, as previously shown by other groups **(73)**; (2) the cohorts had similar MAFs for risk variants identified in the PSP case–control GWAS **(25)**; and (3) there was 100% concordance between clinical and pathological diagnoses in the clinical cohort for the subset of patients that had undergone post-mortem examination.

MAGMA analysis confirmed signals in *TRIM11*, *TRIM17*, and *HIST3H3* genes. There was evidence for differential regional brain expression of TRIM11 and TRIM17 in reference databases, with my RT-PCR work providing independent verification of TRIM11 expression in human brain. Both human and mouse RNA-Seq data revealed high levels of neuronal TRIM11 expression. In addition, it is likely that this GWAS was significantly underpowered to detect signals in the eQTL colocalization analyses.

It is important to note that SNPs within this genomic locus are in high LD, as evidenced by the spread of genome-wide significant SNPs identified in this GWAS. Therefore, it is challenging to know which gene is driving the association signal. However, the localization of the lead SNP in the dataset and the gene expression profiles described above suggest that *TRIM11* is the most likely candidate gene at the chromosome 1q42.13 locus, although *TRIM17*, is also a candidate. The eQTL profile of the significant SNPs in GTEx, when analysed individually, was particularly interesting. The strong association between several SNPs in high LD with rs564309 and decreasing TRIM11 and TRIM17 expression in non-brain tissues highlights the concept of tissue/region/cell-specific expression of transcripts potentially being determined by disease state and at specific time points in development or ageing **(103)**. However, my data do not exclude potentially important functional roles for the other transcripts within this locus.

The major limitation of this study is that the cohort size was relatively small compared to case–control GWAS in PSP (25) and other neurodegenerative diseases (104) (105). Since lamda genomic inflation factors are affected by sample size, previous studies have found it informative to report the lambda1000 value, which represents the inflation factor for an equivalent study of 1000 cases and 1000 controls. This can be achieved by re-scaling lambda: 1 + ((lambda - 1) \* (1 / case + 1 / control) \* 500) (106). In contrast to the initial lambda value of 1.05, the lamba1000 value of this study was 1.26, suggesting genomic inflation. Further replication of my findings in larger cohorts is desirable, including other non-

PSP-RS phenotypes such as PSP-F, to confirm the role of TRIM11 and identify genetic determinants of clinical phenotype in PSP at other loci. Furthermore, the lead SNP is an intronic variant that does not pass the false discovery rate threshold for being a brain eQTL at the genes in the locus of interest, and the only coding variants that it is in LD with are from a gene (OBSCN) that is unlikely to be of biological relevance to PSP pathology. This is a common dilemma, as a majority of risk variants identified in GWAS over the past two decades are not associated with coding changes in expressed proteins (107). Furthermore, disease-associated intronic SNPs can regulate the expression of more distant genes. When referring to BRAINEAC and GTEx, rs564309 was found not to be a brain eQTL at distant genes outside of the region of interest. Functional impacts of intronic variants may arise in modes other than gene expression, including via splicing and methylation patterns of targeted transcripts and proteins. It remains a challenge to understand the functional consequences of non-coding genetic variation linked to phenotype, and so functional studies are vital. However, gene expression studies in post-mortem disease tissue can be challenging to interpret because of the confounding effects of changes on cell populations (108).

TRIM proteins are biologically plausible candidates as determinants of clinical phenotype in PSP, and are promising targets for follow-up functional studies. The TRIM family of proteins, most of which have E3 ubiquitin ligase activities, have various functions in cellular processes including intracellular signalling, development, apoptosis, protein quality control, autophagy, and carcinogenesis (109). A recent study has shown that TRIM11 has a critical role in the clearance of misfolded proteins via the ubiquitin proteasome system (UPS), in this case pathogenic fragments of both ataxin-1 (Atxn1 82q) and huntingtin protein (Httex1p 97QP) (110). Other groups have shown that lysine residues of tau are targets for polyubiquitination, which induces proteolytic degradation of tau via the UPS (111). Furthermore, tau accumulation has been associated with decreased proteasome activity in mouse tauopathy models, suggesting a feedback loop between impaired protein degradation, aggravated by a

protein aggregate-based impairment of proteostasis (112). These findings coincide with previous studies that have identified the UPS as a potential drug target in the treatment of neurodegenerative conditions (113) (114). Overexpression of TRIM17, partly controlled by glycogen synthase kinase 3 pathways, has been shown to initiate neuronal apoptosis in cell models (115). This was later shown to be mediated by increased degradation of the anti-apoptotic protein, myeloid cell leukaemia 1 (Mcl-1), via the UPS (116).

I hypothesize that common variation at the chromosome 1q42.13 locus modifies the function of TRIM11 to varying degrees in specific brain regions. In the more slowly progressing non-PSP-RS syndromes, an increase in TRIM11 function may lead to increased degradation of toxic tau species via the UPS, therefore protecting against rapid accumulation of tau pathology. Conversely, a decrease in protein function in the brainstem is more likely to promote rapid accumulation of tau aggregates, manifesting as the malignant PSP-RS phenotype of PSP.

In summary, the results of this study suggest that common variation at the *TRIM11* locus may be a genetic modifier of clinical phenotype in PSP. My findings add further evidence for the UPS playing a key role in tau pathology and therefore representing a potential target for disease-modifying therapies. Further GWAS with larger cohorts to confirm my findings and identify other genetic signals, screening of whole genome/exome sequencing data for rare variants in *TRIM11*, and follow-up functional studies at this locus are priorities.

### Chapter 6 – Progressive supranuclear palsy survival genome-wide association study

### 6.1 Abstract

The genetic basis of variation in the progression of primary tauopathies has not been determined. Here, I have used a genome-wide association study to identify genetic determinants of survival in 1,001 progressive supranuclear palsy (PSP) cases, derived from two independent cohorts.

In the whole-cohort analysis I found a genome-wide significant association with survival at chromosome 12 (lead SNP rs2242367 –  $p=7.5\times10^{-10}$ , hazard ratio=1.42). I obtained further replication for rs2242367 in a third independent cohort of 238 PSP cases (p=0.049, hazard ratio=1.22).

eQTL analyses revealed that rs2242367 is associated with LRRK2 expression in whole blood (eQTLGen database, p=6.7x10<sup>-58</sup>). I found that eQTLs for the long intergenic non-coding RNA, LINC02555, in the eQTLGen dataset colocalised with the PSP survival signal (PP4=0.77), and LINC02555 may control LRRK2 expression.

In conclusion, I have shown that there is shared genetic risk between Parkinson's disease and PSP. Modulation of LRRK2 may have a therapeutic role in multiple neurodegenerative diseases.

#### 6.2 Introduction

Progressive supranuclear palsy (PSP) is a rapidly progressive neurodegenerative tauopathy. In the classical form, PSP-Richardson syndrome (PSP-RS), patients develop imbalance and frequent falls, bulbar failure and dementia, and on average survive 6.9 years from symptom onset (2). More recently defined PSP subtypes, such as PSP-Parkinsonism (PSP-P) and PSP-Progressive Gait Freezing (PSP-PGF), are associated with a slower rate of progression (81). I have previously shown that the PSP phenotype is determined by variation at the *TRIM11/17* locus (71).

The pathology of PSP involves the deposition of insoluble hyperphosphorylated tau in neurons and astrocytes in sub-cortical and cortical brain regions. Recent work in animal models has shown that PSP-tau pathology is transmissible, in extracts from human PSP brain, inoculated into mouse brain leads to PSP-type pathology in the recipient which recapitulates the morphology and immunohistochemical features of the human tauopathy (117). Clinical disease progression relates to the sequential involvement of brain areas and systems. This may relate to differential neuronal susceptibility/resistance (cell-autonomous) or cell to cell spread (non-cell-autonomous) factors (118).

Genome-wide studies in neurodegeneration have focussed on case-control status which have provided powerful insights into the aetiology of neurodegenerative disease. In PSP, common variants at the *MAPT*, *MOBP*, *EIF2AK3* and *STX6* loci are associated with PSP risk **(25)**. However, therapeutic efforts focus on developing therapies which slow or halt disease progression, thus improving survival following a clinical diagnosis. Previous trials of therapeutics targeting microtubule dysfunction and tau hyperphosphorylation have shown no benefit in slowing disease progression in PSP patients **(16) (17)**. Current clinical trials in both PSP and Alzheimer's disease (AD) patients aim to prevent cell-to-cell spread of tau by using neutralising antibodies against tau in the extracellular space **(18)**.

The genetic determinants of clinical disease progression and survival for neurodegenerative diseases are largely unexplored and are likely to provide important clues to biology and effective therapy. A recent disease progression genome-wide association study (GWAS) in Huntington's disease identified a functional variant in *MSH3*, a DNA-repair gene, associated with disease progression, based on longitudinal change in motor, cognitive and imaging measures **(119)**.

Here, I have used a Cox-proportional hazards survival model based GWAS to identify genetic determinants of survival (from motor symptom onset to death) in PSP cases of European ancestry.

### 6.3 Methods

#### Study samples

**Discovery cohort:** I identified cases with a neuropathological diagnosis of PSP in two independent cohorts ("UCL PSP cohort" and "2011 GWAS cohort", with cases obtained from UK, US and German brain banks. All brain banks had approved ethics to cover the analysis of human tissue/DNA samples, and associated clinical data, for research projects, in compliance with the Declaration of Helsinki. Cases with a clinical diagnosis of PSP in the "UCL PSP cohort" were identified from the Progressive Supranuclear Palsy-Cortico-Basal Syndrome-Multiple System Atrophy (PROSPECT) study, a UK-wide longitudinal study of patients with atypical parkinsonian syndromes (Queen Square Research Ethics Committee 14/LO/1575) **(81)**. Individual written consent was obtained from all subjects or their next of kin. A subset of deceased PROSPECT study PSP cases had post-mortem neuropathological confirmation of diagnosis at Queen Square and Cambridge brain banks.

**Replication cohort:** Cases with a neuropathological diagnosis of PSP were identified from Mayo brain bank by Prof Dennis Dickson and Dr Shunsuke Koga. The study was approved by the Mayo Clinic Institutional Review Board. Importantly, unlike Mayo brain bank cases in the discovery cohort, these additional cases had not been included in the original PSP casecontrol GWAS (25).

In both the discovery and replication cohorts, I considered all PSP cases for inclusion in this study, provided they had an adequate depth of phenotype data available for clinical characterisation, and passed genotype quality control steps (see below).

#### **Clinical characterisation**

In total, 1,448 cases (discovery cohort, n = 1,033; replication cohort, n = 415) were considered for inclusion in this study. All PSP cases were assigned a Movement Disorder Society PSP diagnostic criteria phenotype based on clinical features that were present in the first three years from motor symptom onset, using consensus criteria on application of the diagnostic criteria (11) (54). Cases were subsequently stratified into PSP-RS and non-PSP-RS groups, where non-PSP-RS consisted of PSP-parkinsonism and PSP-progressive gait freezing phenotypes. PROSPECT study PSP cases were assigned a baseline PSP phenotype on entry to the study using the same criteria as above. Of note, all included PROSPECT study PSP cases were at least three years into their disease course (from motor symptom onset) and fulfilled at least "possible" criteria for a PSP phenotype. In addition, the following clinical data was collected for all cases: sex; age at motor symptom onset; disease duration from motor symptom onset to death, or motor symptom onset to date of censoring (01/12/2019) for living PROSPECT study PSP cases. The above approach was applied retrospectively using detailed case notes for brain bank cases (UCL cohort cases by myself; 2011 GWAS cohort cases by Prof Dennis Dickson (Mayo brain bank), Prof Günter Höglinger and Dr Gesine Respondek (DZNE, Munich); Mayo cohort cases by Prof Dennis Dickson and Dr Shunsuke Koga). Cases were excluded from the study if there was inadequate clinical data to accurately assign both PSP phenotype and disease duration, respectively.

#### Genotype data

Discovery cohort: All brain bank cases had DNA extracted from frozen brain tissue (cerebellum or frontal cortex). PROSPECT study cases had DNA extracted from whole blood. For the UCL PSP cohort, I was responsible for carrying out DNA sample quality control and preparation for genotyping by doing gel electrophoresis and Qubit fluorometry at the UCL Institute of Neurology. I subsequent delivered plates of DNA samples to the UCL Institute of Child Health where they underwent genotyping using the Illumina NeuroChip (65). I subsequently screened all of these cases for known pathogenic MAPT, LRRK2, and DCTN1 mutations, which are directly genotyped on the Illumina NeuroChip. DNA sample quality control and genotyping of 2011 GWAS cohort cases using the Illumina Human 660W-Quad Infinium chip (25) was led by Prof Günter Höglinger (DZNE, Munich) and Prof Gerry Schellenberg (University of Pennsylvania). For each cohort, using Plink v1.9 I carried out standard steps for data quality control including sex checking, heterozygosity, identity by descent to exclude related individuals and a principal components analysis (PCA) merged with the HapMap reference dataset to exclude all cases of non-European ancestry (see Chapter 5, Figure 5.1). I then merged the two independent cohort datasets and approximately 230K SNPs which were present on both genotyping platforms were extracted and used for SNP imputation against the Haplotype Reference Consortium v1.1 panel (https://imputation.sanger.ac.uk/). I used post-imputation data quality control steps in Plink v1.9 to filter out SNPs with: INFO score <0.70, posterior probability <0.90, genetic missingness >0.05, Hardy-Weinberg equilibrium p-value <1.0x10<sup>-7</sup> and minor allele frequency <0.01. Imputed SNP positions were based on Genome Reference Consortium Human 37/human genome version 19 (GRCh37/hg19).

**Replication cohort:** Genomic DNA was isolated from frozen cerebellar brain tissue from PSP cases using Autogen 245T (Holliston, MA) methods by Dr Owen Ross and Dr Rebecca Valentino (Mayo brain bank). rs2242367 was genotyped by Dr Rebecca Valentino using a TaqMan Allelic Discrimination Assay (assay ID: ANH6JHT) on a QuantStudio 7 Flex Real-

Time PCR System (Applied Biosystems, Foster City, CA) and genotypes were called using QuantStudio Real-Time PCR Software (version 1.1; Applied Biosystems). No evidence of deviation from the Hardy-Weinberg equilibrium was observed. Additionally, 136/238 (57%) of the cases used for analysis had rs2242367 genotype results available from whole-genome sequencing by which showed 100% concordance with TaqMan assay genotypes.

### eQTL data

I assessed the expression quantitative trait loci (eQTL) signals of all significant SNPs from the GWAS using FUMA (https://fuma.ctglab.nl), a platform to annotate, prioritise, visualise and interpret GWAS results (120). The following eQTL data sources are used as part of FUMA's eQTL analysis, with a more detailed description of tissue types and sample sizes available at https://fuma.ctglab.nl/tutorial#snp2gene: GTEx version 6, Blood eQTL browser, BIOS QTL browser, BRAINEAC, GTEx version 7, MuTHER, xQTLServer, CommonMind Consortium, eQTLGen, PsychENCODE, DICE, van der Wijst et al. scRNA eQTLs and GTEx version 8. Additionally, the North American Brain Expression Consortium (NABEC) eQTL dataset was analysed (dbGaP Study Accession: phs001300.v1.p1).

### **Colocalisation analyses**

Miss Regina Reynolds and Dr Mina Ryten (UCL Institute of Neurology) applied coloc version 3.1 (96) to evaluate the probability of colocalisation between PSP survival loci and eQTLs from eQTLGen (121) and PyschENCODE (122) datasets, which represent the largest human blood and brain expression datasets, respectively (eQTLGen, sample size = 31,684 individuals; PyschENCODE, sample size = 1,387 individuals). Notably, only *cis*-eQTLs from both datasets were used. They ran coloc using default parameters and priors, and restricted analyses to genes within 1Mb of a significant region of interest, as defined by PSP survival p <  $5 \times 10^{-8}$ ). All loci in which PP3 + PP4 < 0.8 were excluded, to exclude loci which were underpowered to detect colocalisation. Loci with PP4 ≥ 0.75 were considered colocalised

due to a single shared causal variant (PP4), as opposed to two distinct causal variants (PP3).

### WGS data

A subset (n=140) of PROSPECT study PSP cases of European ancestry underwent wholegenome sequencing (WGS) at the National Institutes of Health (NIH), primarily led by Dr Sonja Scholz and Dr Bryan Traynor. Blood-derived DNA samples were diluted to a target concentration of 50 ng/µl in Low-EDTA TE buffer (Quality Biological). PCR-free, paired-end, non-indexed libraries were constructed by automated liquid handlers using the Illumina TruSeq chemistry. DNA libraries were denatured and diluted to a final DNA concentration of 20 pM, followed by 'single library' - 'single lane' clustering on patterned flow cells using the Illumina cBot system. The libraries were sequenced on an Illumina HiSeq X Ten sequencer (v2.5 chemistry, Illumina) using 150 bp, paired-end cycles.

Raw genome data in FASTQ file format were transferred to Google Cloud Storage. Pairedend sequences were processed in accordance with the pipeline standard developed by the Centers for Common Disease Genomics (CCDG; https://www.genome.gov/27563570/). Sequence alignment to the human reference genome (GRCh38) was achieved using the Burrows-Wheeler Aligner (BWA-MEM) **(123)** and was followed by post-alignment processing for variant discovery using the Genome Analysis Toolkit Best Practices pipeline (GATK v4.1.2.0) **(124)**. The average sequencing read-depth after filtering by alignment quality was 35x. Single nucleotide and short indel variants were called from the processed WGS data using GATK HaplotypeCaller to generate individual gVCF files for each subject and followed by joint genotype calling of merged gVCFs to generate a multi-sample VCF file. Standard variant filtering was then performed using GATK4 Variant Quality Score Recalibration (VQSR) tools VariantRecalibrator and ApplyRecalibration.

For sample-level quality control checks, genomes with a high contamination rate (>5% based on VerifyBamID freemix metric) were excluded **(125)**. Next, they converted the multi-sample VCF-file to a PLINK 2.0 binary file (cog-genomics.org/plink/2.0) **(126)** and removed samples with an excessive heterozygosity rate (exceeding +/- 0.15 F-statistic), a sample call rate <95%, discordance between reported sex and genotypic sex, and duplicate samples (determined by pi-hat statistic). For variant-level quality control, they excluded variants with an overall missingness rate of >5%, non-autosomal variants and variants that significantly departed from Hardy-Weinberg equilibrium (p<1.0x10<sup>-10</sup>).

### **Statistical analyses**

In the discovery cohort I used a cox-proportional hazards survival model GWAS that adjusted for sex, age at motor symptom onset, PSP phenotype (PSP-RS or non-PSP-RS) and the first three principal components derived from PCA. Code for the cox-proportional hazards survival model GWAS was written by myself and Dr Maryam Shoai (UCL Institute of Neurology) and is available on GitHub (https://github.com/huw-morris-lab/PSP-survivalgwas). All reported significant SNPs had a Cox.zphP value >0.05, indicating that the model adhered to the assumption of proportional hazards. I conducted these analyses and created (with assistance from Dr Maryam Shoai and Miss Regina Reynolds) Figures 16a, 17 and 20 in R version 3.3.2. Figure 6.2b was created by Dr Maryam Shoai using LocusZoom version 0.10. Figure 6.4 was created by myself using Stata version 15. The Bonferroni-corrected threshold for genome-wide significance in the discovery cohort was set at  $p<1.0x10^{-8}$ . Dr Raph Gibbs (NIH, Bethesda) used GCTA-COJO to assess the independence of GWAS signals (127). In the replication cohort, I used the same cox-proportional hazards survival model as in the discovery cohort, with significance at p<0.05 as only one SNP (rs2242367) was analysed. I conducted the meta-analysis using the "metan" command in Stata version 15, which used the rs2242367 hazard ratio and 95% confidence interval from each cohort to give an overall hazard ratio, with heterogeneity defined as an I-squared value >50%, p<0.05.

### 6.4 Results

### **Discovery cohort**

In total, 1,033 cases were considered for inclusion in the study. Following post-imputation data quality control, 4,817,946 common (minor allele frequency  $\geq$ 1%) SNPs were used for analyses. 32 cases were excluded due to either insufficient clinical data or failing genotype data quality control, including one case that was found to have a previously described *MAPT* L284R mutation **(86)**. This resulted in 1,001 PSP cases of European ancestry (confirmed by genetic principal components analysis) for analyses, derived from two independent cohorts, with pathological confirmation of PSP achieved in 100% of deceased cases that had undergone post-mortem (n=841) **(Table 6.1)**. 895/1001 (89.4%) of the whole cohort were deceased at the date of censoring, with significant differences in disease duration observed between PSP-RS and non-PSP-RS cases, consistently across both cohorts **(Table 6.2)**.

Sample source Number		% of samples with post-mortem		
	samples	confirmation of PSP pathology		
UCL PSP cohort				
Queen Square Brain Bank, UK	248	100		
PROSPECT study, UK	180	11*		
Cambridge Brain Bank, UK	54	100		
Johns Hopkins Brain Bank, USA	37	100		
MRC London Neurodegenerative	25	100		
Diseases Brain Bank, UK				
Multiple Sclerosis and Parkinson's	21	100		
Brain Bank, UK				
Newcastle Brain Bank, UK	12	100		
2011 GWAS cohort				
Mayo Clinic Brain Bank, USA	340	100		
Munich Brain Bank, Germany	84	100		

## Table 6.1: PSP sample sources for survival GWAS

\*=74/180 PROSPECT study cases were deceased at the date of censoring. 20/74 had a

post-mortem examination, of which 20/20 had a pathological diagnosis of PSP.

20/180=11%.

### Table 6.2: Clinical characteristics and rs2242367 association statistics from PSP

	Whole cohort		UCL cohort		2011 GWAS cohort	
	PSP-	non-PSP-	PSP-	non-PSP-	PSP-	non-PSP-
	RS	RS	RS	RS	RS	RS
No. of subjects	720	281	395	182	325	99
% male subjects	59.2	51.2	62.3	51.1	55.4	51.5
Mean age at motor symptom	67.83	66.79	68.14	66.69	67.45	66.97
onset, years [range]	[45-90]	[45-86]	[45-87]	[45-86]	[46-90]	[48-83]
No. of cases deceased at the	652/720	243/281	327/395	144/182	325/325	99/99
date of censoring*, (%)	(90.6)	(86.5)	(82.8)	(79.1)	(100)	(100)
Mean disease duration in	6.38	9.81	6.06	9.35	6.69	10.48
deceased cases, years [SD]	[2.10]	[3.50]	[1.86]	[3.22]	[2.27]	[3.79]
rs2242367 risk allele	24.2	27.4	23.5	28.0	24.9	26.3
frequency**, %						
rs2242367 hazard ratio (95%	1.42 (1.26-1.60)		1.46 (1.25-1.71)		1.39 (1.19-1.63)	
CI), coefficient [SE]	0.35 [0.06]		0.38 [0.08]		0.33 [0.08]	
rs2242367 p-value	7.5x10 <sup>-10</sup>		1.4x10 <sup>−6</sup>		6.7x10⁻⁵	

### survival GWAS, stratified by cohort and PSP phenotype

Clinical characteristics of cases included in the PSP survival GWAS. Results are accompanied by either standard deviation (SD) or standard error (SE) measurements. \* = Date of censoring was 01/12/2019. \*\* = rs2242367 risk allele frequency in 7,692 reference controls of non-Finnish European ancestry was 28.2%, taken from <a href="https://gnomad.broadinstitute.org">https://gnomad.broadinstitute.org</a> on 10/01/2020.

### **Survival GWAS**

Using a cox proportional hazards survival model, I found no association between survival and previously identified individual PSP risk variants (p>0.05) after adjusting for sex, age at motor symptom onset and the first three principal components. There was an association between the *TRIM11/17* locus and survival (rs564309, p = 0.01, hazard ratio = 0.85), but this signal was attenuated when phenotype was added in as a binary (PSP-RS or non-PSP-RS) covariate. I then conducted a survival GWAS using sex, age at motor symptom onset, PSP phenotype and the first three principal components as covariates. The relationship between the covariates and survival are shown below **(Figure 6.1)**. The genomic inflation factor (lambda) was 1.05.





**Figure 6.1** – Kaplan-Meier survival curves for covariates (hazard ratio (95% confidence interval), p-value) used in the PSP survival GWAS. Top left = Sex (hazard ratio = 0.82 (0.72-0.94), p<0.01); Top right = PSP phenotype (hazard ratio = 3.39 (2.87-4.00), p<0.001); Bottom = Age at motor symptom onset (hazard ratio = 1.74 (1.33-2.27), p<0.001).

In the whole-cohort analysis I found a genome-wide significant association signal at chromosome 12 (Figure 6.2a). The lead SNP at this locus, rs2242367 (GRCh37 chr12:40413698), was associated with PSP survival –  $p = 7.5 \times 10^{-10}$ , hazard ratio = 1.42 (Table 6.2) (Table 6.3) (Figure 6.2b), with the minor allele associated with worsening survival (Figure 6.3). The association between rs2242367 and survival was replicated in each independent cohort (Table 6.2). A conditional analysis that controlled for rs2242367 genotype did not show other independent association signals. rs2242367 was not associated with PSP risk or PSP phenotype (Table 6.4).

 Table 6.3: Association statistics of genome-wide significant SNPs in the PSP survival

 GWAS

SNP	Genotyped or	Genotyping	Hazard ratio	p-value
	Imputed	success rate		
rs2242367	Genotyped	100%	1.42	7.5x10 <sup>-10</sup>
rs1542594	Imputed	N/A	1.42	7.5x10 <sup>-10</sup>
rs2128276	Imputed	N/A	1.42	8.9x10 <sup>-10</sup>
rs11174918	Imputed	N/A	1.40	2.4x10⁻⁰
rs10878029	Imputed	N/A	1.40	2.4x10⁻⁰
rs7967822	Imputed	N/A	1.39	5.9x10⁻⁰
rs1871895	Imputed	N/A	1.39	6.6x10⁻⁰
rs11564279	Imputed	N/A	1.39	8.3x10⁻⁰





highlighting LRRK2 association with PSP survival. a, Manhattan plot of whole-cohort PSP

survival GWAS, highlighting a genome-wide significant signal at chromosome 12. The red line indicates the Bonferroni-corrected threshold for genome-wide significance (p<1.0x10<sup>-8</sup>). **b**, Regional association plot of whole-cohort PSP survival GWAS, identifying rs2242367 as the lead SNP with associated eQTL signals for LRRK2. SNP positions, recombination rates and gene boundaries are based on GRCh37/hg19.



**Figure 6.3** – Kaplan-Meier survival curve for rs2242367 genotypes, highlighting differences in survival in PSP, comparing cases carrying rs2242367 GG genotype (red line) vs cases carrying rs2242367 AG and AA genotypes (green line).

### Table 6.4: rs2242367 association statistics in PSP case-control and phenotype GWAS'

	PSP case-control GWAS*	PSP phenotype GWAS**	
Odds ratio	0.95	0.80	
P-value	0.35	0.21	

\*= From phase 1 (pathologically diagnosed PSP cases) of Hoglinger *et al.,* 2011 (25) \*\*= From Jabbari *et al.,* 2018 (71)

### **Replication cohort and meta-analysis**

An additional 415 pathologically diagnosed PSP cases of European ancestry from Mayo brain bank were genotyped for rs2242367 using a TaqMan assay. 238/415 (57%) of these cases had adequate phenotype data to accurately assign a PSP phenotype and disease duration using the same methods as in the discovery cohort and were therefore analysed (53% male; mean age at onset = 67.6 years; mean disease duration = 6.5 years). Using the same Cox proportional hazards survival model as in the discovery cohort, I found that rs2242367 was associated with PSP survival — p = 0.049, hazard ratio 1.22. A meta-analysis of the rs2242367 association statistics in the three independent cohorts revealed an overall hazard ratio (95% CI) of 1.37 (1.25-1.51), with no evidence of heterogeneity between the cohorts (Figure 6.4). There appeared to be an additive allele / survival relationship in deceased cases across all three cohorts, and in both PSP-RS and non-PSP-RS cases (Table 6.5).

#### rs2242367 association statistics



**Figure 6.4** – Meta-analysis of rs2242367 association statistics in UCL, 2011 GWAS and Mayo cohorts. Individual cohort boxes indicate hazard ratio (size of box relates to weight (%) which is derived from size of cohort and 95% confidence interval of the hazard ratio) and horizontal lines indicate 95% confidence interval of the hazard ratio.

# Table 6.5: Impact of rs2242367 genotype on disease duration in deceased PSP cases

	UCL cohort		2011 GWAS cohort		Mayo cohort	
	PSP-RS	non-	PSP-RS	non-	PSP-RS	non-
		PSP-RS		PSP-RS		PSP-RS
No. of subjects	395	182	325	99	220	18
rs2242367 genotype						
counts (GG/AG/AA)	227/150/18	94/74/14	177/134/14	50/46/3	127/75/18	11/5/2
Mean disease	6.35	10.15	7.07	11.14	6.47	10.55
duration in deceased	[1.93]	[3.07]	[2.39]	[4.16]	[2.31]	[3.75]
cases with						
rs2242367 GG						
genotype, years [SD]						
Mean disease	5.73	8.72	6.24	9.85	5.88	9.00
duration in deceased	[1.73]	[3.08]	[2.07]	[3.26]	[2.25]	[2.92]
cases with						
rs2242367 AG						
genotype, years [SD]						
Mean disease	5.47	7.86	6.29	9.33	5.89	6.50
duration in deceased	[1.44]	[3.89]	[1.86]	[4.04]	[1.81]	[2.12]
cases with						
rs2242367 AA						
genotype, years [SD]						

rs2242367 major allele = G, rs2242367 minor allele = A, SD = standard deviation.

### Structure of the region

rs2242367 lies within intron 3 of *SLC2A13*, 190Kb from the common *LRRK2* risk SNP for Parkinson's disease (PD), rs76904798 (GRCh37 chr12:40614434, PD association statistics:  $p = 1.5x10^{-28}$ , odds ratio = 1.15) (128). The two variants are in low r<sup>2</sup> linkage disequilibrium in non-Finnish European populations, but with a high D' (r<sup>2</sup>=0.05, D'=0.90) (LDlink v3.8). Of note, rs2242367 had association statistics that approached genome-wide significance in the most recent PD case-control GWAS (p =  $1.6x10^{-7}$ , odds ratio = 0.94) (128). This region also contains association signals for two immune/inflammatory disease associations, Crohn's disease (129) and leprosy (130). None of the *LRRK2* locus lead SNPs from the PD, Crohn's disease and leprosy case-control GWAS' had significant association statistics in the PSP survival GWAS (Table 6.6). 

 Table 6.6: PSP survival GWAS association statistics for LRRK2 locus lead SNPs from

 Parkinson's disease, Crohn's disease and leprosy case-control GWAS'

	PD-associated	CD-associated SNP	Leprosy-associated		
	SNP (rs76904798)	(rs11175593)	SNP (rs1873613)		
Minor allele frequency (%)*	12.8	2.1	28.8		
r <sup>2</sup> with rs2242367**	0.05	0.01	0.09		
D' with rs2242367**	0.90	1.0	0.32		
PSP survival GWAS hazard ratio	0.96	1.11	NA		
PSP survival GWAS p-value	0.60	0.57	NA		

PD = Parkinson's disease, CD = Crohn's disease, NA = Not available

\*= Data taken from non-Finnish European cases in gnomAD browser

(https://gnomad.broadinstitute.org/)

\*\*= Data taken from non-Finnish European cases in LDlink

(https://ldlink.nci.nih.gov/?tab=ldpair)

#### eQTL datasets

There are no coding variants in linkage disequilibrium with the lead SNP, rs2242367, as defined by the region encompassed by variants with an r<sup>2</sup>>0.3 (Figure 6.5). Using FUMA (120), I found that rs2242367, and its associated tagging intronic SNPs, are strong eQTLs for LRRK2 expression in whole blood and oesophagus in eQTLGen and GTEXv8 and datasets (Figure 6.2b). Of note, the minor allele at rs2242367, which I show to be associated with a reduction in survival in PSP, was associated with an increase in LRRK2 expression. rs2242367 was also an eQTL for the long intergenic non-coding RNA (lincRNA) LINC02555 (ENSG0000260943, p= $3.3 \times 10^{-310}$ ) and AC079630.4 (ENSG00000223914, p= $5.8 \times 10^{-22}$ ) in the eQTLGen whole blood dataset (121). Additionally, I interrogated the NABEC dataset (213 human frontal cortex samples) and found no significant eQTL signals associated with rs2242367.



**Figure 6.5** – LD structure around rs2242367. Plot from LDlink (https://ldlink.nci.nih.gov/) highlighting that there are no coding variants in linkage disequilibrium with the lead SNP, rs2242367, as defined by the region encompassed by variants with an  $r^2>0.3$ . Blue dot = lead SNP rs2242367, yellow dots = non-coding SNPs, red dots = coding SNPs. Regulatory potential of each SNP (based on RegulomeDB scores -

http://www.regulomedb.org/help#score) indicated by numerical value between 1 (high) to 7 (low).

### **Colocalisation analyses**

As most GWAS loci are thought to operate by regulating gene expression (131) (132), a colocalisation analysis was conducted to evaluate the probability that the same causal SNP was responsible for modifying PSP survival and modulating gene expression. eQTLs were obtained from eQTLGen and PsychENCODE, the largest available human blood and brain eQTL datasets, respectively. The analyses identified a strong colocalisation between PSP survival loci and blood-derived eQTLs regulating the expression of the lincRNA, *LINC02555*, PP4 = 0.77 (Figure 6.6). No significant associations between PSP survival loci and eQTLs from PsychENCODE were found.



**Figure 6.6** – Colocalisation of blood-derived eQTLs regulating *LINC02555* expression and overlapping PSP survival loci. Plot of  $-\log_{10}(p\text{-values})$  for the region surrounding *LINC02555* shows colocalisation of blood-derived *LINC02555* eQTLs (top) and PSP survival signal (bottom) (PP4 = 0.77).

### WGS data

A subset of the UCL PSP cohort cases (n=140) underwent WGS. In these cases, the allele calls of the significant GWAS SNPs had 100% concordance between WGS and chip-based imputed datasets. I ran the Cox-proportional hazards survival model on variants with a minor allele frequency  $\geq 0.1\%$  in the region of interest (GRCh38 chr12:39225001-41369277) and

found 214 variants with a stronger association signal than rs2242367 (p<0.01). All of these were non-coding variants in linkage disequilibrium with rs2242367 ( $r^2$ =0.30-0.35, D'=0.96-1.0) and were strong eQTLs for LRRK2 expression in whole blood (p<1.0x10<sup>20</sup>) in the eQTLGen dataset. None of the variants were significant (FDR-corrected) eQTLs for LRRK2 expression in brain.

#### 6.5 Discussion

I have shown that common variation at the *LRRK2* locus is a genetic determinant of survival in the primary tauopathy, PSP. The replication of this signal in three independent cohorts (total n = 1,239) suggests that this finding is robust. A particular strength of this study is the fact that the outcome measure was disease duration that captures the entire clinical disease course, from motor symptom onset to death. This is in contrast to other neurodegenerative disease progression GWAS that have used longitudinal change in clinical rating scale scores (119) (133), which captures a specific time-point in the disease course and may be subject to inter- and intra-rater variability. Additionally, this study included patients across the phenotypic spectrum of PSP, i.e. those with the classical PSP-RS subtype as well as non-classical subtypes, and so I used subtype as a covariate in the model. Diagnostic accuracy was not a limitation in this study as a majority (87%) of PSP cases I included had pathological confirmation of PSP pathology at post-mortem. Of note, PSP risk loci from previous case-control GWAS did not show up as genome-wide significant signals in this PSP survival GWAS, a result which was been noted in PD (133) and sporadic amyotrophic lateral sclerosis (134).

*LRRK2* is an established major risk factor for PD with common (rs76904798) and rare (G2019S) variants associated with disease **(128)**. G2019S can cause familial PD, although often with reduced penetrance. Recently, the *LRRK2* PD risk-associated SNP, rs76904798, has also been nominally associated (p<0.05) with the rate of progression of PD **(133)**. My analyses suggest that the genetic signal at the *LRRK2* locus associated with PSP survival,

although in high D' LD, is distinct from the *LRRK2* locus signals identified in both PD and Crohn's disease case-control GWAS'. The allele frequency of rs2242367 is 28% and rs76904798 is 13% (GnomAD v2.1.1) so the *LRRK2* PD risk SNP may define a sub-haplotype of the ancestral PSP survival risk haplotype. Additionally, data from the subset of cases with WGS validates the genotype data used for this GWAS and suggests that the association signal at rs2242367 may be part of a larger haplotype block, paving the way for further studies using larger WGS datasets to clarify this.

LRRK2 is expressed in multiple human tissues including brain and whole blood. In brain, it is expressed ubiquitously across all regions and is found in neurons, astrocytes, microglia and oligodendroglia (135). Pathogenic mutations in *LRRK2* lead to phosphorylation of a subset of Rab proteins which have important roles in the formation and trafficking of intracellular vesicles (136). This in turn may affect proteostasis and the inflammatory response, both of which may be important in mediating disease progression in PSP.

There is also an established link between LRRK2 and tau pathology, which was reported in chromosome 12 linked PD families before the identification of the *LRRK2* gene (137). In rare cases, *LRRK2* mutations have been identified in patients with PSP-tau pathology at post-mortem (138). Recently, the link between LRRK2 and tau has been explored in cell and animal models, identifying dysregulation of actin and mitochondrial dynamics, and the impairment of tau degradation via the proteasome (139) (140) (141). The lead SNP from this GWAS, rs2242367, is an eQTL for controlling LRRK2 expression. However, the colocalisation analyses using multiple eQTL datasets from different tissue types did not show that LRRK2 eQTLs colocalised with the PSP survival signal. Instead, I found a colocalisation signal for the lincRNA, LINC02555, which is <500Kb proximal to *LRRK2*. It is well established that some lincRNAs control the expression of LRRK2 (142), and silencing of LINC02555 in papillary thyroid carcinoma cells has been shown to decrease LRRK2 expression and enhance autophagy in association with reduced tumour formation (143).

However, it should be noted that coloc assumes at most one "causal" variant per trait and when this assumption does not hold, coloc effectively tests whether the strongest signals for the two traits colocalise, which may be conservative **(144)**. In comparison to the eQTLGen dataset, the PSP survival GWAS is significantly underpowered and this is reflected in the disparity in y-axis scales between the two datasets **(Figure 6.6)**.

The strongest evidence for the effect on expression comes from blood rather than bulk RNA analysis from brain. This may relate to the cell types present and sample size differences in blood and brain expression datasets. In the PsychENCODE bulk RNA dataset, most of the power for detecting eQTL signals in the brain comes from astrocytes and neurons, with microglia representing only 5% of normalised cell fractions (122). Alternatively, the predominant effect on expression in blood may relate to peripheral immune response-driven neuroinflammation (145) or be due to a specific effect in monocyte/microglial lineage cells. The advent of expression analysis in defined brain cell subpopulations will clarify this issue. Increased LRRK2 expression may result in a reactive microglia-induced pro-inflammatory state which drives ongoing accumulation of misfolded tau protein and clinical disease progression (146) (147). This hypothesis is supported by *in vivo* positron emission tomography evidence of a pattern of microglial activation which correlates with disease severity in PSP (148) (149), and co-localises with the tauopathy of PSP and other forms of frontotemporal lobar degeneration tauopathy (150). In mouse models, microglial inflammatory responses are attenuated by LRRK2 inhibition (147) and this strategy is currently under investigation in PD as a disease-modifying therapy.

Future studies on the *LRRK2* locus as a potential genetic determinant of disease progression and survival in related tauopathies, namely AD, frontotemporal degeneration and corticobasal degeneration, are of great importance. Additionally, this study paves the way for further functional studies assessing the impact of LRRK2 on tau aggregation and exploration of LRRK2 inhibition as a therapeutic approach in patients with tauopathies.

### Chapter 7 – Conclusions and future directions

Neurodegenerative diseases are a common and debilitating cause of morbidity and mortality in the United Kingdom, and represent a significant economic cost to the National Health Service. Prior to this PhD, based on studying only its classical clinical phenotype, it was estimated that PSP had a prevalence of 5-7 per 100,000 of the population **(1) (2)**, with healthcare costs previously estimated at £43,000 per patient/per year **(151)**. Through my work on the PROSPECT study, I have shown that up to 50% of PSP cases present with nonclassical phenotypes such as PSP-P and therefore, although up-to-date epidemiological studies are required, we can infer that PSP may be twice as common as previously thought. The importance of PSP research is further highlighted by the fact that it shares clinical, genetic and pathological features with more common neurodegenerative diseases, AD and FTD. Indeed, PSP is often seen as a "testbed" for AD, with respect to tau directed therapies. In comparison to AD, due to its relatively rapid rate of disease progression, clinical trials in PSP are likely to require smaller numbers of participants and a shorter duration of time to see a positive therapeutic effect.

The first aim of this PhD was to enhance the early diagnosis of PSP by prospectively characterising the profiles of various PSP subtypes using data from the PROSPECT study (Chapter 2). By stratifying patients into PSP subtypes using the 2017 MDS diagnostic criteria, I found that subcortical presentations of PSP had a more benign clinical trajectory in comparison to cortical presentations of PSP and PSP-RS. When considering markers that allow early differentiation of PSP-subcortical from PD (a commonly encountered clinical dilemma in movement disorders clinics), I found that serum NF-L and the degree of cognitive impairment using the MoCA were especially helpful. The early differentiation of PSP-cortical from AD and FTD is an important future study that was outside of the scope of this PhD. A promising potential marker of this is midbrain atrophy, which for the first time was shown to be present across the phenotypic spectrum of PSP in comparison to controls, although I was

unable to assess midbrain volumes in PD and other neurodegenerative diseases. The limited pathological data from PROSPECT study cases that I had access to suggest that non-classical PSP presentations, as defined in the MDS diagnostic criteria, predict underlying PSP pathology. As the number of cases coming to post-mortem increase in the near future, the PROSPECT study is ideally placed to validate the MDS PSP diagnostic criteria. Additionally, as more longitudinal PROSPECT study data is gathered, I expect the diagnostic accuracy of PSP to further improve with a well powered multivariate approach, including cross-validated machine learning algorithms.

The second aim of this PhD was to enhance PSP clinical trial design through the development of effective diagnostic biomarkers, and biomarkers that predict and track disease progression. To date, there have been no disease-modifying therapies for PSP patients which have been shown to be effective in clinical trials. The following factors impact on trial outcomes and can be enhanced by the use of robust biomarkers in conjunction with clinical assessment: 1) diagnostic accuracy of cases; 2) diagnosis of cases in early disease stage when therapeutics are more likely to be effective; 3) minimising heterogeneity within groups, e.g. fast vs slow progressors; 4) the use of objective biomarkers that change in conjunction with clinical disease progression and in response to a therapeutic effect.

With regards to diagnostic accuracy and early-stage diagnosis, there have been many examples of previous studies which have shown that the levels of blood and CSF NF-L reliably differentiate PSP from PD. However, the levels of NF-L do not differentiate PSP from related atypical parkinsonian disorders, CBD and MSA. My work in chapters 2 and 3 validate the above profile of blood and CSF NF-L in atypical parkinsonian disorders and PD. However, in chapter 3 I also identified potential novel fluid biomarkers of inflammation associated with atypical parkinsonian disorders using ultra-sensitive proximity extension
assay (PEA). These findings, and the assay itself, will need to replicated/validated in independent cohorts which also compare PEA to single molecule array and standard ELISA. Although I was underpowered to detect disease-specific markers, as recruitment continues, the PROSPECT study will be ideally placed and adequately powered to meet this challenge. A particular strength of the PROSPECT study is its inclusion of an "indeterminate" group of cases that have features of an atypical parkinsonian syndrome but do not fulfil diagnostic criteria on entry to the study. In the future, studying the baseline biomarker profiles of indeterminate cases that go on to fulfil diagnostic criteria has the potential to discover markers of early stage disease. As genetic markers can be thought of as static, my findings in chapter 4 suggest that a genetic risk score, based on known PSP risk variants, may be able to differentiate PSP from PD at all disease stages. However, the diagnostic utility of this cannot be fully assessed until we are able to measure the PSP genetic risk score in CBD and MSA cases, and measure a PD genetic risk score in PSP cases.

With regards to predicting and tracking disease progression, my work on NF-L in chapters 2 and 3 show that baseline levels in both blood and CSF are able predict survival in PSP and may also differentiate fast and slow progressing subtypes, i.e. PSP-RS vs PSP-P. The application of this to clinical trials has the potential to increase the power of trials with more homogenous disease groups. Again, as more serial samples are collected, the PROSPECT study is ideally placed to investigate the longitudinal change of NF-L levels and its correlation with clinical disease progression. In the meantime, we eagerly await the results of recent tau-antibody trials (18) which, although shown to have no effect on disease progression, measured NF-L at baseline and 1 year in a subset of patients. As the current era of potential novel therapeutics for neurodegenerative diseases continues to develop, I envisage that in the near future blood NF-L will be readily available as a quick and noninvasive marker of treatment response. The potential of this approach has already been explored in multiple sclerosis (44).

The third aim of this PhD was to discover the genetic determinants of PSP phenotype and survival using a hypothesis-free GWAS approach, which has the potential to identify disease modifying therapeutic targets. The logic behind conducting the phenotype GWAS in chapter 5 was that PSP phenotype can be considered a surrogate marker of rate of progression, as we've seen in the PROSPECT study where the subcortical subtype of PSP progresses at a slower rate vs PSP-RS. Therefore, I focused on PSP-RS, PSP-P and PSP-PGF phenotypes and found that the TRM11/17 locus was a genetic determinant of phenotype. This adds further evidence for the role of the ubiquitin proteasome system in tau pathology, which is currently being explored as target for disease modifying therapies (152). Although further replication of this finding with larger sample sizes is required, it is especially pleasing that further work on this locus is currently being undertaken by Prof Karen Duff's group at the UCL Dementia Research Institute. This work will explore the functional effects of rs564309 as these are currently unknown - my study did not identify any significant coding variants or effects on gene expression in brain. In addition, the impact of the TRIM11/17 locus on phenotype/progression has been recently explored in a small number of CBD cases (153), and I would expect a similar approach to be used in AD and FTD in the near future, although it should be noted that in-depth phenotype data is required to carry out such studies accurately.

Undoubtedly, the work I am most proud of is the PSP survival GWAS in chapter 6. I felt it was important to take a different approach to the phenotype GWAS and study survival (defined as motor symptom onset to death) as this represented a "hard" end-point of disease progression that was quantifiable, not subject to inter- or intra-rater variability and captured the entire disease course. I discovered that the *LRRK2* locus is a genetic determinant of PSP survival in three independent cohorts (consisting of a total of 1,239 cases). This finding shows that there is shared genetic risk between Parkinson's disease and PSP and builds on previous work that has shown a link between *LRRK2* mutations and tau pathology. In the

future, as with the findings from the phenotype GWAS, it will be very important to assess the impact of the *LRRK2* locus on the progression of other tauopathies such as AD, FTD and CBD.

The strongest evidence for an effect on LRRK2 expression came from whole blood as opposed to bulk RNA analysis from brain. To answer the question of whether this is simply a reflection of the large difference in blood vs brain eQTL dataset sample sizes, single cell RNA analysis from brain will be vital to look for cell-specific (e.g. neuronal vs glial) effects on gene expression. If the effect on LRRK2 expression genuinely is restricted to blood then this may impact on the activation of monocyte-derived microglia, and builds on existing evidence that reactive microglia-induced neuroinflammation contributes to the progression of neurodegenerative diseases **(154)**. To investigate this further, I plan on measuring CSF levels of sTREM2 (a marker of microglial activation) **(155)** in PROSPECT study PSP cases, comparing fast vs slow progressors. Additionally, assessing the functional impact of LRRK2 modulation on tau pathology in induced pluripotent stem cell models will be extremely important, as will assessing the impact of LINC02555 on LRRK2 expression – I intend on contributing to the writing of a research grant to carry out these experiments. Ultimately, these follow-up studies may provide evidence for modulation of LRRK2 as a disease modifying therapy in PSP and other tauopathies, and pave the way for future clinical trials.

## **References**

(1) Schrag A, Ben-Shlomo Y, Quinn NP. Prevalence of progressive supranuclear palsy and multiple system atrophy: a cross-sectional study. *Lancet* 1999; 354:1771-1775.

(2) Coyle-Gilchrist IT, Dick KM, Patterson K, et al. Prevalence, characteristics, and survival of frontotemporal lobar degeneration syndromes. *Neurology* 2016; 86:1736-1743.

(3) Spillantini MG, Goedert M. Tau pathology and neurodegeneration. *Lancet Neurol* 2013;12:609-622.

(4) Walsh DM, Selkoe DJ. A critical appraisal of the pathogenic protein spread hypothesis of neurodegeneration. *Nat Rev Neurosci* 2016; 17:251-260.

(5) Steele JC, Richardson JC, Olszewski J. 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; 10:333-359.

(6) Litvan I, Agid Y, Calne D, et al. Clinical research criteria for the diagnosis of progressive supranuclear palsy (Steele-Richardson-Olszewski syndrome): report of the NINDSSPSP international workshop. *Neurology* 1996; 47:1-9.

(7) Nath U, Ben-Shlomo Y, Thomson RG, et al. Clinical features and natural history of progressive supranuclear palsy: a clinical cohort study. *Neurology* 2003; 60:910-916.
(8) Williams DR, de Silva R, Paviour DC, et al. Characteristics of two distinct clinical phenotypes in pathologically proven progressive supranuclear palsy: Richardson's syndrome and PSP-parkinsonism. *Brain* 2005; 128:1247-1258.

(9) Williams DR, Holton JL, Strand C, et al. Pure akinesia with gait freezing: a third clinical phenotype of progressive supranuclear palsy. *Mov Disord* 2007; 22:2235-2241.

(10) Williams DR, Holton JL, Strand C, et al. Pathological tau burden and distribution distinguishes progressive supranuclear palsy parkinsonism from Richardson's syndrome. *Brain* 2007; 130:1566-1576.

(11) Hoglinger GU, Respondek G, Stamelou M, et al. Clinical diagnosis of progressive supranuclear palsy: the Movement Disorder Society Criteria. *Mov Disord* 2017; 32:853-864.
(12) Hughes AJ, Daniel SE, Ben-Shlomo Y, et al. The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. *Brain* 2002; 125:861-870.
(13) Whitwell JL, Hoglinger GU, Antonini A, et al. Radiological biomarkers for diagnosis in PSP: Where are we and where do we need to be? *Mov Disord* 2017; 32:955-971.

(14) Golbe LI, Ohman-Strickland PA. A clinical rating scale for progressive supranuclear palsy. *Brain* 2007; 130:1552-1565.

(15) Stamelou M, Giagkou N, Hoglinger GU. One decade ago, one decade ahead in progressive supranuclear palsy. *Mov Disord* 2019; 34:1284-1293.

(16) Boxer AL, Lang AE, Grossman M, et al. Davunetide in patients with progressive supranuclear palsy: a randomised, double-blind, placebo-controlled phase 2/3 trial. *Lancet Neurol* 2014; 13: 676-685.

(17) Tolosa E, Litvan I, Höglinger GU, et al. A phase 2 trial of the GSK-3 inhibitor tideglusib in progressive supranuclear palsy. *Mov Disord* 2014; 29: 470-478.

(18) Boxer AL, Qureshi I, Ahlijanian M, et al. Safety of the tau-directed monoclonal antibody BIIB092 in progressive supranuclear palsy: a randomised, placebo-controlled, multiple ascending dose phase 1b trial. *Lancet Neurol* 2019; 18: 549-558.

(19) Im SY, Kim YE, Kim YJ. Genetics of progressive supranuclear palsy. *J Mov Disord* 2015; 8:122-129.

(20) Stanford PM, Halliday GM, Brooks WS, 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; 123:880-893.

(21) Zimprich A, Biskup S, Leitner P, et al. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. *Neuron* 2004; 44:601-607.

(22) Stamelou M, Quinn NP, Bhatia KP. "Atypical" atypical parkinsonism: new genetic conditions presenting with features of progressive supranuclear palsy, corticobasal degeneration, or multiple system atrophy—a diagnostic guide. *Mov Disord* 2013; 28:1184-1199.

(23) Ros R, Gomez Garre P, Hirano M, et al. Genetic linkage of autosomal dominant progressive supranuclear palsy to 1q31.1. *Ann Neurol* 2005; 57:634-641.

(24) Baker M, Litvan I, Houlden H, et al. Association of an extended haplotype in the tau gene with progressive supranuclear palsy. Hum Mol Genet 1999; 8:711-715.
(25) Hoglinger GU, Melhem NM, Dickson DW, et al. Identification of common variants influencing risk of the tauopathy progressive supranuclear palsy. *Nat Genet* 2011; 43:699-705.

(26) Kouri N, Ross OA, Dombroski B, et al. Genome-wide association study of corticobasal degeneration identifies risk variants shared with progressive supranuclear palsy. *Nat Commun* 2015; 6:7247.

(27) Chen JA, Chen Z, Won H, et al. Joint genome-wide association study of progressive supranuclear palsy identifies novel susceptibility loci and genetic correlation to neurodegenerative diseases. *Mol Neurodegener* 2018; 13:41.

(28) Bruch J, Xu H, Rosler TW, et al. PERK activation mitigates tau pathology in vitro and in vivo. *EMBO Mol Med* 2017; 9:371-384.

(29) Halliday M, Hughes D, Mallucci GR. Fine-tuning PERK signalling for neuroprotection. *J Neurochem* 2017; 142:812-826.

(30) Meredith Jr. JE, Sankaranarayanan S, Guss V, et al. Characterization of novel CSF tau and ptau biomarkers for Alzheimer's Disease. *PLoS One* 2013; 8:10.

(31) Jabbari E, Zetterberg H, Morris HR. Tracking and predicting disease progression in progressive supranuclear palsy: CSF and blood biomarkers. *J Neurol Neurosurg Psychiatry* 2017; 88:883-888.

(32) Süssmuth SD, Uttner I, Landwehrmeyer B, et al. Differential pattern of brain specific CSF proteins tau and amyloid- $\beta$  in parkinsonian syndromes. *Mov Disord* 2010; 25:1284-1288.

(33) Hall S, Öhrfelt A, Constantinescu R, et al. Accuracy of a panel of 5 cerebrospinal fluid biomarkers in the differential diagnosis of patients with dementia and/or parkinsonian disorders. *Arch Neurol* 2012; 69:1445.

(34) Olsson B, Lautner R, Andreasson U, et al. CSF and blood biomarkers for the diagnosis of Alzheimer's disease: a systematic review and meta-analysis. *Lancet Neurol* 2016; 15:673-684.

**(35)** Wagshal D, Sankaranarayanan S, Guss V, et al. Divergent CSF τ alterations in two common tauopathies: alzheimer's disease and progressive supranuclear palsy. *J Neurol Neurosurg Psychiatry* 2015; 86:244-250.

(36) Zetterberg H, Wilson D, Andreasson U, et al. Plasma tau levels in Alzheimer's disease. *Alzheimers Res Ther* 2013; 5:9.

(37) Gaye B, Sikkema D, Lee TN. Development of an ultra-sensitive single molecule counting assay for the detection of interleukin-13 as a marker for asthmatic severity. *J Immunol Methods* 2015; 426:82-85.

(38) Assarsson E, Lundberg M, Holmquist G, et al. Homogenous 96-plex pea immunoassay exhibiting high sensitivity, specificity, and excellent scalability. *PLoS One* 2014; 9:e95192.
(39) Blennow K, Hampel H, Weiner M, et al. Cerebrospinal fluid and plasma biomarkers in alzheimer disease. *Nat Rev Neurol* 2010; 6:131-144.

(40) Hall S, Öhrfelt A, Constantinescu R, et al. Accuracy of a panel of 5 cerebrospinal fluid biomarkers in the differential diagnosis of patients with dementia and/or parkinsonian disorders. *Arch Neurol* 2012; 69:1445.

(41) Hansson O, Janelidze S, Hall S, et al. Blood-based NfL: A biomarker for differential diagnosis of parkinsonian disorders. *Neurology* 2017; 88:930-937.

(42) Meeter LH, Dopper EG, Jiskoot LC, et al. Neurofilament light chain: a biomarker for genetic frontotemporal dementia. *Ann Clin Transl Neurol* 2016; 3:623-636.

(43) Byrne LM, Rodrigues FB, Blennow K, et al. Neurofilament light protein in blood as a potential biomarker of neurodegeneration in Huntington's disease: a retrospective cohort analysis. *Lancet Neurol* 2017; 16:601-609.

(44) Kuhle J, Kropshofer H, Haering DA, et al. Blood neurofilament light chain as a biomarker of MS disease activity and treatment response. *Neurology* 2019; 92:e1007-e1015.

(45) Magdalinou NK, Paterson RW, Schott JM, et al. A panel of nine cerebrospinal fluid biomarkers may identify patients with atypical parkinsonian syndromes. *J Neurol Neurosurg Psychiatry* 2015; 86:1240-1247.

(46) Rojas JC, Karydas A, Bang J, et al. Plasma neurofilament light chain predicts progression in progressive supranuclear palsy. *Ann Clin Transl Neurol* 2016; 3:216-225.
(47) Testa D, Monza D, Ferrarini M, et al. Comparison of natural histories of progressive supranuclear palsy and multiple system atrophy. *Neurol Sci* 2001; 22:247-251.

(48) Armstrong MJ, Litvan I, Lang AE, et al. Criteria for the diagnosis of corticobasal degeneration. *Neurology* 2013; 80:496-503.

(49) Ling H, O'Sullivan SS, Holton JL, et al. Does corticobasal degeneration exist? A clinicopathological re-evaluation. *Brain* 2010; 133:2045-2057.

**(50)** Alexander SK, Rittman T, Xuereb JH, et al. Validation of the new consensus criteria for the diagnosis of corticobasal degeneration. *J Neurol Neurosurg Psychiatry* 2014; 85:925-929.

(51) Gilman S, Wenning GK, Low PA, et al. Second consensus statement on the diagnosis of multiple system atrophy. *Neurology* 2008; 71:670-676.

(52) Malek N, Swallow DMA, Grosset KA, et al. PRoBaND. Tracking Parkinson's: study design and baseline patient data. *J Parkinsons Dis* 2015; 5:947-959.

**(53)** Hughes AJ, Daniel SE, Lees AJ. Improved accuracy of clinical diagnosis of Lewy body Parkinson's disease. *Neurology* 2001; 57:1497-1499.

**(54)** Grimm MJ, Respondek G, Stamelou M, et al. Movement Disorder Society-endorsed PSP Study Group. How to apply the movement disorder society criteria for diagnosis of progressive supranuclear palsy. *Mov Disord* 2019; 34:1228-1232.

**(55)** Wenning GK, Tison F, Seppi K, et al. Multiple System Atrophy Study Group. Development and validation of the Unified Multiple System Atrophy Rating Scale (UMSARS).

Mov Disord 2004; 19:1391-1402.

(56) Goetz CG, Tilley BC, Shaftman SR, et al. Movement Disorder Society UPDRS Revision Task Force. Movement Disorder Society–sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): scale presentation and clinimetric testing results. *Mov Disord* 2008; 23:2129-2170.

(57) Schwab JF, England AC. Projection technique for evaluating surgery in Parkinson's disease. In: Gillingham FJ, Donaldson MC, eds. *Proceedings of the Third Symposium on Parkinson's Disease* 1969; Edinburgh, UK: Churchill Livingstone; 1969.

**(58)** Nasreddine ZS, Phillips NA, Bedirian V, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J Am Geriatr Soc* 2005; 53:695-699.

(59) Mioshi E, Dawson K, Mitchell J, et al. The Addenbrooke's Cognitive Examination Revised (ACE-R): a brief cognitive test battery for dementia screening. *Int J Geriatr Psychiatry* 2006; 21:1078-1085.

(60) Abrahams S, Newton J, Niven E, et al. Screening for cognition and behaviour changes in ALS. *Amyotroph Lateral Scler Frontotemporal Degener* 2014; 15:9-14.

(61) Kuhle J, Barro C, Andreasson U, et al. Comparison of three analytical platforms for quantification of the neurofilament light chain in blood samples: ELISA, electrochemiluminescence immunoassay and Simoa. *Clin Chem Lab Med* 2016; 54:1655-1661.

**(62)** Faull M, Ching SY, Jarmolowicz AI, et al. Comparison of two methods for the analysis of CSF Aβ and tau in the diagnosis of Alzheimer's disease. *Am J Neurodegener Dis* 2014; 3:143-151.

(63) Borroni B, Premi E, Agosti C, et al. CSF Alzheimer's disease-like pattern in corticobasal syndrome: evidence for a distinct disorder. *J Neurol Neurosurg Psychiatry* 2011; 82:834-838.
(64) Ritchie C, Smailagic N, Noel-Storr AH, et al. CSF tau and the CSF tau/ABeta ratio for the diagnosis of Alzheimer's disease dementia and other dementias in people with mild cognitive impairment (MCI). *Cochrane Database Syst Rev* 2017; 3:CD010803.

(65) Blauwendraat C, Faghri F, Pihlstrom L, et al. NeuroChip, an updated version of the NeuroX genotyping platform to rapidly screen for variants associated with neurological diseases. *Neurobiol Aging* 2017; 57:e9-e247.

(66) Guo Y, He J, Zhao S, et al. Illumina human exome genotyping array clustering and quality control. *Nat Protoc* 2014; 9:2643-2662.

(67) Reed E, Nunez S, Kulp D, et al. A guide to genome-wide association analysis and postanalytic interrogation. *Stat Med* 2015; 34:3769-3792.

(68) Rohrer JD, Nicholas JM, Cash DM, et al. Presymptomatic cognitive and neuroanatomical changes in genetic frontotemporal dementia in the Genetic Frontotemporal dementia Initiative (GENFI) study: a cross-sectional analysis. *Lancet Neurol* 2015; 14:253-262.

(69) Fischl B, Salat DH, Busa E, et al. Whole brain segmentation: automated labelling of neuroanatomical structures in the human brain. *Neuron* 2002; 33:341-355.

(70) Iglesias JE, Van Leemput K, Bhatt P, et al. Bayesian segmentation of brainstem structures in MRI. *Neuroimage* 2015; 113:184-195.

(71) Jabbari E, Woodside J, Tan MMX, et al. Variation at the *TRIM11* locus modifies progressive supranuclear palsy phenotype. *Ann Neurol* 2018; 84:485-496.

(72) Dutt S, Binney RJ, Heuer HW, et al; AL-108-231 investigators. Progression of brain atrophy in PSP and CBS over 6 months and 1 year. *Neurology* 2016; 87:2016-2025.
(73) Respondek G, Stamelou M, Kurz C, et al. Movement Disorder Society–Endorsed PSP Study Group. The phenotypic spectrum of progressive supranuclear palsy: a retrospective multicentre study of 100 definite cases. *Mov Disord* 2014; 29:1758-1766.

(74) Santangelo G, Cuoco S, Pellecchia MT, et al. Comparative cognitive and neuropsychiatric profiles between Parkinson's disease, multiple system atrophy and progressive supranuclear palsy. *J Neurol* 2018; 265:2602-2613.

(75) Dickson DW, Ahmed Z, Algom AA, et al. Neuropathology of variants of progressive supranuclear palsy. *Curr Opin Neurol* 2010; 23:394-400.

(76) Seppälä TT, Nerg O, Koivisto AM, et al. CSF biomarkers for Alzheimer disease correlate with cortical brain biopsy findings. *Neurology* 2012; 78:1568-1575.

(77) Lagerstedt L, Egea-Guerrero JJ, Rodríguez-Rodríguez A, et al. Early measurement of interleukin-10 predicts the absence of CT scan lesions in mild traumatic brain injury. *PLoS One* 2018; 13:e0193278.

(78) Tighe PJ, Ryder RR, Todd I, et al. ELISA in the multiplex era: potentials and pitfalls. *Prot Clin Appl* 2015; 9:406-422.

**(79)** Alcolea D, Irwin DJ, Illan-Gala I, et al. Elevated YKL-40 and low sAPPβ:YKL-40 ratio in antemortem cerebrospinal fluid of patients with pathologically confirmed FTLD. *J Neurol Neurosurg Psychiatry* 2019; 90:180-186.

(80) Carmona S, Zahs K, Wu E, et al. The role of TREM2 in Alzheimer's disease and other neurodegenerative disorders. *Lancet Neurol* 2018; 17:721-730.

**(81)** Jabbari E, Holland N, Chelban V, et al. Diagnosis across the spectrum of progressive supranuclear palsy and corticobasal syndrome. *JAMA Neurol* 2020; 77:377-387.

(82) Athauda D, Maclagan K, Skene SS, et al. Exenatide once weekly versus placebo in Parkinson's disease: a randomised, double-blind, placebo-controlled trial. *Lancet* 2017; 390:1664-1675.

(83) Hochberg Y, Benjamini Y. More powerful procedures for multiple significance testing. *Stat Med* 1990; 9:811-818.

(84) De Strooper B, Annaert W, Cupers P, et al. A presenilin-1-dependent gammasecretase-like protease mediates release of Notch intracellular domain. *Nature* 1999; 398:518-522.
(85) Morris HR, Osaki Y, Holton J, et al. Tau exon 10 +16 mutation FTDP-17 presenting clinically as sporadic young onset PSP. *Neurology* 2003; 61:102-104.

(86) Rohrer JD, Paviour D, Vandrovcova J, et al. Novel L284R MAPT mutation in a family with an autosomal dominant progressive supranuclear palsy syndrome. *Neurodegener Dis* 2011; 8:149-152.

(87) Imrie J, Heptinstall L, Knight S, et al. Observational cohort study of the natural history of Niemann-Pick disease type C in the UK: a 5-year update from the UK clinical database. *BMC Neurol* 2015; 15:257.

(88) Hernandez DG, Reed X, Singleton AB. Genetics in Parkinson's disease: Mendelian vs. non-Mendelian inheritance. *J Neurochem* 2016; 139:59-74.

(89) Mehanna R, Moore S, Hou G, et al. Comparing clinical features of young onset, middle onset and late onset Parkinson's disease. *Parkinsonism Relat Disord* 2014; 20:530-534.

(90) Batla A, De Pablo-Fernandez E, Erro R, et al. Young-onset multiple system atrophy: Clinical and pathological features. *Mov Disord* 2018; 33:1099-1107.

(91) Donker Kaat L, Boon AJ, Azmani A, et al. Familial aggregation of parkinsonism in progressive supranuclear palsy. *Neurology* 2009; 73:98-105.

(92) Kempster PA, O'Sullivan SS, Holton JL, et al. Relationships between age and late progression of Parkinson's disease: a clinico-pathological study. *Brain* 2010; 133:1755-1762.
(93) Robinson JL, Lee EB, Xie SX, et al. Neurodegenerative disease concomitant proteinopathies are prevalent, age-related and APOE4-associated. *Brain* 2018; 141:2181-2193.

(94) de Leeuw CA, Mooij JM, Heskes T, et al. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol* 2015; 11:e1004219.

(95) Maston GA, Evans SK, Green MR. Transcriptional regulatory elements in the human genome. *Annu Rev Genomics Hum Genet* 2006; 7:29-59.

(96) Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PLoS Genet* 2014; 10:e1004383.

(97) Ramasamy A, Trabzuni D, Guelfi S, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. *Nat Neurosci* 2014; 17:1418-1428.

(98) Hawrylycz MJ, Lein ES, Guillozet-Bongaarts AL, et al. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature* 2012; 489:391-399.

(99) Zhang Y, Sloan SA, Clarke LE, et al. Purification and characterisation of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron* 2016; 89:37-53.

(100) Saunders A, Macosko E, Wysoker A, et al. A single-cell atlas of cell types, states, and other transcriptional patterns from nine regions of the adult mouse brain. *bioRxiv* doi: https://doi.org/10.1101/299081.

(101) Marone M, Mozzetti S, De Ritis D, et al. Semiquantitative RT-PCR analysis to assess the expression levels of multiple transcripts from the same sample. *Biol Proced Online* 2001; 3:19-25.

(102) Young P, Ehler E, Gautel M. Obscurin, a giant sarcomeric Rho guanine nucleotide exchange factor protein involved in sarcomere assembly. *J Cell Biol* 2001; 154:123-136.
(103) Murthy MN, Blauwendraat C, Guelfi S, et al. Increased brain expression of GPNMB is associated with genome wide significant risk of Parkinson's disease. *Neurogenetics* 2017; 18:121-133.

(104) Lambert JC, Ibrahim-Verbaas CA, Harold D, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nat Genet* 2013; 45:1452-1458.

(105) Chang D, Nalls MA, Hallgrimsdottir IB, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet* 2017; 49:1511-1516.

(106) de Bakker PIW, Ferreira MAR, Jia X, et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet* 2008; 17(R2):R122-128.

(107) Coetzee SG, Pierce S, Brundin P, et al. Enrichment of risk SNPs in regulatory regions implicate diverse tissues in Parkinson's disease aetiology. *Sci Rep* 2016; 6:30509.

(108) Galvin JE. Neurodegenerative diseases: pathology and the advantage of single-cell profiling. *Neurochem Res* 2004; 29:1041-1051.

(109) Hatakeyama S. TRIM family proteins: roles in autophagy, immunity and carcinogenesis. *Trends Biochem Sci* 2017; 42:297-311.

(110) Chen L, Brewer MD, Guo L, et al. Enhanced degradation of misfolded proteins promotes tumorigenesis. *Cell Rep* 2017; 18:3143-3154.

(111) Kontaxi C, Piccardo P, Gill AC. Lysine-directed post-translational modifications of tau protein in Alzheimer's disease and related tauopathies. *Front Mol Biosci* 2017; 4:56.

(112) Myeku N, Clelland CL, Emrani S, et al. Tau-driven 26S proteasome impairment and cognitive dysfunction can be prevented early in disease by activating cAMP-PKA signalling. *Nat Med* 2016; 22:46-53.

(113) Ying Z, Wang H, Wang G. The ubiquitin proteasome system as a potential target for the treatment of neurodegenerative diseases. *Curr Pharm Des* 2013; 19:3305-3314.

(114) Gong B, Radulovic M, Figueiredo-Pereira ME, et al. The ubiquitin proteasome system: potential therapeutic targets for Alzheimer's disease and spinal cord injury. *Front Mol Neurosci* 2016; 9:4.

(115) Lassot I, Robbins I, Kristiansen M, et al. Trim17, a novel E3 ubiquitin-ligase, initiates neuronal apoptosis. *Cell Death Differ* 2010; 17:1928-1941.

(116) Magiera MM, Mora S, Mojsa B, et al. Trim17-mediated ubiquitination and degradation of Mcl-1 initiate apoptosis in neurons. *Cell Death Differ* 2013; 20:281-292.

(117) Clavaguera F, Akatsu H, Fraser G, et al. Brain homogenates from human tauopathies induce tau inclusions in mouse brain. *Proc Natl Acad Sci* 2013; 110:9535-9540.

(118) Mudher A, Colin M, Dujardin S, et al. What is the evidence that tau pathology spreads through prion-like propagation? *Acta Neuropathol Commun* 2017; 5:99.

(119) Moss DJH, Pardiñas AF, Langbehn D, et al. Identification of genetic variants associated with Huntington's disease progression: a genome-wide association study. *Lancet Neurol* 2017; 16:701-711.

(120) Watanabe K, Taskesen E, van Bochoven A, et al. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017; 8:1826.

(121) Võsa U, Claringbould A, Westra H-J, et al. Unravelling the polygenic architecture of complex traits using blood eQTL metaanalysis. *bioRxiv* 2018; p.447367. Available from: https://www.biorxiv.org/content/10.1101/447367v1.

(122) Wang D, Liu S, Warrell J, et al. Comprehensive functional genomic resource and integrative model for the human brain. *Science* 2018; 362:eaat8464.

(123) Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* 2009; 25:1754-1760.

(124) DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genetics* 2011; 43:491-498.

(125) Jun G, Flickinger M, Hetrick KN, et al. Detecting and estimating contamination of human DNA samples in sequencing and array-based genotype data. *Am J Hum Genet* 2012; 91:839-848.

(126) Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 2015; 4: 7.

(127) Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genetics* 2012; 44:369-375.

(128) Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol* 2019; 18:1091-1102.

(129) de Lange KM, Moutsianas L, Lee JC, et al. Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat Genet* 2017; 49:256-261.

(130) Zhang FR, Huang W, Chen SM, et al. Genomewide association study of leprosy. *N Engl J Med* 2009; 361:2609-2618.

(131) Nicolae DL, Gamazon E, Zhang W, et al. Trait-associated SNPs are more likely to be eQTLs: annotation to enhance discovery from GWAS. *PLoS Genet* 2010; 6:e1000888.

(132) Li YI, van de Geijn B, Raj A, et al. RNA splicing is a primary link between genetic variation and disease. *Science* 2016; 352:600-604.

(133) Iwaki H, Blauwendraat C, Leonard HL, et al. Genetic risk of Parkinson disease and progression: An analysis of 13 longitudinal cohorts. *Neurol Genet* 2019; 5:e348.

(134) Fogh I, Lin K, Tiloca C, et al. Association of a locus in the CAMTA1 gene with survival in patients with sporadic amyotrophic lateral sclerosis. *JAMA Neurol* 2016; 73:812-820.

(135) Miklossy J, Arai T, Guo J-P, et al. LRRK2 expression in normal and pathologic human brain and in human cell lines. *J Neuropathol Exp Neurol* 2006; 65:953-963.

(136) Alessi DR, Sammler E. LRRK2 kinase in Parkinson's disease. *Science* 2018; 360:36-37.

(137) Funayama M, Hasegawa K, Kowa H, et al. A new locus for Parkinson's disease (PARK8) maps to chromosome 12p11.2-q13.1. *Ann Neurol* 2002; 51:296-301.

(138) Sanchez-Contreras M, Heckman MG, Tacik P, et al. Study of *LRRK2* variation in tauopathy: progressive supranuclear palsy and corticobasal degeneration. *Mov Disord* 2017; 32:115-123.

(139) Bardai FH, Ordonez DG, Bailey RM, et al. Lrrk promotes tau neurotoxicity through dysregulation of actin and mitochondrial dynamics. *PLoS Biol* 2018; 16:e2006265.

(140) Guerreiro PS, Gerhardt E, Lopes da Fonseca T, et al. LRRK2 Promotes Tau Accumulation, Aggregation and Release. *Mol Neurobiol* 2016; 53:3124-3135.

(141) Nguyen APT, Daniel G, Valdés P, et al. G2019S LRRK2 enhances the neuronal transmission of tau in the mouse brain. *Hum Mol Genet* 2018; 27:120-134.

(142) Elkouris M, Kouroupi G, Vourvoukelis A, et al. Long non-coding RNAs associated with neurodegeneration-linked genes are reduced in Parkinson's disease patients. *Front Cell Neurosci* 2019; 13:58.

(143) Zhao Y, Zhao L, Li J, et al. Silencing of long noncoding RNA RP11-476D10.1 enhances apoptosis and autophagy while inhibiting proliferation of papillary thyroid carcinoma cells via microRNA-138-5p-dependent inhibition of LRRK2. *J Cell Physiol* 2019; 234:20980-20991.

(144) Gusev A, Ko A, Shi H, et al. Integrative approaches for large-scale transcriptome-wide association studies. *Nat Genet* 2016; 48:245-252.

(145) Cao W, Zheng H. Peripheral immune system in aging and Alzheimer's disease. *Mol Degen* 2018; 13:51.

(146) Maphis N, Xu G, Kokiko-Cochran ON, et al. Reactive microglia drive tau pathology and contribute to the spreading of pathological tau in the brain. *Brain* 2015; 138:1738-1755.

(147) Moehle MS, Webber PJ, Tse T, et al. LRRK2 inhibition attenuates microglial inflammatory responses. *J Neurosci* 2012; 32:1602-1611.

(148) Gerhard A, Trender-Gerhard I, Turkheimer F, et al. In vivo imaging of microglial activation with [11C](R)-PK11195 PET in progressive supranuclear palsy. *Mov Disord* 2006; 21:89-93.

(149) Passamonti L, Vazquez-Rodriguez P, Hong YT, et al. [11C]PK11195 binding in Alzheimer disease and progressive supranuclear palsy. *Neurology* 2018; 90:e1989-1996.

(150) Bevan-Jones WR, Cope TE, Jones PS, et al. Neuroinflammation and protein aggregation co-localize across the frontotemporal dementia spectrum. *Brain* 2020; 143:1010-1026.

(151) McCrone P, Payan CA, Knapp M, et al. The economic costs of progressive supranuclear palsy and multiple system atrophy in France, Germany and the United Kingdom. *PLoS One* 2011; 6:e24369.

(152) Kargbo RB. Treatment of Alzheimer's by PROTAC-Tau protein degradation. *ACS Med Chem Lett* 2019; 10:699-700.

(153) Ling H, Gelpi E, Davey K, et al. Fulminant corticobasal degeneration: a distinct variant with predominant neuronal tau aggregates. *Acta Neuropathol* 2020; 139:717-734.

(154) Bachiller S, Jimenez-Ferrer I, Paulus A, et al. Microglia in neurological diseases: A roadmap to brain-disease dependent-inflammatory response. *Front Cell Neurosci* 2018; 12:488.

(155) Woollacott IOC, Nicholas JM, Heslegrave A, et al. Cerebrospinal fluid soluble TREM2 levels in frontotemporal dementia differ by genetic and pathological subgroup. *Alzheimers Res Ther* 2018; 10:79.