Identifying genetic determinants of progression in Parkinson’s disease

Manuela Mei Xuan Tan

UCL Queen Square Institute of Neurology

For the award of Doctor of Philosophy

Declaration

I, Manuela Mei Xuan Tan, 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.

Acknowledgements

I would like to thank Parkinson's UK for funding this project. I would also like to thank all the people with Parkinson's who have taken part in these research studies and helped to shape this research. I am also very grateful to key members of the Royal Free Patient and Public Involvement group, who gave me invaluable advice on my funding applications.

I would first like to thank my primary supervisor, Prof Huw Morris, without whom this PhD would not have been possible. I am incredibly grateful for his guidance, support, and the opportunities he has provided me which have shaped my future career.

I would also like to thank my secondary supervisors, Dr Maryam Shoai, Prof Anette Schrag, and Dr Rimona Weil for their support and expert input at various points during this project. Thank you in particular to Maryam for your statistical advice, support and mentoring.

A huge thank you to the Morris lab group and other people I have worked with during the course of my PhD: Alyssa Costantini, Miriam Pollard, Dr Edwin Jabbari, Laura Smith, Soraya Rahall, Dr Philip Campbell, Dr Etienne Laverse, Lesley Wu, Dr Richard Rees, Dr Raquel Real, John Harvey, Dr Marco Toffoli, and Dr Aaron Bradshaw. I am incredibly lucky to have you as friends, not just colleagues. Thank you for all your support.

I would like to thank Dr Donald Grosset and Dr Michele Hu for allowing me to access and analyse data from the Tracking Parkinson's and Oxford Discovery cohorts. I am also grateful for their active feedback on the analyses, results, presentations, and publications from my work, and the opportunity to share my research at funder and study meetings. Thank you also to Sofia Kanavou, Dr Michael Lawton, and Dr Leon Hubbard who prepared and cleaned the clinical and genetic datasets.

I would also like to thank other colleagues at UCL who have helped with various aspects of the project: Prof Mina Ryten, Regina Reynolds, Dr Raf Ferrari, David Zhang, Dr Demis Kia, Hallgeir Jonvik, David Murphy, and Catherine Storm.

I am grateful to Dr Andrew Singleton for the opportunity to visit the National Institutes of Health in 2019. This was an invaluable learning experience for me. In particular, I would like to thank Dr Cornelis Blauwendraat and Dr Hirotaka Iwaki for their support while I was at NIH and beyond my short visit.

I would like to thank Prof Nigel Williams for his expert genetic advice on my project, and also for the opportunity to visit his laboratory and work with Dr Leon Hubbard in Cardiff.

I would also like to thank Prof Yoav Ben-Shlomo and Dr Michael Lawton for their ongoing statistical advice and input, and for the opportunity to visit their lab in Bristol.

Finally, thanks to my friends and family who have been on this journey with me and for your encouragement.

Abstract

Parkinson's disease (PD) is a progressive neurodegenerative condition for which there are currently no treatments to stop or slow disease progression. A number of genome-wide association studies (GWASs) of PD patients compared to controls have identified genetic variants associated with disease risk, however these cannot inform us about the genetic factors and biology underpinning progression.

The aim of this PhD is to identify genetic variants associated with disease progression. I first examined the frequency and baseline clinical features of patients carrying rare pathogenic Mendelian mutations (including variants in *LRRK2*, *SNCA*, *Parkin*, and *PINK1*) in the Tracking Parkinson's cohort. I showed that *Parkin* and *PINK1* carriers had better cognition than other early-onset patients at baseline despite having longer disease duration, suggesting slower progression. In analysis of longitudinal data, I also showed that *GBA* carriers appeared to have more rapid motor and cognitive progression than non-carriers.

Prior to conducting GWASs, I sought to understand the clinical predictors of progression and showed that age at onset and gender were associated with progression to clinical milestones.

Following a new method from the Huntington's disease progression GWAS, I used principal components analysis (PCA) to combine multiple motor and cognitive scales in PD to create composite progression scores. I showed that *APOE* ε4 was strongly associated with cognitive progression, and identified a novel signal in *ATP8B2* which was nominally associated with motor progression.

Finally, I conducted large-scale GWASs of survival to clinical milestones: mortality, Hoehn and Yahr stage 3, and dementia, using data from Tracking Parkinson's, Oxford Discovery, Parkinson's Progression Markers Initiative, Queen Square Brain Bank, UK Biobank, and Calypso studies. I identified loci in or near *APOE*, *ADRA2A,* and *SH3GL2* which were nominally associated with progression to mortality. I also showed that the *APOE* ε4 variant, rs429358, was strongly associated with progression to dementia.

Impact statement

In this thesis, I have studied genetic variants associated with Parkinson's disease progression. I hope this work will build on the knowledge available in the scientific community about the biology of disease progression, and that the wider body of research will facilitate the development of new disease-modifying therapies.

In addition, I believe the work from this thesis can be used to improve prediction of progression. At present, clinical trials of potential therapies are hampered by the lack of individual prediction of progression, as patient trajectories are so heterogeneous. Integrating clinical and genetic factors to better predict individual trajectories will enable more efficient clinical trials and potentially stratification of patients for targeted trials.

During the course of this thesis, I have genotyped all the Queen Square Brain Bank pathologically-confirmed Parkinson's disease samples that had DNA available in the UCL Queen Square Institute of Neurology laboratory. The raw genotype data as well as the processed NeuroChip array data that I have generated has been stored at the Queen Square Institute of Neurology and will be made available to other researchers on request. I hope this will be helpful for further research on the relationship between genotype and pathology.

I have published or plan to publish all of the work in this thesis in peer-reviewed journals. The PCA GWAS is currently under review at *Movement Disorders* and I have presented the results through poster presentations at national and international conferences, including as a top abstract at the recent MDS Congress.

Some of the results from my thesis have been used to support funding applications for further related research, such as the MRC grant and the ASAP progression grant.

I have made my code and analysis pipelines all publicly available on the platform GitHub (https://github.com/huw-morris-lab). This can be used by others to replicate my methods in their own cohorts. I have also written more general instructional pipelines for GWAS and other genetic techniques which I hope can aid the learning process for others in my lab group and in the wider scientific community.

Outside academia, I have had opportunities to engage with people with Parkinson's, funders, and other organisations. I presented my research to Parkinson's UK for the annual review of the Tracking Parkinson's study in 2019. I have also presented my research findings at the annual Tracking Parkinson's meeting in Glasgow to study coordinators, nurses, and study members. I participated in a Parkinson's UK and Health Data Research UK workshop which brought together people with Parkinson's, researchers, and other representatives to design collaborative projects with the aim to bring health care benefits to people affected by Parkinson's. I have also participated in Parkinson's UK and Cure Parkinson's Trust patient engagement activities and events, such as an online Q&A panel and video about taking part in research (https://www.parkinsons.org.uk/research/take-part-research). I hope that these activities will encourage people to see the value of taking part in research studies and to show the results and questions that can be answered from the data collected in these studies.

Abbreviations

AD – Alzheimer’s Disease

FUMA – Functional Mapping and Annotation of GWAS

GO – Gene Ontology

GRS – Genetic Risk Score

GWAS – Genome Wide Association Study

HD – Huntington’s Disease

HES – Hospital Episode Statistics

LDSC – Linkage Disequilibrium Score Regression

MAGMA – Multi-marker Analysis of GenoMic Annotation

MDS-UPDRS – Movement Disorders Society Unified Parkinson’s Disease Rating Scale

MLPA – Multiplex ligation-dependent probe amplification

MoCA – Montreal Cognitive Assessment

PCA – Principal Components Analysis

PD – Parkinson’s Disease

PIGD – postural instability gait difficulty (motor subtype)

PPMI – Parkinson’s Progression Markers Initiative

PSP – Progressive Supranuclear Palsy

RBD – Rapid Eye Movement Sleep Behaviour Disorder

REM – Rapid Eye Movement

List of tables

List of figures

Table of Contents

Chapter 1 : Introduction

Overview

Parkinson's disease (PD) is a progressive neurodegenerative disease. There are currently no treatments that stop or slow the progression of PD. There is strong evidence that genetic variants contribute to disease risk, and more recently there is growing evidence that genetics contribute to the rate of disease progression. The relationship between intrinsic disease heterogeneity, risk, and progression is complex. The aim of this project is to identify genetic determinants of progression in PD. This work will help us to better understand the biology of progression and potentially identify drug targets for new disease-modifying treatments.

Epidemiology, diagnosis, and pathology of PD

PD was first described by James Parkinson in 1817 in *An Essay on the Shaking Palsy* [1]. PD is characterised by a collection of features known as parkinsonism – bradykinesia (slowness of movement and decrease in amplitude or speed of repeated movements), in combination with either rest tremor or rigidity [2,3]. The diagnosis of PD can be supported by the presence of postural instability, as included in the original Queen Square Brain Bank clinical diagnosis criteria, but this usually occurs in the later stages of PD [3].

Pathologically, PD manifests with the selective loss of dopamine neurons in the pars compacta of the substantia nigra. The pathological hallmark of PD and gold standard for diagnosis is the presence of Lewy bodies, which are abnormal deposits of the protein α-synuclein [4]. These are found in surviving neurons in the substantia nigra as well as surrounding brain areas of patients with PD [2,5]. The Braak staging hypothesis, based on post-mortem examination of donors with different disease extent, suggests that Lewy body pathology begins in the brainstem and spreads progressively to other areas of the brain: the pons, substantia nigra, limbic system, temporal cortex, and finally to multiple regions in the cortex [6]. This may be mediated through cell to cell spread of α-synuclein pathology [7,8].

PD affects up to 1,903 out of 100,000 people aged over 80 [9,10]. Although the prevalence of PD increases with age, it is not just a disease of the elderly; it affects 41 in 100,000 people aged between 41 to 50 [9]. It is estimated that 6.2 million people worldwide are affected by PD currently, and this is expected to increase to up to 12.9 million in the next 20 years as the population ages and life expectancy increases [11].

Clinical progression in PD

PD progression is heterogeneous

The clinical progression of PD seems to follow a general pattern [8]. Diagnosis is generally made on the basis of motor symptoms, including bradykinesia, rigidity, and tremor. In the early stages after diagnosis and with treatment, these may be accompanied by treatment complications, such as fluctuations and dyskinesia (involuntary movements). Other motor milestones, such as dysphagia (swallowing difficulties), postural instability, and falls, develop in later disease stage [8]. Major non-motor milestones, including dementia and hallucinations, may develop later in the disease course. However, other non-motor symptoms may present earlier in disease, such as mild cognitive impairment. In addition, there are some non-motor symptoms which may be present long before the diagnosis of PD and initial motor symptoms. These can include constipation, hyposmia, and Rapid Eye Movement (REM) sleep behaviour disorder [12].

Our knowledge on the disease course of PD comes from longitudinal studies with extensive periods of follow-up. The Sydney Multicentre Study is a long-running study of L-dopa naïve, newly diagnosed, idiopathic PD patients, initially in a randomised trial of low-dose L-dopa compared to low-dose bromocriptine [13,14]. This study is widely reported as indicating the long-term outcomes in patients with PD. 38% of patients had died within 10 years [14]. By 15 years, 94% of patients had experienced dyskinesia, 56% had experienced dystonia, 50% had experienced dysphagia (choking), 81% had experienced falls, 81% had experienced freezing, 48% dementia, 36% had mild cognitive impairment, 50% hallucinations, autonomic failure (35% hypotension, 41% urinary incontinence), and 65% had died [15]. By 20 years, 74% of patients had died, most patients were no longer independent, 87% had experienced falls, autonomic dysfunction was common, 74% had experienced hallucinations, and 83% had dementia [16]. This suggests that most patients who had died were bedridden, and more severe disease at baseline (including factors such as dementia, Hoehn and Yahr stage 3, early development of instability) was predictive of mortality [15]. This suggests that there is a general progression towards common disease milestones, and most patients seem to follow this pattern.

Although the Sydney Multicentre Study indicates the overall long-term outcomes, other longitudinal studies show that clinical progression is heterogeneous. The CamPaIGN study (Cambridgeshire Parkinson's Incidence from GP to Neurologist) showed that in an incident cohort of newly diagnosed patients, outcomes were very heterogenous after 10 years of follow-up [17]. By 10 years, 45% of patients had died, 50% of patients had developed dementia, and 68% had reached Hoehn and Yahr stage 3 with postural instability. However, there was a proportion of patients (23%) who still had a good outcome after 10 years, surviving with no postural instability, and intact cognition [17]. The Sydney Multicentre Study similarly found a proportion of patients (10%) who, after 10 years of follow-up, still were in early Hoehn and Yahr stages, did not have troublesome fluctuations, dementia, or hallucinations, and were very responsive to treatment [14]. Data from these studies shows that although there seems to be a pattern for the development of key symptoms and milestones, there is substantial variability in the rate of disease progression between patients.

The heterogeneity of clinical progression is further confirmed in studies identifying subtypes of patients with different rates of progression. There have been many subtype studies using a variety of clustering methods, and each classifying slightly different subgroups of PD patients. However, taken together, these studies show that the rate of disease progression is heterogeneous, and can often be clustered according to groups of patients defined by baseline characteristics. Using a data driven approach, Lewis et al. identified 4 subgroups of patients: young onset patients with slower disease progression, tremor dominant, non-tremor dominant, and rapid motor disease progression without cognitive impairment [18]. Here, progression was evaluated from a single observation (rather than longitudinal data) by dividing the Unified Parkinson’s Disease Rating Scale (UPDRS) Part III motor score by years of disease duration. However, this has been supported by other studies showing that subtypes defined using baseline features have different rates of progression in longitudinal assessments [19,20].

The pathology of progression

It is possible that clinical disease progression is related to the extent and severity of neuropathology. Several PD studies suggest that this may be the case; for instance, dementia and cognitive impairment are correlated with Lewy body neuropathologic stage [21], and Lewy body load in the neocortex and other regions [22,23]. A recent systematic review of postmortem PD cases suggested that α-synuclein pathology had the strongest association with dementia in PD [24].

However, there are many other studies which conflict with these suggestions that Lewy body pathology is correlated with clinical progression. Some studies have reported cortical Lewy body pathology in individuals without a neurological diagnosis and PD patients without cognitive impairment [25,26]. Neuropathological assessment in the Sydney Multicentre Study showed that pathological staging was not consistent in all groups of patients. In young onset PD patients with typical long disease duration, Lewy body pathology seemed to follow the hierarchical pattern predicted by Braak staging and this correlated with clinical progression [27]. However, some patients had rapidly progressing disease with dementia, short survival, and high neocortical Lewy body loads. The last group of patients had older onset, more complex disease, and shorter survival; these patients had diffuse Lewy body loads and often co-occuring amyloid plaque pathology. These distinct groups suggest that the progression of pathology does not follow the same pattern in all patients, and that other neuropathological substrates may contribute to progression [27]. The systematic review by Smith et al. also found that co-occuring amyloid pathology was common in PD cases with dementia [24].

Taken together, these studies suggest that Lewy body load is not the sole driver of clinical progression and dementia. Other types of pathology, such as amyloid, tau, and vascular disease, may drive progression [16,24,28,29] and may indeed be correlated and interact with Lewy body pathology [22,30].

Measuring clinical progression

The lack of clear pathological markers of progression means there is no gold standard for measuring clinical progression in PD. The Movement Disorders Society Unified Parkinson’s Disease Rating Scale (MDS-UPDRS) Part III and Part II are most commonly used in clinical trials. These measure motor symptoms in a clinician examination (Part III) and the patient-reported motor experiences of daily living (Part II) [31].

However, there are many other ways of assessing symptom severity in PD as well as other aspects, such as quality of life and impact on patients' daily experiences. In addition, motor assessments are affected by medication and can be conducted when the patient is either in an 'off' or 'on' state, with regard to whether their PD medication is in effect or not. In order to measure progression accurately, scales must be sensitive to change over time otherwise they do not provide enough information about variability and progression. The scales that are the most sensitive to change over time over a 1 year period (with the largest change scores) are the Hoehn and Yahr scale, the UPDRS Part II, and the UPDRS Part III [32]. In a population-based sample assessed with remote questionnaires, the most sensitive measures were the Schwab & England Activities of Daily Living Scale, the activities of daily living section of the Parkinson's Disease Questionnaire (PDQ) 39, and the visual analogue scale found in the quality of life instrument EQ-5D [32]. Overall, the measures of impairment and disability were more sensitive to change over time than the quality of life scales, possibly because quality of life is a subjective report and may adapt to change over time.

This study clearly shows that different scales measure slightly different aspects of PD clinical signs and progression – whether disability/impairment, 'objective' symptoms, or quality of life. There is no clear answer of which scale is best, as different scales may be better suited to measuring different things, for example the Hoehn and Yahr scale may be more sensitive to detecting symptoms which are not responsive to treatment (such as axial symptoms) whereas the UPDRS motor assessment may not change as much because treatment has been optimised [32].

The question of whether mortality is the gold standard of PD progression also cannot be clearly answered. Unlike in other rapidly progressing diseases, such as Progressive Supranuclear Palsy (PSP), where time from disease onset to death is a good marker for rate of progression, PD has a long disease duration. The cause of death may not be PD or its complications, unlike other diseases. The Sydney Mutlicenter Study found that PD contributed to death in only 53% of patients [15]. In addition, survival time in PD may not reflect rate of progression, for example, some patients live for longer but have very poor quality of life towards the end of life.

In addition, there are many different ways of analysing clinical progression – whether absolute symptom severity, time to clinical milestones, or change from baseline scores. Again, there is no clear consensus as to which method is the best, and this is compounded by the fact that there is no gold standard of progression by which to compare different methods.

To summarise, the nature of PD progression, measurement, and analysis of progression, is complex. However, this work is essential because clinical trials aim to test new therapies that could potentially stop or slow disease progression, often after the point of PD diagnosis. In order to develop new disease modifying treatments, we need to better understand the biology of disease progression. One way to do this is to study genetic factors.

The role of genetics in PD: Rare variants

Genetics have already shaped our understanding of PD risk and biology. Evidence from family based studies has shown that genetic factors influence both PD risk as well as clinical features and progression.

Mutations in *SNCA*, *LRRK2*, *PARK2 (parkin)*, *PINK1*, *DJ1*, and a few other genes, have been shown to increase the risk of developing PD in a small proportion of patients [33]. Importantly, identification of these genetic factors have highlighted a number of pathways and potential targets that are important for neurodegeneration in PD. For example, *SNCA* mutations, in particular whole-gene multiplications, suggest that the overproduction of α-synuclein is a key process for neurodegeneration in PD and could be targeted as a therapeutic approach [34]. *LRRK2* mutations are the most common cause of PD, and while there is still much that is not known about the activity of *LRRK2* and its role in the pathology of PD, there is evidence that mutations in *LRRK2* are associated with increased kinase activity [35,36]. This has led to trials of Lrrk2 kinase inhibitors as a potential therapeutic strategy for PD [37].

SNCA mutations, particularly whole-gene triplications, are associated with a more severe phenotype including more severe dementia, rapid progression, hallucinations, and autonomic dysfunction [38–41]. *LRRK2* mutations are suggested to be associated with milder disease, less cognitive impairment, and slower motor progression than idiopathic PD [42–44]. However, other studies have not confirmed these findings [45].

PARK2 (*parkin*) and *PINK1* are recessive genes typically associated with early-onset PD/parkinsonism. Both these genes are important for regulating mitochondrial function and mitophagy, which is the clearance of damaged mitochondria from the cell [46,47], and disruption of this pathway may be an important mechanism in the pathogenesis of PD [48]. The mitophagy pathway has been being targeted as another potential therapeutic strategy.

Parkin and *PINK1* mutations are generally associated with younger age at onset, slower disease progression, dystonia in some cases, good response to L-dopa, and less cognitive impairment [49–57].

The glucocerebrosidase (*GBA*) gene is a risk loci for PD and is associated with a smaller increase in disease risk than mutations in other genes – approximately 5-fold increase in the risk of PD [58]. Although the risk conferred is not as large as the other Mendelian genes, *GBA* mutations are more common in PD; prevalence ranges from 2% to 30% depending on the population [58–61].

Exactly how *GBA* mutations contribute to the development of PD is still unknown, however they have highlighted the potential role of the glucocerebrosidase enzyme (GCase) and the lysosomal pathway as a pathogenic mechanism [62]. *GBA* mutations are associated with earlier onset, more rapid progression to mortality, motor impairment, and dementia, as well as more frequent neuropsychiatric symptoms [63–71].

Although these genetic factors are rare, there is early evidence that the affected pathways are also involved in idiopathic PD patients [72]. In addition, there is overlap in the genes that contain pathogenic rare variants and those that contain common variants that increase risk for idiopathic PD, as discussed in the next section [73] . Therefore, it is hoped that potential therapies that target these biological pathways will be important not just for patients carrying mutations in these genes but also for other PD patients.

Common variants in PD risk

In recent years, there have been a number of loci identified which are associated with disease. This is based on the common disease/ common variant hypothesis, which suggests that a common, complex disease such as PD is likely to be caused by multiple common genetic variants which are present in the general population [74]. Each variant confers only a small increase in risk but collectively they can contribute to the development of disease.

A number of genome-wide association studies (GWASs) in PD have identified loci associated with disease risk [75–80]. These are conducted by comparing large numbers of PD patients and healthy controls. The most recent meta-analysis of PD GWAS analysed over 37,000 PD cases and 1.4 million controls [75]. This identified 90 independent signals across 78 loci that were associated with PD risk.

These GWASs have identified hits in *LRRK2* and *SNCA*, which have also been implicated in autosomal dominant PD. This confirms that these genes are important for disease risk, both in rare high-risk penetrant mutations, and common variants that contribute small amounts of risk [81]. Secondly, case-control GWASs have contributed to evidence that certain pathways are important for PD risk, such as lysosomal function [75]. Finally, these GWASs have identified novel genes and pathways that are being investigated as potential targets for new therapies.

Importantly, these GWASs have large numbers of PD patients and controls, but do not have detailed clinical data and so cannot identify common variants that are associated with PD phenotypes or progression. This is the primary aim of this project.

Genetics of PD progression: Candidate gene studies

There is a growing body of evidence that genetic factors are not only important for the development of PD, but are also associated with clinical features and progression.

This is evident for rare variants in Mendelian genes, as discussed previously. Most of these studies have been carried out in clinical referral series, and have not been systematically studied in a large-scale population-based cohort. Additionally, these pathogenic mutations are present in only a small proportion of the PD population (< 10%).

I hypothesise that other genetic factors contribute to the variability in disease progression. Understanding this association will be crucial to the development of new therapies to stop or slow PD progression.

There have been several candidate-gene studies of common variants associated with PD progression. The most frequently studied are the *MAPT* haplotypes and *APOE* ε4 genotypes. Studies in two longitudinal cohorts have suggested that the *MAPT* H1/H1 haplotype is associated with more rapid cognitive progression and dementia [17,82–84]. Other cross-sectional studies show that *MAPT* H1/H1 carriers have worse performance in cognitive tasks [85] and have more frequent dementia [86]. However, these findings have not been confirmed in other large studies, including longitudinal cohorts [87–89].

The *APOE* ε4 allele has also been extensively studied in PD, and is an important risk factor for Alzheimer's disease (AD). Several longitudinal and cross-sectional studies have shown that the ε4 allele is associated with the rate of cognitive decline in a range of scales [87–89] and is more frequent in PD patients with dementia [90]. However, other longitudinal incident cohorts have not replicated this finding [91,92]. A recent meta-analysis found a small, significant overrepresentation of *APOE* ε4 carriers in PD dementia cases (odds ratio 1.74), however suggested that this may be confounded by heterogeneity between studies, small sample sizes, and publication bias [92].

Latourelle and colleagues used machine learning methods in a relatively small study cohort (N = 312) to show that genetic variation was a predictive marker of motor progression. Progression was defined as the rate of change in the MDS-UPDRS Part II and III combined. In particular, two variants rs9298897 (located in the intron of the *LINGO2* gene) and rs17710829 were associated with progression [93]. Further replication of these results in independent cohorts is needed.

In addition, the PD Genetic Risk Score (GRS) has been suggested to be associated with motor and cognitive progression in PD in two studies [94,95]. The GRS is an individual score based on the cumulative weighted number of PD risk variants, from case-control GWASs. However, these studies have been conducted with small sample sizes (285 and 336 patients respectively), and with earlier versions of the GRS based on fewer loci. These findings need to be confirmed in large-scale studies.

The problem with candidate gene studies is that they are subject to confirmation bias and cannot identify new variants associated with progression. However, these studies have contributed to our knowledge that genetics plays a role in PD phenotypes and progression.

Genome-wide association studies of progression

GWASs of phenotypes and progression are relatively new, not just in PD but also in other diseases. However, analysing phenotypes, either in discrete or quantitative traits, is important to identify genetic variants that influence heterogeneity of phenotypes within patients only. The benefits of genome-wide approaches to disease phenotypes can be seen in well-powered studies of PD age at onset [96], and progression in other diseases such as Huntington's disease (HD).

A recent genome-wide meta-analysis of 28,568 PD cases showed that loci in *SNCA* and *TMEM175*/ *GAK* were associated with age at onset in PD [96]. These two loci are both well-established PD risk loci from case-control GWASs. In addition, this study also confirmed previous candidate gene studies showing that *GBA* variants (N370S, E326K, and T369M) are associated with younger age at onset.

Importantly, the results of this study suggest there is partial but not complete overlap between the genetics of PD risk and PD age at onset. The GRS was associated with age at onset, however, other well-established PD risk loci were not associated with age at onset, including *GCH1* and *MAPT*.

This study highlights the differences in the genetic architecture underlying PD risk and that of PD phenotypes, such as age at onset and disease progression. Therefore, therapies that target pathways involved in PD risk may not be effective at slowing progression in individuals who already have PD, as different pathways and mechanisms may be more important for the progression of disease.

GWASs in other diseases have successfully identified genetic determinants of progression and pointed to particular mechanisms that could be targeted. One of these is the GWAS of HD progression [97]. This study of approximately 2,000 HD patients clearly showed that a locus overlapping the *MSH3* gene was associated with disease progression. *MSH3* is involved in the DNA mismatch repair pathway and has been implicated in the pathogenesis of HD through somatic expansion of the CAG repeat. This study has drawn attention to *MSH3* and this pathway as a potential therapeutic target in HD.

There are clear differences between the analyses of progression in HD and PD. HD is a more homogeneous population with a single gene cause and more predictable progression. In addition, progression in different domains are well correlated in HD. The PD population is more heterogenous, with a number of genetic causes in a small proportion of cases, and more variability both between and within subjects in terms of progression. It is clear that accurate measures of disease progression will be needed, and likely larger sample sizes to overcome heterogeneity in PD cohorts.

Recently, the first large-scale GWAS of PD progression was conducted [98]. This examined a range of progression measures, including change in the MDS-UPDRS, Hoehn and Yahr staging, the Mini Mental State Examination (MMSE), and Montreal Cognitive Assessment (MoCA) in 12 longitudinal cohorts. Overall, analyses included 4,093 PD patients with 25,254 follow-up visits, although not all patients had data on all the measures so individual GWAS numbers are smaller. However, this study has demonstrated a number of key points.

Firstly, it shows that there are single variants and loci that can be detected in sample sizes of approximately 4,000 PD cases. While this number is still relatively small for a GWAS, there were two loci that reached genome-wide significance.

Secondly, the study identified a novel locus, rs382940, associated with more rapid progression to Hoehn and Yahr stage 3 in survival analysis. This variant is in an intronic region of *SLC44A1*, solute carrier family 44 member 1. This gene has not been previously reported in PD, and this finding needs to be replicated in independent datasets.

Thirdly, this GWAS has confirmed previous findings from candidate gene studies. In targeted analyses (not genome-wide significant), *GBA* variants (T369M and E326K) were associated with more rapid motor and cognitive progression, and the *APOE* ε4 tagging variant, rs429358, was associated with lower MoCA and MMSE scores [98].

Finally, the results suggest that the loci underpinning PD risk are, to a large extent, distinct from those that are involved in PD progression. Iwaki et al. tested 88 risk variants from the most recent PD case-control GWAS, of which 10 passed analysis-wide significance ($p < 0.002$). Certain PD risk variants were associated with clinical features, including cognition, motor progression, and daytime sleepiness [98]. However, the majority of PD risk variants were not associated with any markers of progression. Further studies are needed to replicate these results and identify whether the genetic architecture of PD risk differs from that of PD progression.

The challenges addressed by this PhD/ Aims

This PhD addresses the following challenges:

1. Establishing the frequency and baseline clinical characteristics of pathogenic Mendelian mutations in PD in a large UK cohort (Chapter 3)
2. Understanding the clinical predictors of progression (Chapter 4)
3. Conducting genome-wide association studies using composite scores of motor, cognitive, and cross-domain progression (Chapter 5)
4. Conducting genome-wide association studies of survival to clinical milestones in PD: mortality, Hoehn and Yahr stage 3 or greater, and dementia (Chapter 6)

Chapter 2 : Methods

Cohorts: Recruitment, inclusion/exclusion criteria, and clinical assessments

Tracking Parkinson's

Tracking Parkinson's is a multi-centre observational study recruiting patients from 72 centres across the UK. Patients with a clinical diagnosis of PD, meeting the UK Brain Bank diagnostic criteria, were recruited [99]. Ethics approval was provided by West of Scotland Research Ethics Service. The study was carried out in accordance with the Declaration of Helsinki and is registered as NCT02881099 at ClinicalTrials.gov.

Recent onset PD

Patients who were diagnosed with PD within 3.5 years of study entry were recruited as recent onset participants. These participants were assessed every 18 months with detailed, standardised clinical assessments, including motor, cognitive, and other non-motor assessments.

Established young onset PD

Patients with age at diagnosis ≤ 50 years and with time from diagnosis > 3.5 years were recruited as young onset participants. These patients were not assessed longitudinally so were only included for baseline analyses.

Oxford Discovery

The Oxford Parkinson's Disease Centre Discovery study (Oxford Discovery) is another UK observational multi-centre study. PD patients were recruited from neurology clinics in the Thames Valley area. Patients who met the UK Brain Bank diagnostic criteria for PD and were diagnosed within the last 3 years were recruited to the study [100]. Ethical approval for the study was granted by the Berkshire Regional Ethics Committee. Participants were excluded if they had non-idiopathic parkinsonism, dementia preceding PD by one year, or cognitive impairment which meant informed consent could not be obtained [100]. Participants were assessed every 18 months using standardised clinical assessments similar to Tracking Parkinson's.

Parkinson's Progression Markers Initiative (PPMI)

Patients with PD were recruited at multiple centres across Europe, America, and Australia if they met the following inclusion criteria [101] (https://www.ppmi-info.org/):

1. Asymmetric resting tremor or asymmetric bradykinesia or two of bradykinesia, resting tremor, and rigidity
2. Diagnosis within 2 years
3. Hoehn and Yahr Stage I or II at baseline
4. Untreated for PD, and not expected to require PD medication within 6 months at baseline
5. Dopamine transporter (DAT) imaging showing DAT deficit
6. 30 years or older at time of PD diagnosis

Participants were assessed every 3 months in the first year, then every 6 months until the end of the fifth year and every year following. Cognitive assessments were only performed at yearly visits. The Montreal Cognitive Assessment (MoCA) was performed at the screening visit and not at baseline, so the screening assessment and corresponding disease duration was used for analysis. For motor assessments, the annual assessments were conducted in the “practically defined off” state, where the participant did not take their PD medications since the night before the visit, and for at least 12 hours prior to the visit. As the cognitive assessments and “practically defined off” motor assessments were conducted at annual visits, I only included data for annual visits in the analysis. I downloaded the PPMI clinical data on 14/08/2019 and performed all the data cleaning and merging.

Queen Square Brain Bank

Patients with a pathologically confirmed diagnosis of PD, regardless of clinical diagnosis, and that had DNA available at the UCL Queen Square Institute of Neurology were included for analyses. Summary clinical data was queried and provided by Mr Hallgeir Jonvik (UCL). This included age at onset, gender, age at death, clinical diagnosis, and pathological diagnosis.

Calypso

Patients were recruited between 2006 and 2008 through 3 methods: a community-based prevalence study in Cardiff, referrals from neurologists, geriatricians, and PD nurses, and self-referral [102]. Patients met the UK Brain Bank diagnostic criteria for PD and provided consent for review of their medical records. Vital status was obtained from the NHS Spine by Ms Miriam Pollard in June 2020.

UK Biobank

The UK Biobank is large, prospective, population-based study and an open-access resource of phenotypic, genotypic, and health record data for 500,000 participants [103,104]. Participants aged between 40 and 69 were recruited to one of 22 centres across the UK [104]. Access to the UK Biobank data was through Application 46450.

PD cases were identified from hospital episode statistics (HES) (ICD10 code, in either the primary or secondary position), self-report, or death. Data was downloaded on 13/06/2020, after the death register records were updated (up to April 2020).

PD patients were classed as either prevalent or incident cases following the 'Definitions of Parkinson's Disease and the major causes of Parkinsonism: UK Biobank Phase 1 Outcomes Adjudication' document (version 1.0, March 2018; http://biobank.ctsu.ox.ac.uk/showcase/showcase/docs/alg_outcome_pdp.pdf).

Briefly, prevalent cases were defined as patients who had the first PD ICD code (ICD10 code G20) date prior to the baseline assessment, or self-reported PD at the baseline assessment. Incident cases were defined as patients with PD detected by HES with the PD ICD code date after the date of baseline assessment. Patients with PD coded in any position in the death register records, but who did not have PD in the HES records at any point, were also defined as incident cases.

The date of PD diagnosis was defined according to UK Biobank guidelines, using the earliest date of the PD code (HES, self-report, or death register).

Patients who did not self-report PD at baseline but at a follow-up visit, and who did not have PD in any HES records or death register records, were not classified as either prevalent or incident. I assigned these patients as a separate 'undefined' category.

This is in line with the UK Biobank guidelines, which only included PD self-report at the baseline assessment.

Genotyping

Tracking Parkinson's

At study entry, blood samples were collected from every participant and DNA was extracted from an ethylene diamine tetraacetic acid sample by Ms Catherine Bresner and Prof Nigel Williams' team (Cardiff University). DNA samples were genotyped using the Illumina HumanCore Exome array with custom content. This covered approximately 250,000 common variants, 250,000 rare variants, and over 27,000 custom variants that have been implicated in neurological and psychiatric disorders [99]. Genotyping was performed by Ms Catherine Bresner (Cardiff University) and Dr Leon Hubbard (Cardiff University). Genotype data was in genome build hg19/GRCh37.

PD gene sequencing and genotyping

Almost all samples were genotyped for the *LRRK2* G2019S mutation using the 'Kompetitive' allele-specific polymerase chain reaction (KASP) assay (LGC Genomic Solutions). Subsets of samples were also screened for mutations in *GBA*, *Parkin,* and *PINK1* with Sanger sequencing.

Whole exome sequencing

Whole exome sequencing was performed in a subset of young-onset and familial patients (N=489). Exome sequencing was performed by Macrogen (http://www.macrogen.com/) using the Agilent SureSelect capture kit (Santa Clara, CA, USA). Quality control and annotation was performed by Dr Alan Pittman (UCL). Variant calls and individual genotypes that did not meet quality filters were excluded. Samples were aligned to the human genome (build hg19). The Genome Analysis Toolkit (GATK) was used for local realignment, base quality score recalibration, and multi-sample variant calling (Unified Genotyper). GATK Variant Quality Score Recalibration and recommended GATK training sets were used to create a high-quality set of variant calls [105]. ANNOVAR was used to annotate variants with information on functional consequence, minor allele frequency (MAF), variant type and previous reporting [106]. I screened the annotated exome sequencing data for pathogenic variants in *SNCA*, *LRR2K*, *Parkin*, *PINK1*, *DJ-1* and *VPS35.*

Multiplex ligation-dependent probe amplification (MLPA)

MLPA was performed to detect and confirm copy number variation in *Parkin*, *PINK1*, *DJ1* and *SNCA* . MLPA was performed in a subset of young-onset patients and familial patients by Ms Theresita Joseph. It was conducted with the MRC Holland SALSA MLPA P051 Parkinson kit (version D1), according to the manufacturer's instructions.

Oxford Discovery

DNA samples from Oxford Discovery were genotyped on the Illumina HumanCore Exome-12 v1.1 and the Illumina InfiniumCoreExome-24 v1.1 arrays. Each of these arrays included approximately 500,000 variants, half of which were exome variants. Genotype data was in genome build hg19/GRCh37. Genotype quality control and imputation was conducted by Dr Stephanie Miller (University of Oxford).

Parkinson's Progression Markers Initiative

Whole genome sequencing (WGS) data from the PPMI was used for all analyses. Only variants that passed filters in the joint calling process were included. Data was merged and filtered by Dr Hirotaka Iwaki (NIH). Data was in genome build hg38.

Queen Square Brain Bank

DNA samples were genotyped on the Illumina NeuroChip array version 1.1. This is a custom array containing a tagging variant backbone (the Illumina Infinium HumanCore-24 array) of approximately 300,000 variants together with approximately 180,000 manually curated custom variants implicated in neurological diseases [107]. I conducted DNA sample quality control and preparation for genotyping by performing Qubit fluorometry (to determine the concentration of stock DNA samples) and dilution at the UCL Institute of Neurology. I delivered plates of DNA samples to UCL Genomics (based at the UCL Institute of Child Health and the UCL Zayed Centre for Research) where they were genotyped with the NeuroChip array.

I conducted genotype calling from raw intensity data using Illumina GenomeStudio v2.0. Raw data from different genotyping batches were combined to improve accuracy of clustering of the intensity data. Together with Ms Lesley Wu (UCL), I performed quality control steps in GenomeStudio. I manually reclustered common variants (MAF > 1%) with GenTrain score between 0.4 and 0.7 (inclusive) on the autosomal chromosomes. I did not manually recluster rare variants (MAF < 1%) and variants on the sex chromosomes as these were later excluded in PLINK. Samples with a call rate < 90% were excluded, and variants with GenTrain score < 0.4 were excluded. All SNPs with AA T mean > 0.3, BB T mean < 0.7 or cluster separation < 0.3 were excluded in GenomeStudio before exporting to PLINK format.

Calypso

DNA samples from the Calypso study were genotyped as part of the Wellcome Trust Case Control Consortium 2 GWASs [80,108]. Samples were genotyped on the Illumina BeadArray Human660-Quad array by the Wellcome Trust Sanger Institute (WTSI), Cambridge.

UK Biobank

DNA samples from UK Biobank were genotyped on the Applied Biosystems UK BiLEVE Axiom Array by Affymetrix and UK Biobank Axiom Array [103]. I used version 2 of the genotype data (available as genotype calls). I performed quality control and imputation on the subset of PD cases separately.

Statistical methods

Correlation

Correlation is a method to describe the relationship between two variables. The correlation coefficient indicates the strength of the linear relationship between two variables – how closely the individual points fall to the line of best fit. For continuous variables, this is indicated by the Pearson product-moment correlation. For each point, the difference of the x from the x mean is multiplied by the difference of the y value from the y mean; this is then summed for all the points and divided by the sums of squares (indicated by the formula below):

$$r = \frac{\sum(x-\bar{x})(y-\bar{y})}{\sqrt{\sum(x-\bar{x})^2 \sum(y-\bar{y})^2}}$$

The r value can vary from -1 to 1. An r of 1 or -1 indicates a perfect correlation, where none of the observed data points deviates from the straight line. An r coefficient of greater than 0.5 suggests a large correlation, a coefficient of 0.3 to 0.5 suggests a moderate correlation, and a coefficient of 0.1 to 0.3 suggests a weak correlation. Importantly, the correlation coefficient does not give information about the slope of the line of best fit.

Linear regression

A linear regression examines the relationship between a dependent or outcome variable, and one or more predictor/ independent/ explanatory, variables. This assumes there is a linear relationship between the outcome variable and the explanatory variables. It can be used for prediction, whereas correlation cannot.

A line of best fit is derived using least squares, whereby the sum of squared distances from each data point to the line is minimised (also known as the residuals). The beta coefficient of the slope (β_1) indicates the increase in the dependent variable (Y) associated with each unit increase in the independent variable. The intercept (β_0) indicates the mean of the dependent variable when the value of the independent variable/s is 0.

$$\boldsymbol{Y = \beta_0 + \beta_1 . X}$$

Linear regression is intended for use when the dependent variable is continuous. When the dependent variable is binary and categorical, logistic regression can be used. In this case, the exponential of the slope beta coefficient gives the change in odds of the dependent variable.

The fit of the model can be assessed using R^2, which is the proportion of the variability explained by the regression model. R^2 is calculated as the sum of squared errors of the proposed model divided by the sum of the squared errors of the null model (the mean), subtracted from 1. The adjusted R^2 is more useful when there are multiple independent variables, as it is adjusted for the number of predictors in the model and only increases if the additional variables reduce the overall error of predictions.

Mixed effects models

Mixed effects models, also known as linear mixed models, are an extension of simple linear models. These can be used to analyse data that are hierarchical or correlated, for example, the observations from the same individual over time which are likely to be correlated. Mixed effects models consist of fixed effects and random effects. The fixed effects are the same as the explanatory variables in simple regression; these are variables that are expected to have an effect on the outcome/dependent variable. The random effects refer to group-specific variation that we try to control for. Random effects models allow us to estimate the associations between the predictors and outcomes while accounting for the correlation between observations from the same group (e.g. one individual, or one class). Random effects can be included for the intercept and the slope. The random effect for the intercept means that the intercept varies across clusters, but the slope is the same (e.g. individual variation in baseline performance). The random effect slope allows for variation around the population average slope (e.g. individual-specific variation in the rate of progression). I conducted mixed effects models using the ‘lme4’ package in R.

Survival (time to event) analysis

Survival analysis, or time to event analysis, is useful when analysing duration times (as these are always positive), and when there is censoring, for example, when an individual withdraws from the study or completes the last follow-up visit before the outcome of interest is observed [109].

The survival curve/function indicates the probability of surviving or meeting the outcome beyond a timepoint *t*. It can be estimated by the Kaplan-Meier curve. Each drop/step on the Kaplan-Meier curve indicates the proportion of individuals who have met the outcome at that timepoint, and the bars indicate individuals who have been censored (observation has stopped before the individual has met the outcome) [110]. The Kaplan-Meier curve estimates the probability of surviving to the end of a given time period, conditional on surviving up to the beginning of that time period [110,111]. The log rank test can be used to compare the survival curves of two groups.

While the Kaplan-Meier curve uses a step function to estimate the survival function, the Weibull model estimates the survival function using a smooth line, based on the Weibull distribution. This model can include covariates.

The Cox Proportional Hazards model is another way of estimating the survival function, and can also analyse the effect of covariates on the outcome. It is semi-parametric, whereas the Weibull model is parametric, meaning that the Cox model has less strict assumptions about the distribution of the time to event data. As it is more flexible, the Cox model is the most widely used in survival analyses. Both the Weibull and Cox models assume proportional hazards, meaning that the hazard ratios for the independent variables are constant over time. Censoring should be independent from the outcome of interest. The regression coefficients from the Cox model are on a log scale and the exponents of the coefficients gives the hazard ratio. I conducted survival analyses using the ‘survival’ and ‘survminer’ packages in R.

The hazard ratio is a ratio of two hazard functions (hazard function for group 1 divided by the hazard function for group 2) [110] . This is different to a relative risk (or risk ratio), which is a ratio of two probabilities (probability of the event in group 1 divided by the probability of the same event in group 2) [112]. This is again different from the odds ratio. Odds refer to the probability of an event occurring in a group divided by the probability of the event not occurring in the same group. The odds ratio is the odds of an event for group 1 divided by the odds of the same event in group 2 [112]. Importantly, the hazard ratio is related to the timing of the event, with the hazard rate at any given time interval (almost an instantaneous rate) [111]. In contrast, the odds ratio and relative risk only refer to the cumulative probabilities and total number of events over an entire study using a defined, single endpoint.

Another important point to highlight is that the hazard ratio does not give an indication of how much faster or slower the event occurs between the groups, only the probability of meeting the outcome. To estimate the time-based parameters (how much faster or slower the groups progress to the event), the mean and median times to the event can be compared [111].

There are several methods to estimate the R^2 in survival models, which are used to approximate the R^2 from linear regression models. These are often called pseudo R^2. One common method is the Nagelkerke R^2, which is a method to estimate the improvement of the fitted model from the null model [113].

Meta-analysis

Meta-analysis is a set of methods to combine data from multiple studies. It is useful for GWASs to increase power to detect association signals, as individual studies may be small and underpowered. In addition, meta-analysis only requires summary statistics and not individual level data from each GWAS, thus increasing the ability to share data from multiple studies. I conducted meta-analysis of GWASs using METAL [114]. In R, meta-analysis was conducted using the 'meta' package.

There are two common methods for meta-analysis: fixed effects and random effects. Fixed effects meta-analysis assumes that the true effect size is the same every study. The random effects meta-analysis assumes that each study is estimating different (though similar) effects, sampled from a distribution and variability that depends on the variance between studies [115,116]. The aim of random-effects meta-analysis is to understand the distribution of effects across studies. If there is no between-study heterogeneity, then the fixed and random effects models produce the same results. If there is greater between-study heterogeneity, the random effects estimates usually have larger variance.

There are two commonly used statistics to describe heterogeneity in meta-analysis. Cochran's Q is calculated as the weighted sum of squared differences between individual study effects and the summary effect [115]. It is based on a chi-squared distribution with the degrees of freedom based on the number of studies ($n – 1$). This is used to determine if there is statistically significant heterogeneity. If the number of studies is small, this test may be underpowered to detect heterogeneity so a threshold of $p < 0.1$ is recommended [115,117].

The I^2 statistic estimates the percentage of variability in the results that is due to real differences and not to chance. Unlike Cochran's Q, it accounts for the number of studies. It is calculated by dividing the Q minus degrees of freedom by Q itself. When I^2 is 0%, the variability in effects between the studies can be explained by chance. If it is over 50%, it indicates there is substantial heterogeneity between the studies that is not due to chance alone.

Principal Components Analysis

Principal Components Analysis (PCA) is a method of data reduction. This is helpful when there are many variables present which may be related to each other. PCA is a way of simplifying the dataset to reveal the underlying structure as Principal Components which explain the most variability. It is useful to reduce dimensionality of the data without losing information (e.g. by removing variables). The eigenvector is the direction of the Principal Component, while the corresponding eigenvalue is the variance in the data in that direction. The most significant variance is found on the first component, and each subsequent component is orthogonal to the last (linearly independent and uncorrelated) and has less variance. The number of eigenvectors and eigenvalues from PCA corresponds to the number of dimensions (the number of input variables) in the dataset.

Genetic methods

Genome-wide association studies

A GWAS is an unbiased search across the genome for common variants associated with disease status (in case-control studies) or phenotypes. The common disease, common variant hypothesis suggests that complex traits are linked to multiple, numerous common variants. In PD, it is likely that both the common disease common variant hypothesis and the multiple rare variant hypothesis can be true at the same loci [118]. It is likely that both common and rare risk alleles exist at a single locus.

A GWAS is conducted in unrelated individuals, in contrast to traditional linkage studies which analyse family members. GWASs focus on single nucleotide polymorphisms (SNPs), usually those with minor allele frequency (MAF) greater than 1% or 5%. Each variant is statistically tested for association with disease (in the case-control studies) or a phenotype of interest. In order to correct for multiple testing, a p-value threshold of 5×10^{-8} is typically used in GWASs. Power in these studies is affected by the effect size of the loci, the allele frequency, sample size, and the standard deviation of the characteristic/ trait. GWASs can be conducted under different genetic models; the most common is the additive model which assumes that risk increases with each minor allele (0, 1, and 2). The alternative genetic models are the dominant model (e.g. {AA, AT} vs. TT) and the recessive model (e.g. AA vs. {AT, TT}). GWASs can be conducted with a variety of statistical models, such as logistic regression (in case-control studies), linear regression (for quantitative phenotypes) and others models. For linear or logistic regression models, I conducted GWASs in rvtests using the single variant Wald test [119]. This fits the alternative model (as opposed to the null model) and estimates the effect size, and can be used for both binary and continuous traits. For other statistical models, such as survival analysis, I used R v3.6 on the UCL kronos High Performance Computing (HPC) cluster to conduct GWASs.

A quantile-quantile (QQ) plot is a plot of the observed vs. expected p-values, or the corresponding χ^2 test statistics, from an association study. Deviation from the null (the expected values) through the entire distribution suggests there may be a systematic source of spurious association - likely population stratification or cryptic relatedness [120]. Deviation just at the tail end of the distribution indicates true associations at susceptibility loci [120]. Related to the QQ plot, the lambda value (also known as the genomic inflation factor) is calculated by dividing the median of the observed χ^2 distribution by the median of the expected χ^2 distribution.

Annotation of GWAS results

I used the online platform Functional Mapping and Annotation (FUMA) of Genome-Wide Association Studies (https://fuma.ctglab.nl/) (version 1.3.6) to annotate GWAS summary statistics [121]. FUMA can be used to help prioritise genes which may be relevant to disease by adding information from biological datasets and tools. FUMA identifies independent SNPs and risk loci, annotates them with their functional consequences, and maps them to genes.

FUMA can perform both positional and eQTL mappings to map SNPs to genes (SNP2GENE process). In positional mapping, SNPs are mapped to genes based on physical position (up to 10kb away). Gene annotation is performed using Ensembl genes (build 85). SNPs are mapped to rsIDs using the dbSNP build 146. The 1000 Genomes Phase3 EUR reference panel is used to account for the LD structure.

I ran FUMA with standard settings, with the exception of performing eQTL mapping in addition to positional mapping. Here, FUMA can map SNPs to genes that they are likely to affect expression of, up to 1 Mb away (cis-eQTLs), using data from GTEx v6 and v7.

From the mapping, FUMA generates a list of prioritised genes which are used for the GENE2FUNC process. This provides data on gene expression in different tissues, tissue specificity (whether there is a difference between the genes of interest and pre-defined differentially expressed genes in each tissue type), and gene sets (whether the genes of interest are overrepresented in any pre-defined gene sets including those from the Molecular Signatures Database [MsigDB], and Gene Ontology [GO]) [121]. In particular, I looked for enrichment of gene-sets or pathways in Gene Ontology (GO; MsigDB c5), Reactome (MsigDB c2), and the Kyoto Encyclopedia of Genes and Genomes (KEGG; MsigDB c2).

Multi-marker Analysis of GenoMic Annotation (MAGMA)

Multi-marker Analysis of GenoMic Annotation (MAGMA) is a tool for gene and gene-set analysis of GWAS results [122]. It can be run through the FUMA online platform. Gene and gene-set analysis can be more powerful than the single SNP vs. phenotype analysis that is conducted in GWAS [122].

In gene analysis, MAGMA aggregates SNP-level p-values to the level of the whole gene. It quantifies the association that each gene has with the phenotype of interest. SNPs are mapped to genes based on physical position. The standard setting in FUMA is without a window around the gene, however I also ran MAGMA in FUMA with a window of 35kb upstream and 10kb downstream of each gene, as most transcriptional regulatory elements fall within this interval. Gene locations for protein coding genes are based on the National Centre for Biotechnology Information (NCBI) build 37.3 definitions (genome build GRCh37/hg19) (https://ctg.cncr.nl/software/magma).

The MAGMA analysis is run independently of the SNP2GENE positional mapping described above, hence the window size for mapping SNPs to genes can be different.

The MAGMA gene analysis is independent from and differs from the GENE2FUNC tests in that it tests all the SNP p-values, whereas the GENE2FUNC annotation in FUMA only tests for enrichment of prioritised genes.

In gene-set analysis, MAGMA tests whether the genes in a gene-set are associated with the phenotype of interest, and whether these are more strongly associated than other genes [122]. Gene-set analysis is performed for curated gene sets (c2.all) and GO terms obtained from MsigDB v6.2 (biological processes [c5.bp], cellular components [c5.cc], and molecular functions [c5.mf]). Bonferroni correction is applied for the number of gene sets that are tested (n = 10678 in FUMA v1.3.6).

Code availability

I have made all my analysis scripts and code publicly available at https://github.com/huw-morris-lab.

Chapter 3 : Association of rare Mendelian mutations with clinical features

Introduction

PD affects approximately 140 in 100,000 people within the UK [123]. It is caused by genetic mutations in *LRRK2*, *SNCA*, *Parkin* (*PARK2*), and *PINK1* in up to 10% of patients according to previous studies [124–126]. These genetic factors also influence clinical features of the disease, such as age at onset [66,67,124,127,128], motor features, presenting symptoms, disease progression [69] and cognition [63,68,129].

However, many previous studies have focussed on highly selected cohorts recruited from specialist clinics. This is likely to lead to bias in both estimates of frequency and clinical characteristics associated with specific genetic mutations.

In order to overcome some of these issues, I analysed data from Tracking Parkinson's, a large-scale, population-based prospective cohort study of recently diagnosed and early onset PD patients in the UK. It is the largest single cohort study of PD and is relatively unbiased. Analysis of this cohort is important to: 1) develop more accurate estimates of genetic risk and the likelihood of a known genetic cause overall, as well as in specific patient sub-groups; 2) estimate the likelihood of further high risk genes that have not yet been identified, and 3) understand the contribution of Mendelian gene variation to the phenotype of PD.

Several studies have examined the frequency of gene mutations in early onset PD patients [130,131]. However, some mutations, such as those in *LRRK2*, are also present at a significant rate in non-familial late onset PD patients [132]. Previous studies have also sometimes used single techniques such as partial Sanger sequencing, which are not able to detect copy number variation (which is common in *Parkin*) and less common point mutations. In my analysis, mutations were comprehensively identified using a range of different genetic screening methods, including whole-exome sequencing, Multiplex Ligation-dependent Probe Amplification (MLPA), and Sanger sequencing.

The aim of this study is to describe the frequency of pathogenic Mendelian gene variants in the general PD population and in specific disease sub-groups. In addition, I sought to understand the relationship between Mendelian mutations and clinical phenotype at presentation, with some preliminary longitudinal analysis of *GBA* and *LRRK2* carriers.

Methods

PD patients were recruited to Tracking Parkinson's from sites across the UK. Patients were required to have a clinical diagnosis if PD fulfilling Queen Square Brain Bank criteria [133].

Patients with disease duration of less than 3.5 years at time of diagnosis were recruited as 'recent onset' participants. Patients with disease duration of greater than 3.5 years at time of diagnosis and age at onset ≤ 50 years were recruited as 'established young onset' participants. Patients were recruited regardless of ethnicity, including Jewish ethnicity. Full eligibility criteria, exclusion criteria and methods of recruitment have been described previously [99]. Importantly, unlike most studies of this type, patients were recruited irrespective of any prior information on genetic status.

Participants' motor features and non-motor features were assessed using standardised and validated scales.

Pathogenic mutations in the studied genes were defined according to MDSGene (http://www.mdsgene.org) [51,134], and the Parkinson Disease Mutation Database (PDmutDB; http://www.molgen.vib-ua.be/Parkinson's diseaseMutDB/). Variants that did not meet pathogenicity criteria according to MDSGene (variants classified as 'benign') were not reported.

Genetic analysis of PD gene mutations

Point mutations in *Parkin*, *PINK1*, and *GBA* were identified with Sanger sequencing. The full results of *GBA* sequencing and analysis of baseline clinical features have been reported separately [135], however here I conducted a preliminary analysis of *GBA* carriers longitudinally.

Whole exome sequencing was performed at UCL in a subset of young-onset and familial patients (N=489). Exome sequencing data was screened for pathogenic variants in *SNCA*, *LRRK2*, *Parkin*, *PINK1*, *DJ-1* and *VPS35*.

2,106 patients with Parkinson's disease were genotyped for the *LRRK2* G2019S mutation using the 'Kompetitive' allele-specific polymerase chain reaction (KASP) assay (LGC Genomic Solutions).

SNP array genotyping was completed for 2116 samples. Samples were genotyped using the Illumina HumanCore Exome array supplemented with custom content. Imputation was performed by Dr Leon Hubbard (Cardiff University). Genotypes were aligned to the 1000 Genomes Phase 3 v5 mixed population reference panel [136] (build hg19/ GRCh37) and imputed using Minimac3 [137] on the Michigan Imputation Server.

Genotyping in young-onset patients

Patients with age at onset ≤ 50 were screened for point mutations in *Parkin* and *PINK1* using Sanger sequencing. MLPA was performed to detect and confirm copy number variation in *Parkin*, *PINK1*, *DJ1* and *SNCA*. Of 424 patients, 291 (68.7%) were successfully genotyped for *Parkin* and *PINK1* with both MLPA and Sanger sequencing. Eleven patients were screened for copy number variants using MLPA but were not Sanger sequenced. Exome sequencing was performed in 269 patients.

For our final phenotype-genotype analyses, I included young-onset patients if MLPA had been completed, and either Sanger sequencing or exome sequencing, or both, had been completed. The combination of these methods was selected in order to detect both copy number variants and point mutations in *Parkin* and *PINK1*. In total, 302 patients with age at onset ≤ 50 were included for final analysis.

Genotyping in late-onset patients

Exome sequencing was performed in 219 late-onset patients with a positive family history of PD and 1 patient with missing AAO and a positive family history.

In late-onset patients with 2 or more additional family members affected by PD, MLPA was performed in 65 of 74 (87.8%) patients.

For the final phenotype-genotype analyses, I included late-onset patients if either *LRRK2* KASP genotyping or exome sequencing had been successfully completed. In total, 1701 late-onset patients were included for final analysis, as well as 2 patients with missing AAO.

In total, 2005 patients with PD were included for final analysis (302 young-onset, 1701 late-onset, 2 missing AAO).

Statistical analysis

Demographic characteristics were compared using t-tests, Fisher's exact tests for proportions, or two-sample proportion tests. Linear regression was used for comparisons of demographic characteristics with covariate adjustment. To assess the association between clinical outcomes and genetic status, I used linear regressions of continuous scores against gene status (mutation positive or mutation negative) adjusting for age at assessment, disease duration at study entry, sex and LEDD. Hoehn and Yahr stage, MoCA subdomain and dystonia comparisons were conducted using ordered logistic regression. Motor subtype was analysed using multinomial logistic regression with the tremor dominant group as the comparator. All p-values were 2-tailed. I applied the Bonferroni correction for multiple testing for the number of independent tests in Table 3.5 and 3.7. Analysis was conducted using version 1 (31/05/3019) of the Tracking Parkinson's clinical dataset. Statistical analysis was conducted using STATA (version 14, StataCorp, Texas, USA) and R (version 3.5.1).

Prevalence estimates

I estimated the absolute numbers of PD patients with a Mendelian genetic cause in the UK using the following approach. I used age-specific prevalence rates from a previous UK meta-analysis [123] and applied the rates to the Office of National Statistics Great Britain mid-2016 population estimates [138] to derive an approximate number of all PD patients. The age distribution of the PD population (as a percentage) was used to standardise the rates of genetic PD within our cohort (per 100,000). From this, I derived the new age-standardised rate of genetic PD. I applied this age-standardisation method because our over-sampling of young onset cases has resulted in a non-representative age-distribution of patients. This new rate was then applied to the total PD population to estimate the absolute number of patients with a Mendelian genetic cause in the UK population. It is important to note that as I have derived the rates from our incident cases (excluded established young onset cases), I have assumed that the rates are representative of all prevalent cases. This may not be true if these Mendelian forms of PD are associated with better or worse survival, in which case our estimates will be either an under- or over-estimate of the true numbers. 95% confidence intervals were calculated using the Poisson distribution.

Longitudinal analysis

I conducted preliminary longitudinal analysis of *GBA* and *LRRK2* carriers. Only recent onset PD patients were followed longitudinally, and not established early onset patients, so not all mutation carriers had longitudinal data available. There was only one recent-onset PD patient with a *Parkin* mutation, so longitudinal analysis was not possible. Longitudinal change in the MDS-UPDRSIII and the MoCA was analysed using mixed effects models, adjusting for age at onset, gender, and their interactions with years from the baseline visit. This analysis was conducted with the lme4 and lmerTest packages in R. Version 2 (17/06/2020) of the Tracking Parkinson's clinical dataset was used for longitudinal analysis.

GBA variants were classified using the same criteria as previously published in this cohort [135] and similar studies [71]. Group 1 are variants that are pathogenic for Gaucher's Disease (GD) in the homozygous state and associated with PD risk in the heterozygous state, including L444P and N370S. Group 2 are variants that have been linked to GD when occurring with other variants and are also associated with PD risk, including E326K and T369M. Group 3 are variants of unknown significance. Patients carrying these variants (and no other Group 1 or Group 2 variants) were grouped with patients who were screened and negative for *GBA* mutations. I compared patients carrying any pathogenic *GBA* variant (Group 1 and 2 combined) to non-carriers, as well as stratified analysis of Group 1 and Group 2 variants separately.

Results

Table 3.1 shows the baseline demographics for participants that met PD diagnostic criteria. Data are presented separately for three groups below, according to inclusion criteria for recruitment. Early-onset patients were separated into recently diagnosed and established PD patients, as only the recent onset patients represent an incident, largely population-based cohort. For this reason, only recent onset patients were used to estimate the prevalence of genetic forms of PD in the UK.

1. Recent late onset Parkinson's disease patients (AAO > 50, disease duration ≤ 3.5 years at time of diagnosis),
2. Recent early onset Parkinson's disease patients (AAO ≤ 50, disease duration ≤ 3.5 years at time of diagnosis)
3. Established early onset Parkinson's disease patients (AAO ≤ 50, disease duration > 3.5 years at time of diagnosis).

37 patients received a revised alternative diagnosis other than PD or had conflicting dopamine transporter (DaT) scan results and were excluded from further analysis. On rare occasions, *LRRK2* mutations may be present in progressive supranuclear palsy or atypical parkinsonian patients [139,140], however I did not identify any pathogenic mutations in these patients. None of the rediagnosed patients carried a *GBA* mutation.

Summary of genotyping

For young-onset patients, I included samples for final analysis if MLPA had been completed, and either Sanger sequencing or exome sequencing or both had been successfully completed. In total, 302 patients with age at onset ≤ 50 were included for final analysis of *Parkin* and *PINK1*.

For late-onset patients, I included patients for final analysis if the samples had been genotyped with the *LRRK2* KASP assay for G2019S, and/or exome sequencing. In total, 1701 late-onset patients were included for final analysis, as well as 2 patients with missing age at onset.

In total, 2005 PD patients with were included for final analysis (302 early-onset, 1701 late-onset, 2 missing age at onset).

Table 3.1. Baseline demographics for all PD patients with known age at onset in Tracking Parkinson's.

	Recent, late onset patients (AAO>50, ≤3.5 years from diagnosis) N=1799	**Recent, early onset patients (AAO≤50, ≤3.5 years from diagnosis) N=197**	**Established early onset patients (AAO≤50, >3.5 years from diagnosis) N=227**	**Total N=2223**
Age at recruitment (years)	69.3 (7.5)	48.8 (6.2)	54.5 (7.7)	66.0 (10.2)
Age at onset (years)	66.4 (7.7)	43.7 (5.6)	41.1 (7.1)	61.8 (12.1)
Disease duration at diagnosis (years)	1.3 (0.9)	1.4 (1.0)	11.4 (6.4)	2.4 (3.8)
Disease duration at entry (years)	2.9 (2.1)	5.2 (6.6)	13.1 (7.4)	4.0 (4.6)
Family history (n, (%))				
No family history	1442 (80.2%)	145 (73.6%)	166 (73.1%)	1753 (78.9%)
1 additional affected family member	267 (14.8%)	41 (20.8%)	47 (20.7%)	355 (16.0%)
2 additional affected family members	59 (3.3%)	8 (4.1%)	8 (3.5%)	75 (3.4%)
3 additional affected family members	11 (0.6%)	2 (1.0%)	4 (1.8%)	17 (0.8%)
4 or more additional affected family members	4 (0.2%)	0 (0.0%)	1 (0.4%)	5 (0.2%)
Consistent with dominant inheritance	305 (17.0%)	49 (24.9%)	57 (25.1%)	411 (18.5%)
Consistent with recessive inheritance	36 (2.0%)	2 (1.0%)	3 (1.3%)	41 (1.8%)
Consanguinity				
Non-consanguineous	1741 (96.8%)	191 (97.0%)	220 (96.9%)	2152 (96.8%)
Consanguineous	16 (0.9%)	2 (1.0%)	2 (0.9%)	20 (0.9%)
Ethnicity				
White	1742 (96.8%)	188 (95.4%)	211 (93.0%)	2141 (96.3%)
Asian or Asian British	16 (0.9%)	3 (1.5%)	8 (3.5%)	27 (1.2%)
Black or Black British	10 (0.6%)	3 (1.5%)	2 (0.9%)	15 (0.7%)
Chinese	0 (0.0%)	0 (0.0%)	2 (0.9%)	2 (0.1%)
Mixed	4 (0.2%)	0 (0.0%)	0 (0.0%)	4 (0.2%)
Other	2 (0.1%)	1 (0.5%)	0 (0.0%)	3 (0.1%)
Sex				
Male	1181 (65.7%)	124 (62.9%)	149 (65.6%)	1454 (65.4%)

AAO = Age at onset

Consistent with dominant inheritance=family members from multiple generations affected.

Consistent with recessive inheritance=family members only from the same generation affected.

Summary of mutations identified

I identified 14 different pathogenic mutations in *LRRK2*, *SNCA*, *Parkin,* and *PINK1* in 29 out of 2005 patients (1.4%, 95% CI 0.9-2.0%) (Tables 3.2 and 3.3). This estimate is conservative as not all samples were comprehensively tested, therefore the true mutation rate may be higher.

18 patients carried a mutation in *LRRK2*, 1 patient carried a *SNCA* mutation, 8 patients carried biallelic *Parkin* mutations and 2 patients carried biallelic *PINK1* mutations. No patients were found carrying pathogenic mutations in *VPS35* or *DJ1*. No patient carried pathogenic mutations in more than one gene. 3 patients carried the *LRRK2* G2019S mutation and additionally one or more mutations in *GBA* (p.E326K and p.P122H). The mean age at onset for patients carrying mutations in both *LRRK2* and *GBA* mutations was 43.2 years (SD=5.1), compared to an onset age of 56.5 years (SD=12.9) for *LRRK2* mutation carriers without *GBA* mutations.

I identified 9 patients carrying single heterozygous pathogenic mutations in *Parkin* and *PINK1*. Previous analysis of this cohort showed no differences between carriers of single heterozygous *Parkin* mutations (including mutations of uncertain pathogenicity) and non-carriers other than in olfaction [141], therefore patients with single heterozygous mutations in recessive genes were analysed as non-carriers. One patient carried 3 pathogenic mutations in *Parkin*.

Mutations were common in patients with very early onset and patients with multiple family members also affected by PD. 18.8% (3/16; 95% CI 6.6 – 43.0%) of patients with onset ≤ 30 carried pathogenic mutations. In early-onset patients, 18.2% (4/22; 95% CI 7.3 – 38.5%) of patients with 2 or more additional affected family members carried pathogenic mutations. In late-onset patients, 4.2% (3/72; 95% CI 1.4-11.5%) of patients with 2 or more additional affected family members carried pathogenic mutations.

Notably, the *LRRK2* G2019S mutation was more common in early-onset patients (2.2%, 9/408; 95% CI 0.7 – 3.6%, Table 3.4) than in later onset patients (0.4%, 7/1701; 95% CI 0.1 – 0.7%), p=0.001 (Fisher's exact test, OR = 5.5, 95% CI 1.8-17.3). In addition, early onset patients were equally likely to have recessive (2.5%, 10/408) and dominant pathogenic mutations (2.2%, 9/408). Pathogenic mutations were only identified in patients reporting ‘White’ ethnicity (N=2005 genotyped).

IBD analysis was conducted based on 25,781 SNPs in linkage equilibrium. This showed that none of the mutation carriers were related to each other (pi-hat <0.1 for all, indicating no closer relations than third-degree relatives).

Table 3.2. Overall frequency of dominant gene mutation carriers for known pathogenic variants in successfully genotyped patients.Percentages and 95% CIs are shown in brackets.

	Early onset N=408	**Late onset N=1701**	**All N=2003**
LRRK2	9 (2.2%; 0.8-3.6%)	9 (0.5%; 0.2-0.9%)	18 (0.9%; 0.5-1.3%)
SNCA	0 (0%; 0.0 – 0.9%)	1 (0.06%; 0.01-0.3%)	1 (0.05%; 0.04-0.1%)
All autosomal dominant (*LRRK2* and *SNCA*)	9 (2.2%; 0.8-3.6%)	10 (0.6%; 0.2-1.0%)	19 (0.9%; 0.5-1.4%)

Table 3.3. Overall frequency of biallelic recessive gene mutation carriers for known pathogenic variants in successfully genotyped early onset patients (age at onset ≤ 50).Percentages and 95% CIs are shown in brackets.

Parkin	**Early onset N = 302**
Homozygous	0 (0%; 0.0-0.1.3%)
Compound heterozygous	8 (2.6%; 0.8-4.5%)
PINK1	
Homozygous	1 (0.3%; 0.06-1.9%)
Compound heterozygous	1 (0.3%; 0.06-1.9%)
All autosomal recessive (*Parkin* and *PINK1* biallelic mutations)	10 (3.3%; 1.3-5.3%)

Table 3.4. Rate of known dominant pathogenic mutations based on clinical presentation.

	LRRK2 **N=18**	***SNCA*** **N=1**	**Rate of all pathogenic dominant mutations**
Age at onset			
≤20 years (N=4)	0/4 (0%)	0/4 (0%)	0/4 (0%)
≤30 years (N=18)	0/18 (0%)	0/18 (0%)	0/18 (0%)
≤40 years (N=118)	2/118 (1.7%)	0/118 (0%)	2/118 (1.7%)
≤50 years (N=408)	9/408 (2.2%)	0/408 (0%)	9/408 (2.2%)
≤60 years (N=784)	10/784 (1.3%)	1/784 (0.1%)	11/784 (1.4%)
≤70 years (N=1552)	17/1552 (1.1%)	1/1552 (0.06%)	18/1552 (1.2%)
≤80 years (N=2050)	18/2050 (0.9%)	1/2050 (0.05%)	19/2050 (0.9%)
All (N=2109)	18/2109 (0.9%)	1/2109 (0.05%)	19/2109 (0.9%)
Mean age of onset in years (SD)	54.3 (12.9)		54.1 (12.6)
Family history			
No other family members affected	8/1658 (0.5%)	0/1658 (0%)	8/1658 (0.5%)
1 other family member affected	7/344 (2.0%)	0/344 (0%)	7/344 (2.0%)
2 other family members affected	1/72 (1.4%)	1/72 (1.4%)	2/72 (2.8%)
3 other family members affected	2/17 (11.8%)	0/17 (0%)	2/17 (11.8%)
4 or more family members affected	0/5 (0%)	0/5 (0%)	0/5 (0%)

LRRK2

I identified 18 patients carrying heterozygous *LRRK2* mutations, either G2019S (N=16) or R1441C (N=2). 55.6% (10/18) carriers reported a positive family history of PD.

Both *LRRK2* R1441C carriers reported a family history of PD. As the R1441C mutation was only screened through exome sequencing in familial and/or early-onset patients, these results for R1441C cannot be used to compare familial vs. non-familial patients.

I only included *LRRK2* G2019S mutation carriers for the analysis of family history. G2019S mutations were more common among patients with a positive family history (1.9%, 95% CI 0.5-3.1%) than patients without a family history of PD (0.5%, 95% CI 0.1-0.8%), p=0.009 (Fisher's exact test, OR = 3.9, 95% CI 1.3-11.8). However, within the G2019S carriers, 50% had a positive family history and 50% did not have a family history of PD (50%, 95% CI 25.5-74.5%).

LRRK2 mutation carriers (G2019S and R1441C carriers together) had an earlier mean onset (54.3 years, 95% CI 47.9-60.7) compared to non-carriers (61.7 years, 95% CI 61.2-62.2; p=0.01). Age at onset for *LRRK2* carriers ranged from 35.2 to 78.7 years. *LRRK2* mutations were more frequent in early onset (2.2%, 95% CI 1.0-4.2%) compared to late onset patients (0.5%, 95% CI 0.2-1.0%), p=0.003 (Fisher's exact test, OR = 4.2, 95% CI = 1.5-12.1).

Clinical features of *LRRK2* carriers compared to non-carriers are presented in Table 3.5 (excluding patients with recessive gene mutations). I did not include the *SNCA* carrier in this analysis given that previous literature suggests that *LRRK2* and *SNCA* mutation carriers have different clinical features [142]. I did not find any differences in clinical features between *LRRK2* carriers and non-carriers.

SNCA

SNCA copy number variants were screened with MLPA in 65 patients with familial PD with 2 or more family members affected. One patient (1.5%) carried a heterozygous whole gene duplication was identified, who reported 2 additional family members affected by PD. I was unable to compare the clinical features of *SNCA* carriers to non-carriers given that only one carrier was identified.

Table 3.5. Comparison of motor features, fluctuations and non-motor features by LRRK2 mutation status (LRRK2 carriers vs. non-carriers).Patients carrying biallelic recessive mutations and one patient carrying a SNCA mutation were excluded from analyses. Scores in the first 2 columns are means (SD), except for Hoehn and Yahr stage, symptoms present at diagnosis and motor subtype which are shown as N or proportions (%). Increasing scores and increasing beta values for motor and non-motor variables are associated with worse symptoms, with the exception of the MoCA test scores. Increasing scores and increasing beta values for the MoCA test are associated with better cognition.

Variable	**Mutation negative N=2082**	***LRRK2* positive N=18**	**Beta (95% CI) *LRRK2* carriers vs. non-carriers**	**p-value[a]**
Age at entry, years	66.0 (10.1)	60.1 (10.4)	-5.2 (-9.9, -0.5)	0.030[b]
Age at onset, years	61.8 (11.9)	54.3 (12.9)	-5.2 (-9.9, -0.5)	0.030[b]
Disease duration, years	4.0 (4.4)	5.2 (4.5)	0.7 (-1.3, 2.8)	0.482[c]
Delay to diagnosis - time from symptom onset to diagnosis, years	1.8 (2.9)	1.5 (1.3)	-0.4 (-1.8, 1.0)	0.580[c]
Motor features				
MDS-UPDRS III total score	23.4 (12.7)	28.6 (15.2)	6.7 (0.1, 13.3)	0.047
Severity score MDS-UPDRS-III/years from symptom onset	10.4 (11.8)	9.4 (7.3)	0.6 (-5.7, 6.8)	0.862[d]
Upper limb score, max 56	10.7 (6.3)	12.1 (6.3)	2.1 (-0.9, 5.1)	0.163
Lower limb score, max 32	5.1 (3.9)	6.8 (5.5)	1.7 (-0.2, 3.6)	0.085
Gait and freezing, max 8	1.1 (1.1)	1.6 (1.7)	0.4 (-0.1, 0.9)	0.097
Hoehn and Yahr stage			0.3 (-0.7, 1.2)	0.595
0-1.5 (%)	950 (46.0%)	7 (38.9%)		
2 or 2.5 (%)	957 (46.3%)	10 (55.6%)		
3+ (%)	160 (7.7%)	1 (5.6%)		
Symptoms present at diagnosis (%)				
Tremor	1499/2017 (74.3%)	13/18 (72.2%)	0.3 (-0.8, 1.6)	0.586
Rigidity	1410/1925 (73.2%)	13/18 (72.2%)	-0.08 (-1.2, 1.2)	0.891
Bradykinesia	1554/1966 (79.0%)	12/18 (66.7%)	-0.8 (-1.8, 0.3)	0.121
Postural problems	363/1898 (19.1%)	4/18 (22.2%)	0.009 (-1.5, 1.2)	0.989
Other	456/1827 (25.0%)	4/16 (25 %)	0.2 (-1.1, 1.3)	0.731

Motor subtype (%)				
Tremor dominant	835/1892 (44.1%)	7/17 (41.2%)		
Non-tremor dominant/ PIGD	813/1892 (43.0%)	10/17 (58.8%)	-2.8 (-0.5, 1.8)	0.246
Mixed	244/1892 (12.9%)	0/17 (0%)	-8.7 (NA)*	NA*
Motor complications				
MDS-UPDRS-IV total score	1.3 (2.8)	2.8 (3.3)	0.1 (-0.9, 1.2)	0.794
Dyskinesias - MDS-UPDRS IV part 1 and 2 sum, max 8	0.3 (1.0)	0.4 (0.9)	-0.2 (-0.5, 0.1)	0.259
Fluctuations - MDS-UPDRS IV part 3, 4 and 5 sum, max 12	0.9 (1.9)	2.1 (2.6)	0.3 (-0.4, 1.1)	0.408
Dystonia, max 4	0.2 (0.6)	0.3 (0.6)	0.01 (-0.2, 0.3)	0.915
Non-motor features				
Cognition - total MoCA score, max 30	25.2 (3.5)	25.4 (3.2)	-0.2 (-1.9, 1.4)	0.761
Visuospatial, max 5	4.3 (1.1)	4.2 (1.2)	-0.2 (-0.7, 0.3)	0.359
Naming, max 3	2.9 (0.3)	2.9 (0.3)	-0.05 (-0.2, 0.1)	0.535
Attention, max 6	5.2 (1.0)	5.3 (0.8)	0.1 (-0.4, 0.6)	0.690
Language, max 3	2.4 (0.8)	2.4 (0.7)	-0.03 (-0.4, 0.3)	0.865
Abstraction, max 2	1.6 (0.6)	1.7 (0.7)	0.003 (-0.3, 0.3)	0.983
Recall, max 5	2.7 (1.6)	2.9 (1.8)	0.05 (-0.7, 0.8)	0.898
Orientation, max 6	5.8 (0.5)	5.8 (0.5)	-0.03 (-0.2, 0.2)	0.756
LADS Anxiety score, max 18	4.5 (3.8)	5.8 (3.8)	0.9 (-0.8, 2.6)	0.287
LADS Depression score, max 18	4.5 (3.3)	5.1 (3.3)	0.3 (-1.2, 1.8)	0.706
Sleep disturbance -ESS score	7.1 (4.8)	9.7 (6.8)	1.6 (-0.7, 3.8)	0.173
REM Sleep Behaviour Disorder Screening Questionnaire score	4.8 (3.2)	6.4 (3.5)	1.0 (-0.5, 2.5)	0.191
Autonomic function: SCOPA total score	9.3 (5.8)	10.8 (6.4)	2.6 (-1.1, 6.3)	0.170

SD = standard deviation; CI = confidence interval; MDS-UPDRS = Movement Disorder Society Unified Parkinson’s Disease Rating Scale; PIGD = postural instability gait difficulty; MoCA= Montreal Cognitive Assessment; LADS = Leeds Anxiety and Depression Scale; ESS= Epworth Sleep Scale; RBDSQ = Rapid Eye Movement Sleep Behaviour Disorder Screening Questionnaire; SCOPA = SCales for Outcomes in PArkinson's disease.

[a] P value of clinical features of LRRK2 carriers together compared to non-carriers, excluding patients with recessive gene mutations and one patient with SNCA mutation. Adjusting for age at entry, gender, disease duration at entry/assessment and LEDD total, unless otherwise specified. [b] Adjusting for gender and disease duration at entry. [c] Adjusting for gender and age at entry. [d] Adjusting for age, gender and LEDD total. *Insufficient count to fit model

Early-onset patients

I identified 19/302 (6.3%) early-onset patients carrying pathogenic mutations in both dominant and recessive genes. The proportions of mutation carriers by age at onset and family history are presented in Table 3.6. Recessive gene mutation carriers had an earlier mean onset (32.7 years) compared to non-carriers (41.1 years), p<0.001, excluding dominant mutation carriers.

When considering all early-onset mutation carriers (*Parkin*, *PINK1*, *LRRK2* and *SNCA*) mutation carriers, the mean onset was also younger than non-carriers (37.5 vs. 41.1 years; p=0.02). Mutations were more frequent in patients with a positive family history (11.0%) than in patients with no family history of PD (4.2%), p=0.04 (Fisher's exact test, OR = 2.8, 95% CI 1.0-8.1).

Table 3.6. Cumulative rate of pathogenic mutations based on clinical presentation in successfully genotyped early onset PD patients (age at onset ≤ 50), N=302.

	PINK1 **(biallelic) N=2**	***Parkin*** **(biallelic) N=8**	**All recessive gene mutations N=10**
Age at onset			
≤20 years (N=4)	0/4 (0%)	2/4 (50%)	2/4 (50%)
≤30 years (N=18)	0/16 (0%)	3/16 (18.8%)	3/16 (18.8%)
≤40 years (N=118)	1/110 (0.9%)	6/110 (5.5%)	7/110 (6.4%)
≤50 years (N=408)	2/302 (0.7%)	8/302 (2.6%)	10/302 (3.3%)
Mean age of onset in years (SD)	42.3 (5.5)	30.3 (11.5)	
Family history			
No other family members affected	1/213 (0.5%)	4/213 (1.9%)	5/213 (2.3%)
1 other family member affected	1/67 (1.5%)	1/67 (1.5%)	2/67 (3.0%)
2 other family members affected	0/15 (0%)	3/15 (20%)	3/15 (20%)
3 other family members affected	0/6 (0%)	0/6 (0%)	0/6 (0%)
4 or more other family members affected	0/1 (0%)	0/1 (0%)	0/1 (0%)

Parkin

Of all early-onset patients that were successfully genotyped for *Parkin*, biallelic pathogenic *Parkin* mutations were present in 2.6% (8/302, 95% CI 0.8-4.4%). No *Parkin* carriers had homozygous mutations; all mutations were present in compound heterozygous state.

Parkin mutations were present in 20% (3/15, 95% CI 7.0-45.2%) of early onset patients with 2 additional family members affected by PD. However, there was no significant difference in the frequency of mutations in early onset patients with a positive family history (4.2%, 95% CI 0.2-8.4%) and without a family history (1.9%, 95% 0.05-3.7%), p>0.2 (Fisher's exact test, OR = 2.3, 95% CI 0.4-12.9). Early-onset patients from large PD families (2 or more additional family members affected) were more likely to carry a *Parkin* mutation (13.6%) than early onset patients with 1 or no additional family members affected (1.6%), p=0.01 (Fisher's exact test, OR = 8.5, 95% CI 1.2-47.9).

The clinical features of *Parkin* and *PINK1* mutation carriers compared to early-onset non-carriers are presented in Table 3.7. *Parkin* carriers had younger onset than early onset patients with *LRRK2* mutations (42.9 years, 95% CI 39.3-46.6), p=0.009. There was no difference in age at onset of *Parkin* and *PINK1* carriers, p>0.2.

PINK1

Bi-allelic *PINK1* mutations were present in 0.7% (2/302, 95% CI 0.2-2.4%) of all screened early-onset patients. Mutations were present in 1.1% (1/89) of early-onset patients with a positive family history and 0.5% (1/213) of patients with no family history. Mutations were not more frequent with patients with a positive family history, p=0.50 (Fisher's exact test, OR = 2.4, 95% CI 0.03-189.7).

Parkin and *PINK1* mutation carriers had earlier age at study entry and earlier age at onset than other early-onset non-carriers, adjusting for gender and disease duration (Table 3.7). They also had longer disease duration than non-carriers, adjusting for age at entry and gender (Table 3.7).

Parkin and *PINK1* mutation carriers also reported more postural problems at diagnosis than non-carriers and tended to report a higher rate of dyskinesias, after adjusting for age at entry, gender, disease duration and LEDD total, although this did not survive correction for multiple testing. They also tended to have more gait and freezing problems at assessment, after adjusting for age, gender, disease duration and LEDD total (p=0.021), although this was not significant after correction for multiple testing.

Finally, *Parkin* and *PINK1* carriers had better cognition than non-carriers as assessed by the MoCA, even after adjusting for age, gender, disease duration and LEDD (p=0.007). This appears to be driven by better performance in the attention subdomain (p=0.004) though one must be cautious in interpreting the sub-domains as they may be overly simplistic.

Table 3.7. Comparison of motor features, fluctuations and non-motor features of early onset patients by recessive gene status (*Parkin* and *PINK1* carriers vs. non-carriers), excluding patients carrying dominant gene mutations.Scores in the first 4 columns are means (SD), except for Hoehn and Yahr stage, symptoms present at diagnosis and motor subtype which are shown as N or proportions (%). Increasing values and increasing betas for motor and non-motor variables are associated with worse symptoms, with the exception of the MoCA test scores. Increasing values and increasing betas for the MoCA test are associated with better cognition. Cells with only a single case are indicated with brackets (N=1).

Variable	**Mutation negative**	**Mutation positive (bi-allelic)**			**Beta (95% CI)**	**p-value**[a]
	N=292	Total N=10	*Parkin* N=8	*PINK1* N=2	Carriers vs. non-carriers	
Age at entry, years	51.9 (8.1)	50.9 (11.1)	51.8 (12.2)	47.5 (5.9)	-7.0 (-10.9, -3.1)	0.001[b]
Age at onset, years	41.1 (6.2)	32.7 (11.5)	30.3 (11.5)	42.3 (5.5)	-7.0 (-10.9, -3.1)	0.001[b]
Disease duration, years	10.4 (7.6)	18.2 (14.4)	21.9 (14.4)	5.2 (0.4)	8.9 (5.0, 12.7)	<0.001[c]
Delay to diagnosis, years	2.4 (4.2)	4.5 (4.1)	5.2 (4.4)	2.2 (0.1)	2.2 (-0.6, 5.1)	0.123[c]
Motor features						
MDS-UPDRS-III total score	26.1 (14.9)	29.0 (24.0)	33.0 (23.6)	5.0 (N=1)	-3.3 (-14.4, 7.8)	0.564
Severity score MDS-UPDRS-III/years from symptom onset	4.1 (6.8)	2.4 (2.9)	2.7 (3.1)	0.9 (N=1)	-2.5 (-7.7, 2.8)	0.356[d]
Upper limb score, max 56	11.6 (6.7)	13.9 (8.8)	15.3 (8.7)	8.5 (9.2)	-1.1 (-5.5, 3.3)	0.621
Lower limb score, max 32	6.2 (4.4)	7.7 (5.6)	8.5 (6.0)	4.5 (3.5)	-0.1 (-3.1, 3.0)	0.973
Gait and freezing, max 8	1.6 (1.5)	3.2 (1.9)	3.6 (1.7)	1.5 (2.2)	1.1 (0.03, 2.1)	0.043
Hoehn & Yahr stage					1.8 (0.1, 3.6)	0.049
0-1.5 (%)	107 (36.7%)	1 (11.1%)	1 (12.5%)	0 (0%)		
2 or 2.5 (%)	140 (48.1%)	4 (44.4%)	3 (37.5%)	1 (100%)		
3+ (%)	44 (15.1%)	4 (44.4%)	4 (50%)	0 (0%)		
Symptoms present at diagnosis						
Tremor	188/263 (71.5%)	7/10 (70.0%)	6/8 (75.0%)	1/2 (50.0%)	-0.9 (-2.4 0.8)	0.275
Rigidity	204/255 (80%)	8/9 (88.9%)	6/7 (85.7%)	2/2 (100%)	0.7 (-1.2, 3.7)	0.561
Bradykinesia	209/257 (81.3%)	9/10 (90.0%)	7/8 (87.5%)	2/2 (100%)	15.1 (-55.4, NA)	0.986
Postural problems	39/252 (15.5%)	6/9 (66.7%)	6/7 (85.7%)	0/2 (0%)	2.3 (0.7, 4.0)	0.005
Other	54/229 (23.6%)	3/9 (33.3%)	3/7 (42.9%)	0/2 (0%)	0.4 (-1.6, 2.0)	0.684

Motor subtype (%)						
Tremor dominant	79/257 (30.7%)	2/8 (25.0%)	1/6 (16.7%)	1/2 (50%)		
Non-tremor dominant/ PIGD	150/257 (58.4%)	6/8 (75.0%)	5/6 (83.3%)	1/2 (50%)	0.4 (-1.4, 2.3)	0.646
Mixed/ Indeterminate	28/257 (10.9%)	0/8 (0%)	0/6 (0%)	0/2 (0%)	-9.5 (NA, NA)	>0.1
Motor complications						
MDS-UPDRS-IV total score	5.0 (4.9)	6.2 (5.7)	6.1 (6.3)	6.5 (3.5)	2.3 (-0.5, 4.5)	0.105
Dyskinesias (presence and severity; max 8)	1.3 (1.9)	2.3 (2.5)	2.1 (2.8)	3.0 (1.4)	1.2 (0.03, 2.3)	0.04
Fluctuations, max 12	3.0 (2.9)	3.3 (4.0)	3.4 (4.3)	3.0 (4.2)	0.9 (-0.8, 2.6)	0.309
Dystonia, max 4	0.7 (1.1)	0.6 (1.3)	0.6 (1.4)	0.5 (0.7)	0.1 (-0.7, 0.8)	0.891
Non-motor features						
Cognition - total MoCA score, max 30	25.6 (3.6)	27.6 (2.2)	27.4 (2.3)	29.0 (N=1)	3.0 (0.8, 5.2)	0.007
Visuospatial, max 5	4.4 (1.1)	4.3 (0.5)	4.4 (0.5)	4.0 (N=1)	0.07 (-0.6, 0.8)	0.847
Naming, max 3	2.9 (0.3)	2.9 (0.3)	2.9 (0.4)	3.0 (0.0)	0.08 (-1.2, 0.3)	0.441
Attention, max 6	5.1 (1.0)	5.6 (0.5)	5.5 (0.5)	6.0 (0.0)	0.9 (0.3, 1.6)	0.004
Language, max 3	2.5 (0.7)	2.3 (0.8)	2.4 (0.7)	2.0 (1.4)	-0.07 (-0.5, 0.4)	0.767
Abstraction, max 2	1.7 (0.6)	1.6 (0.7)	1.6 (0.7)	1.5 (0.7)	0.09 (-0.4, 0.5)	0.704
Recall, max 5	3.1 (1.6)	4.2 (1.3)	4.3 (1.4)	4.0 (1.4)	0.9 (-0.2, 2.0)	0.116
Orientation, max 6	5.7 (0.7)	6.0 (0.0)	6.0 (0.0)	6.0 (0.0)	0.3 (-0.08, 0.6)	0.131
LADS Anxiety score, max 18	6.6 (4.2)	6.1 (2.6)	6.3 (2.8)	5.5 (2.1)	-0.4 (-3.3, 2.4)	0.763
LADS Depression score, max 18	5.8 (3.5)	5.8 (2.3)	6.4 (1.8)	3.5 (3.5)	-0.2 (-2.7, 2.4)	0.901
Sleep disturbance, ESS score	9.0 (5.7)	8.5 (7.6)	9.5 (8.3)	4.5 (2.1)	-0.1 (-4.2, 4.0)	0.961
REM Sleep Behaviour Disorder Screening Questionnaire score	5.8 (3.4)	4.3 (2.5)	4.4 (2.8)	4.0 (0.0)	-1.2 (-3.6, 1.1)	0.307
Autonomic function: SCOPA total score	10.8 (6.9)	12.3 (7.4)	9.5 (4.8)	20.5 (9.2)	0.1 (-5.0, 5.3)	0.959

SD = standard deviation; CI = confidence interval; MDS-UPDRS = Movement Disorder Society Unified Parkinson’s Disease Rating Scale; PIGD = postural instability gait difficulty; MoCA= Montreal Cognitive Assessment; LADS = Leeds Anxiety and Depression Scale; ESS= Epworth Sleep Scale; RBDSQ = Rapid Eye Movement Sleep Behaviour Disorder Screening Questionnaire; SCOPA = SCales for Outcomes in PArkinson's disease.

[a] *P* value of clinical features of *Parkin* and *PINK1* carriers together compared to non-carriers, excluding patients with dominant gene mutations. Adjusting for age at entry, gender, disease duration at entry/assessment and LEDD total, unless otherwise specified. [b] Adjusting for gender and disease duration at entry. [c] Adjusting for gender and age at entry. [d] Adjusting for age, gender and LEDD total.

Prevalence

In the recent onset cohort (both early-onset and late-onset), the rate of pathogenic mutations was 1.0% (17/1787). This is a large-scale cohort unselected for age at onset, family history and genetic status. I used this to estimate the frequency of pathogenic mutations in the general UK PD population. The crude prevalence rate of genetic forms of PD is 951 per 100 000 (95% CI 892-1013, using the Poisson distribution). Age specific rates are presented in Table 3.8. The age-standardised rate of genetic forms of PD was 708 per 100 000 (95% confidence interval 657-762 per 100 000), standardised to the mid-2016 Great Britain population. This provides an estimate of approximately 725 genetic PD patients in a total of 102,403 patients in the UK currently living, using estimates from a meta-analysis [123] and the Office of National Statistics Great Britain population estimates for mid-2016 [138] assuming these genes do not impact on survival. A recent report from Parkinson's UK using primary care diagnosis estimated a larger number PD patients in the UK (145,519) in 2018 [143]. If this figure is more accurate, then the number of genetic PD cases would be larger (estimated at 1030).

Table 3.8. Age specific and crude prevalence rate of genetic forms of PD, using data from **recent onset** patients only.

Age	**Parkinson's disease genetic patients in cohort**	**Total number of Parkinson's disease patients in cohort (screened)**	**Age specific rates per 100,000 Parkinson's disease patients**
0-29	0	0	0
30-39	1	11	9091
40-49	4	58	6897
50-59	4	219	1826
60-69	5	728	687
70-79	2	633	316
≥80	1	138	725
Total	**17**	**1787**	
Crude prevalence per 100,000 Parkinson's disease patients	951 (525-1442)		
Age adjusted prevalence per 100,000 Parkinson's disease patients*	708 (612-713)		

*Age distribution derived from age-specific PD rates [123] applied to the UK mid-2016 population estimates [138].

Longitudinal analysis of *GBA* carriers

In the longitudinal dataset, an additional 9 patients were rediagnosed with a non-PD condition. None of these patients carried a pathogenic mutation. These patients were removed from analysis, leaving 1,960 PD patients with longitudinal data. The mean follow-up time was 3.6 years (median 3.4 years, SD = 2.1 years).

44 patients carried a GD-pathogenic mutation (Group 1) and 115 patients carried a PD-pathogenic mutation (Group 2; Table 3.9). 28 patients carried *GBA* mutations of uncertain significance; these were grouped with non-carriers. In total, 159 patients carried a *GBA* mutation (GD or PD-pathogenic) and 1,623 patients were screened and negative for *GBA* pathogenic mutations. The mean follow-up time for *GBA* carriers was 3.6 years (SD 2.1 years) compared to 3.7 years for non-carriers (SD 2.1 years). Mean disease duration at baseline was 3.0 years (SD 2.3 years) in *GBA* carriers compared to 3.2 years in non-carriers (SD 3.0 years).

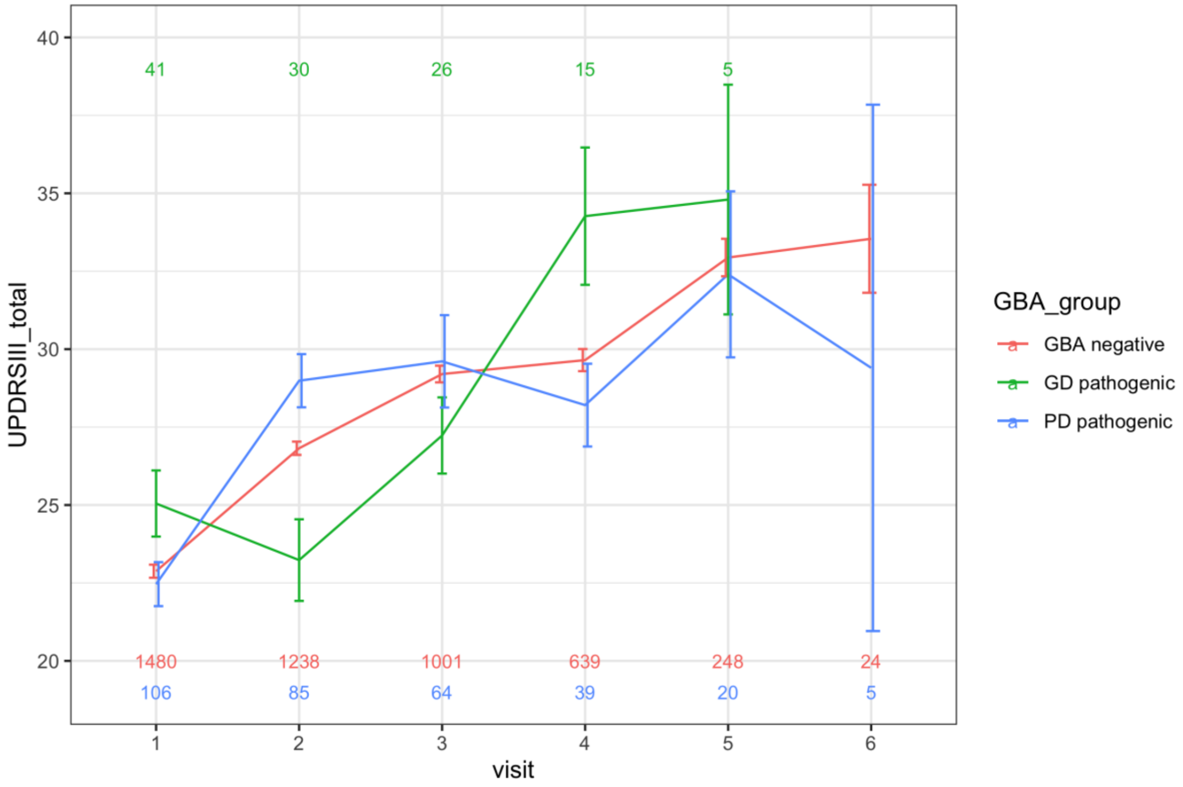
Table 3.9. Classification and frequency of *GBA* variants in 1,782 patients with longitudinal data that were screened for *GBA*.

Cases, n (%)	**Recognised GD pathogenic mutations (Group 1)**	**PD-associated non-GD variants (Group 2)**	**Rare variants of unknown significance (Group 3)**
29	p.L444P		
10	p.N370S		
5	p.R463C		
2	p.G202R		
2	p.R359S		
83		p.E326K	
35		p.T369M	
1 for each variant (0.06%)			p.D409H, p.F213I, p.G189V, p.G377S, p.K157Q, p.L383Xfs, p.L66P, p.M123T, p.N382Xfs, p.R163S, p.R257Q, p.S173S, p.E481Xfs, p.G10S, p.G325W, p.R170H, p.T323I, p.L175I, p.P55S, p.R262H, p.R329H, p.R395C, p.T267I, p.L268L
6			p.A456P
6			p.V460V
2			p.D140H
2			p.I308T
2			Ex4 hemizygous deletion

Based on classification used in Malek et al. [135] in the same cohort. Note that some patients carried more than one variant.

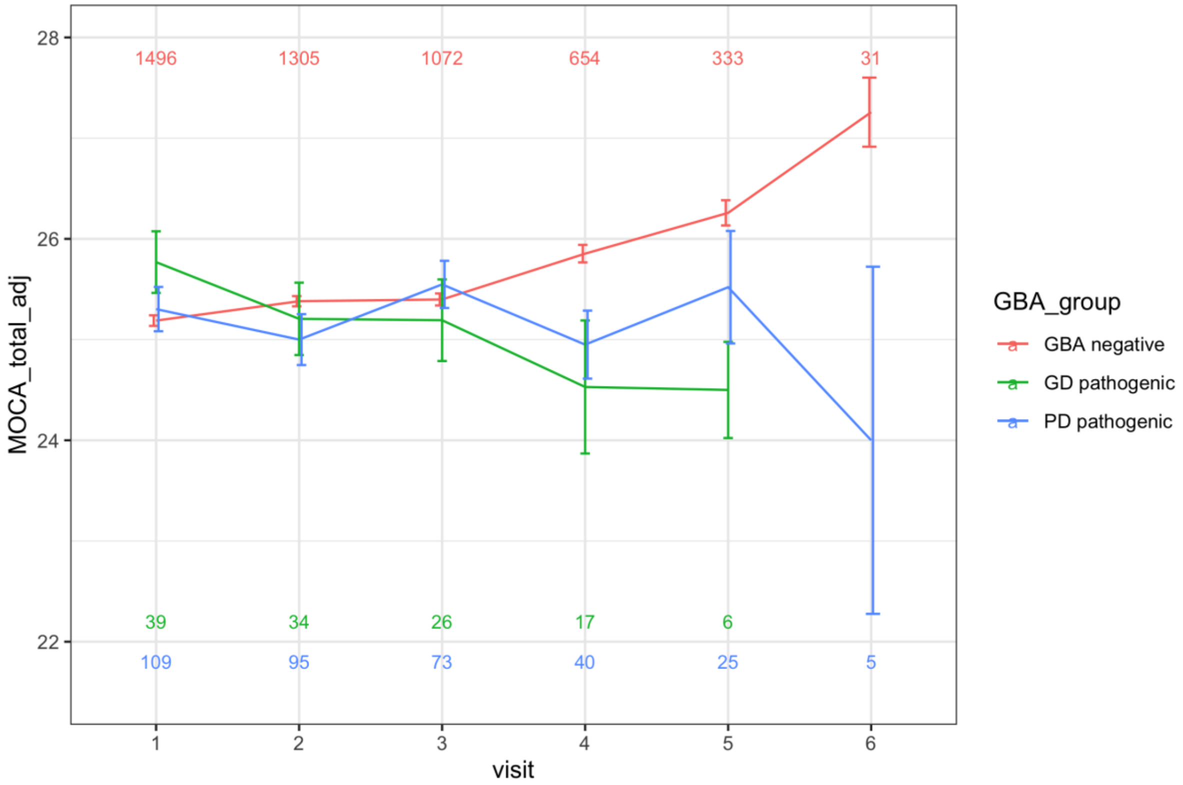
GBA carriers (GD- and PD-pathogenic variants combined) tended to progress more rapidly in the MDS-UPDRSIII compared to non-carriers, after adjusting for age at onset and gender (beta = 0.7, p = 0.04) (Figure 3.1). When analysing GD-pathogenic carriers separately compared to non-carriers, there was no difference in the MDS-UPDRSIII change (beta = 0.5, p = 0.41) but there was a nominal difference between the PD-pathogenic carriers compared to non-carriers (beta = 0.7, p = 0.06).

Figure 3.1. Means (+/- standard error) of the MDS-UPDRSIII total score by *GBA* status.Any data points with < 5 individuals were removed. The number of patients in each group at each timepoint is annotated with the corresponding colour label. Visits are at 1.5 year intervals, with visit 1 being the baseline visit.



GBA carriers also had worse decline in the MoCA compared to non-carriers (beta = -0.3, p = 0.001). Here, negative effect sizes (betas) indicate worse cognitive decline as higher scores in the MoCA indicate better cognition (Figure 3.2). This effect appeared to be larger in the GD-pathogenic variant carriers compared to non-carriers (beta = -0.4, p = 0.004) than in the PD-pathogenic variant carriers compared to non-carriers (beta = -0.2, p = 0.04).

Figure 3.2. Means (+/- standard error) of the MoCA total score by *GBA* status.Any data points with < 5 individuals were removed. The number of patients in each group at each timepoint is annotated with the corresponding colour label. Visits are at 1.5 year intervals, with visit 1 being the baseline visit.



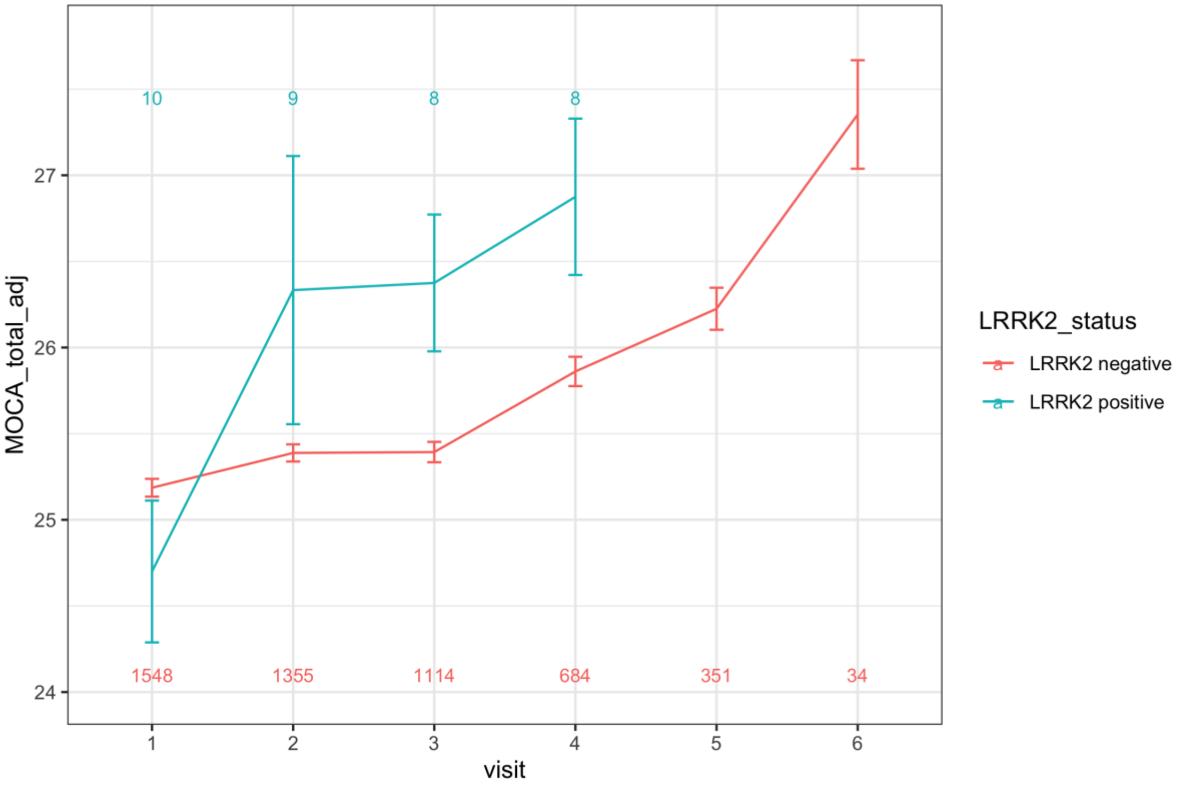
Longitudinal analysis of *LRRK2* carriers

There were 12 *LRRK2* carriers with longitudinal data available. One patient carried a *LRRK2* mutation and a *GBA* mutation (G2019S and E326K); this patient was excluded from analysis, Other *GBA* carriers were also excluded. After excluding *GBA* carriers, there were 11 (0.6%) *LRRK2* carriers and 1,685 patients screened and negative for *LRRK2.*

Mean follow-up was 4.5 years in *LRRK2* carriers (SD 2.6 years) compared to 3.7 years in non-carriers (SD 2.1 years). The mean disease duration at baseline was 3.0 years in *LRRK2* carriers (SD 1.6 years) compared to 3.1 years in non-carriers (SD 3.0 years).

There was no difference in progression in the MDS-UPDRSIII between *LRRK2* carriers and non-carriers (beta = -1.4, p = 0.2). In the MoCA, there no association between *LRRK2* status and cognitive progression (beta = 0.4, p = 0.1) (Figure 3.3).

Figure 3.3. Means (+/- standard error) of the MoCA total score by *LRRK2* status, excluding *GBA* carriers.Any data points with < 5 individuals were removed. The number of patients in each group at each timepoint is annotated with the corresponding colour label. Visits are at 1.5 year intervals, with visit 1 being the baseline visit.



Discussion

This study is the largest study examining the rate of known PD gene mutations. I report an overall rate of mutations of 1.4% (29/2005), across both early-onset and late-onset patients. In combination with *GBA* gene analysis in the same cohort [135], my results suggest that up to 10% of PD patients carry a known genetic variant that could potentially be targeted by new drug therapies. For instance, G2019S and other mutations in the *LRRK2* gene have been shown to increase kinase activity, and *LRRK2* kinase inhibitors that counteract this activity are currently being tested in phase 1 clinical trials as a potential therapeutic target (reviewed in [37,144,145].

Firstly, I showed that there are systematic clinical differences at baseline between *Parkin* and *PINK1* mutation carriers compared to other early-onset non-carriers. *Parkin* and *PINK1* had longer disease duration at baseline. These patients had more postural problems at diagnosis and better cognition than other early-onset patients, even after adjusting for age, disease duration, gender, and LEDD.

Secondly, this has enabled more accurate estimation of the prevalence of known pathogenic mutations in the general PD UK population, assuming there are no survival effects. I show clearly that *LRRK2* mutations are present at a significant rate in patients with onset under 50 years (2.2%), and that *SNCA* mutations are present in 1.5% of patients with a strong family history of PD (2 or more additional family members affected). In addition, my results highlight the importance of systematically screening for copy number variants in *Parkin*, *PINK1,* and *SNCA*, as these may be missed with methods such as exome sequencing. However, overall these mutations are rare in the PD population.

The strengths of this study lie in the relatively unbiased, population-based patient ascertainment. This increases the generalisability of our findings, in particular the prevalence estimates of PD patients carrying pathogenic mutations. A further strength of this study is inclusion of both early and late-onset patients, where previous genetic studies have tended to focus on early-onset patients.

LRRK2* and *SNCA

Mutations in *LRRK2* (PARK8, dardarin) were first identified in autosomal dominant, mostly late-onset families with Parkinson's disease [146–148]. The frequency of *LRRK2* mutations varies widely; mutations are more common in familial PD (5-6%) [149,150] than in sporadic disease (~1%) [151,152], but are present at higher frequencies in Ashkenazi Jewish (up to 28%) and North African patients (up to 41%) [43,125,153–156]. I found that *LRRK2* mutations were present at a rate of 0.9% overall, most commonly the G2019S mutation. This is comparable with a previous community-based cohort in the UK [154] and other Caucasian North American and UK cohorts with estimates between 0.4 and 1.7% [152,154,157–159].

R1441C mutations were present in 0.4% of early-onset and familial patients. This is in keeping with other studies showing the rarity of *LRRK2* R1441C mutations in Caucasian populations, with previous studies reporting frequencies between 0% and 0.3% [158,160,161].

Almost half of the *LRRK2* carriers did not report a family history of PD, in keeping with other studies [151,156]. This is likely because *LRRK2* mutations have incomplete penetrance, which is strongly age-dependent [43,132,156,162]. As the population ages, it is likely that increasing numbers of relatives carrying *LRRK2* will develop PD, and the prevalence of this form of PD will increase in the UK.

I did not find any differences in baseline motor or non-motor features between *LRRK2* carriers and non-carriers. There was also no evidence of differences in longitudinal progression, however this analysis only included 11 *LRRK2* carriers. This study is limited by the small number of *LRRK2* carriers, and larger sample sizes may reveal differences in progression. Previous cross-sectional studies suggest that *LRRK2* mutations are associated with less severe clinical symptoms [150], lower risk of cognitive impairment and better cognitive performance [43,163,164].

Recently, the first longitudinal study of *LRRK2* found that carriers had slower motor progression in the UPDRSIII, and nominally slower cognitive progression the MoCA, though this did not reach significance [44]. That cohort of 144 *LRRK2* carriers were later in disease stage (8 years at baseline) whereas in this study the mean disease duration of *LRRK2* carriers was 4.5 years at baseline. A potential explanation is that differences in *LRRK2* carriers are only apparent later disease stages, and this may explain why I did not clinical differences at baseline or in longitudinal analysis. Further longitudinal follow-up of *LRRK2* carriers in large case series is needed.

Furthermore, Saunders-Pullman et al. analysed Ashkenazi Jewish PD patients with and without *LRRK2* mutations [44]. It remains to be seen whether clinical progression is different in Ashkenazi Jewish PD *LRRK2* carriers compared to other populations. It is possible that the genetic background on which *LRRK2* mutations occur may modify the effects on progression, and some of this may differ with Ashkenazi Jewish ancestry. One study has shown that the PD Genetic Risk Score influences the penetrance of the *LRRK2* G2019S mutation [165]. Smaller studies have suggested that variants in *DNM3* [166] and *SNCA* [167] may influence *LRRK2* penetrance and age at onset, though these findings have not been consistently replicated.

SNCA mutations were first identified in large PD families with an autosomal dominant pattern of inheritance [34,41,168]. *SNCA* mutations are rare in studies of Caucasian patients [169–171]. I found one patient carrying a heterozygous duplication, comprising 1.5% of patients reporting 2 or more additional family members affected by PD. This is in line with previous studies reporting a mutation prevalence of 1.7% to 5.8% in familial PD patients [40,172–174].

It has previously been reported that *SNCA* mutation carriers have more frequent and more severe dementia, rapid progression, hallucinations and autonomic dysfunction [38–41,125,164,174–176]. *SNCA* triplications cause a more severe phenotype while duplications tend to cause more ‘typical’ Parkinson’s disease [173,177,178]. I was not able to compare clinical features in this cohort due to the rarity of *SNCA* mutations.

Early-onset PD

I found pathogenic mutations in 6.3% (19/302) of early-onset patients, including mutations in both dominant and recessive genes. These are comparable to the frequencies previously reported in other early-onset cohorts [130,131,179]. In accordance with previous studies [130,180], I showed that mutations were more common in patients with earlier onset.

Compound heterozygous *Parkin* mutations were identified in 2.6% of early-onset patients. While this is lower than other prevalence estimates in Caucasian populations [53,55,181,182], these findings are in accordance with a previous UK community-based study which found that *Parkin* mutations accounted for 3.7% of patients with onset under 45 years [131]. Mutations tended to be more common in familial (4.2%) than in sporadic patients (1.9%), and 20% of patients with 2 additional family members affected carried *Parkin* mutations. Previous studies suggest that *Parkin* mutations are more common in familial patients [130].

PINK1 mutation carriers were present in 0.7% of early-onset patients. This is comparable to the rate reported in a previous community-based study [131]. Mutations are more common in Asian and Italian patients [50,56,183–185], reflecting population-specific allele frequencies. Our findings are consistent with the low prevalence estimates in Northern Europe and North American patients [186,187]. However contrary to previous reports [131], I did not find that mutations were more frequent in patients with a family history of PD (1.1%) compared to sporadic patients (0.5%). This may be due to the small number of *PINK1* carriers.

After controlling for age and disease duration, I found that *Parkin* and *PINK1* carriers had earlier onset, reported more postural symptoms at diagnosis and had better cognition compared to other early-onset patients. This suggests that *Parkin* and *PINK1* carriers have slower progression, despite longer disease duration at study entry. I was not able to confirm differences in progression as only recent-onset patients were followed longitudinally.

This baseline data is consistent with previous studies showing that *Parkin* and *PINK1* mutations are generally associated with slower disease progression and less cognitive impairment [49,50,53,54,56,57,164,175,179,185]. Some studies have suggested that atypical features, such as dystonia, and psychiatric symptoms may be more common in *PINK1* and *Parkin* carriers [50,164,188], however I did not find evidence to support this. There is also substantial variability of the frequency of these symptoms in previous reports [164]. My findings are in line with a recent MDSGene systematic review, which suggested that recessive gene mutation carriers have less common cognitive decline, good treatment response and otherwise clinically typical disease [51]. While a few conflicting reports suggest there are no clinical differences between

Parkin carriers and non-carriers [189], my findings in a large population-based study suggest that there are clinical differences between mutation carriers and non-carriers. This may be associated with the lack of Lewy body pathology in the brain at post-mortem [190,191], although there are small numbers of *Parkin* cases with pathological data and there is variability in findings [176,192].

GBA

Analysis of this same cohort at baseline showed that GBA carriers had earlier age at onset, more advanced Hoehn and Yahr stage, and more frequent PIGD motor subtype than non-carriers [135]. However, there were no differences in baseline cognition in the MoCA or motor severity in the MDS-UPDRSIII.

In my longitudinal analysis, I found that *GBA* carriers had more rapid motor and cognitive progression than non-carriers. This contributes to evidence from both cross-sectional and longitudinal studies that *GBA* carriers have more rapid clinical progression and more severe clinical phenotypes that non-carriers. Cross-sectional studies indicate that *GBA* carriers have more severe cognitive impairment, motor impairment, neuropsychiatric symptoms and autonomic dysfunction [60,65,68,129]. Longitudinal studies show that *GBA* carriers have more rapid progression to dementia, motor impairment, and mortality [64,66,69,71,193].

This study found differences in longitudinal progression, but not baseline symptom severity, of *GBA* carriers, and this is likely because patients were assessed earlier in disease stage at baseline (mean disease duration 3 years) when compared to other cross-sectional studies (ranging from 6 to 9 years). This suggests that clinical differences in *GBA* carriers may only emerge later in disease course.

Studies also suggest that different *GBA* mutations have different effects on symptom severity and progression. Patients carrying ‘severe’ *GBA* mutations including L444P and N370S progress more rapidly than patients with ‘milder’ mutations such as E326K and T369M, though all carriers still progressed more rapidly than non-carriers [71,194]. In this study, there was some evidence that the GD-associated mutations were linked with more rapid cognitive progression, but not motor progression, than PD-associated *GBA* mutations. This analysis is still limited by the relatively small number of *GBA* carriers, especially when divided into subgroups by mutation type.

This may explain why I did not see stronger effects for severe vs. mild mutation carriers. In addition, the N370S mutation has sometimes been classified as a mild mutation because it is associated with non-neuropathic GD [66,193,195]. However, overall, my results are consistent with previous studies showing that *GBA* carriers have more rapid progression than non-carriers.

Limitations

This cohort was predominantly Caucasian and no pathogenic mutations were identified in non-Caucasian groups. Therefore, these results have limited application in other populations. Further studies are needed to establish the prevalence and clinical features of mutation carriers in other ethnic groups. For instance, previous studies have shown that *PINK1* mutations are more common in Asian patients [56].

These results are also limited by the lack of complete screening of all cases. Exome sequencing, MLPA, and *Parkin* and *PINK1* sequencing of all patients was not feasible due to cost limitations and the size of the cohort. Recessive gene mutations are rare in patients with older onset [130,131], however *Parkin* mutations have been found in late-onset patients with onset up to 78 years [196,197]. Therefore, there may have been a small number of mutation carriers that were not detected with these screening methods. This data therefore represents a minimal estimate of the frequency of pathogenic mutations, and the true numbers may be slightly higher. In addition, the genetic rates are based on both incident and prevalent cases. This is based on the assumption that survival and hence prevalence is not influenced by these genes, but if some genes (e.g. *Parkin* and *PINK1*) are associated with better survival then I may have underestimated the number of cases in the general population.

A further limitation is that, while this is a large cohort study, the rarity of pathogenic mutations means that our group difference comparisons may be under-powered to detect modest phenotypic differences.

Finally, this cohort is likely to still have some biases in it, given that this was not a rigorous community-based study collecting all cases of the condition. PD patients were recruited from specialist clinics at secondary care centres [99], however this included geriatric and general medicine clinics as well as neurology clinics. This goes some way to reduce recruitment bias where other studies have recruited from more specialist neurology clinics only. A previous community-based study of early-onset PD in Cardiff found no mutation carriers compared to a higher rate of carriers in referral-based series from neurologists and PD specialists in Wales and the UK [131]. However, this study only identified 14 early-onset patients in the community-based cohort and concluded that these types of studies may be underpowered for genetic epidemiology [131]. They suggest that the similar rate of mutations in the Wales and the UK cohort indicate that referral bias is not affecting estimates of mutation prevalence, although both these cohorts were still recruited from consultant neurologists and PD specialists, so may be missing PD cases in the community. It is also well-known that PD patients in specialist clinics are not fully representative of those in the community and general population [198], and this may affect my conclusions about the prevalence and clinical features of pathogenic mutations.

Conclusions

I show that Mendelian gene mutations are a rare but important cause of PD. Patients carrying *Parkin* or *PINK1* differ from other early-onset patients in baseline clinical features and potentially disease progression. It is likely that the progression of PD is determined by a range of genetic variants including common and rare variants. Though rare variants may have larger effects on progression, common variants are likely to be important for a larger number of patients and will be the main focus for the remainder of this thesis.

Chapter 4 : Clinical predictors of progression

Introduction

Prior to conducting GWASs to identify genetic determinants of progression, I first analysed the clinical variables that were associated with progression. This can be helpful in determining which clinical factors to include as covariates in the genetic models, as well as more broadly predicting the outcome in individual incident PD patients.

There is a clear difference between modelling progression for the purpose of prediction, compared to identifying relevant biological factors underpinning progression. When studying the biology of progression, it may not be appropriate to adjust for factors that may be intermediate phenotypes/markers or on the causal pathway between genotype and progression, as doing so may mask true associations [102,199]. The scenario may be different when trying to predict progression more accurately, as including more covariates/predictors may improve the accuracy of prediction. For example, REM sleep behaviour disorder (RBD) is associated with progression to dementia and cognitive impairment [200–202] and so might assist in an algorithm to predict dementia risk in early-stage PD patients, but it would not be appropriate to include as a covariate in a PD-dementia GWAS as it would reduce the power of the study. Here, my aim is to study the biological factors underpinning progression, however these genetic factors could later be incorporated into predictive models of PD progression.

Associated clinical factors may relate to co-pathology (likely important in ageing), aspects of the disease process (e.g. visual hallucinations reflecting cortical involvement in PD dementia) or disease heterogeneity/subtype.

Many previous studies have examined clinical predictors of progression in PD (summarised in Table 4.1). Age of onset, motor subtype or non-tremor dominant presentation, baseline impairment, and early cognitive impairment or dementia, are among clinical variables that have been shown in multiple studies to be associated with progression – whether mortality, disability, or motor or cognitive progression measured in various ways (reviewed in [203–206]). There are also other factors suggested to predict progression, such as gender, urate, and increased levodopa responsiveness, which have not been robustly replicated across studies [205,207]. Some of these studies have been very large retrospective studies using health record data but with only basic demographic variables available for analysis (such as Willis et al. [208]), whereas other studies are prospective cohort studies with in-depth clinical data collection but smaller sample sizes (Table 4.2).

Here, I tested a set of pre-specified clinical variables for their association with clinical milestones, based on previous literature: age at onset, gender, motor subtype, baseline severity, estimated progression prior to study entry, disease duration at study entry, and education. In addition, I examined disease progression in the first year of follow-up to determine whether this was predictive of later disease progression.

Table 4.1. Summary of studies analysing clinical predictors of different outcomes in PD: mortality, motor progression (such as Hoehn and Yahr stage, MDS-UPDRS/UPDRS scores), cognitive progression (such as MoCA scores, classification of dementia), and other markers of disability (such as the Schwab and England Activities of Daily Living scale, nursing home placement). Studies are listed more than once if multiple outcomes (e.g. mortality and dementia) were investigated. Only longitudinal prospective or retrospective studies are included. Studies are listed in no particular order. Each variable in a study is reported as predictive if it was significant in multiple regression models only (if performed; some studies did univariate analysis only).

	MORTALITY		**MOTOR PROGRESSION**		**COGNITIVE PROGRESSION**		**DISABILITY**	
Variable	**Predictive**	**Not predictive**	**Predictive**	**Not predictive**	**Predictive**	**Not predictive**	**Predictive**	**Not predictive**
Older age at onset/ diagnosis	Auyeung2012		Alves2005		Schrag2016	Keener2018	Alves2005	
	Keener2018		Zhao2010		Levy2000		Hely1999	
	Marras2005		Williams-Gray2013		Willis2012			
	Hely1999				Cereda2016			
	Willis2012				Pigott2015			
	Williams-Gray2013				Williams-Gray2013			
	Hely2005				Domellöf2015			
					Liu2017			
					Pedersen2013			
Male gender	Marras2005	Keener2018		Zhao2010	Levy2000	Keener2018		Hely1999
	Willis2012	Hely1999		Williams-Gray2013	Willis2012	Schrag2016		
		Hely2005			Cereda2016	Pedersen2013		
		Williams-Gray2013			Pigott2015	Williams-Gray2013		
					Liu2017	Domellöf2015		
Baseline motor severity	Keener2018	Hely1999	Zhao2010	Williams-Gray2013	Levy2000	Keener2018		Hely1999
	Marras2005	Hely2005			Pigott2015	Schrag2016		
	Levy2002	Williams-Gray2013			Pedersen2013			
					Williams-Gray2013			
					Liu2017			
					Domellöf2015			
Prestudy motor progression	Marras2005	Hely2005					Hely1999	
	Hely1999							
Disease duration		Keener2018	Alves2005	Burn2006	Cereda2016	Keener2018	Alves2005	Hely1999
		Hely1999	Zhao2010			Levy2000		
						Schrag2016		
						Burn2006		
						Pedersen2013		
						Pigott2015		
		Keener2018	Alves2005	Williams-Gray2013		Keener2018		Alves2005

Levodopa dose		Williams-Gray2013				Williams-Gray2013		
Baseline cognition/ dementia	Auyeung2012	Hely2005	Alves2005	Burn2006	Schrag2016	Keener2018	Alves2005	Hely1999
	Keener2018	Williams-Gray2013		Williams-Gray2013	Pigott2015	Burn2006		
	Hely1999				Pedersen2013			
	Willis2012				Williams-Gray2013			
					Liu2017			
					Domellöf2015			
Incident dementia (after baseline)	Levy2002							
	Willis2012							
Depression	Keener2018	Levy2002		Burn2006	Liu2017	Keener2018		Alves2005
		Williams-Gray2013		Williams-Gray2013		Schrag2016		
						Pigott2015		
						Pedersen2013		
						Domellöf2015		
						Williams-Gray2013		
Education	Keener2018				Levy2000	Keener2018		
					Cereda2016	Schrag2016		
					Liu2017	Pedersen2013		
						Domellöf2015		
Motor subtype	Auyeung2012	Williams-Gray2013	Vu2012	Burn2006	Burn2006	Keener2018		
	Keener2018		Williams-Gray2013			Schrag2016		
						Williams-Gray2013		
						Domellöf2015		
Smell					Schrag2016			
Sleep disorder e.g. RBDSQ					Schrag2016			
					Marion2008			

Table 4.2. Sample sizes (number of PD patients) in each study by outcome. Studies listed for each outcome correspond to Table 4.1.

Outcome	Mortality studies	N	Motor progression studies	N	Cognitive progression studies	N	Disability studies	N
Study name, year, [reference no.]	Auyeung2012 [209]	171	Alves2005 [210]	232	Schrag2016 [202]	390	Alves2005 [210]	232
	Keener2018 [211]	242	Zhao2010 [212]	695	Marion2008 [201]	65	Hely1999 [14]	126
	Marras2005 [203]	800	Vu2012 [213]	795	Willis2012 [208]	138728		
	Hely1999 [14]	130	Burn2006 [214]	35	Levy2000 [215]	173		
	Willis2012 [208]	138728	Williams-Gray2013 [17]	142	Burn2006 [214]	35		
	Levy2002 [216]	180			Cereda2016 [217]	6599		
	Hely2005 [15]	130			Pigott2015 [218]	141		
	Williams-Gray2013 [17]	142			Pedersen2013 [219]	182		
					Williams-Gray2013 [17]	142		
					Domellöf2015 [220]	115		
					Liu2017 (discovery cohort) [221]	1350		

Methods

Cohorts

Data from the Tracking Parkinson’s, Oxford Discovery, PPMI, QSBB, Calypso, and UK Biobank (UKB) cohorts were included for the analysis of mortality. Version 2 (17/06/2020) of the Tracking Parkinson’s clinical dataset was used for this analysis. Only the clinical cohorts (Tracking Parkinson’s, Oxford Discovery, and PPMI) were used for analysis of survival to other clinical milestones: Hoehn and Yahr stage 3 or more, and dementia (MoCA ≤ 21 or withdrawal due to dementia). Across all three clinical studies, patients who received alternative diagnoses during follow up or had neuroimaging results conflicting with a PD diagnosis were excluded from analyses.

Related/ duplicated individuals and ancestry outliers were removed, based on genetic quality control steps (see Chapter 6). In total, clinical data was available for 5,309 patients.

Statistical analysis

Cox proportional hazard models were used to analyse the association between clinical predictors and clinical milestones in each cohort: mortality, Hoehn and Yahr stage 3 or greater, and dementia (defined as MoCA ≤ 21 or withdrawal due to reported dementia or cognitive problems). This cutoff for dementia using the MoCA has been used in previous studies [135,222]. Time was measured from PD symptom onset, or estimated PD diagnosis in the UK Biobank cases. Time to event was taken as the first visit where the outcome was met. Individuals who were missing data at all timepoints for the clinical outcome being assessed were excluded (e.g. if Hoehn and Yahr stage data was missing at all visits for analysis of progression to Hoehn and Yahr stage 3+).

Cox proportional hazard models were conducted in each cohort separately. Random-effects meta-analysis using the inverse variance method was used to pool effect estimates across all cohorts.

To assess pre-study progression, I created a variable for the baseline MDS-UPDRSIII divided by the years of disease duration at baseline. Previous studies have found that this was associated with mortality [203]. I also did the same for the number of incorrect items in the MoCA (reverse scored out of 30).

To assess early stage progression in the first year, I created annual progression scores for the first year of follow-up (first follow-up visit score minus baseline visit score divided by number of years from baseline to first follow-up visit), for both the MDS-UPDRSIII and MoCA (after reverse scoring). Higher scores indicate more rapid progression in the first year.

Bonferroni correction for the number of univariate tests in each clinical milestone was applied (0.05/13 = 0.0038).

Multiple regression models

Multiple regression models were performed in each cohort separately and pooled using random-effects meta-analysis.

Only age at onset and gender were available in all datasets. I first performed multiple regression models for mortality in all datasets with age at onset and gender as predictors.

Secondly, I conducted multiple regression models in the three cohorts with detailed clinical data: Tracking Parkinson's, Oxford Discovery and PPMI. As some variables are strongly correlated and derived from each other (such as pre-study trajectory which is calculated from the baseline MDS-UPDRSIII or MoCA and disease duration), I analysed three models to avoid overadjusting for related variables. The first multiple regression model includes raw baseline variables only (baseline MDS-UPDRSIII and MoCA scores), the second with calculated pre-study trajectory variables (such as baseline MDS-UPDRSIII divided by disease duration), and the third with 1 year progression variables (annual change in the MDS-UPDRSIII and MoCA in the first year). All these symptom score variables aim to estimate the same latent progression, and it may be the case that the 1 year progression variables are stronger/more accurate predictors than the baseline scores alone.

Sensitivity analysis

Though most studies conduct survival/time-to-event analysis using PD onset or diagnosis as the starting timepoint [221], this creates a potential bias as there is a period between PD onset and study entry in which patients cannot meet the outcome as they are not being assessed. More rapidly progressing patients who reach clinical milestones before study entry may be less likely to join clinical cohort studies. This bias may be particularly evident in analysis of Hoehn and Yahr stage and dementia Therefore, as a sensitivity analysis, I analysed time measured from study entry where this data was available in a subset of cohorts to determine if this changed results. This was conducted in the Tracking Parkinson's, Oxford Discovery, PPMI, UK Biobank prevalent cases, and Calypso cohorts, where prevalent PD cases were recruited to a prospective study.

Results

Table 4.3 shows the baseline demographics and the number of patients meeting each outcome in each cohort.

Table 4.3. Demographics at baseline and the number of patients meeting each clinical milestone/outcome in each cohort. Means (SD) are shown unless otherwise indicated.

Demographics	**Tracking Parkinson's**	**Oxford Discovery**	**PPMI**	**QSBB**	**UKB PD incident**[§]	**UKB PD prevalent**	**Calypso WTCCC2**
Number of PD patients overall	1963	985	413	339	1157	914	196
Number of PD patients with mortality data after QC	1779	780	356	285	970	820	180
Male (%)	65.1%	64.2%	65.4%	60.7%	60.8%	62.4%	66.3%
Age at onset, years	64.5 (9.8)	64.5 (9.8)	59.5 (10.0)	61.8 (10.1)	NA	NA	59.8 (10.0)
Age at diagnosis, years	66.3 (9.3)	66.1 (9.6)	61.0 (9.7)	NA	69.5 (5.7)	57.4 (7.2)	61.5 (9.8)
Age at study entry, years	67.6 (9.3)	67.4 (9.6)	61.5 (9.8)	NA	63.9 (5.4)	62.8 (5.5)	67.5 (9.4)
Disease duration at baseline - time from symptom onset to study entry, years	3.2 (3.0)	2.9 (1.9)	2.0 (2.0)	NA	NA	NA	7.7 (5.2)
Time from diagnosis to study entry, years	1.3 (0.9)	1.3 (0.9)	0.5 (0.5)	NA	NA	5.4 (4.8)	5.8 (4.8)
Number of patients died (%)	133 (7.5%)	53 (6.8%)	15 (4.2%)	285 (100%)	370 (38.1%)	294 (35.9%)	121 (67.2%)
Time from PD onset to death, years	6.7 (4.5)	6.6 (2.6)	5.4 (2.6)	15.8 (7.8)	2.7 (2.3)	13.9 (6.1)	15.6 (6.1)
Time from PD onset to censoring/last follow-up in surviving cases, years	7.8 (3.4)	7.4 (2.7)	8.0 (2.5)	NA	5.5 (1.9)	16.1 (4.4)	19.6 (4.5)
Number of patients meeting H&Y≥3[^]	511 (28.8%)	181 (23.2%)	72 (16.8%)	NA	NA	NA	NA
Number of patients meeting dementia criteria[^]	470 (26.7%)	241 (31.3%)	75 (20.2%)	NA	NA	NA	NA

PPMI = Parkinson’s Progression Markers Initiative; QC = Quality Control; QSBB = Queen Square Brain Bank pathologically-confirmed PD cases; UKB = UK Biobank PD cases (including prevalent, incident, and undefined cases).

Percentages are shown of the total number of PD cases in the whole cohort, as the final number included in each analyses varied. Not all patients had all clinical data available (e.g. age at onset, gender, clinical outcomes) and these patients were excluded depending on the outcome of interest and which covariates were included in the model.

^ Shown as a percentage of people with data for at least one timepoint. Individuals who were missing data for the outcome of interest at all timepoints were excluded.

§ Note that this number excludes PD incident cases who were only identified through death records.

Mortality

5,170 individuals had data for mortality/survival. 1,408 (27.2%) died with mean time to death 10.1 years (SD = 7.8 years). 3,901 patients did not die with mean follow-up time 8.7 years (SD 4.7 years). The median time to death was 8.8 years.

Univariate associations are shown in Table 4.4. Older age at onset, male gender, PIGD motor subtype, baseline motor severity, more rapid pre-study motor and cognitive progression, more rapid progression in the MoCA in the first year, and shorter disease duration at study entry, were all associated with greater risk of mortality.

For mortality, the proportional hazard assumption was met in almost models in all cohorts ($p > 0.05$), except for the association with gender in the UKB incident cohort, and disease duration in PPMI.

On visual inspection of the Kaplan-Meier curves for the UKB incident patients, the effect of gender on mortality appeared to change in later disease stages, with men progressing more rapidly than women until approximately 7.5 years from onset. This is a potentially interesting finding, but was not seen in any of the other cohorts. The mean time to death was much shorter in UKB incident patients (2.5 years for men, 3.2 years for women) than in other cohorts (12.6 for men, 14.2 for women). This may relate to the identification of PD cases in the UKB incident cohort from HES, which suggests they are not a truly incident cohort but more rapidly progressing because they are identified from hospital visits and PD onset is likely many years before presentation at secondary care. I therefore restricted analysis of the UKB incident cohort to a maximum 7.5 years from PD onset/diagnosis and used these estimates for meta-analysis. The same restriction was used for the PPMI cohort in analysis of disease duration.

Table 4.4. Pooled effect estimates (from random-effects meta-analysis) for univariate associations between clinical predictors and mortality in univariate Cox proportional hazard analyses in each cohort.The hazard ratios are for a one unit increase in the variable of interest for numeric variables. P-values in bold indicate tests that passed Bonferroni correction for the number of tests for each outcome (0.05/13 = 0.0038).

Variable	**Time from PD onset**	
	HR	**p value**
Age at onset	**1.1**	**2.7×10^{-16}**
Gender – male	**1.6***	**8.4×10^{-10}**
Motor subtype – TD ref		
PIGD	**1.8**	**0.0003**
Indeterminate	1.3	0.3
Baseline severity		
BL HY2+	**1.8**	**0.0001**
MDS-UPDRSIII	**1.04**	**1.3×10^{-14}**
MOCA	0.9	0.12
MDS-UPDRSIII/disease duration at entry	**1.02**	**1.4×10^{-22}**
MOCA/disease duration at entry	**1.09**	**3.7×10^{-11}**
MDS-UPDRSIII annual change in year 1	1.01	0.36
MOCA annual change in year 1	**1.2**	**1.3×10^{-7}**
Disease duration at study entry	**0.8**§	**1.4×10^{-11}**
Education – more than 12 years or higher education	0.7	0.17

*UKB incident cases restricted to 7.5 years of follow-up to meet proportional hazards assumption.

§PPMI cases restricted to 7.5 years of follow-up to meet proportional hazards assumption.

In multiple regression analyses of mortality against age at onset and gender in each cohort, both older age at onset (HR = 1.1) and male gender (HR = 1.6) were significantly associated with more rapid progression to death (pooled effect estimates from meta-analysis, $p < 1.5 \times 10^{-15}$). The Nagelkerke pseudo R^2 ranged between 0.04 to 0.49.

In multiple regression models in the cohorts with detailed clinical data (Tracking Parkinson's, Oxford Discovery and PPMI), age of onset and disease duration at study entry were consistently associated with mortality (Table 4.5). In the baseline and pre-study trajectory models, PIGD subtype, male gender, and higher baseline MDS-UPDRSIII scores or pre-study trajectory in the MDS-UPDRSIII were associated with more rapid progression. However, in the multiple regression model with 1 year progression variables, progression in the MoCA but not MDS-UPDRSIII appeared to be predictive of mortality, and gender was no longer associated. The Variance Inflation Factors (VIFs) were all less than 5, indicating no multicollinearity between the predictor variables.

Table 4.5. Associations between clinical predictors and mortality in multiple regression analyses, in Tracking Parkinson's, Oxford Discovery, and PPMI only.Separate models were performed for baseline variables, pre-study trajectory variables, and 1 year progression variables. The hazard ratios (HRs) and p-values from random effects meta-analysis are reported.

Baseline variables only		
Variable	**HR**	**p value**
Age at onset	**1.1**	**5.5 x 10^{-25}**
Gender – male	**1.6**	**0.02**
Motor subtype – TD ref		
PIGD	**1.5**	**0.03**
Indeterminate	1.1	0.63
Baseline HY2+	0.97	0.88
Baseline MDS-UPDRSIII	**1.03**	**0.0003**
Baseline MOCA	1.02	0.78
Disease duration at study entry	**0.7**	**9.3 x 10^{-7}**
Education – more than 12 years or higher education	0.97	0.90
Pre-study trajectory variables		
Age at onset	**1.2**	**5.5 x 10^{-33}**
Gender – male	**1.7**	**0.01**
Motor subtype – TD ref		
PIGD	**1.4**	**0.03**
Indeterminate	1.3	0.37
Baseline HY2+	0.97	0.93
Education – more than 12 years or higher education	0.93	0.77
MDS-UPDRSIII/disease duration at entry	**1.02**	**0.002**
MOCA/disease duration at entry	0.97	0.49
1 year progression variables		
Age at onset	**1.1**	**3.6 x 10^{-17}**
Gender – male	1.3	0.48
Motor subtype – TD ref		
PIGD	1.5	0.07
Indeterminate	1.2	0.69
Baseline HY2+	1.3	0.25
Disease duration at study entry	**0.7**	**0.003**
Education – more than 12 years or higher education	0.9	0.75
MDS-UPDRSIII annual change in year 1	1.0	0.96
MOCA annual change in year 1	**1.2**	**0.007**

HY3+

2,912 individuals had data available for Hoehn and Yahr stage. 764 individuals met the outcome of Hoehn and Yahr stage 3 or greater, with mean time to event 5.7 years (SD = 3.1 years). 2,148 individuals did not meet the outcome with mean follow-up 7.5 years (SD = 3.2 years). The median time to Hoehn and Yahr stage 3 or greater was 5.3 years.

For progression to Hoehn and Yahr stage 3 or greater, there were multiple predictors across cohorts in which the proportional hazards assumption was not met, except for age at onset. I therefore stratified the time interval to 0 to 5 years, 5 to 10 years, and more than 10 years from PD onset, based on visual inspection of Kaplan-Meier curves.

Older age at onset, PIGD motor subtype, greater baseline motor and cognitive severity, progression prior to study entry in the MDS-UPDRSIII and MoCA, pre-study trajectory in the MDS-UPDRSIII, first year progression in the MDS-UPDRSIII, and disease duration at study entry, were all associated with more rapid progression to Hoehn and Yahr stage 3 or greater in univariate analysis (Table 4.6). However, many of these variables had slightly different effects at different time intervals, though mostly consistent in direction. Interestingly, disease duration at study entry appeared to have a different direction of effect according to time interval – associated with more rapid progression up to 10 years, but then with a protective effect after 10 years from PD onset. This is likely due to patients with long disease duration at study entry, who cannot be observed to meet the study outcome between onset and entry into the study, and are also likely to be slower progressing.

Table 4.6. Pooled effect estimates (from random-effects meta-analysis) for univariate associations between clinical predictors and progression to Hoehn and Yahr stage 3 or greater.This was stratified into intervals as the proportional hazards assumption was not satisfied for most variables in the full time period: ≤5 years, >5 and ≤10 years, and >10 years from PD onset. The hazard ratios are for a one unit increase in the variable of interest for numeric variables. P-values in bold indicate tests that passed Bonferroni correction for the number of tests for each outcome (0.05/13 = 0.0038).

Variable	**0 to 5 years**		**5 to 10 years**		**> 10 years**	
	HR	**p value**	**HR**	**p value**	**HR**	**p value**
Age at onset	1.03	0.14	**1.06**	**1.9 x 10^{-11}**	**1.08**	**1.7 x 10^{-5}**
Gender – male	0.9	0.41	0.96	0.76	0.8	0.49
Motor subtype – TD ref						
PIGD	**3.1**	**3.9 x 10^{-16}**	**2.6**	**7.7 x 10^{-15}**	**2.9**	**0.002**
Indeterminate	1.5	0.09	**1.7**	**0.002**	2.6	0.03
Baseline severity						
BL HY2+	2.1	0.17	**1.7**	**0.001**	0.8	0.56
MDS-UPDRSIII	**1.04**	**1.6 x 10^{-7}**	**1.03**	**0.0002**	1.01	0.27
MOCA	0.97	0.17	**0.91**	**0.003**	0.96	0.40
MDS-UPDRSIII/disease duration at entry	1.01	0.01	1.01	0.22	**1.3**	**1.5 x 10^{-5}**
MOCA/disease duration at entry	1.00	0.96	1.05	0.17	1.6	0.05
MDS-UPDRSIII annual change in year 1	1.01	0.10	**1.03**	**0.0001**	0.99	0.73
MOCA annual change in year 1	1.04	0.62	1.1	0.005	0.99	0.99
Disease duration at study entry	1.1	0.25	**1.1**	**0.0001**	**0.8**	**0.0003**
Education – more than 12 years or higher education	**0.7**	**0.0005**	0.7	0.01	0.5	0.05

In multiple regression models, also stratified by time period, age at onset, gender, PIGD subtype, and baseline MDS-UPDRSIII were consistently associated with progression to Hoehn and Yahr stage 3 or greater (Table 4.7).

Table 4.7. Associations between clinical predictors and progression to Hoehn and Yahr stage 3 or greater in multiple regression analyses, in Tracking Parkinson's, Oxford Discovery, and PPMI only.Separate models were performed for baseline variables, pre-study trajectory variables, and 1 year progression variables. The hazard ratios (HRs) and p-values from random effects meta-analysis are reported in stratified time intervals from PD onset, as the proportional hazards assumption was consistently not satisfied in models with the full time period. To correct for multiple testing in the three models in three different time periods, only p-values < 0.0056(0.05/9) were considered significant.

Baseline variables only						
Variable	**0 to 5 years**		**5 to 10 years**		**> 10 years**	
	HR	**p value**	**HR**	**p value**	**HR**	**p value**
Age at onset	1.02	0.23	**1.06**	**2.4×10^{-13}**	1.07	0.01
Gender – male	0.9	0.36	0.8	0.17	0.8	0.65
Motor subtype – TD ref						
PIGD	**2.9**	**6.3×10^{-12}**	**2.5**	**1.4×10^{-12}**	**3.8**	**0.001**
IND	1.4	0.14	1.6	0.02	3.5	0.02
Baseline HY2+	1.6	0.01	1.02	0.93	0.7	0.5
Baseline MDS-UPDRSIII	**1.03**	**4.4×10^{-9}**	**1.03**	**0.0002**	1.01	0.44
Baseline MOCA	0.99	0.83	0.97	0.37	1.02	0.76
Disease duration at study entry	**0.8**	**0.004**	1.09	0.04	0.8	0.01
Education – more than 12 years or higher education	0.9	0.53	0.8	0.12	0.5	0.12
Pre-study trajectory variables						
	HR	**p value**	**HR**	**p value**	**HR**	**p value**
Age at onset	1.03	0.15	**1.06**	**8.2×10^{-14}**	1.1	0.03
Gender – male	0.9	0.44	0.9	0.52	0.7	0.41
Motor subtype – TD ref						
PIGD	**2.9**	**3.4×10^{-12}**	**2.5**	**1.6×10^{-12}**	**3.6**	**0.002**
IND	1.4	0.18	1.7	0.007	3.4	0.1
Baseline HY2+	1.8	0.13	1.4	0.13	0.7	0.43
Education – more than 12 years or higher education	0.9	0.44	0.8	0.08	0.5	0.13
MDS-UPDRSIII/disease duration at entry	**1.02**	**1.3×10^{-5}**	1.0	0.66	1.2	0.01
MOCA/disease duration at entry	0.95	0.17	1.0	0.97	0.7	0.36

Table 4.7 (cont).

1 year progression variables						
	HR	**p value**	**HR**	**p value**	**HR**	**p value**
Age at onset	1.02	0.10	**1.07**	**1.2 x 10^{-8}**	1.06	0.05
Gender – male	0.95	0.76	0.9	0.46	0.7	0.49
Motor subtype – TD ref						
PIGD	2.5	0.006	**2.6**	**1.2 x 10^{-10}**	**4.7**	**0.003**
IND	1.5	0.14	1.6	0.02	3.6	0.06
Baseline HY2+	2.0	0.05	1.6	0.02	0.9	0.91
Disease duration at study entry	1.01	0.87	1.1	0.01	0.9	0.14
Education – more than 12 years or higher education	0.7	0.06	0.9	0.29	0.3	0.03
MDS-UPDRSIII annual change in year 1	1.03	0.02	**1.03**	**0.0009**	0.99	0.88
MOCA annual change in year 1	1.03	0.72	1.08	0.07	0.9	0.55

Dementia

2,887 individuals had data for dementia. 783 individuals met the dementia outcome with mean time to dementia 4.8 years (SD = 3.1 years). 2,104 individuals did not meet the outcome of dementia, with mean follow-up 7.7 years (SD = 3.2 years). The median time to dementia was 4.2 years.

All clinical variables tested, with the exception of 1 year progression in the MDS-UPDRSIII, were associated with progression to dementia (Table 4.8). Some models did not meet the proportional hazards assumption in one or more cohorts, so these were analysed in stratified time intervals: 0 to 5 years, 5 to 10 years, and more than 10 years from PD onset (Table 4.8).

The proportional hazard assumption appeared to be not met most frequently in baseline symptom scale variables and disease duration at study entry, and there were still some models that did not meet the assumption after stratification by time. It may be that time from study entry, rather than PD onset, may be more appropriate to analyse in these cases, due to the wide range of disease durations at study entry and potential bias where some patients with recent onset PD are recruited into the study and assessed, whereas other patients with long duration PD have a long period where they are not observed to meet the outcome. It may also be helpful to analyse time from PD diagnosis, although this is highly correlated with age at onset.

Table 4.8. Pooled effect estimates (from random-effects meta-analysis) for univariate associations between clinical predictors and dementia in univariate Cox proportional hazard analyses in each cohort.The hazard ratios are for a one unit increase in the variable of interest for numeric variables. P-values in bold indicate tests that passed Bonferroni correction for the number of tests for each outcome (0.05/13 = 0.0038).

Variable	**Full time period**		**0 to 5 years**		**5 to 10 years**		**>10 years**	
	HR	**p value**	**HR**	**p value**	**HR**	**p value**	**HR**	**p value**
Age at onset	**1.09**	**4.5 x 10^{-33}**						
Gender – male	**1.6**	**2.5 x 10^{-6}**						
Motor subtype – TD ref								
PIGD	**1.7**	**0.0002**						
IND	**1.4**	**0.002**						
Baseline HY2+*	**1.6**	**0.0002**	1.3	0.13	1.9	0.02	1.7	0.13
Baseline MDS-UPDRSIII	**1.03**	**7.7 x 10^{-23}**						
Baseline MoCA*	**0.7**	**2.0 x 10^{-35}**	0.8	3.8 x 10^{-75}	0.7	1.1 x 10^{-54}	0.7	0.02
MDS-UPDRSIII/ disease duration at entry*	**1.03**	**5.0 x 10^{-5}**	1.01	0.04	1.01	0.07	1.1	0.29
MoCA/disease duration at entry*	**1.2**	**1.3 x 10^{-5}**	1.1	0.005	1.2	6.4 x 10^{-6}	6.5	3.5 x 10^{-7}
MDS-UPDRSIII annual change in year 1*	1.02	0.22	0.99	0.85	1.01	0.13	1.06	0.04
MoCA annual change in year 1*	**1.2**	**5.7 x 10^{-5}**	1.02	0.47	1.3	7.9 x 10^{-5}	0.95	0.71
Disease duration at study entry*	**0.8**	**1.3 x 10^{-19}**	0.9	0.12	1.3	1.2 x 10^{-11}	0.9	0.005
Education – more than 12 years or higher education	**0.5**	**1.2 x 10^{-12}**						

* Did not meet proportional hazard assumption in at least one cohort, so analysis was stratified into different time intervals.

TD = Tremor Dominant subtype; PIGD = Postural Instability Gait Disorder, IND = Indeterminate or mixed subtype; MoCA = Montreal Cognitive Assessment

In multiple regression models, age at onset, disease duration, and baseline MoCA were consistently associated with progression to dementia (Table 4.9).

Table 4.9. Associations between clinical predictors and progression to dementia in multiple regression analyses, in Tracking Parkinson's, Oxford Discovery, and PPMI only.Separate models were performed for baseline variables, pre-study trajectory variables, and 1 year progression variables. The hazard ratios (HRs) and p-values from random effects meta-analysis are reported in stratified time intervals from PD onset, as the proportional hazards assumption was consistently not satisfied in models with the full time period. To correct for multiple testing in the three models in three different time periods, only p-values < 0.0056 (0.05/9) were considered significant.

Baseline variables only						
Variable	**0 to 5 years**		**5 to 10 years**		**> 10 years**	
	HR	**p value**	**HR**	**p value**	**HR**	**p value**
Age at onset	**1.03**	**9.7 x 10^{-5}**	**1.05**	**3.2 x 10^{-8}**	1.07	0.10
Gender – male	1.1	0.40	1.9	0.03	1.2	0.83
Motor subtype – TD ref						
PIGD	0.95	0.82	1.3	0.16	**3.6**	**0.007**
IND	1.1	0.67	1.08	0.72	2.0	0.27
Baseline HY2+	0.99	0.97	1.09	0.80	11.9	0.07
Baseline MDS-UPDRSIII	0.99	0.09	**1.02**	**0.004**	0.9	0.01
Baseline MOCA	**0.7**	**2.5 x 10^{-18}**	**0.7**	**6.6 x 10^{-27}**	**0.5**	**0.001**
Disease duration at study entry	**0.5**	**4.8 x 10^{-7}**	1.1	0.27	**0.8**	**0.008**
Education – more than 12 years or higher education	0.9	0.7	**0.6**	**0.002**	1.5	0.40
Pre-study trajectory variables						
	HR	**p value**	**HR**	**p value**	**HR**	**p value**
Age at onset	**1.04**	**5.5 x 10^{-7}**	**1.05**	**5.2 x 10^{-9}**	1.09	0.02
Gender – male	1.06	0.64	**2.1**	**0.003**	1.3	0.57
Motor subtype – TD ref						
PIGD	0.9	0.63	1.5	0.04	**3.8**	**0.004**
IND	1.3	0.08	1.2	0.33	2.4	0.14
Baseline HY2+	1.4	0.009	2.2	0.08	1.9	0.24
Education – more than 12 years or higher education	**0.7**	**0.002**	0.7	0.008	1.4	0.48
MDS-UPDRSIII/disease duration at entry	0.97	0.20	0.97	0.02	0.87	0.34
MOCA/disease duration at entry	1.2	0.03	1.3	0.007	**9.7**	**8.6 x 10^{-5}**

Table 4.9 (cont).

1 year progression variables						
	HR	**p value**	**HR**	**p value**	**HR**	**p value**
Age at onset	**1.04**	**0.002**	**1.08**	**1.2 x 10^{-23}**	1.2	0.006
Gender – male	1.2	0.53	**1.7**	**5.2 x 10^{-5}**	1.9	0.05
Motor subtype – TD ref						
PIGD	0.97	0.90	1.4	0.02	**2.9**	**0.0003**
IND	1.07	0.82	1.4	0.11	14.6	0.09
Baseline HY2+	1.95	0.16	1.2	0.32	0.7	0.72
Disease duration at study entry	0.9	0.19	0.9	0.06	0.9	0.41
Education – more than 12 years or higher education	0.6	0.09	0.7	0.12	**0.4**	**0.005**
MDS-UPDRSIII annual change in year 1	0.99	0.34	1.02	0.39	1.02	0.38
MOCA annual change in year 1	1.08	0.19	**1.2**	**0.0002**	1.3	0.28

Sensitivity analysis – time from study entry

When analysing progression using time from study entry, rather than time from PD onset, the results were very similar. There was a strong correlation between age at PD onset and age at study entry ($r = 0.93$, $p < 2.2 \times 10^{-16}$) in the subset of prevalent PD patients in prospective cohorts.

In univariate analyses, the same predictors were significantly associated with progression mortality, Hoehn and Yahr stage 3 or greater, and dementia. The effect sizes were consistent in direction and size to the main analysis, with the exception of disease duration at study entry. When analysing time from study entry, longer disease duration at study entry was associated with more rapid progression to mortality (HR = 1.04, p = 1.4 x 10^{-6}), but not significant for Hoehn and Yahr stage 3+ (HR = 1.04, p = 0.06), and dementia (HR = 1.003, p = 0.86).

In multiple regression models for mortality, using time from study entry, the same predictors were significant and with effects approximately the same magnitude as in the analysis using time from PD onset. The only change was that pre-study trajectory in the MDS-UPDRSIII was no longer significantly associated with mortality.

In multiple regression models for Hoehn and Yahr 3 and dementia, using time from study entry, there were some differences in the significance of predictors though the direction of effects were generally the same.

Discussion

In this study, I examined the clinical predictors of progression measured in time to clinical milestones: mortality, Hoehn and Yahr stage 3 or greater, and dementia.

To summarise the key findings:

1. Mortality: Age at onset, gender, motor subtype, disease duration at study entry, baseline MDS-UPDRSIII (including pre-study trajectory), and 1 year progression in the MoCA were associated with mortality in multiple regression analyses.

2. Hoehn and Yahr stage 3+: the Cox proportional hazards assumption was not satisfied for most predictors, except age at onset. In multiple regression analysis, older age at onset, PIGD subtype, higher baseline scores and pre-study trajectory in the MDS-UPDRSIII, and disease duration at study entry were associated with more rapid progression, though some were not consistently associated at different disease stages.

3. Dementia: older age at onset, PIGD subtype, baseline MoCA and pre-study trajectory in the MoCA, and disease duration at study entry were consistently associated with more rapid progression in multiple regression models. There was some evidence that PIGD subtype and more rapid 1 year progression in the MoCA were also associated with dementia, but these were not consistent at different stages.

Age at onset

I showed that age at onset is a strong predictor of all clinical milestones. This association has been robustly reported in many previous studies and systematic reviews [204–206,223], although the review by Marras et al. suggest that older age at onset was only robustly associated with disability and not motor progression [204].

Only a handful of studies have not confirmed this association between age at onset and disease progression [224–226]. This may be because age at onset was stratified into two groups and these were not sufficient to capture the variation in progression.

However, when looking at prediction of mortality, it is important to consider the age-standardised mortality rate as older patients have an increased all-cause mortality. This study did not include comparison to healthy controls, but previous studies have shown that the association between age of onset and mortality remains even after comparing with age-matched controls [227,228]. Even so, this analysis is complex because healthy controls surviving to very old ages may not be fully representative of the general population.

My results replicate many previous studies, including population-based studies and community cohorts, showing that older age of onset is associated with greater risk of mortality [203,227–229]. However, when considering age-specific life expectancy, patients with younger onset have a greater reduction in life expectancy than patients with older onset [230].

Gender

I found that male gender was associated with progression to mortality. The majority of previous studies of mortality and survival in PD show that men have increased risk of mortality than women [208], but when comparing this to the standardised mortality ratios of men and women, there does not appear to be a sex difference in PD mortality relative to the general population [203,227–229].

Interestingly, male gender was associated with more rapid progression to death, but not Hoehn and Yahr stage 3 in either univariate or multiple regression analysis. Other studies have generated conflicting results on the association between sex and motor progression. In the Sydney Multicentre Study, Hely et al. found that after 10 years, women had higher Hoehn and Yahr scores than men, although there was no difference in the rate of increase in other disability scales [14]. In the same cohort of patients at 5 years of follow-up, there was no difference between men and women in development of balance disorder or progression in the Columbia scale [223]. Some studies have reported no association between sex and motor progression

[212,224,225,231], while only a few suggest that men have more rapid motor progression [232].

Male gender was associated with more rapid progression to dementia in univariate analysis, but not in multiple regression models. Several previous studies suggest that men have increased risk of dementia in PD [217,218,233]. A recent study in the PPMI cohort found that men had lower MoCA scores at 2 year follow-up (indicating worse cognitive performance) in univariate analysis, but this was not significant in multiple regression models after backwards stepwise elimination [202]. There was also no association between gender and classification of cognitive impairment [202].

Overall, this evidence from this and other studies suggests that gender is not an important predictor for progression to Hoehn and Yahr stage 3 or greater or dementia, after accounting for other baseline demographic and symptom severity scales. It is possible that men have more severe symptoms at study entry but the rate of progression is the same.

Motor subtype

I found that baseline motor subtype was a strong predictor of all clinical milestones, with the PIGD subtype progressing more rapidly than the TD subtype. This replicates the results of previous studies showing that the PIGD subtype is associated with worse prognosis, including mortality [209,211,234], and dementia [213,214], while the TD subtype is more benign.

One limitation is that subtype classification can change over time, typically from TD to PIGD as disease progresses [213,235]. This may add noise to the classification of motor subtype in the Tracking Parkinson’s and Oxford Discovery cohorts, where patients were recruited slightly later in disease stage, so the classification at study entry may not reflect presentation at symptom onset.

Baseline severity and pre-study trajectory

Baseline severity and estimated pre-study trajectories (baseline score divided by disease duration) in the MDS-UPDRSIII and MoCA were associated with progression to death, Hoehn and Yahr stage 3, and dementia. A few previous studies have shown that estimated pre-study trajectory was associated with mortality [14,203,236], and a number of studies indicate that baseline symptom severity is associated with later outcomes. A previous progression GWAS included classification of Hoehn and Yahr stage 2 or greater as a covariate [98], but I found this was not associated with progression after accounting baseline MDS-UPDRSIII and MoCA scores. However, not all studies will have detailed clinical assessments available.

There are a number of limitations to consider with baseline severity and estimated pre-study trajectories. Firstly, studies have used different scales to determine baseline severity or pre-study progression rate, and this may explain some of the differences between results. Secondly, in this analysis, Tracking Parkinson's and Oxford Discovery were not incident cohorts so there was a wide range of disease durations at study entry and the calculation of these scores may be skewed. In addition, there may be a ceiling effect in the MoCA as this is only a brief screening instrument and may not be as sensitive as detailed neuropsychological assessments [237]. Finally, as Marras highlighted, there may be recall bias such that patients with particular initial symptoms may recall onset more accurately than patients with more subtle initial symptoms, and this would affect disease duration estimates [203].

1 year progression

I found some evidence that progression in the first year was predictive of later progression. Progression after 1 year in the MoCA, but not baseline MoCA scores, were associated with mortality. This result could be due to two possible explanations. Firstly, it could be that there is less measurement error in the MoCA in the 1 year progression variable as this is more accurate at estimating latent progression than the baseline MoCA score. Secondly, 1 year progression may be a stronger predictor because the rate of decline is non-linear.

In contrast, baseline performance in the MDS-UPDRSIII but not 1 year progression was associated with mortality. I hypothesise that this is due to a stronger ceiling effect in the MoCA at baseline (scored out of 30), whereas the motor MDS-UPDRSIII is on a larger scale. In addition, patients are diagnosed by motor symptoms so by the point of study entry there is more likely to be observed motor impairment than cognitive impairment. In this case, the baseline MDS-UPDRSIII may be more accurate at estimating latent progression.

Other studies have shown that 1 year progression [238], as well as early cognitive impairment and incident dementia is a predictor of progression [214,216]. The results of the current study suggest that 1 year progression is associated with later progression, and further analysis using predictive modelling methods is needed.

Disease duration

I found that disease duration at study entry was predictive of progression to all milestones. However this may be biased, as disease onset was used as the starting time point in survival analysis so patients with longer disease duration have a longer period where they cannot be observed to meet the outcome as they are not actively assessed in a study. In addition or alternatively, there may be a bias whereby patients with longer disease duration at study entry are likely to be more slower progressing than typical patients with the same disease duration, who may be too disabled or unwilling to take part in intensive research studies. Results from a sensitivity analysis using time from study entry suggests the first scenario is likely to be the case. These results showed either no effect or the opposite effect for disease duration in univariate models. However, the results for other predictors in multivariable models were generally similar.

There is mixed evidence from other studies of the association between baseline disease duration and mortality [14,211], motor progression [210], cognitive progression [211], and other markers of disability [14,210]. Further studies of incident PD patients in population-based settings is needed to fully answer this question.

Limitations

One limitation of this study is that detailed clinical data was not available in the majority of cohorts in the analysis of mortality. The prospective cohorts that had systematic data available followed patients earlier in disease stage, and therefore included a smaller proportion of patients who died. This points to an important trade-off in clinical cohorts, where the more covariates/predictors that are assessed and included in statistical models, the fewer cases will have full clinical data and can be included in analyses, especially for more intensive clinical scales.

A second limitation is that the proportional hazard assumption does not appear to hold true for many predictors in progression to Hoehn and Yahr stage 3 and dementia. To overcome this, I analysed time from PD onset in three intervals, however even in the stratified analysis there were a few variables which did not meet the proportional hazards assumption. This indicates that there may be some interaction between the clinical predictors and time. Other models for these predictors and outcomes may be needed, such as the accelerated failure time model.

When reviewing previous literature, it becomes clear that comparisons between studies are very difficult. Firstly, studies use different measures of motor and cognitive progression – some use change in different scales, whereas others use milestones or cut-off points in clinical scales. While various scales within the same domain (e.g. motor scales) are likely to be correlated, they measure different symptoms and aspects of progression. Furthermore, each study includes different predictors/covariates in multiple regression models, and this can dramatically affect the results of the model. Studies also use different methods for models – some use multiple regression while others use stepwise models, and each with slightly different criteria for selection. This can also dramatically influence which predictors are evaluated, and consequently the results of the model. Finally, due to these issues listed, it can be difficult to interpret whether a clinical predictor is important for progression if it is significant in univariate but not multiple regression or stepwise models. One would suggest that greater standardisation is needed between studies, however this issue is complex as there is no gold standard for measuring or assessing progression, so the nature of this research is exploratory.

In summary, I found robust evidence that age at onset, baseline severity, and disease duration at study entry were associated with progression. There was moderate evidence that gender is associated with progression but not in multiple regression analysis. Some of these variables, such as baseline severity, may be on the causal pathway (intermediary markers) between genotype and progression. In the GWAS chapters, I included age at onset and gender as covariates.

Chapter 5 : Genome-wide association studies using a principal components analysis approach

Introduction

Progression in PD is heterogeneous, with some patients progressing rapidly while others remain relatively stable over time [17]. There is a clear need to identify genetic variants that affect symptom progression in PD.

GWASs in PD have identified 90 independent loci associated with disease risk [75]. However, the majority of PD GWASs have compared cases to healthy controls to identify variants linked to disease status. In order to identify variants that are associated with disease progression, it is necessary to compare phenotypes within patients.

Progression of clinical signs in PD can be measured in different ways [239] and there is no gold standard measure of progression, although the MDS-UPDRS Part III and Part II are commonly used in clinical trials. Individual scales, including the MDS-UPDRS, are affected by measurement error particularly for change over time [240], including rater subjectivity and practice effects in cognitive assessments. Therefore, combining multiple measures may improve the accuracy of measuring progression [202,241].

Principal Components Analysis (PCA) is commonly used in clinical studies to combine multiple scales and identify latent components that explain the most variability in the data. PCA was used successfully in a GWAS of Huntington's disease (HD) progression to combine multiple motor, cognitive, and brain atrophy variables, to create a composite progression score [97].

The aim of this approach is to improve the phenotypic measure of progression and increase power in genetic studies. This is particularly useful as our sample sizes for progression studies is still relatively small for GWASs.

In this study, I analysed data from three large-scale, prospective, longitudinal studies: Tracking Parkinson's, Oxford Parkinson's Disease Centre Discovery (Oxford Discovery), and PPMI. Following a similar approach in HD, I combined multiple measures of motor and cognitive progression using PCA to create progression scores. These scores were analysed in GWASs to identify variants associated with composite (cross-domain), motor, and cognitive progression in PD.

Methods

Cohorts

For this study, I analysed genome-wide SNP array data from Tracking Parkinson's and Oxford Discovery, and whole genome sequencing data from PPMI.

Clinical data from PPMI was downloaded on 14/08/2019. Only data from the annual visits was analysed from PPMI, as the motor assessments were performed in the technically defined 'off' state at these visits.

All studies used the same Queen Square Brain Bank diagnostic criteria for PD. Patients who received alternative diagnoses during follow-up or had neuroimaging results conflicting with a PD diagnosis were excluded from analyses.

Genotyping

Genotyping arrays are described in Chapter 2 (Methods). Standard quality control procedures were conducted in PLINK v1.9. I conducted quality control steps and imputation on the Tracking Parkinson's data, while the Oxford Discovery data was filtered and imputed by Dr Stephanie Millin (University of Oxford). The cohorts were genotyped and filtered separately, but following the same quality control steps and filters. Individuals with low overall genotyping rates (< 98%), related individuals (Identity-By-Descent PIHAT > 0.1) and heterozygosity outliers (> 2 SDs away from the mean) were removed, as were individuals whose clinically reported biological sex did not match the genetically determined sex.

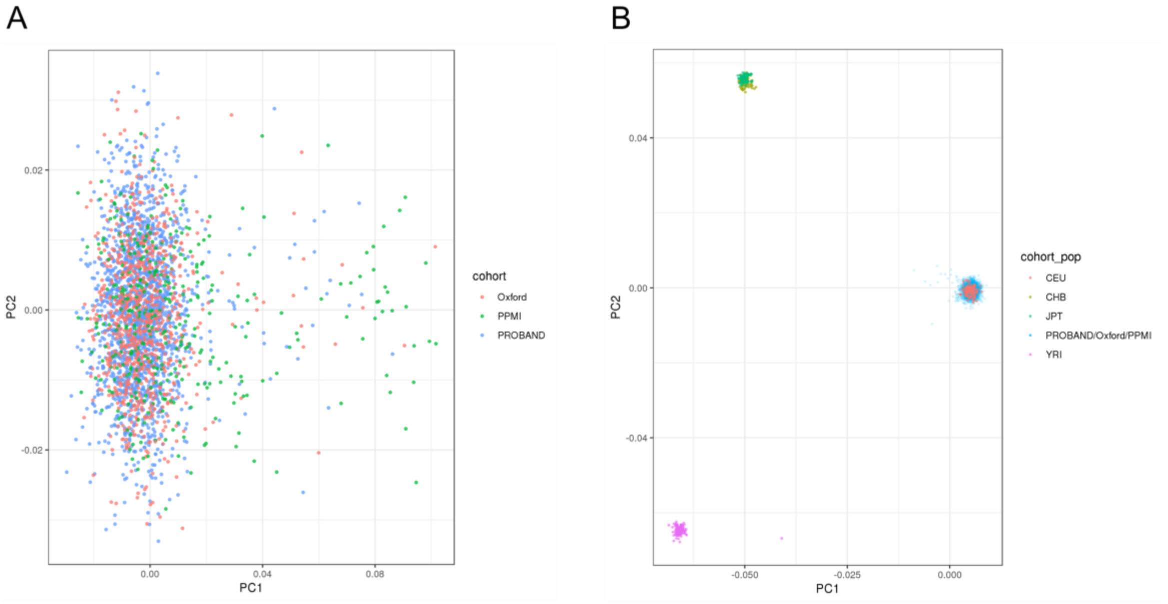
PCA was conducted on a linkage disequilibrium (LD) pruned set of variants (removing SNPs with an $r^2 > 0.05$ in a 50kb sliding window shifting 5 SNPs at a time) after merging with European samples from the HapMap reference panel. Individuals who were > 6 SDs away from the mean of any of the first 10 principal components were removed.

Variants were removed if they had a low genotyping rate (< 99%), Hardy-Weinberg Equilibrium p-value < 1 x 10^{-5}, and minor allele frequency < 1%.

Following quality control, genotypes for Tracking Parkinson's and Oxford Discovery were imputed separately to the 1,000 Genomes Project reference panel (phase 3 release 5)[136] using the Michigan Imputation Server (https://imputationserver.sph.umich.edu). I performed post-imputation filtering on both the Tracking Parkinson's and Oxford Discovery datasets. Only variants with imputation quality >0.8 were retained, to keep only high quality calls to merge across the cohorts. Tracking Parkinson’s and Oxford Discovery data was lifted over to genome build hg38 using liftOver (https://genome.ucsc.edu/cgi-bin/hgLiftOver) to merge with PPMI. PPMI data was not imputed as this was whole-genome sequencing data.

The three datasets were then merged, with only shared variants retained. Twenty genetic principal components were generated from a linkage-pruned SNP set (removing SNPs with an r^2 > 0.02 in a 1000kb sliding window shifting 10 SNPs at a time). The first 2 components were plotted to check that there were no differences between the cohorts. I removed extreme outliers from the first 5 principal components (> 6 SDs away from the mean). The genetic principal components were then recalculated after removing outliers, as extreme outliers can substantially affect the calculation of genetic principal components. These first 5 new principal components were included as covariates in the GWASs to adjust for population substructure. Additional outliers who were > 6 SDs away from the mean of any of the first 5 principal components were excluded. The plot of the first 2 principal components after removing outliers is shown in Figure 5.1A, and when merged with the HapMap reference samples (Figure 5.1B).

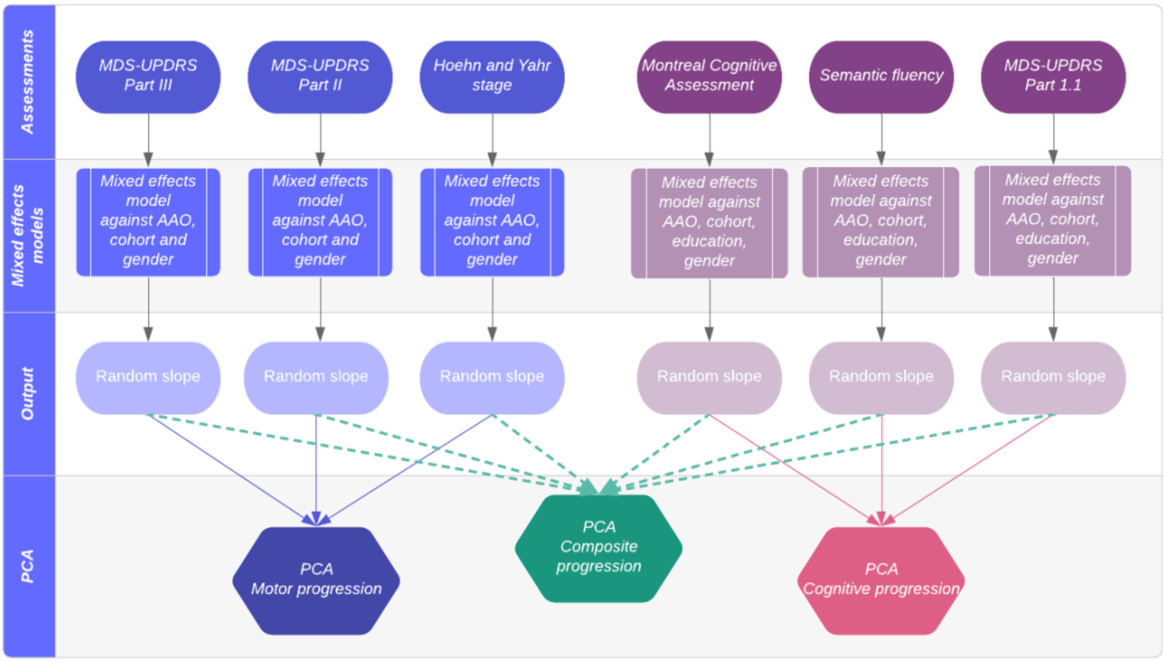
Figure 5.1. A) First two genetic principal components plotted for each cohort, from the final genetic PCA (after removing outliers in two passes). B) First two principal components, generated from a PCA merged with HapMap data.



Clinical outcome measures

Individual-level data from the three cohorts was merged. In order to increase power and the accuracy of the final progression scores, I performed all transformations and created progression scores from the merged dataset as follows (Figure 5.2).

Figure 5.2. Steps to create composite, motor, and cognitive progression scores.


Assessments
Mixed effects models
Output
PCA
MDS-UPDRS Part III
MDS-UPDRS Part II
Hoehn and Yahr stage
Montreal Cognitive Assessment
Semantic fluency
MDS-UPDRS Part 1.1
Mixed effects model against AAO, cohort and gender
Mixed effects model against AAO, cohort and gender
Mixed effects model against AAO, cohort and gender
Mixed effects model against AAO, cohort, education, gender
Mixed effects model against AAO, cohort, education, gender
Mixed effects model against AAO, cohort, education, gender
Random slope
Random slope
Random slope
Random slope
Random slope
Random slope
PCA Motor progression
PCA Composite progression
PCA Cognitive progression

AAO = Age at onset; MDS-UPDRS = Movement Disorders Society Unified Parkinson’s Disease Rating Scale; PCA = Principal Components Analysis.

The motor and cognitive measures were chosen prior to the analysis. Only assessments conducted in all cohorts were included. I selected measures shown to rate motor and cognitive function semi-objectively, in an attempt to minimise observer bias. I did not include scales which may have been affected by a combination of motor, cognitive, and other non-motor symptoms, such as the Schwab and England Activities of Daily Living Scale.

Motor progression was assessed using the MDS-UPDRS Part III (clinician-assessed movement examination), MDS-UPDRS Part II (patient-reported motor experiences of daily living), and Hoehn and Yahr stage (clinician-assessed rating of impairment and disability). In PPMI, I used the motor assessments conducted in the 'off' medication state, as these patients were treatment-naive at study entry.

Cognitive progression was assessed using the Montreal Cognitive Assessment (MoCA), semantic fluency, and item 1.1 of the MDS-UPDRS (cognitive impairment based on patient and/or caregiver report).

To ensure the different measures were comparable, I first converted raw scores into a percentage of the maximum score for that scale, with higher scores indicating worse symptoms. The MoCA was reverse scored to count the number of incorrect items out of the maximum score of 30. Semantic fluency was reverse scored out of the highest individual score at baseline in each cohort.

Each measure was then standardised to the population baseline mean and standard deviation within each cohort, to ensure that measures are on the same scale and to adjust for any differences in the scales or task instructions between cohorts. By standardising to the baseline mean and SD (rather than the mean and SD of each visit), I also preserved data on longitudinal change which is important for analysis of progression.

Analysis of progression scores

I derived severity scores from mixed effects regression models using follow-up data up to 72 months. Each variable was regressed on age at onset, sex, cohort, and their interactions with time from disease onset. For the cognitive measures, I also included the following education variables as covariates: the number of years of education before higher education, and whether higher education was undertaken (yes/no). This was the format that education data was collected in Tracking Parkinson's and Oxford Discovery. In each model, I included terms for subject random effects to account for individual heterogeneity in the intercept (baseline values) and slope (rate of progression).

The random effect slope values were used as the measure of 'residual' progression not predicted by age at onset, cohort, gender, and education, for each individual. I performed PCA on these values, with the input variables zero centred and scaled to have unit variance.

Removal of non-PD cases

Any patients that were diagnosed with a different condition during follow-up were removed from analyses. I also conducted sensitivity analyses to remove any cases which may have non-PD conditions but an alternative diagnosis had not yet been confirmed. Firstly, I repeated analyses after removing patients in Tracking Parkinson's and Oxford Discovery who had a clinician-rated diagnostic certainty of PD of less than 90%. This cutoff has been used in previous studies of these cohorts [20,242]. Secondly, I removed the fastest and slowest progressors in the top and bottom 5% of the distribution, to address the possibility of confounding by misdiagnosis with more benign (e.g. essential tremor) or more malignant (e.g. MSA, PSP) conditions.

GWAS

Statistical analysis was conducted in R v3.4.1 (https://www.r-project.org/)[243]. Clinical data from Tracking Parkinson's and Oxford Discovery was managed and cleaned by Dr Michael Lawton and Ms Sofia Kanavou (Bristol University) using STATA (version 15.1, StataCorp, Texas, USA).

For each GWAS, I included the following covariates: cohort (to adjust for any differences in genotyping data and measurement error) and the first 5 genetic principal components generated from the merged genotype dataset (to adjust for population substructure). I conducted all GWASs in rvtests [119] using the single variant Wald test. Genome-wide Complex Trait Analysis conditional and joint analysis (GCTA-COJO) was used to identify independent signals [244,245]. Individuals carrying rare variants in *GBA*, *LRRK2* or other PD genes were not excluded from the GWASs. I also performed sex-stratified analysis to identify if there are different genetic associations in men and women.

Functional Mapping and Annotation of GWAS (FUMA; https://fuma.ctglab.nl/) was used with standard settings to annotate, prioritise, and visualize GWAS results[121]. Gene-based and gene-set analyses were conducted in FUMA with MAGMA. I looked for enrichment of gene-sets or pathways in Gene Ontology (GO; MsigDB c5), Reactome, and the Kyoto Encyclopedia of Genes and Genomes (KEGG). GTEx (https://gtexportal.org/) and the eQTLGen Consortium (http://www.eqtlgen.org/index.html) were used to look up expression quantitative trait loci (eQTLs). LDlink (https://ldlink.nci.nih.gov/) was used to calculate linkage between SNP pairs (using LDpair) in European populations excluding the Finnish population.

Genetic Risk Scores (GRS) were calculated using the 90 loci from the most recent and largest PD case-control GWAS meta-analysis [75]. The association between the genetic risk score and each progression score was assessed using linear regression, adjusting for cohort and the first 5 genetic principal components. LD Score regression (LDSC) [246,247] was used to estimate the genetic correlation between the progression GWASs and the PD case-control GWAS using summary statistics excluding 23andMe samples [75].

GBA

I analysed *GBA* rare variant carriers compared to non-carriers in a subset of patients, using Sanger sequencing data from Tracking Parkinson's and whole genome sequencing data from PPMI. In PPMI, only the following *GBA* variants were covered: N370S, T369M, E326K, and R463C. I classified patients as carrying a pathogenic *GBA* variant, including Gaucher's Disease variants and variants associated with PD but excluding novel variants, following previous studies [135,248]. I analysed *GBA* status in relation to the progression scores using linear regressions, adjusting for cohort and the first 5 genetic principal components.

Levodopa-equivalent Daily Dose (LEDD)-adjusted sensitivity analyses

Medication may affect MDS-UPDRSIII scores, in particular in Tracking Parkinson's and Oxford Discovery where patients were assessed in the 'on' state. To address this, I performed a sensitivity analysis adjusting for LEDD, as described in a previous study, where I estimated the effect of levodopa on the MDS-UPDRSIII [242]. Merely adjusting for treatment as a covariate is not adequate, as therapy is not a simple confounder but a direct outcome of the underlying symptom – individuals who have more severe symptoms are more likely to be treated [199], and most likely with higher doses. Using the recommended method in previous studies [199], I added a sensible constant to the MDS-UPDRSIII scores to estimate what they would be if the patients were untreated, according to LEDD at each timepoint. I used data from the ELLDOPA study (personal communications) [249]. First, I converted UPDRS values from the ELLDOPA study to the MDS-UPDRS equivalent differences [250]. Second, I used a square root regression model at each timepoint to estimate the effect of different levodopa doses on the MDS-UPDRSIII [242]. This was only performed as a sensitivity analyses, as it involves extrapolation and the range of LEDD in these studies exceeds that from the ELLDOPA data.

Results

I included clinical data for 3,364 PD patients with 12,144 observations (Table 5.1). The mean follow-up time was 4.2 years (SD = 1.5 years), and mean disease duration at study entry was 2.9 years (SD = 2.6 years). 79.7% of patients had completed the 72-month follow-up visit.

Table 5.1. Cohort demographics at baseline. Means (SD) are shown unless otherwise indicated.

Demographics at baseline	**Tracking Parkinson's**	**Oxford Discovery**	**PPMI**
Number of PD patients	1966	985	413
Total number of visits analysed	5980	3137	3066
Mean length of follow-up (years)	3.8 (1.4)	4.3 (1.7)	5.4 (1.2)
Male (%)	65.2%	64.2%	65.4%
Age at onset (years)	64.4 (9.8)	64.5 (9.8)	59.5 (10.0)
Age at diagnosis (years)	66.3 (9.3)	66.1 (9.6)	61.0 (9.7)
Age at study entry (years)	67.6 (9.3)	67.4 (9.6)	61.5 (9.8)
Disease duration - time from symptom onset to assessment (years)	3.2 (3.0)	2.9 (1.9)	2.0 (2.0)
Time from diagnosis to assessment (years)	1.3 (0.9)	1.3 (0.9)	0.5 (0.5)
MDS-UPDRS Part III	22.9 (12.3)	26.8 (11.1)	20.7 (8.8)
MDS-UPDRS Part II	9.9 (6.6)	8.9 (6.2)	5.8 (4.1)
Hoehn and Yahr stage mean*	1.8 (0.6)	1.9 (0.6)	1.6 (0.5)
Hoehn and Yahr stage proportions*			
0 to 1.5 (%)	48.1%	23.2%	44.8%
2 to 2.5 (%)	45.1%	68.8%	54.7%
3+ (%)	6.8%	8.1%	0.5%
MoCA total (adjusted for education)	24.9 (3.6)	24.5 (3.5)	27.1 (2.3)
Semantic fluency$^+$	21.8 (6.9)	34.7 (9.0)	21.0 (5.4)
MDS-UPDRS Part I.1	0.5 (0.7)	0.5 (0.6)	0.3 (0.5)

* Tracking Parkinson's used the modified Hoehn and Yahr stage scale, while Oxford Discovery and PPMI used the original scale. Hoehn and Yahr stage proportions are shown as a total of the number of people with non-missing Hoehn and Yahr ratings at baseline.

$^+$ Instructions and timing for the semantic fluency task was slightly different between cohorts (completed within 60 seconds or 90 seconds). To account for these differences, I standardised all scales within each cohort separately (see Methods).

Within the motor progression PCA, using the MDS-UPDRS Part III, Part II, and Hoehn and Yahr Stage, the first principal component explained 61.0% of the total variance (Figure 5.3). Within the cognitive domain PCA, using the MoCA, semantic fluency, and MDS-UPDRS 1.1, the first principal component explained 59.8% of the total variance (Figure 5.4).

Figure 5.3. Scree plot and plot showing the proportion of variance explained in the motor progression principal component analysis.

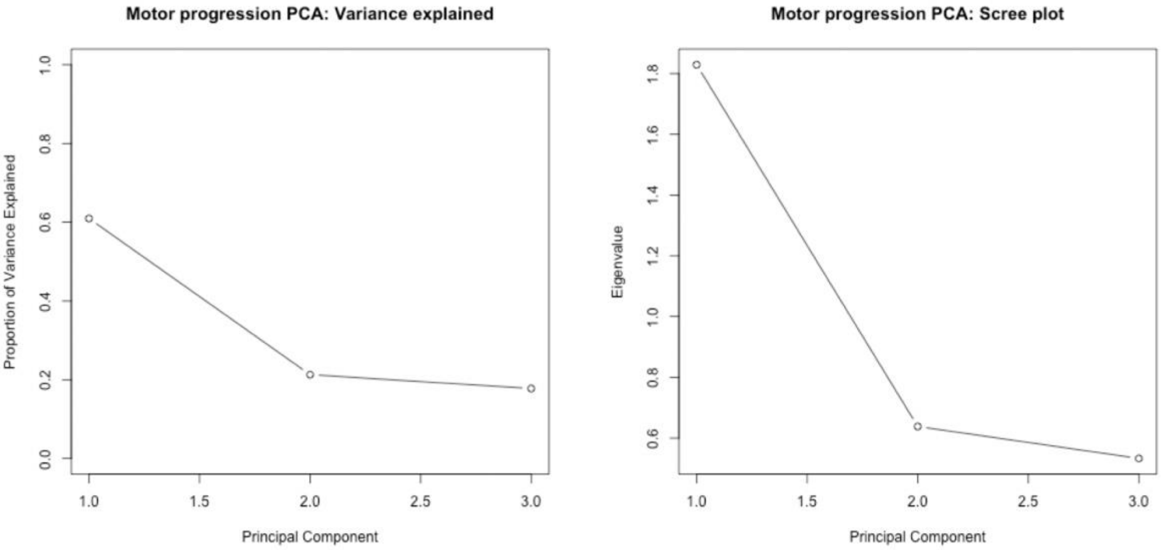
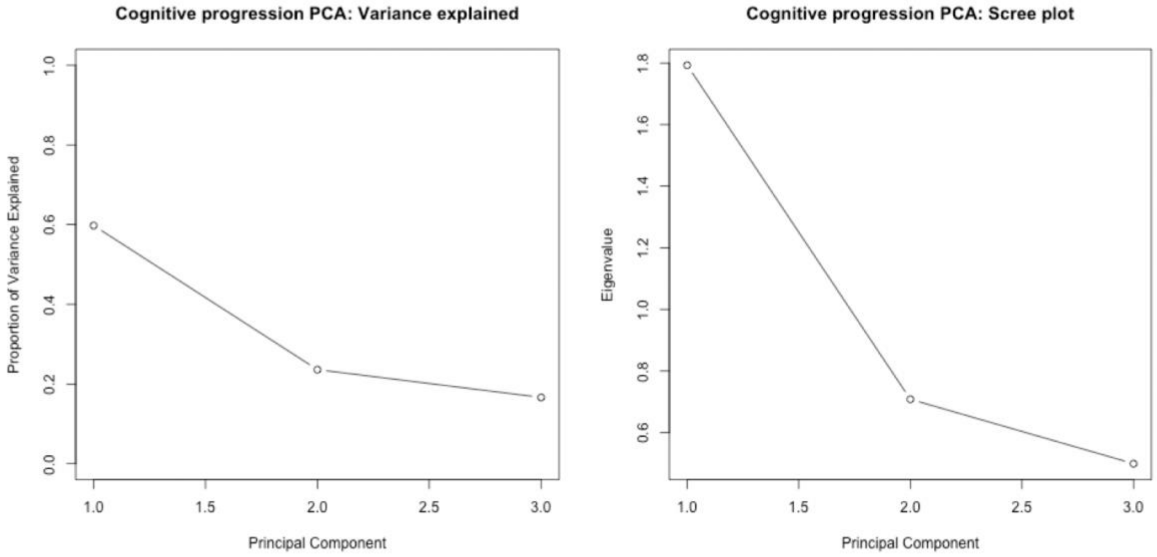


Figure 5.4. Scree plot and plot showing the proportion of variance explained in the cognitive progression principal component analysis.



I found that the first principal components for motor and cognitive progression were moderately correlated (r = -0.35, $p < 2.2 \times 10^{-16}$). The components from the PCA are latent components that explain the most variability in the data, and cannot therefore be interpreted as directional.

Because of the correlation between the motor and cognitive components, I therefore conducted a PCA combining all motor and cognitive measures, to create a composite progression score. The first principal component from this cross-domain PCA accounted for 41.0% of the joint variance. Table 5.2 shows that all the raw scales were well correlated with the final composite progression score and that there was no single scale which did not correlate as well as the others. None of the composite, motor, or cognitive principal components were associated with cohort (all p-values > 0.9).

Table 5.2. Correlation between first principal component from combined progression PCA and random slopes from individual measures. Pearson's r is reported.

	PC1	**PC2**	**PC3**	**PC4**	**PC5**	**PC6**
MDS-UPDRSIII	0.65	-0.48	0.12	-0.33	-0.46	0.11
MDS-UPDRSII	0.72	-0.28	-0.29	-0.30	0.40	-0.27
Hoehn and Yahr	0.57	-0.54	0.30	0.51	0.17	0.08
MoCA total	0.67	0.48	0.07	0.22	-0.26	-0.45
Semantic fluency	0.56	0.53	0.50	-0.23	0.22	0.23
MDS-UPDRS 1.1	0.66	0.30	-0.54	0.18	-0.05	0.37
% of variance explained	*41.0%*	*20.1%*	*12.3%*	*9.9%*	*8.6%*	*8.1%*

GWAS of composite progression

After quality control, imputation, and merging, 5,918,868 variants were available for analysis. 2,755 PD patients had composite progression scores and passed genetic quality control. The GWAS lambda was 1.02. One variant rs429358 in Chromosome 19 passed genome-wide significance (p=1.2 x 10^{-8}, Figure 5.5, Table 5.3). This variant tags the *APOE* ε4 allele. In the gene-based test, *APOE, TOMM40* and *APOC1* reached significance (p < 2.8 x 10^{-6}, correcting for the number of mapped protein coding genes). When I performed conditional analysis on the top SNP rs429358, there were no other SNPs that passed significance in this region. In the FUMA GENE2FUNC analysis of the prioritised genes, the Reactome pathway cytosolic sulfonation of small molecules pathway was significantly enriched (p = 6.9 x 10^{-6}). There were no pathways in the MAGMA gene-set analysis that passed Bonferroni correction (N = 15,496 gene sets tested) but the top 10 pathways are shown in Table 5.4.

Figure 5.5. Manhattan plot for GWAS of composite progression.The red dashed line indicates the genome-wide significance threshold p-value 5 x 10^{-8}.

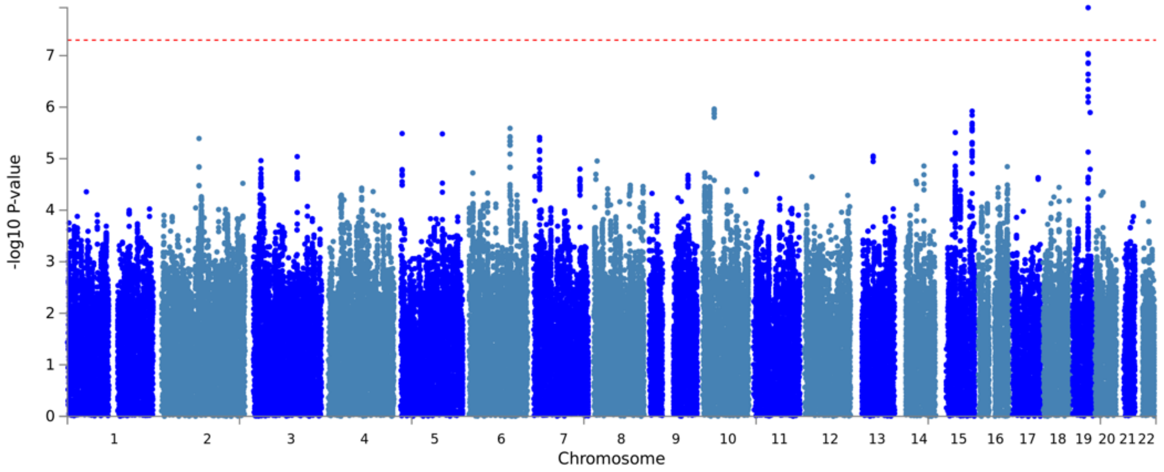


Table 5.3. Top 10 independent SNPs from the GWAS of composite progression.

Chr	Position (GRCh38)	SNP	Effect allele (minor)	Ref allele	Effect allele freq	Nearest gene	Distance to gene (kb)	Beta	SE	p value original	p value conditional (COJO)
19	44908684	rs429358	C	T	0.14	APOE	0	0.35	0.06	1.17E-08	1.07E-08
10	33942102	rs224750	T	C	0.37	PARD3	167458	-0.21	0.04	1.09E-06	1.20E-06
15	94318611	rs11634227	C	T	0.41	MCTP2	0	-0.21	0.04	1.19E-06	1.32E-06
19	50760039	rs4802739	C	A	0.40	GPR32	10425	0.20	0.04	1.27E-06	1.05E-06
6	119112570	rs79987229	T	A	0.01	FAM184A	0	0.85	0.18	2.57E-06	1.21E-06
15	45744252	rs17554587	C	G	0.22	SQRDL	52958	0.24	0.05	3.11E-06	3.39E-06
5	4699328	rs62343939	T	C	0.05	ADAMTS16	441002	0.43	0.09	3.25E-06	3.52E-06
5	122191027	rs17367669	T	G	0.22	LOC100505841	8364	0.23	0.05	3.31E-06	3.59E-06
7	17673826	rs10253857	T	C	0.22	SNX13	116935	-0.23	0.05	3.86E-06	4.19E-06
2	108292945	rs13424530	A	G	0.44	SULT1C2	0	0.20	0.04	4.06E-06	3.25E-06

Table 5.4. Top 10 pathways from MAGMA gene-set analysis for composite progression.This includes curated gene sets and GO terms from MsigDB. No gene sets/pathways passed Bonferroni correction for the number of tested gene sets (N = 15,496).

FULL_NAME	NGENES	BETA	BETA_STD	SE	P_unadjusted
Curated_gene_sets:scibetta_kdm5b_targets_dn	74	0.37	0.02	0.09	4.23E-05
Curated_gene_sets:boylan_multiple_myeloma_d_dn	67	0.36	0.02	0.10	1.57E-04
Curated_gene_sets:reactome_runx3_regulates_notch_signaling	12	0.90	0.02	0.26	2.21E-04
GO_bp:go_regulation_of_protein_k63_linked_ubiquitination	10	0.80	0.02	0.24	3.41E-04
Curated_gene_sets:martinez_tp53_targets_dn	534	0.11	0.02	0.03	4.68E-04
GO_bp:go_zymogen_activation	48	0.39	0.02	0.12	5.59E-04
Curated_gene_sets:gavin_il2_responsive_foxp3_targets_dn	5	1.21	0.02	0.37	5.78E-04
Curated_gene_sets:reactome_attachment_of_gpi_anchor_to_upar	7	1.30	0.03	0.40	6.18E-04
GO_bp:go_attachment_of_gpi_anchor_to_protein	6	1.43	0.03	0.45	6.63E-04
GO_bp:go_protein_quality_control_for_misfolded_or_incompletely_synthesized_proteins	24	0.48	0.02	0.15	7.98E-04

GWAS of motor progression

2,848 PD patients had motor progression scores and genotype data. The lambda was 1.02. No variants passed genome-wide significance (Figure 5.6, Table 5.5). However, in the gene-based test, *ATP8B2* in Chromosome 1 was associated with motor progression ($p = 5.3 \times 10^{-6}$), although this did not reach significance correcting for the number of mapped genes ($p = 2.8 \times 10^{-6}$). There was no enrichment of any gene sets or pathways in either the FUMA GENE2FUNC or MAGMA gene set analysis, but the top MAGMA gene sets are shown in Table 5.6.

I performed follow-up analyses to confirm that the results in the top SNPs were not driven by a single cohort, or a single scale. I conducted GWASs in each cohort separately (Table 5.7) and each motor scale separately (without combining in PCA). These results show that associations are strengthened with the PCA approach (Table 5.8).

The top variant in Chromosome 1, rs35950207, was associated with motor progression, $p = 5.0 \times 10^{-6}$. I examined the associations for our top SNPs in the previous progression GWAS [98] (https://pdgenetics.shinyapps.io/pdprogmetagwasbrowser/); rs35950207 was not significantly associated with binomial trait analysis of Hoehn and Yahr stage 3 or more at baseline (beta = 0.27, p = 0.03).

Table 5.5. Top 10 independent SNPs from the GWAS of motor progression.

Chr	Position (GRCh38)	SNP	Effect allele (minor)	Ref allele	Effect allele freq	Nearest gene	Distance to gene (kb)	Beta	SE	p value original	p value conditional
5	122193658	rs5870994	C	CTT	0.23	LOC100505841	10995	0.21	0.04	1.36E-06	1.49E-06
9	8454921	rs7870456	T	C	0.22	PTPRD	0	0.21	0.04	1.53E-06	1.68E-06
15	94320087	rs72767442	A	T	0.41	MCTP2	0	-0.18	0.04	1.69E-06	1.85E-06
2	23493673	rs6741991	G	A	0.26	KLHL29	0	0.20	0.04	2.91E-06	3.17E-06
1	154319482	rs35950207	T	C	0.31	AQP10	1585	-0.18	0.04	5.01E-06	5.40E-06
6	119067987		T	TAAAC	0.01	FAM184A	0	0.70	0.15	5.03E-06	5.40E-06
12	5829410	rs74709761	C	G	0.04	ANO2	0	-0.41	0.09	6.42E-06	8.72E-06
11	114821560	rs4436579	T	C	0.29	NXPE2	114443	0.18	0.04	7.47E-06	8.02E-06
12	12677103	rs12813102	C	A	0.04	GPR19	0	0.43	0.10	7.70E-06	1.05E-05
15	71520619	rs4128840	A	G	0.41	THSD4	0	-0.17	0.04	7.95E-06	8.53E-06

Table 5.6. Top 10 pathways from MAGMA gene-set analysis for motor progression.This includes curated gene sets and GO terms from MsigDB. No gene sets/pathways passed Bonferroni correction for the number of tested gene sets (N = 15,496).

FULL_NAME	NGENES	BETA	BETA_STD	SE	P_unadjusted
Curated_gene_sets:biocarta_wnt_lrp6_pathway	6	1.14	0.02	0.28	2.92E-05
GO_mf:go_nedd8_specific_protease_activity	7	1.21	0.02	0.31	3.98E-05
GO_bp:go_formation_of_anatomical_boundary	3	2.41	0.03	0.65	9.99E-05
GO_mf:go_atpase_regulator_activity	39	0.44	0.02	0.12	1.37E-04
GO_mf:go_atpase_activator_activity	24	0.58	0.02	0.16	1.44E-04
GO_bp:go_chaperone_mediated_protein_transport	9	0.92	0.02	0.26	2.43E-04
Curated_gene_sets:park_osteoblast_differentiation_by_phenylamil_dn	6	1.27	0.02	0.37	2.59E-04
Curated_gene_sets:servitja_islet_hnf1a_targets_dn	97	0.28	0.02	0.08	2.64E-04
Curated_gene_sets:watanabe_colon_cancer_msi_vs_mss_dn	55	0.36	0.02	0.11	3.45E-04
GO_bp:go_deoxyribonucleoside_triphosphate_biosynthetic_process	5	1.34	0.02	0.40	4.61E-04

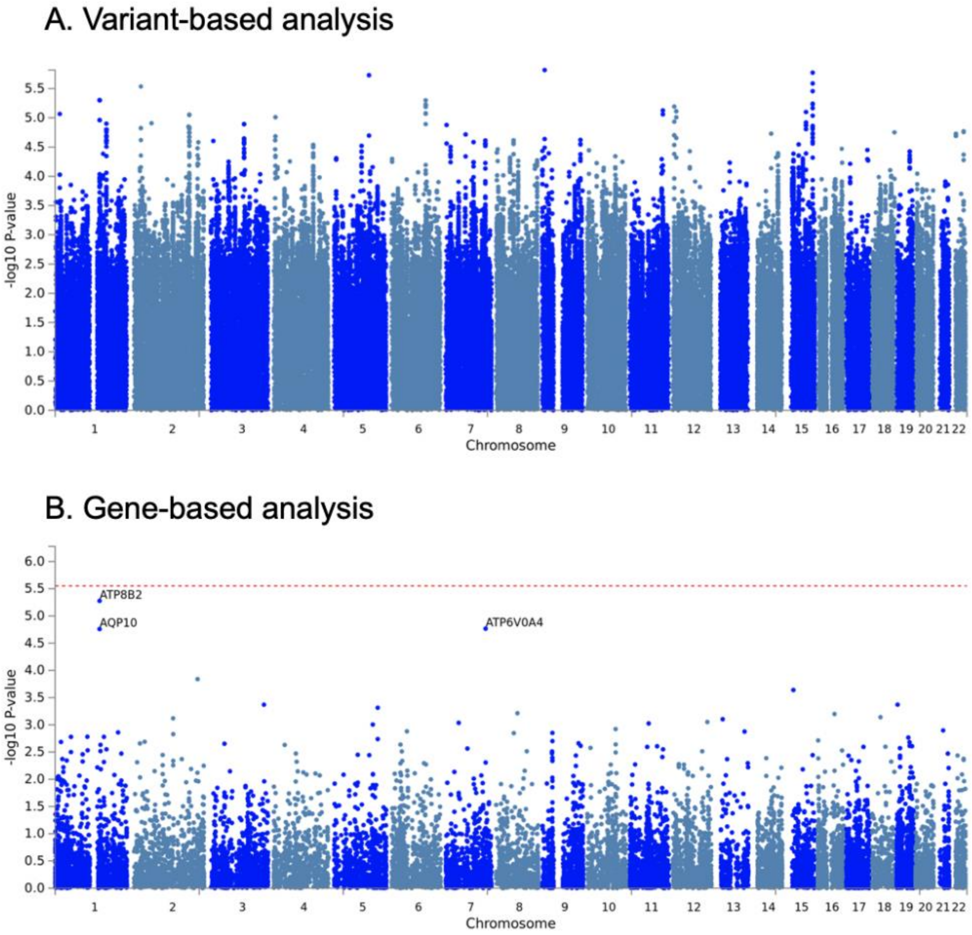
Table 5.7. Motor progression GWAS performed in each cohort separately. Progression scores were created in the merged cohort. The results for the top 5 independent hits from the combined motor progression GWAS are shown here. These show that the effects and allele frequencies are consistent across all three cohorts.

SNP	**Nearest gene**	**Combined**			**Tracking Parkinson's**			**Oxford**			**PPMI**		
		Beta	*p*	*MAF*	*Beta*	*p*	*MAF*	*Beta*	*p*	*MAF*	*Beta*	*p*	*MAF*
rs5870994	LOC100505841	0.21	1.36E-06	0.23	0.20	5.31e-05	0.23	0.17	0.04	0.21	0.30	0.064	0.26
rs7870456	PTPRD	0.21	1.53E-06	0.22	0.12	0.016	0.23	0.32	3.81e-05	0.23	0.36	0.052	0.21
rs72767442	MCTP2	-0.18	1.69E-06	0.41	-0.14	0.001	0.41	-0.15	0.026	0.40	-0.40	0.005	0.41
rs6741991	KLHL29	0.20	2.91E-06	0.26	0.10	0.040	0.25	0.17	0.022	0.26	0.65	7.19e-05	0.27
rs35950207	AQP10	-0.18	5.01E-06	0.31	-0.15	0.0007	0.31	-0.18	0.010	0.30	-0.28	0.080	0.32

Table 5.8. Motor progression GWAS performed for each scale separately. The results for the top 5 independent hits from the combined motor progression GWAS are shown here. The random slope from the mixed effects model for each scale was used as the progression measure. These results show that the effects are consistent across each of the different motor scales.

SNP	**Nearest gene**	**Combined**		**MDS-UPDRSIII random slope**		**MDS-UPDRSII random slope**		**Hoehn and Yahr random slope**	
		Beta	*p*	*Beta*	*p*	*Beta*	*p*	*Beta*	*p*
rs5870994	LOC100505841	0.21	1.36E-06	0.013	2.32e-05	0.012	9.43e-05	0.006	0.0008
rs7870456	PTPRD	0.21	1.53E-06	0.008	0.008	0.008	0.012	0.010	4.74e-10
rs72767442	MCTP2	-0.18	1.69E-06	-0.013	1.42e-06	-0.008	0.003	-0.004	0.006
rs6741991	KLHL29	0.20	2.91E-06	0.011	0.0004	0.010	0.002	0.007	1.08e-05
rs35950207	AQP10	-0.18	5.01E-06	-0.014	3.21e-06	-0.005	0.106	-0.006	2.79e-05

Figure 5.6. Manhattan plots for the GWAS of motor progression. A) Variant-based analysis. Genome-wide significance is the standard p-value 5 x 10^{-8} (not indicated in the figure). B) Gene-based analysis. Genome-wide significance was defined at p = 0.05/17802 (the number of mapped protein coding genes) = 2.81 x 10^{-6}.



rs35950207 is a variant 2kb upstream of *AQP10*. It is an eQTL for *AQP10* in whole blood (GTEx p = 1.7 x 10^{-6}, eQTLGen p = 3.6 x 10^{-139}) and other tissues (subcutaneous adispose, skin, esophagus, testis, and heart). It is also an eQTL for *ATP8B2* in blood (GTEx p = 1.5 x 10^{-5}, eQTLGen p = 7.8 x 10^{-42}) and in the cerebellum (GTEx p = 7.8 x 10^{-5}). *GBA* is also located in Chromosome 1 and *GBA* variants are associated with both PD risk and progression[251]. However, rs35950207 is not in linkage disequilibrium with any of the main *GBA* variants that are implicated in PD (p.E326K, p.N370S, p.L444P, p.T369M).

rs17367669 in Chromosome 5 was the top SNP in the variant-based analysis, but there were no genes in this region that approached significance in the gene-based analysis. This variant is in an intergenic region and is closest to *LOC100505841*, Zinc Finger Protein 474-Like gene. No significant eQTLs were identified for this variant.

GWAS of cognitive progression

2,788 patients had cognitive progression scores and genotype data. The lambda value was 1.02. The top variant was rs429358, which tags the *APOE* ε4 allele ($p = 2.5 \times 10^{-13}$, Figure 5.7, Table 5.9). Figure 5.8 shows that ε4 had more severe cognitive progression. *APOE* was also significantly associated with cognitive progression in the gene-based analysis, in addition to *APOC1* and *TOMM40*. There was no enrichment of any gene-sets or pathways, but the top pathways from the MAGMA gene set analysis are shown in Table 5.10. Follow-up analyses showed that the effects for the top 5 independent SNPs were consistent in each cohort and each scale (Tables 5.11, 5.12).

Figure 5.7. Manhattan plot for the variant-based GWAS of cognitive progression.The red dashed line indicates the genome-wide significance threshold p-value 5×10^{-8}.

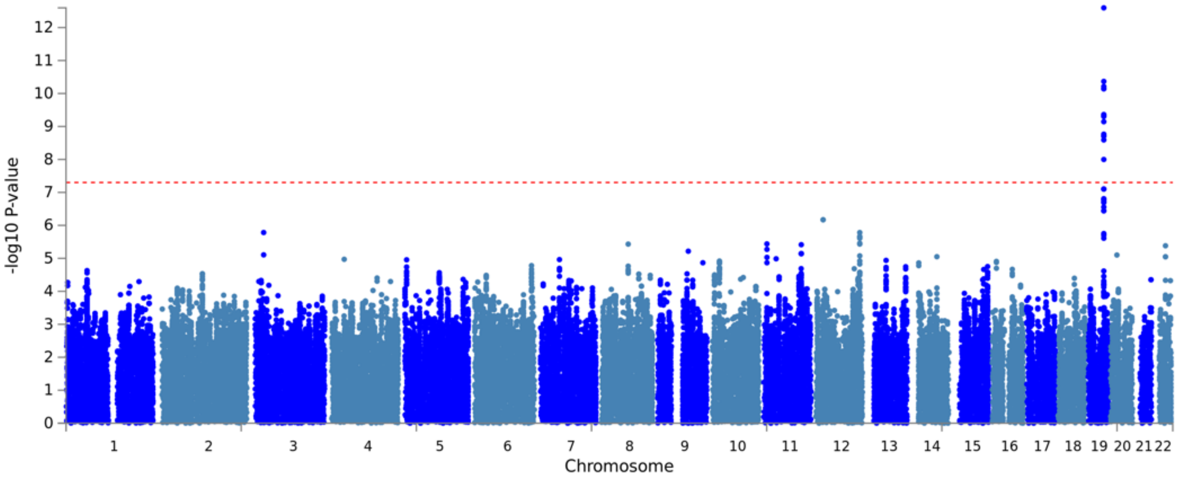
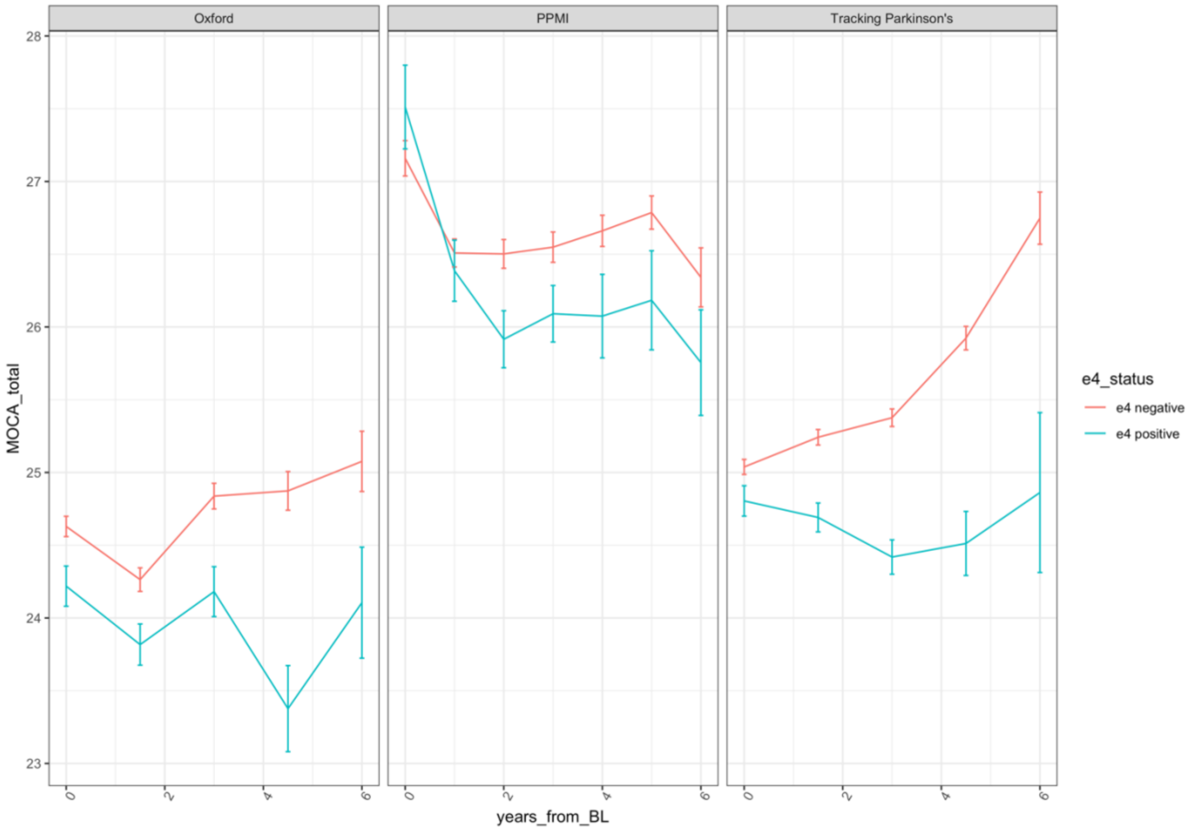


Figure 5.8. Raw MoCA scores in each cohort by *APOE* ε4 status (carriers vs. non-carriers).Note that the cognitive progression score was also based on semantic fluency performance and Part 1.1 of the MDS-UPDRS. Any mean data points with < 5 individuals were removed. Lines show the means ± standard errors of individuals who had data at that timepoint. Some of the means increase over time, likely because of participant drop-out. However, individuals who had data for at least one timepoint were still included in the progression scores and GWAS analysis; this graph is for illustrative purposes and does not capture all the data that was used to create the progression scores. The PPMI cohort were assessed at different timepoints (1 year intervals) than Tracking Parkinson's and Oxford Discovery (1.5 year intervals).



When I performed conditional analysis on the top SNP rs429358, a group of SNPs still passed genome-wide significance, indicating independent signals (Figure 5.9). The top SNP was rs6857 (beta=-0.33, p=4.4 x 10^{-11}). This is a 3' UTR Variant in *NECTIN2*. I also conditioned on the other *APOE* SNP rs7412 in addition to rs429358 (if both rs429358 and rs7412 harbour the C alleles then this codes the ε4 allele). This did not change the results.

When conditioning on both rs429358 and rs6857, there were still several SNPs that passed significance, the top being rs12721051, an intronic variant in *APOC1*.

Figure 5.9. Regional association plots of the Chromosome 19 locus associated with cognitive progression, conditioning on the top SNP rs429358.The recombination rate is shown in the blue line, based on European samples (build GRCh38). Plots were generated using LocusZoom (LocalZoom tool; http://locuszoom.org/). Conditional analysis reveals a group of SNPs that remain significant after removing the effect associated with rs429358. The top SNP is rs6857 (purple diamond).

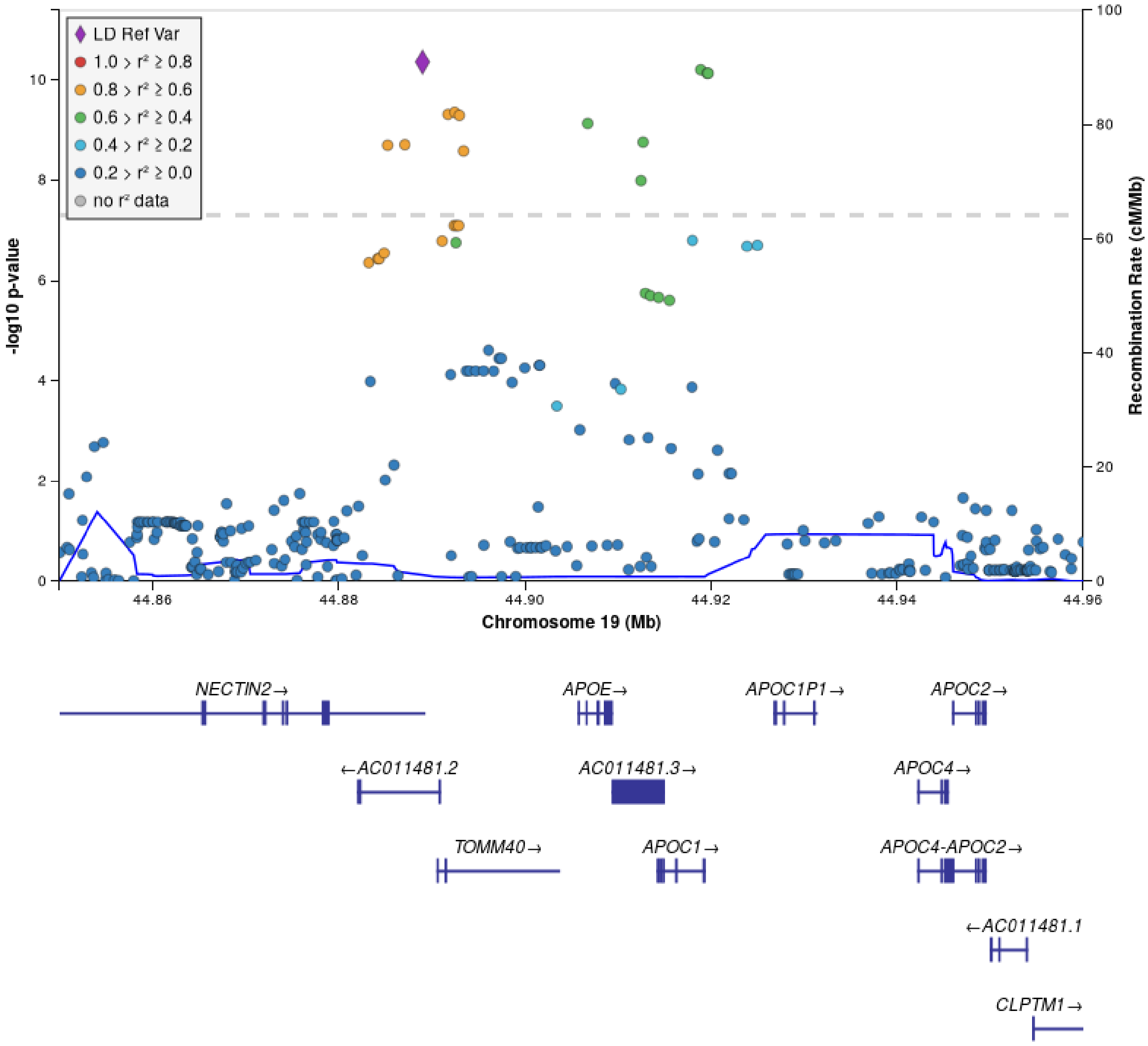


Table 5.9. Top 10 independent SNPs from the GWAS of cognitive progression.

Chr	Position (GRCh38)	SNP	Effect allele (minor)	Ref allele	Effect allele freq	Nearest gene	Distance to gene (kb)	Beta	SE	p value original	p value conditional
19	44908684	rs429358	C	T	0.14	APOE	0	-0.38	0.05	2.53E-13	4.20E-13
12	20812884	rs143371462	G	A	0.02	SLCO1B3	0	-0.64	0.13	6.76E-07	7.53E-07
3	23951314	rs113730632	G	A	0.05	NR1D2	0	0.41	0.09	1.65E-06	6.59E-07
12	125083207	rs6488987	C	T	0.36	AACS	0	0.18	0.04	1.65E-06	1.95E-06
11	8882396	rs34105455	G	A	0.13	ST5	0	-0.25	0.05	3.64E-06	3.94E-06
8	74970819	rs2956605	A	C	0.40	CRISPLD1	13654	-0.17	0.04	3.70E-06	4.02E-06
11	107127349	rs17092224	C	G	0.11	CWF19L2	198996	-0.27	0.06	3.85E-06	4.62E-06
22	34970241	rs5755468	C	T	0.37	ISX-AS1	0	-0.17	0.04	4.16E-06	4.49E-06
9	84627637	rs148603475	T	C	0.08	NTRK2	40821	-0.31	0.07	6.08E-06	8.94E-06
20	18097908	rs1124933	A	G	0.42	PET117	39947	-0.16	0.04	7.94E-06	8.50E-06

Table 5.10. Top 10 pathways from MAGMA gene-set analysis for cognitive progression.This includes curated gene sets and GO terms from MsigDB. No gene sets/pathways passed Bonferroni correction for the number of tested gene sets (N = 15,496).

FULL_NAME	NGENES	BETA	BETA_STD	SE	P_unadjusted
Curated_gene_sets:faelt_b_cll_with_vh3_21_up	38	0.54	0.02	0.12	3.99E-06
Curated_gene_sets:gregory_synthetic_lethal_with_imatinib	127	0.28	0.02	0.07	2.02E-05
Curated_gene_sets:zaidi_osteoblast_transcription_factors	12	1.04	0.03	0.26	4.14E-05
GO_mf:go_amino_acid_transmembrane_transporter_activity	67	0.38	0.02	0.10	1.13E-04
GO_bp:go_pyrimidine_nucleoside_metabolic_process	6	1.58	0.03	0.43	1.30E-04
Curated_gene_sets:yamazaki_tceb3_targets_up	154	0.22	0.02	0.06	1.78E-04
GO_bp:go_very_low_density_lipoprotein_particle_clearance	9	1.03	0.02	0.29	2.33E-04
GO_bp:go_calcium_ion_transmembrane_transport_via_high_voltage_gated_calcium_channel	12	0.87	0.02	0.25	2.95E-04
Curated_gene_sets:gavin_il2_responsive_foxp3_targets_dn	5	1.27	0.02	0.38	3.49E-04
GO_bp:go_positive_regulation_of_cgmp_mediated_signaling	6	0.98	0.02	0.29	3.63E-04

Table 5.11. Cognitive progression GWAS performed in each cohort separately. Progression scores were created in the merged cohort. The results for the top 5 independent hits from the combined cognitive progression GWAS are shown here. These show that the effects and allele frequencies are consistent across all three cohorts.

SNP	**Nearest gene**	**Combined**			**Tracking Parkinson's**			**Oxford**			**PPMI**		
		Beta	*p*	*MAF*	*Beta*	*p*	*MAF*	*Beta*	*p*	*MAF*	*Beta*	*p*	*MAF*
rs429358	APOE	-0.38	2.53E-13	0.14	-0.35	6.88e-09	0.14	-0.45	3.56e-06	0.14	-0.43	0.026	0.13
rs143371462	SLCO1B3	-0.64	6.76E-07	0.02	-0.54	0.0002	0.02	-0.73	0.003	0.02	-1.04	0.049	0.02
rs113730632	NR1D2	0.41	1.65E-06	0.05	0.36	0.0004	0.05	0.19	0.240	0.05	0.80	0.004	0.06
rs6488987	AACS	0.18	1.65E-06	0.36	0.21	1.66e-06	0.37	0.05	0.428	0.36	0.29	0.042	0.34
rs34105455	ST5	-0.25	3.64E-06	0.13	-0.15	0.018	0.13	0.10	0.0002	0.12	-0.39	0.047	0.14

Table 5.12. Cognitive progression GWAS performed for each scale separately. The results for the top 5 independent hits from the combined cognitive progression GWAS are shown here. The random slope from the mixed effects model for each scale was used as the progression measure. These results show that the direction of effects and p values are consistent across each of the different cognitive scales.

SNP	**Nearest gene**	**Combined**		**MoCA random slope**		**Fluency random slope**		**MDS-UPDRS 1.1 random slope**	
		Beta	*p*	*Beta*	*p*	*Beta*	*p*	*Beta*	*p*
rs429358	APOE	-0.38	2.53E-13	0.02	6.84e-13	0.007	4.04e-06	0.02	1.04e-07
rs143371462	SLCO1B3	-0.64	6.76E-07	0.03	0.0001	0.02	0.0001	0.03	9.78e-05
rs113730632	NR1D2	0.41	1.65E-06	-0.02	0.0004	-0.01	1.595e-05	-0.01	0.006
rs6488987	AACS	0.18	1.65E-06	-0.007	0.002	-0.004	9.16e-05	-0.009	2.37e-05
rs34105455	ST5	-0.25	3.64E-06	0.01	2.64e-05	0.006	0.0008	0.009	0.002

LEDD-adjusted analyses

When I performed GWASs of composite progression and motor progression after adjusting for LEDD, I did not find substantial differences. No SNPs passed genome-wide significance. The top SNP for composite progression was still rs429358, and this was in the same direction and similar effect size as in the main analysis (beta = 0.33, p = 8.8x10^{-8}). For motor progression, the top SNP was also the same as in the main analysis, and *ATP8B2* and *AQP10* still the top genes in the MAGMA gene analysis, though not genome-wide significant.

Sex-stratified analyses

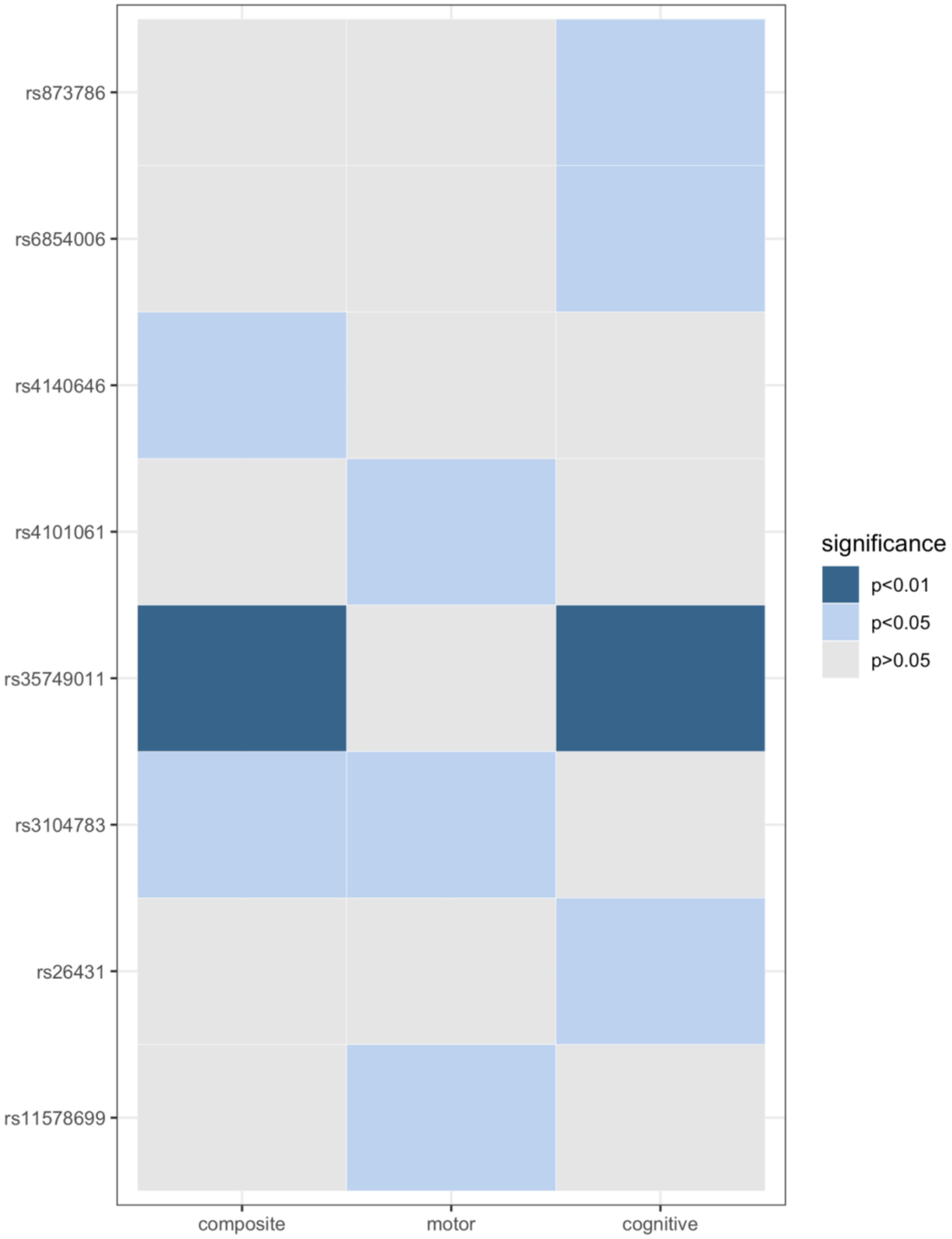
The *APOE* loci passed genome-wide significance only in men for composite progression and cognitive progression (p < 5 x 10^{-8}). Other than this locus, there were no SNPs that passed significance. These analyses are underpowered and sex differences need to be investigated in more detail.

Targeted assessment of PD risk loci

Of the 90 risk variants from the PD case-control GWAS [75], 73 were present in the final dataset, including the *SNCA* and *TMEM175/GAK* variants associated with PD age at onset[96]. I extracted results for these variants from the composite, motor, and cognitive progression GWASs. No variants passed analysis-wide significance (p = 0.05/73). Variants with at least one association p < 0.05 are shown in Figure 5.10.

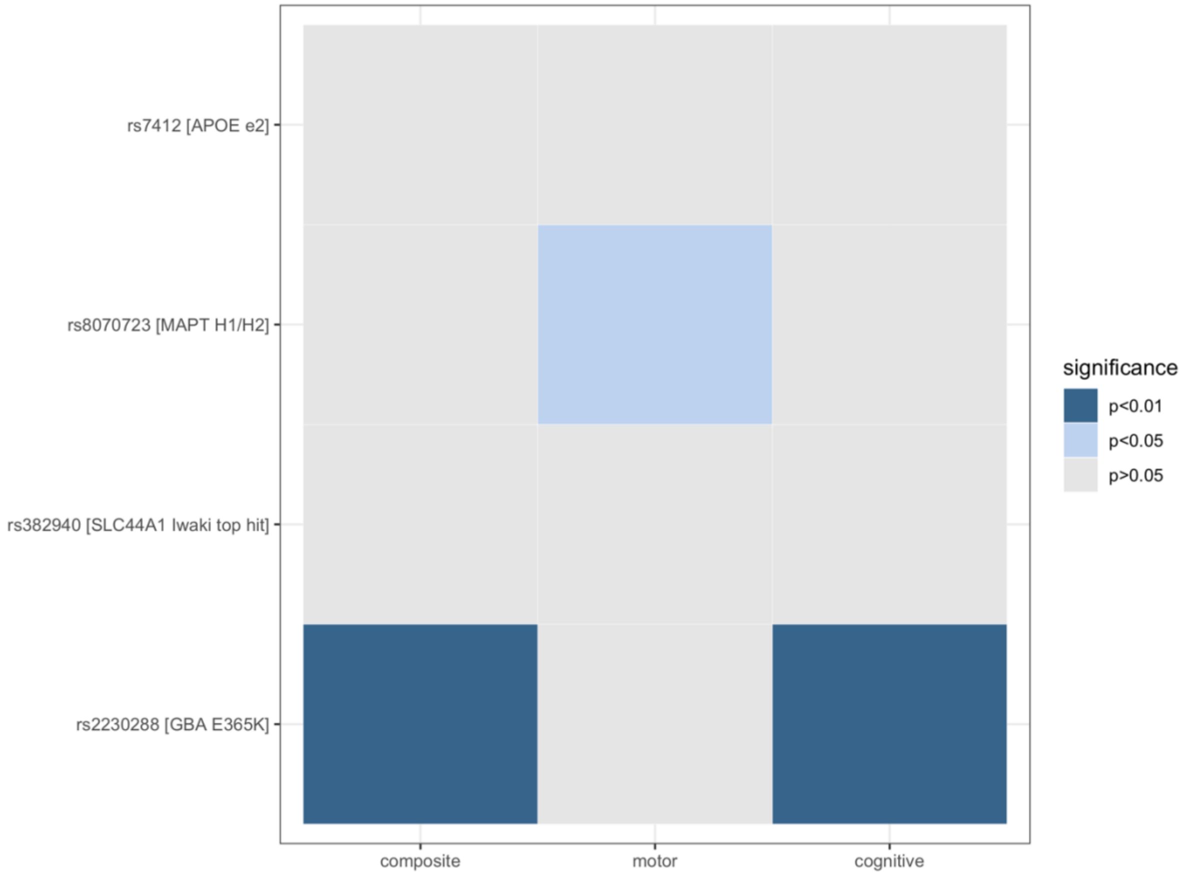
I found that only a small number of risk variants were associated with progression with p-values < 0.05. rs35749011 was associated with both composite progression (beta = 0.40, p = 0.003) and cognitive progression (beta = -0.37, p = 0.002), but not motor progression (beta = 0.20, p = 0.09). This variant is in linkage disequilibrium with the *GBA* p.E326K variant (also known as p.E365K), D'=0.90, R^2=0.78.

Figure 5.10. Heatmap of the PD GWAS risk loci and their association with composite, motor, or cognitive progression. Only variants with at least one association p < 0.05 are shown in the heatmap.



I also extracted results for other candidate variants that have been implicated in PD progression (Figure 5.11). I did not find that the top variant rs382940 in *SLC44A1* that was associated in progression to H&Y stage 3 from the Iwaki GWAS [98] was associated with either composite, motor or cognitive progression in our GWASs (all p-values > 0.05).

Figure 5.11. Heatmap of candidate variants and their association with composite, motor, or cognitive progression.



Overall, I did not find any overlap between the variants associated with PD risk, age at onset, and progression. The LDSC results also suggested very little overlap between the each of the progression GWASs and PD case-control GWAS (all p-values >0.5).

PD Genetic Risk Score

73 PD risk SNPs were present in our genotype data, and 2 proxies were identified for missing variants. There was no association between the standardised GRS and

composite progression (beta = -0.01, p = 0.65), motor progression (beta = 0.0008, p = 0.97), or cognitive progression (beta = 0.02, p = 0.36).

GBA

GBA data was available for 2,020 patients from Tracking Parkinson’s and PPMI. 194 patients (9.6%) carried a pathogenic variant in *GBA* (Table 5.13). *GBA* status was significantly associated with composite progression (beta = 0.40, p = 0.001) and cognitive progression (beta = -0.35, p = 0.0008), but not motor progression (beta = 0.18, p = 0.10).

Table 5.13. *GBA* variants included as pathogenic, and their frequencies.Note that some individuals carried more than one variant. Frequencies are shown as a percentage of the total number of patients in Tracking Parkinson’s and PPMI who were screened for GBA with sequencing (N= 2020).

Variant	Number of carriers (%)
p.E326K	103 (5.1%)
p.L444P	25 (1.2%)
p.N370S	14 (0.7%)
p.T369M	46 (2.3%)
p.G202R	2 (0.10%)
p.R463C	6 (0.30%)
p.D409H	1 (0.05%)
p.F213I	1 (0.05%)
p.G377S	1 (0.05%)
p.R257Q	1 (0.05%)

Removal of non-PD cases

I conducted sensitivity analyses to remove patients with potential non-PD conditions. Removing patients with <90% diagnostic certainty did not substantially affect my results; the top signals had slightly weaker associations in these sensitivity analyses. When I removed the extreme 5% of progressors, the top results from the main GWASs had larger p-values, although the direction of effects were the same (Tables 5.14, 5.15).

Table 5.14. Sensitivity analysis excluding PD cases with less than 90% diagnostic certainty. The top SNPs in the main analysis are shown, with the results from the sensitivity analysis for comparison. 5.2% (51/985) patients were removed from Oxford Discovery, 21.3% (419/1966) patients were removed from Tracking Parkinson's.

SNP	**Nearest gene**	**Results for top SNPs in full dataset**		**Results in PD cases with ≥ 90% diagnostic certainty**		
		Beta	*p*	*Beta*	*p*	*N*
Composite progression						
rs429358	APOE	0.35	1.17E-08	0.34	6.00e-07	2459
rs224750	PARD3	-0.21	1.09E-06	-0.20	5.36e-05	2459
rs11634227	MCTP2	-0.21	1.19E-06	-0.20	2.14e-05	2459
rs4802739	GPR32	0.20	1.27E-06	0.22	2.86e-06	2459
rs79987229	FAM184A	0.85	2.57E-06	0.99	3.64e-07	2459
Motor progression						
rs5870994	LOC100505841	0.21	1.36E-06	0.19	4.19e-05	2496
rs7870456	PTPRD	0.21	1.53E-06	0.19	0.0001	2496
rs72767442	MCTP2	-0.18	1.69E-06	-0.15	0.0001	2496
rs6741991	KLHL29	0.20	2.91E-06	0.19	2.27e-05	2496
rs35950207	AQP10	-0.18	5.01E-06	-0.18	2.13e-05	2496
Cognitive progression						
rs429358	APOE	-0.38	2.53E-13	-0.39	2.05e-12	2474
rs143371462	SLCO1B3	-0.64	6.76E-07	-0.64	3.22e-06	2474
rs113730632	NR1D2	0.41	1.65E-06	0.43	4.34e-06	2474
rs6488987	AACS	0.18	1.65E-06	0.19	1.69e-06	2474
rs34105455	ST5	-0.25	3.64E-06	-0.25	1.79e-05	2474

Table 5.15. Sensitivity analysis excluding fastest and slowest progressing cases (top and bottom 5% of each distribution)

SNP	**Nearest gene**	**Results for top SNPs in full dataset**		**Results in PD cases excluding extreme 5%**		
		Beta	*p*	*Beta*	*p*	*N*
Composite progression						
rs429358	APOE	0.35	1.17E-08	0.17	0.0002	2483
rs224750	PARD3	-0.21	1.09E-06	-0.10	0.002	2483
rs11634227	MCTP2	-0.21	1.19E-06	-0.10	0.003	2483
rs4802739	GPR32	0.20	1.27E-06	0.11	0.004	2483
rs79987229	FAM184A	0.85	2.57E-06	0.42	0.002	2483
Motor progression						
rs5870994	LOC100505841	0.21	1.36E-06	0.12	0.0003	2570
rs7870456	PTPRD	0.21	1.53E-06	0.10	0.002	2570
rs72767442	MCTP2	-0.18	1.69E-06	-0.07	0.012	2570
rs6741991	KLHL29	0.20	2.91E-06	0.08	0.008	2570
rs35950207	AQP10	-0.18	5.01E-06	-0.06	0.029	2570
Cognitive progression						
rs429358	APOE	-0.38	2.53E-13	-0.17	1.28e-05	2511
rs143371462	SLCO1B3	-0.64	6.76E-07	-0.32	0.001	2511
rs113730632	NR1D2	0.41	1.65E-06	0.17	0.006	2511
rs6488987	AACS	0.18	1.65E-06	0.09	0.002	2511
rs34105455	ST5	-0.25	3.64E-06	-0.15	0.0002	2511

Discussion

I used a new method of analysing clinical progression in PD, by combining multiple assessments in a data-driven PCA to derive scores of composite, motor, and cognitive progression.

This study contributes to evidence that improving the phenotypic measure can increase power in genetic studies. I showed that associations at the top signals strengthened when using the combined motor and cognitive progression scores compared to using the scales separately. The HD progression GWAS also showed that motor, cognitive, and brain imaging measures were well correlated, and successfully identified a variant in *MSH3* associated with composite progression[97]. Other studies have shown that the prediction accuracy of PD status or progression (such as development of cognitive impairment) is improved by combining multiple clinical, genetic, and biomarker factors[202,252].

In PD, there are many different scales for assessing symptoms. Each scale has a degree of measurement error [240] and different sensitivity to progression of underlying symptoms[32]. PCA is commonly used with clinical data. It is a data-driven approach that combines multiple measures to identify latent components that explain the most variability in the data, and these components may more accurately reflect disease progression.

My progression GWASs have identified two main findings. Firstly, I replicated previous findings for *APOE* ε4. Many studies have shown that the ε4 allele is associated with cognitive impairment and dementia in PD [85,87,88,92], and possibly in healthy individuals separate from the risk of Alzheimer's disease (AD)[253]. One possible mechanism is that *APOE* is associated with amyloid-β pathology, as comorbid AD pathology is common in PD patients with dementia (PDD) at postmortem[24]. Alternatively, *APOE* may drive cognitive decline independently of amyloid/AD pathology. Recent animal model work has shown that the ε4 allele is independently associated with α-synuclein pathology and toxicity[254]. In addition, the ε4 allele is overrepresented in Dementia with Lewy Body cases with 'pure' Lewy body pathology, compared to PDD cases[255]. A systematic review showed that limbic and neocortical α-synuclein pathology had the strongest association with dementia in PD[24]. Further work is needed to determine the mechanisms by which *APOE* influences cognitive decline.

In the *APOE* locus, there may be multiple independent signals for cognitive progression. This is similar to AD, where there have been multiple risk loci located in Chromosome 19 in addition to *APOE*, including *TOMM40, APOC1*, and more distant genes. This study was not powered to conduct analyses stratified by *APOE* genotypes as has been done in AD[256]. Further work is needed to fine-map this region and determine if there are other genes that contribute to cognitive progression.

I identified a novel signal in *ATP8B2* associated with motor progression in a gene-based analysis. This gene encodes an ATPase phospholipid transporter (type 8B, member 2). Phospholipid translocation may be important in the formation of transport vesicles[257]. This gene has not previously been reported in PD or other diseases, and needs to be tested in other independent cohorts.

Our sensitivity analysis adjusting for LEDD suggests that levodopa may influence the absolute scores in the MDS-UPDRSIII but does not influence the rate of progression, and this has been shown in a previous study[258]. I also found that the mean rate of change in the MDS-UPDRSIII was comparable between Tracking Parkinson's/Oxford Discovery and PPMI, despite the different medication states. Together, these suggest that medication has not influenced our results for motor progression.

Importantly, this study suggests that the genetics of PD risk and progression are largely separate. I performed a targeted analysis of the 90 risk loci identified in PD case-control GWAS [75]. *GBA* p.E326K was nominally associated with composite and cognitive progression. Analysis of sequencing data showed that *GBA* status was strongly associated with composite and cognitive progression, but not motor progression. Previous studies show that *GBA* variants are associated with rapid progression and mortality [63,64,66,68,71,194], however many of these studies have longer follow-up, or patients with longer disease duration at initial examination (6 to 15 years). This may explain why I did not find a strong effect for *GBA*, and is supported by analysis of *GBA* in patients earlier in disease stage [135]. In addition, previous studies have used different methods to measure progression. This unbiased genome-wide search suggests that, in addition to *GBA*, there are potentially other genes that are important for PD progression.

My targeted analysis showed that only a few PD risk variants were nominally associated with progression. This is similar to the findings of the previous PD progression GWAS [98,259]. These results suggest that there is minimal overlap in the genetic architecture of PD risk and PD progression. Similarly, the PD age at onset GWAS showed only a partial overlap with the genetics of PD risk [96]. It is now possible to study progression through the integration of detailed clinical data with genome-wide genetic variation in large-scale studies, and this can improve our understanding of the biology of progression.

I did not replicate the finding for the *SLC44A1* variant that was associated with progression to Hoehn and Yahr stage 3 in a previous PD progression GWAS [98]. I have used different methods and a different phenotype to analyse PD progression. Further progression GWASs are needed to replicate both sets of results, and other metrics for PD progression could be analysed, such as mortality.

While no other large genome-wide GWASs have investigated PD progression, many candidate gene studies have nominated common genetic factors associated with progression. Aside from *APOE*, common variants in *MAPT* [17,82–84], *COMT* [84,85]*, BDNF, MTHFR, and SORL1* [260] have been reported to influence cognitive decline (reviewed in Fagan & Pihlstrom [261]) . For motor progression, other than *GBA*, common variants in *SNCA* have been suggested to influence the rate of decline, although these studies are small and have not been confirmed in large studies [87,262–265]. A small GWAS of motor and cognitive progression identified suggestive loci in *C8orf4* and *CLRN3* [266], although these have not been replicated. A novel machine learning approach found that variation in *LINGO2* was associated with change in the MDS-UPDRS [93], although again this finding needs independent replication. I did not replicate these findings, possibly because the GWAS were underpowered to detect variants with smaller effects, or because I have analysed progression using different methods. However, many of these candidate gene studies are small and some associations have not been convincingly replicated.

This study has some limitations. Follow-up was limited to 72-months, and longer follow-up is needed to detect variants which may influence progression in later disease stages, such as *GBA.*

This study may also be underpowered to detect variants with smaller effects on progression, although this is one of the largest progression GWASs in PD. Although the HD GWAS identified significant signals in smaller samples [97], analysis of PD progression is more complex due to the slower rate of progression, greater heterogeneity in genetic risk and rate of progression between patients, and greater dissociation between motor and cognitive progression. Our findings need to be tested in independent cohorts, and the lack of independent replication is another limitation of this study.

A third limitation is that symptom progression may be influenced by non-SNP variants (such as rare variants or structural variants) and gene-gene interactions that would be missed by GWASs, or environmental factors and comorbidities.

A final limitation is the potential inclusion of patients that have non-PD conditions. I did not find that my results changed substantially when I excluded patients with diagnostic certainty < 90%. However certainty data was not available for PPMI, and abnormal dopamine transporter scans cannot differentiate between PD and other degenerative parkinsonian conditions [267]. Despite this, my sensitivity analyses suggests that our results are not being driven by non-PD conditions. In support of this, my GWASs also did not identify loci that are associated with PSP risk, including *MAPT, MOBP* [268], or the variant rs2242367 near *LRRK2* associated with PSP progression [269].

Many of the top variants had weaker signals when I excluded the 10% fastest and slowest progressing patients. This may be in part due to loss of statistical power. With the duration of follow-up in these studies, it is likely that the majority of majority of non-PD patients have been excluded, as diagnostic accuracy improves after 5 years of disease duration [17,270], however it is possible that some have not been excluded. Analysis of pathologically-confirmed PD cases is needed to resolve this issue. Alternatively, this may indicate that genotypes have different effects in the most extreme progressors. This could be due to co-morbidities such as vascular burden[222], or interactions between synuclein and co-pathologies (such as amyloid, and tau)[271,272] in the rapid progressors which exacerbates clinical progression.

This study is the first to use a PCA data reduction method to assess PD progression for genome-wide analysis, based on a successful approach in HD. I robustly replicated the association between *APOE* ε4 and cognitive progression, and have identified other genes which may be associated with progression. These advances are essential to understand the biology of disease progression and nominate therapeutic targets to stop or slow PD progression.

Chapter 6 : Genome-wide association studies using survival analysis to clinical milestones

Introduction

Previous progression GWAS studies in PD have identified loci associated with clinical progression markers, such as rate of change in the MDS-UPDRS, MoCA, and Hoehn and Yahr staging [98].

However, to date, no GWAS studies in PD have analysed variants that are associated with progression to mortality. Mortality is an objective marker of disease progression, while assessing change in clinical rating scales may be subjective.

Survival analysis is useful as it incorporates data on cases that have met the outcome and those that are still surviving. It may be more sensitive to rapidly progressing patients, for example patients who are unable to complete the full motor and cognitive assessments at follow-up visits.

The aim of this study was to conduct GWASs of survival to key clinical milestones in PD: mortality, Hoehn and Yahr stage 3 or greater, and dementia (defined by a MoCA score ≤ 21 or withdrawal due to dementia). I analysed large cohorts with longitudinal data available, including Tracking Parkinson's, Oxford Discovery, PPMI, QSBB, Calypso, and incident and prevalent PD cases from UK Biobank.

Methods

Cohorts

Data from the Tracking Parkinson's, Oxford Discovery, PPMI, QSBB, Calypso, and UK Biobank (UKB) cohorts were included for the analysis of mortality. Version 2 (17/06/2020) of the Tracking Parkinson's clinical dataset was used for this analysis. Only the clinical cohorts (Tracking Parkinson's, Oxford Discovery, and PPMI) were used for analysis of progression to Hoehn and Yahr stage 3, and dementia. Across all three clinical studies, patients who received alternative diagnoses during follow up or had neuroimaging results conflicting with a PD diagnosis were excluded from analyses.

Genotyping

Genotyping arrays are described in Chapter 2 (Methods). Standard quality control procedures were performed in PLINK v1.9. Genotype data from the six studies were called, genotyped and filtered separately, but following the same quality control steps. Individuals with low overall genotyping rates (<98%), related individuals (Identity-By-Descent PIHAT > 0.1), and heterozygosity outliers (>2SDs away from the mean) were removed, as were individuals whose clinically reported biological sex did not match genetically determined sex.

PCA was conducted on a linkage disequilibrium (LD) pruned set of variants (removing SNPs with an $r^2 > 0.05$ in a 50kb sliding window shifting 5 SNPs at a time) after merging with European (CEU) samples from the HapMap reference panel. Individuals who were more than 6 standard deviations away from the mean of any of the first 10 principal components were removed.

Variants were removed if they had a low genotyping rate (<99%), Hardy-Weinberg Equilibrium p value $< 1 \times 10^{-5}$ and minor allele frequency < 1%.

Following quality control, genotypes were imputed separately using the Michigan Imputation Server [137] (https://imputationserver.sph.umich.edu), using Minimac3 or Mimimac4 and Eagle version 2.4. The Oxford Discovery and Tracking Parkinson's data was imputed to the 1,000 Genomes Project reference panel (phase 3 release 5) [136]. I imputed the QSBB, and UK Biobank data was imputed to the Haplotype Reference Consortium panel (r1.1). Only variants with high imputation quality scores (R2) > 0.8 were retained for analysis, and imputation dosages were converted into hard call genotypes.

In order to remove individuals who were in more than one study and related individuals across different cohorts, I also merged individual level genotype data after imputation and quality control in each cohort. One individual from each pair of related individuals (PIHAT > 0.1) was excluded. In order to identify population outliers, twenty genetic principal components were generated from a linkage-pruned SNP set (removing SNPs with an $r^2 > 0.02$ in a 1000kb sliding window shifting 10 SNPs at a time). The first 2 components were plotted to check that there were no differences between the cohorts. I removed extreme outliers from the first 5 principal components (> 6 SDs away from the mean). The genetic principal components were then recalculated after removing outliers, as extreme outliers can substantially affect the calculation of genetic principal components. These first 5 new principal components were included as covariates in the GWASs to adjust for population substructure. There were no additional outliers who were > 6 SDs away from the mean of any of the first 5 principal components. I did this separately for the datasets in hg19/GRCh37 build (all except PPMI, as not enough individuals met the outcome) for the mortality GWAS.

For the GWASs of HY3 and dementia, I merged individual level data in hg38 build for the Tracking Parkinson's, Oxford Discovery and PPMI cohorts after lifting over the Tracking Parkinson's and Oxford Discovery data from hg19/GRCh37 to hg38 coordinates using liftOver.

UK Biobank

PD cases were defined as either prevalent, incident, or undefined PD as described in Chapter 2. The date of PD diagnosis was defined according to the UK Biobank guidelines, using the earliest PD code date from HES, or self-report.

For cases that were only identified through death records, it was not possible to determine the approximate date of PD diagnosis. The UK Biobank guidelines suggest that the date of death is used as the date of PD diagnosis, however this clearly cannot be used for survival analysis where the outcome is mortality. For this reason, these cases (N = 129) were excluded from analysis.

Death data was downloaded on 13/06/2020. I used the DEATH and DEATH_CAUSE tables which have the most updated death register data, according to guidance in UK Biobank document 'Mortality data: Linkage to death registries' (version 2.0, June 2020; http://biobank.ndph.ox.ac.uk/showcase/showcase/docs/DeathLinkage.pdf). Death data was received every month from NHS Digital and the NHS Central Register (NHSCR) for participants in Scotland.

The last date of death in the full dataset and also in the subset of PD cases was 21/04/2020. It was assumed that any participants who were not registered as dead were still alive. Therefore, the date of last follow-up was set as 01/04/2020, to account for some time delay in registering deaths. This is in line with guidance later released by UK Biobank (http://biobank.ndph.ox.ac.uk/showcase/exinfo.cgi?src=Data_providers_and_dates) which recommended a censoring date of 31/04/2020 for death data up to May.

If there were multiple differing death records, only the cause of death from the first death certificate (ins_index = 0) was used. Only 60 participants in the whole dataset had a second death record.

Genetic principal components were generated on the PD cases only (incident and prevalent cases separately), as opposed to the data that was available on the whole UKB cohort, as the PD participants may have been sampled from a slightly different population. The first 5 genetic principal components in each cohort were included as covariates in the survival analysis.

Clinical outcome measures and statistical analysis

I assessed progression to specific clinical milestones: mortality, Hoehn and Yahr Stage 3 or more, and dementia (MoCA score ≤ 21 or withdrawal due to dementia). This cutoff for dementia using the MoCA has been used in previous studies [135,222]. Time was measured from PD symptom onset, or estimated PD diagnosis in the UK Biobank cases. Time to event was taken as the first visit where the outcome was met. Individuals who were missing data at all timepoints for the clinical outcome being assessed were excluded (e.g. if Hoehn and Yahr stage data was missing at all visits for analysis of progression to Hoehn and Yahr stage 3+).

Cohorts were excluded if less than 20 individuals met the outcome of interest during the follow-up period (or < 5% of the total cohort size). Small numbers can produce unreliable effect size estimates and extremely wide confidence intervals.

Progression was assessed using Cox proportional hazard survival models. I ran GWASs adjusting for age at onset, sex, and the first 5 genetic principal components, to adjust for population stratification. I also carried out GWASs without any clinical covariates, to determine the base model - e.g. if the variants for PD progression overlap with PD age at onset then adjusting for age at onset would lose this signal.

Meta-analysis and annotation

Meta-analysis was performed in METAL, using an inverse variance weighted fixed effect model. Genomic control correction was performed to adjust the overall alpha error. Summary statistics from each cohort were annotated with rsIDs, and only SNPs that were present in all datasets were included in the final results. I excluded variants with p-value < 0.05 for Cochran's Q-test for heterogeneity (HetPVal) and I squared > 80. Variants with MAF variability greater than 15% across the cohorts were also excluded. I considered variants with $p < 5 \times 10^{-8}$ as genome-wide significant, as is the standard threshold for GWAS. Forest plots were generated in R v 3.6.2 using the *meta* package.

GWAS results were uploaded to FUMA (https://fuma.ctglab.nl/) with standard settings to annotate, prioritise, and visualize GWAS results [121]. SNPs were mapped to genes with positional mapping and eQTL mapping. In FUMA, I looked for enrichment of prioritized genes in pre-defined gene-sets or pathways in Gene Ontology (GO; MsigDB c5), Reactome, and Kyoto Encyclopedia of Genes and Genomes (KEGG). MAGMA gene and gene-set analysis was also performed in FUMA. GTEx (https://gtexportal.org/) and the eQTLGen Consortium (http://www.eqtlgen.org/index.html) were used to look up expression quantitative trait loci (eQTLs). LDlink (https://ldlink.nci.nih.gov/) was used to calculate linkage between SNP pairs (using LDpair) in European populations excluding the Finnish population.

Heritability for mortality was estimated using Linkage Disequilibrium Score Regression (LDSC) for summary statistics from meta-analysis [246,247]. Heritability estimation using GCTA was not possible as the phenotype is not a standard quantitative trait. I did not attempt to estimate heritability for the other outcomes as the sample sizes were smaller, and likely too small for LDSC.

Analysis of PD risk variants and GRS

Association results for the 90 PD risk loci [75] were extracted from the meta-analysis results of mortality, Hoehn and Yahr stage 3, and dementia. Bonferroni correction was applied for the number of variants tested (p = 0.05/72).

I also examined candidate variants that have previously been associated with progression, including a locus associated with progression to mortality in PSP. These results were extracted from the meta-analysis. Bonferroni correction was applied for the number of variants tested.

The GRS was calculated for each individual using PLINK, using the 90 loci from the most recent and largest PD case-control GWAS meta-analysis [75]. The association between the standardised GRS and progression to each outcome was assessed using a Cox proportional hazards regression, adjusting for age at onset, gender and the first 5 genetic principal components in each cohort. Results were then meta-analysed using random-effects meta-analysis.

Cause of death

To determine whether the top genetic signals for progression to mortality were related to PD or non-PD causes (e.g. general immunity, or COVID), I conducted a sub-analysis in patients by cause of death. In the QSBB cohort, only primary cause of death data was available. Here, I classified the primary cause of death as either PD-related and end of life related, or ‘interrupted’. Interrupted death causes included: cardiac arrest/ heart failure/ myocardial infarct/ heart disease, carcinoma, glioblastoma, gastric intestinal bleed or perforation, head injury, sudden death, or other accidental death causes.

In the UK Biobank cohort, full primary cause of death and contributory causes of death data was available in ICD10 codes (up to 15 levels). Death cause data is structured as primary (level 1) and contributory (level 2) causes, with the primary cause of death as the disease or condition stated to be the underlying cause of death. Only one primary cause of death was listed for each participant. I classified the cases as PD-related if PD was listed as the primary cause of death. If PD was not listed as a primary cause of death, I classified these cases as interrupted.

I conducted survival analysis on the interrupted vs. non-interrupted PD cases separately to identify if there was a difference in the top GWAS signals, which may suggest the genetic associations with mortality are not PD specific.

AD Polygenic Risk Scores

In order to determine whether the association results for PD progression are due to *APOE* specific effects or AD genetic risk in general, I analysed the association between the AD Polygenic Risk Scores (PRS) and PD progression – specifically mortality and dementia. Summary stats from Jansen et al. (2019) [273] were used to create AD PRSs for each individual. This is the largest AD GWAS conducted to date, with 71,880 AD cases (including 47,793 proxy cases defined as individuals from UK Biobank with one or both parents diagnosed with AD), and 383,378 controls (including 328,320 proxy controls) [273]. I created AD PRSs at different p-value thresholds, as studies have shown that a large proportion of the polygenic risk signal lies below the standard threshold for GWAS significance (5×10^{-8}) [274,275]. Following previous studies in AD and PD [274,275], I created PRSs including SNPs that met pre-defined significance thresholds in the original AD GWAS ($p < 1 \times 10^{-4}$, 0.001, 0.1, and up to 0.5). The *APOE* region (hg19 coordinates 19:45,020,859–45,844,508) [273] was excluded from the PRS, as previous studies have done [275,276], in order to determine whether the GWAS association results were *APOE* specific or due to AD risk in general.

The PRS was created for individuals in each cohort separately and standardised within each cohort. Cox proportional hazards regression models were used to assess the association between the AD PRS and PD progression, adjusting for age at onset, gender, and the first 5 genetic principal components. PRSs were created using PRSice2 [277] with standard LD clumping thresholds (clumping SNPs within a 250 Mb window and $r^2 > 0.1$). To improve the LD estimation for clumping, the 1000 Genomes European samples (N = 503) were used as an external reference panel, as is recommended for small samples in particular [278]. Results from the different cohorts were then meta-analysed in R using random-effects meta-analysis. Bonferroni correction for the number of PRSs tested was applied to correct for multiple testing ($p = 0.05/9 = 0.0056$).

Power calculations

Power calculations specifically for survival/time-to-event outcomes were performed using the R package ‘survSNP’ [279]. The alpha level was set to 5×10^{-8} and the sample size was fixed at the sample size of the mortality GWAS (N = 4831). The median for the survival function in the population was based on the median time to death (6 years from PD onset). Plots were generated to illustrate the power for different effect sizes, allele frequencies, and the proportion of individuals meeting the outcome.

Results

Summary of cohort characteristics

Table 6.1 shows the demographics at baseline and the number of patients who met each clinical milestone in each cohort. No studies had a genomic inflation factor (lambda) of greater than 1.2.

Table 6.1. Demographics at baseline and the number of patients meeting each clinical milestone/outcome in each cohort. Means (SD) are shown unless otherwise indicated.

Demographics	**Tracking Parkinson's**	**Oxford Discovery**	**PPMI**	**QSBB**	**UKB PD incident**[§]	**UKB PD prevalent**	**Calypso WTCCC2**
Number of PD patients overall	1963	985	413	339	1157	914	196
Number of PD patients with clinical mortality data and genetic data after QC[+]	1779	780	356	285	970	820	180
Mean length of follow-up, years	3.8 (1.4)	4.3 (1.7)	5.4 (1.2)	NA	NA	NA	NA
Male (%)	65.1%	64.2%	65.4%	60.7%	60.8%	62.4%	66.3%
Age at onset, years	64.5 (9.8)	64.5 (9.8)	59.5 (10.0)	61.8 (10.1)	NA	NA	59.8 (10.0)
Age at diagnosis, years	66.3 (9.3)	66.1 (9.6)	61.0 (9.7)	NA	69.5 (5.7)	57.4 (7.2)	61.5 (9.8)
Age at study entry, years	67.6 (9.3)	67.4 (9.6)	61.5 (9.8)	NA	63.9 (5.4)	62.8 (5.5)	67.5 (9.4)
Disease duration at baseline - time from symptom onset to study entry, years	3.2 (3.0)	2.9 (1.9)	2.0 (2.0)	NA	NA	NA	7.7 (5.2)
Time from diagnosis to study entry, years	1.3 (0.9)	1.3 (0.9)	0.5 (0.5)	NA	NA	5.4 (4.8)	5.8 (4.8)
MDS-UPDRS Part III	22.9 (12.3)	26.8 (11.1)	20.7 (8.8)	NA	NA	NA	NA
MDS-UPDRS Part II	9.9 (6.6)	8.9 (6.2)	5.8 (4.1)	NA	NA	NA	NA
Hoehn and Yahr stage mean*	1.7 (0.6)	1.9 (0.6)	1.6 (0.5)	NA	NA	NA	NA
0 to 1.5 (%)	48.2%	23.2%	44.8%	NA	NA	NA	NA
2 to 2.5 (%)	45.0%	68.8%	54.7%	NA	NA	NA	NA
3+ (%)	6.8%	8.1%	0.5%	NA	NA	NA	NA

MoCA total (adjusted for education)	25.2 (3.5)	24.5 (3.5)	27.1 (2.3)	NA	NA	NA	NA
Number of patients died (%)	133 (7.5%)	53 (6.8%)	15 (4.2%)	285 (100%)	370 (38.1%)	294 (35.9%)	121 (67.2%)
Time from PD onset to death, years	6.7 (4.5)	6.6 (2.6)	5.4 (2.6)	15.8 (7.8)	2.7 (2.3)	13.9 (6.1)	15.6 (6.1)
Time from PD onset to censoring/last follow-up in surviving cases, years	7.8 (3.4)	7.4 (2.7)	8.0 (2.5)	NA	5.5 (1.9)	16.1 (4.4)	19.6 (4.5)
Number of patients meeting H&Y≥3[^]	511 (28.8%)	181 (23.2%)	72 (16.8%)	NA	NA	NA	NA
Number of patients meeting dementia criteria[^]	470 (26.7%)	241 (31.3%)	75 (20.2%)	NA	NA	NA	NA

PPMI = Parkinson’s Progression Markers Initiative; QC = Quality Control; QSBB = Queen Square Brain Bank pathologically-confirmed PD cases; UKB = UK Biobank PD cases (including prevalent, incident, and undefined cases).

Percentages are shown of the total number of PD cases in the whole cohort, as the final number included in each analyses varied. Not all patients had all clinical data available (e.g. age at onset, gender, clinical outcomes) and these patients were excluded depending on the outcome of interest and which covariates were included in the model.

* Tracking Parkinson’s used the modified Hoehn and Yahr stage scale, while Oxford Discovery and PPMI used the original scale. Hoehn and Yahr stage proportions are shown as a total of the number of people with non-missing Hoehn and Yahr ratings at baseline.

[+] Note that the final number of patients included in the analysis may be slightly less, as some patients were missing data on covariates included in the GWAS.

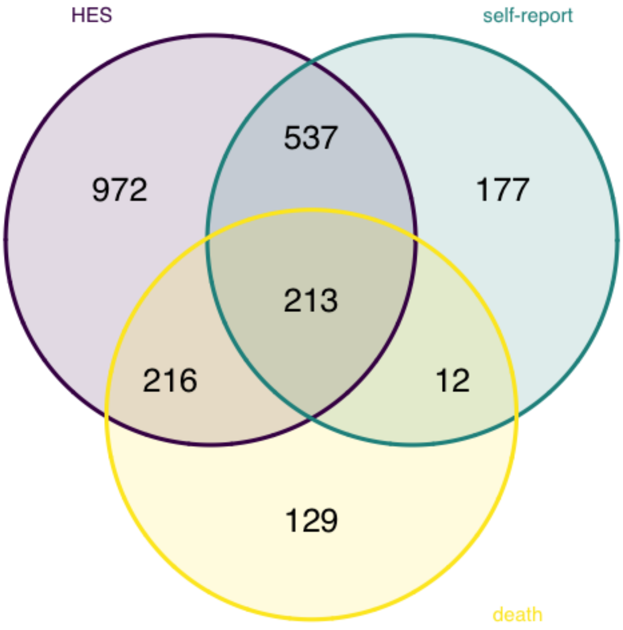
[^] Shown as a percentage of people with data for at least one timepoint. Individuals who were missing data for the outcome of interest at all timepoints were excluded.

[§] Note that this number excludes PD incident cases who were only identified through death records.

UK Biobank summary

PD cases were identified from HES, self-report, and death register records. Figure 6.1 shows the number of PD patients that were identified from each source.

Figure 6.1. Venn diagram showing the number of PD patients identified from each source in UK Biobank. This includes incident, prevalent, and undefined PD cases.



HES = Hospital Episode Statistics

In total, there were 2,256 PD cases. Out of these, 1,286 were classified as incident cases, 914 were classified as prevalent cases, and 56 were undefined (PD was only self-reported at a follow-up, not at baseline).

In total, 884 PD patients had died. Out of these, there were 440 (49.8%) PD cases defined from HES and/or self-report that also had PD listed as a cause of death (ICD10 code G20). There were 315 (35.6%) PD cases identified from HES and/or self-report that had died that did not include PD as a cause of death. The remaining 129 (14.6%) cases were identified as PD only from the death reports.

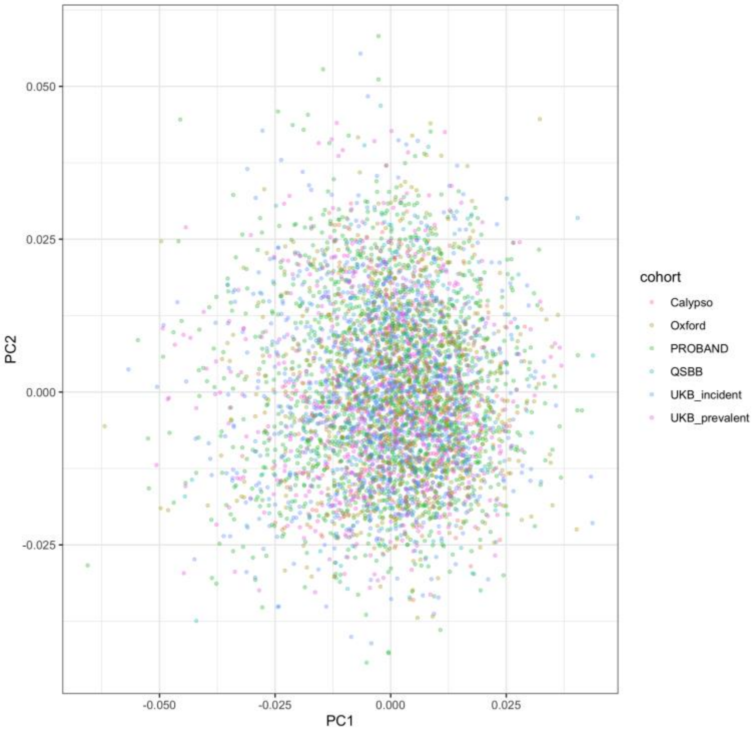
129 individuals who were identified as PD cases only from the death records were excluded from analyses. 2,127 PD participants with clinical data were included for

analyses. 755 PD cases who were identified from HES and/or self-report had died during period of follow-up (last date of death 21/04/2020). The mean time from PD diagnosis to death was 10.1 years (SD = 6.1). 1,372 PD cases did not die during the period of follow-up, with mean follow-up time 7.6 years (SD = 7.0).

GWAS of mortality

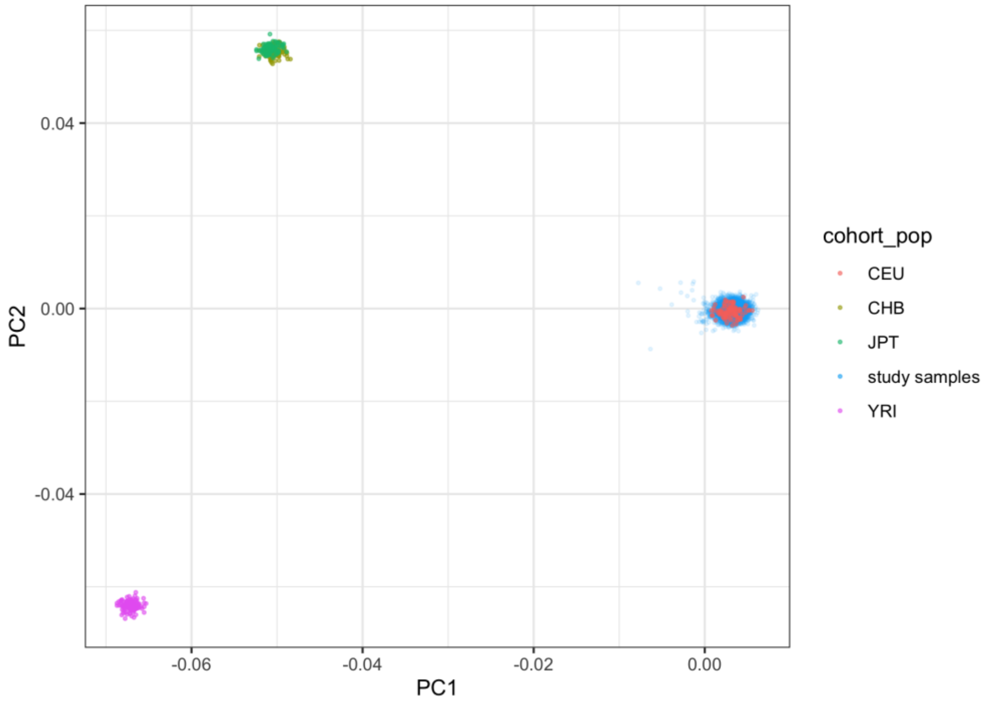
The PPMI cohort was excluded from the meta-analysis of mortality as not enough individuals died during the period of follow-up. 4,831 PD patients with both clinical and genetic data (in hg19/GRCh37 build) after quality control filters were included for analysis, although individuals who were missing event data and/or covariate data were excluded from each GWAS. The plot of the first 2 principal components after removing outliers is shown in Figure 6.2, and when merged with the HapMap reference samples (Figure 6.3).

Figure 6.2. Plot of the first two genetic principal components after removing outliers and related individuals in the merged dataset (excluding PPMI).This shows the cohorts overlap and there are no further population outliers.



PROBAND = Tracking Parkinson’s; QSBB = Queen Square Brain Bank; UKB = UK Biobank.

Figure 6.3. Plot of the first two genetic principal components merged with HapMap data. The samples from the current studies overlap with the European ancestry samples (CEU) from HapMap.



CEU =Utah residents with Northern and Western European ancestry from the CEPH collection; CHB = Han Chinese in Beijing, China; JPT = Japanese in Tokyo, Japan; YRI = Yoruba in Ibadan, Nigeria; PC = Principal Component.

4,814 PD patients were included in the meta-analysis of survival/mortality, adjusting for age at onset (or age at diagnosis in UK Biobank), gender and the first 5 principal components in each cohort.

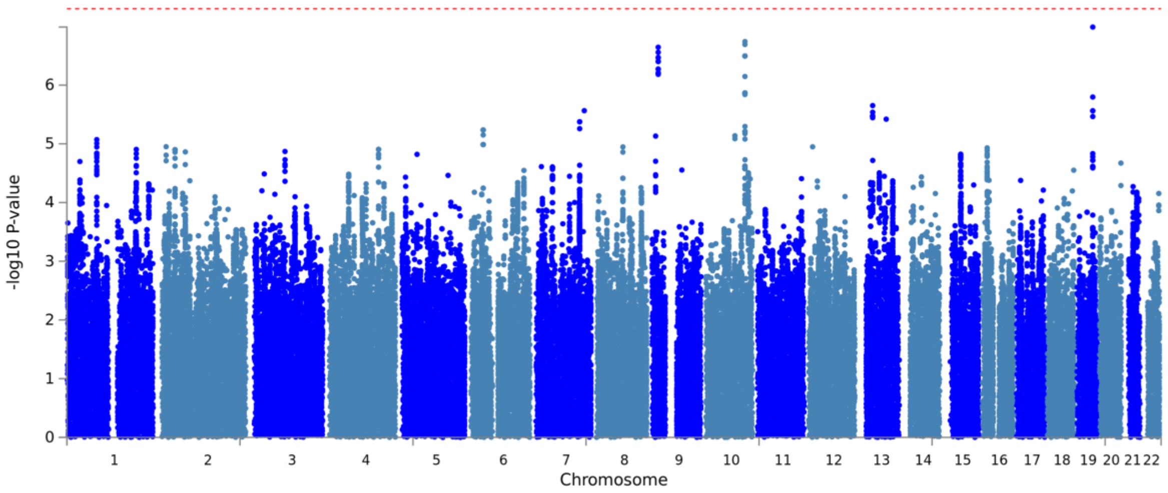
The UK Biobank PD patients were separated into prevalent and incident cases and analysed as separate cohorts. Undefined PD cases from UK Biobank were excluded. 5,099,774 SNPs were present in all 5 cohorts, and 4,872,144 variants passed filtering criteria for heterogeneity and MAF variability. No genome-wide significant SNPs were removed when filtering for heterogeneity and MAF variability. The lambda/genomic inflation factor was 1.01 (after SNP filtering).

Out of the 4,814 patients included in the final meta-analysis, 1,256 (26.1%) patients died. The mean time from PD onset to death was 10.1 years (SD = 7.8 years). The mean follow-up time for individuals who did not die was 8.7 years (SD = 4.8 years). The mean age at onset overall was 64.0 years (65.1 years in patients who died and 63.6 years in patients who did not die during the period of follow-up).

The top SNP was rs429358 (p = 1.0 x 10^{-7}) in Chromosome 19, which is the *APOE* ε4 tagging SNP (Table 6.2). The proportional hazards assumption was met in all cohorts for this SNP.

Another locus in Chromosome 10 approached GWAS significance. The top SNP was rs61871952 (10:113136589, genome build GRCh37/hg19) in Chromosome 10 with p-value = 1.8 x 10^{-7} (Figure 6.4). This is an intergenic variant, closest to the long non-coding RNA *LOC105378484* (also known as ENSG00000227851 or *RP11-381K7.1*). One of the top SNPs in this locus, rs61873401, is an intronic variant in this gene. The proportional hazards assumption was met in all cohorts.

Figure 6.4. Manhattan plot from the meta-analysis of mortality, including Tracking Parkinson’s, Oxford Discovery, QSBB, Calypso, UKB incident PD, and UKB prevalent PD.The PPMI cohort was excluded from this analysis.



PPMI = Parkinson’s Progression Markers Initiative; QSBB = Queen Square Brain Bank; UKB = UK Biobank.

Table 6.2. Top 10 independent SNPs from survival GWAS of mortality.

Chr	Position (GRCh37)	SNP	Effect allele (minor)	Effect allele freq	Ref allele	Nearest gene	Distance to gene (kb)	Beta	SE	p value
19	45411941	rs429358	C	0.15	T	APOE	0	0.30	0.06	1.03E-07
10	113136589	rs61871952	G	0.02	T	ADRA2A	295924	0.69	0.13	1.82E-07
9	17616880	rs3808753	G	0.03	A	SH3GL2	0	0.51	0.10	2.26E-07
13	38091191	rs9547920	G	0.25	C	POSTN	45528	0.22	0.05	2.23E-06
7	139664899	rs144889025	T	0.02	C	TBXAS1	0	0.63	0.13	2.73E-06
13	77237871	rs78017316	T	0.02	C	KCTD12	216433	0.65	0.14	3.80E-06
7	126312844	rs1074728	A	0.43	G	GRM8	0	0.20	0.04	4.21E-06
7	126299008	rs734524	G	0.36	A	GRM8	0	-0.21	0.05	5.51E-06
6	35415880	rs78333619	A	0.02	G	FANCE	4236	0.60	0.13	5.81E-06
10	84149216	rs12266006	G	0.05	C	NRG3	0	0.40	0.09	7.29E-06

Table 6.3. Top 10 pathways from MAGMA gene-set analysis for the GWAS of mortality.This includes curated gene sets and GO terms from MsigDB. No gene sets/pathways passed Bonferroni correction for the number of tested gene sets (N = 15,496).

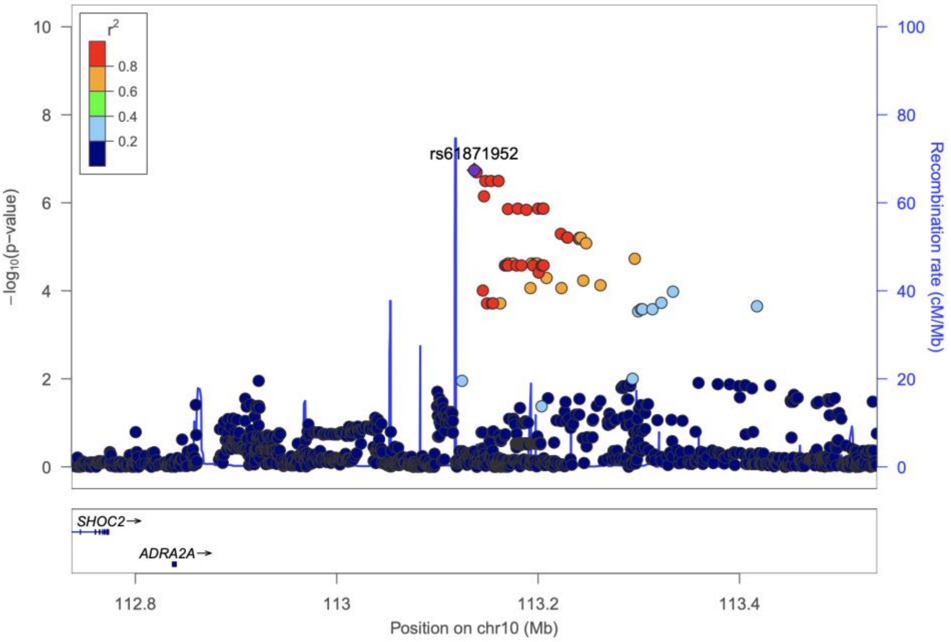
FULL_NAME	NGENES	BETA	BETA_STD	SE	P_unadjusted
GO_bp:go_positive_regulation_of_rna_splicing	29	0.52	0.02	0.15	1.61E-04
Curated_gene_sets:corradetti_mtor_pathway_regulators_dn	5	1.09	0.02	0.31	1.96E-04
Curated_gene_sets:nikolsky_breast_cancer_5p15_amplicon	21	0.92	0.03	0.27	2.82E-04
Curated_gene_sets:pedrioli_mir31_targets_up	181	0.20	0.02	0.06	3.08E-04
Curated_gene_sets:flechner_biopsy_kidney_transplant_ok_vs_donor_up	523	0.12	0.02	0.04	3.13E-04
GO_mf:go_translation_activator_activity	9	1.00	0.02	0.31	6.85E-04
GO_bp:go_triglyceride_rich_lipoprotein_particle_clearance	9	0.90	0.02	0.29	8.42E-04
GO_bp:go_phosphatidylcholine_catabolic_process	8	0.81	0.02	0.26	1.03E-03
Curated_gene_sets:wong_endometrial_cancer_late	6	0.96	0.02	0.31	1.05E-03
GO_bp:go_fructose_6_phosphate_metabolic_process	9	1.13	0.03	0.37	1.21E-03

None of the top SNPs in this region were significant eQTLs in BRAINEAC, GTEx or eQTLGEen (either cis- or trans-eQTLs). *LOC105378484* is expressed in the brain, spleen and testis (https://www.ncbi.nlm.nih.gov/gene/105378484).

The top loci in Chromosome 10 is < 300 kb away from *ADRA2A,* Adrenoceptor Alpha 2A (GRCh37/hg19 position 10:112,836,790-112,840,662; Figure 6.5).

There were no coding variants in linkage disequilibrium with the lead SNP rs61871952. Using LDproxy in European populations excluding Finnish, I searched for the closest SNPs with high regulome (< 3) scores which are more likely to have regulatory potential. This revealed a rare missense variant rs200592713 in *ADRA2A*, however this was far from the lead SNP and not in linkage disequilibrium (D' = 0.49, R^2 = 0.04). There were 2 SNPs in high LD with the lead SNP, rs61870947 (D' = 0.92, R^2 = 0.77) and rs9420101 (D' = 0.91, R^2 = 0.59) which had regulome DB scores of 3a. Finally, there was a common non-coding SNP rs7091217 which is a weak eQTL for *ADRA2A* in temporal cortex (1.3 x 10^{-3}) in BRAINEAC, but this was not in LD with the lead SNP (D' = 0.52, R^2 = 0.03).

Figure 6.5. Regional association plot of the top loci in Chromosome 10 from the PD survival GWAS of mortality.This was generated using the legacy service of LocusZoom (using LD population 1000 Genomes Nov 2014 EUR, hg 19 build).



The top SNP was imputed in all cohorts. Within the top chromosome 10 loci (10:113118502-113295922 as defined in FUMA) there were no directly genotyped SNPs in any of the cohorts, however high imputation quality thresholds had been applied (R2 > 0.8). The alleles matched in all cohorts and the minor allele frequencies ranged between 0.013 and 0.023 across cohorts (mean frequency 0.018, SE = 0.0025).

In the MAGMA gene test, *APOC1* and *APOE* passed genome-wide significance (p < 2.9 x 10^{-6}, correcting for 17,546 mapped protein-coding genes). There was no significant enrichment of any gene-sets in the MAGMA analysis after Bonferroni correction but the top 10 gene sets are shown in Table 6.3.

In the FUMA GENE2FUNC annotation, there was significant enrichment of the 27 prioritised genes in the GO molecular function phosphatidylcholine-sterol O-acyltransferase activator activity, with overlapping genes *APOE* and *APOC1*. There was also enrichment of the MsigDB c2 gene set ‘Roversi glioma copy number up’ (Genes in the most frequently gained loci in a panel of glioma cell lines), with overlapping genes *BCAM, PVRL2, TOMM40, APOE* and *APOC1.*

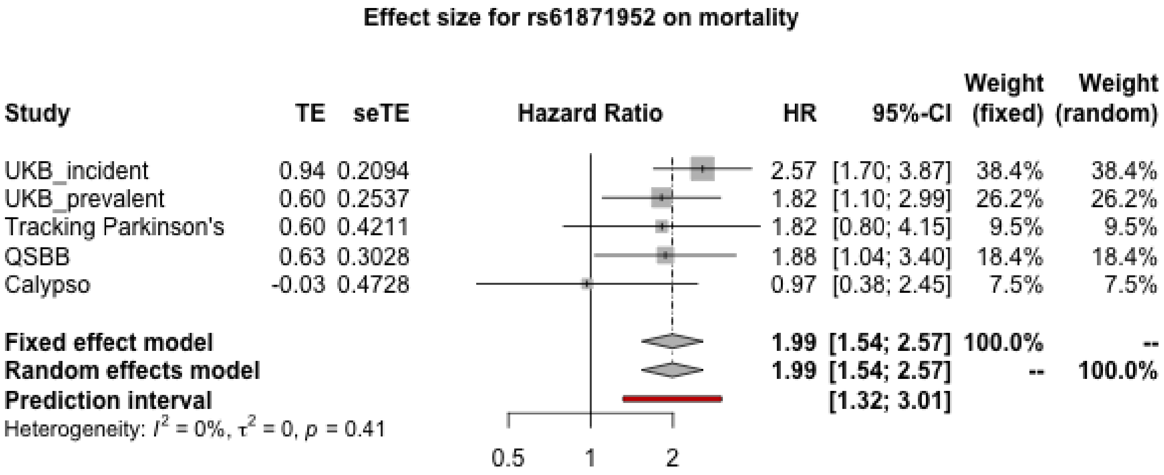
Table 6.4 shows the number of patients carrying the rs61871952 minor allele G, the number of patients that died, and their clinical characteristics. There was only one individual who was homozygous for the alternate allele, so all summary statistics in the table and plots show individuals carrying 1 and 2 minor allele together (dominant model). Figure 6.6 shows the effect size of the SNP in each cohort, suggesting that the effect is being driven by the UK Biobank cohorts and is weaker in the clinical cohorts. Figure 6.7 shows that rs61871952 minor allele carriers have more rapid progression.

Table 6.4. Frequency and clinical characteristics of patients carrying the rs61871952 minor allele, separated into patients who met the outcome (mortality) and those that did not.Means are shown in years (SD). The event mortality = 1 indicates the patients have died.

rs61871952 allele count	Event mortality	N	Time to event (years)	Median time to event	Mean age at onset
0	0	3470	8.7 (4.8)	7.67	63.6 (9.5)
0	1	1194	10.2 (7.9)	9.00	65.1 (9.6)
1 and 2	0	88	9.5 (5.1)	8.54	62.6 (9.7)
1 and 2	1	62	8.4 (6.8)	7.11	66.0 (9.2)

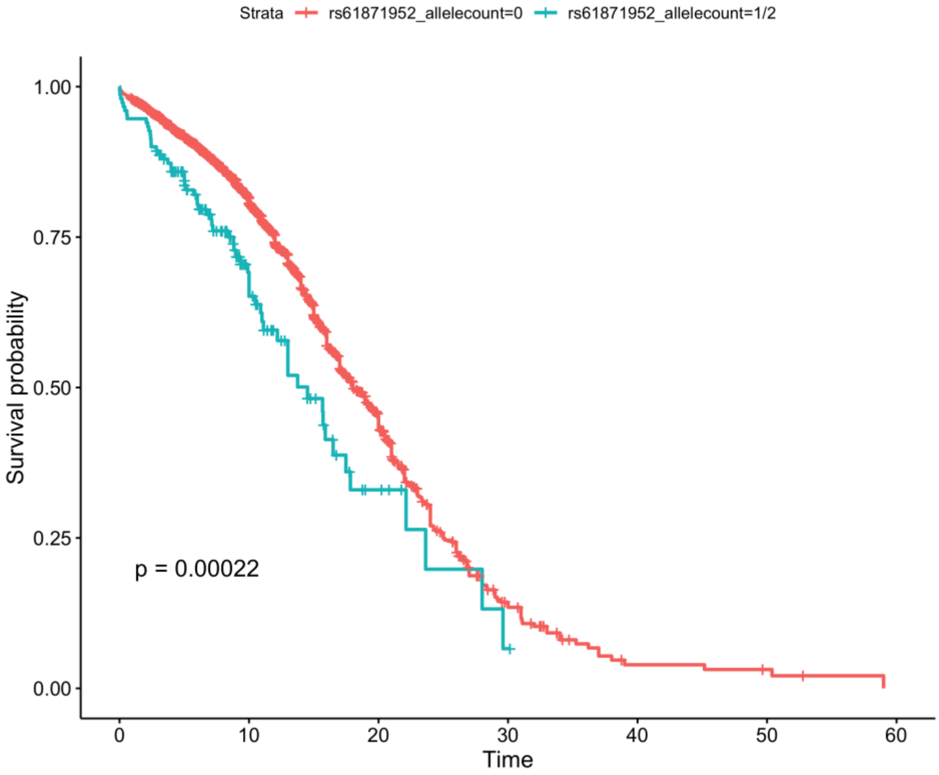
When I conducted GWAS on the merged individual-level dataset, adjusting for cohort, the chromosome 10 signal was diminished. This may be because the Chromosome 10 SNP effect varied between the cohorts. No loci reached genome-wide significance, although the Chromosome 19 and Chromosome 9 signals were nominally significant ($p < 5 \times 10^{-7}$).

Figure 6.6. Forest plot showing the Hazard Ratio of the Chromosome 10 top SNP, rs61871952, in each cohort.The Oxford Discovery cohort was removed from this plot as the confidence intervals were too wide and it was not weighted in the meta-analysis.



QSBB = Queen Square Brain Bank; UKB = UK Biobank

Figure 6.7. Kaplan-Meier survival curve of mortality for rs61871952 genotypes.This highlights differences in survival between rs61871952 TT carriers (red line) vs. cases carrying the rs61871952 GT or GG genotype (blue line). Time in years from PD onset (or PD diagnosis for UKB cases) is shown on the x-axis. The Kaplan-Meier log-rank test p-value is shown in the bottom left.



When analysing mortality without adjusting for age at onset and gender, no SNPs passed genome-wide significance. rs61871952 in Chromosome 10 had a slightly smaller p-value than in the main GWAS ($p = 1.7 \times 10^{-7}$), whereas the *APOE* locus did not pass heterogeneity filters ($I^2 = 57.2$, HetPVal = 0.04).

When attempting to estimate heritability using LDSC, the chi^2 was too low and final heritability estimate was -0.0029. This indicates that there was too little polygenic signal (likely because of low power and small sample size for LDSC) and is suggestive of low heritability. The negative estimate indicates that the true heritability is close to 0 and sampling error has led to an estimate below 0.

Cause of death sub-analysis

In a subset of patients with cause of death data, I conducted stratified analysis to determine if the main genetic signals were related to PD mortality, or were more general.

Many patients in the QSBB dataset were missing primary cause of death data (as our data was extracted from an online database which was set up only recently, coding data from paper records). 64 cases had genetic data after QC and cause of death coded. 47 of these (82.5%) had a PD-related cause of death and 10 (17.5%) were coded as interrupted.

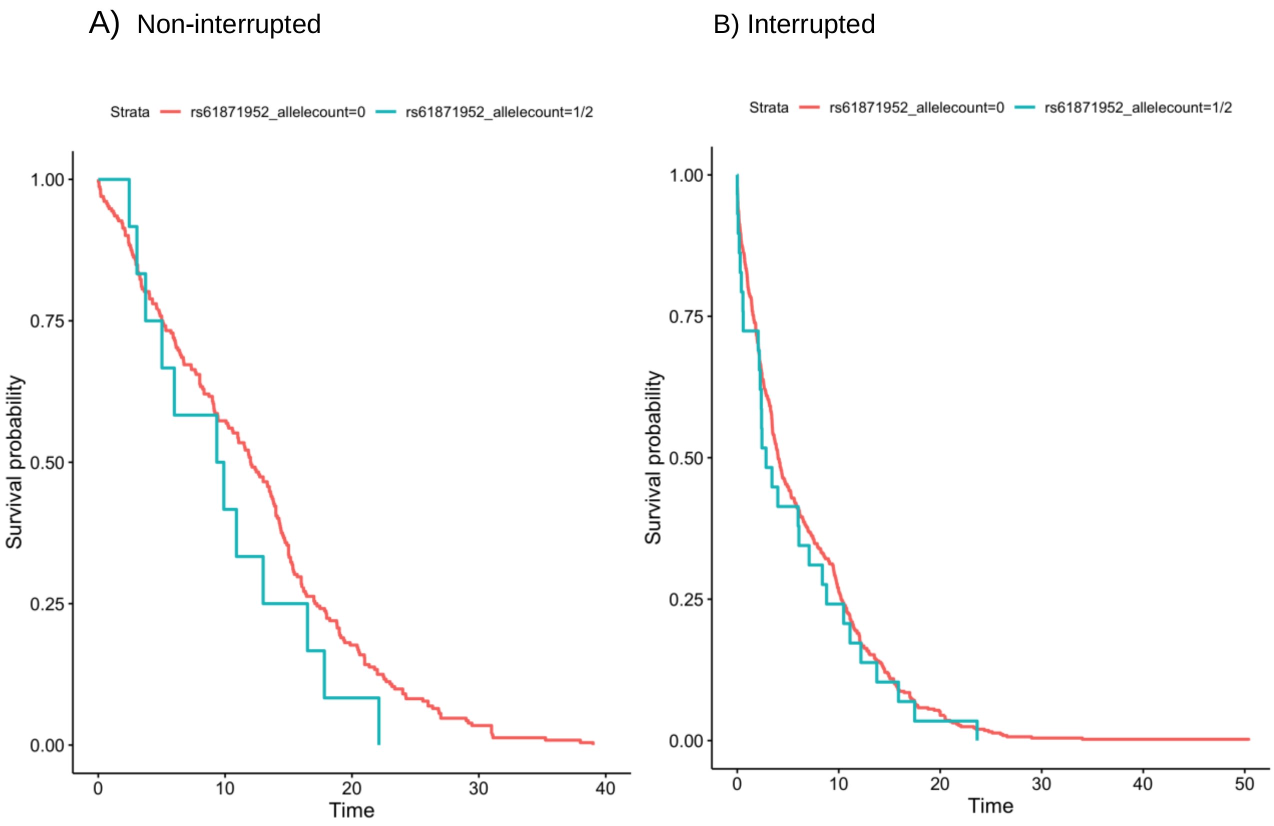
Out of the 370 incident UK Biobank PD cases who had died (excluding those who were only identified as PD from the death records) and had genetic data after QC, 80 (21.6%) had PD listed as a primary cause of death. Out of the 294 prevalent PD cases who had died and had genetic data, 117 (39.8%) had PD listed as the primary cause of death.

In total, after genetic QC, 721 patients out of 1,256 in the total dataset who died (57%) had cause of death data available which limited power for a full GWAS, so only the top 2 GWAS signals were analysed.

244 (33.8% of patients with cause of death data) PD patients who died were classified with a cause of death that was PD-related. In these non-interrupted death cases, there was a nominal effect of the Chromosome 10 SNP rs61871952 on mortality (HR = 1.7, p = 0.046) after adjusting for age at onset and gender.

477 (66.2%) of patients who had cause of death data were classified as ‘interrupted’. In these cases, the minor allele of rs61871952 was not associated with increased risk of mortality (HR = 1.4, p = 0.09). Figure 6.8 shows the Kaplan-Meier survival curves separately for the non-interrupted vs. interrupted cause of death cases.

Figure 6.8. Kaplan-Meier curves in the UK Biobank and QSBB cohorts by rs61871952 genotype, separately for PD cases with A) non-interrupted vs. B) interrupted cause of death.


A) Non-interrupted

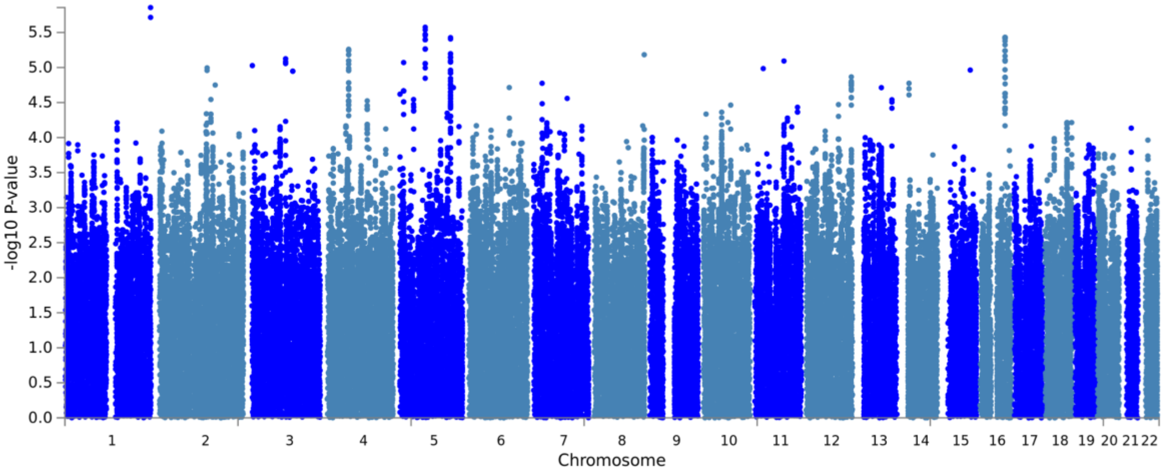
B) Interrupted

The *APOE* SNP rs429358 was associated with increased risk of mortality only in the non-interrupted cases (HR = 1.5, p = 0.002) and not in the interrupted death cases (HR = 1.01, p = 0.9).

GWAS of HY3+

5,193,490 variants were present in all three datasets and 4,940,410 variants remained after heterogeneity filtering. 2,941 individuals from Tracking Parkinson's, Oxford Discovery and PPMI were analysed, after excluding related individuals across cohorts and population outliers. The lambda from the meta-analysis (after filtering) was 0.99. No loci passed genome-wide significance for the meta-analysis of progression to Hoehn and Yahr stage 3 or greater (Figure 6.9, Table 6.5). The top variant was rs72771919 in Chromosome 1, p = 1.4 x 10^{-6}. No genes approached significance in the MAGMA gene analysis.

Figure 6.9. Manhattan plot for the meta-analysis GWAS of progression to Hoehn and Yahr stage 3 or greater.The cohorts included in this analysis were Tracking Parkinson's, Oxford Discovery, and PPMI.



There was no enrichment of any gene-sets in the MAGMA analysis after Bonferroni correction, but the top 10 pathways are shown in Table 6.6. In the FUMA GENE2FUNC, there was enrichment in the MsigDB c2 gene set 'Genes within amplicon 8q23-q24 identified in a copy number alterations study of 191 breast tumor samples (Nikolsky).'

The top variant rs72771919 was not in linkage disequilibrium with any of the common PD-associated *GBA* variants, although the D' was 1.0 for all variants but the R^2 was very low (Table 6.7). This may indicate the common variant is tagging rarer variants, but suggests the common variant itself is not causal. rs72771919 is an intergenic variant closest to *VN1R5*, Vomeronasal 1 Receptor 5.

Table 6.5. Top 10 independent SNPs from survival GWAS of Hoehn and Yahr stage 3+.

Chr	**Position (GRCh37)**	**SNP**	**Effect allele (minor)**	**Effect allele freq**	**Ref allele**	**Nearest gene**	**Distance to gene (kb)**	**Beta**	**SE**	**p value**
1	247425485	rs72771919	G	0.0206	A	VN1R5	5038	0.71	0.15	1.39E-06
5	73272603	rs73118271	T	0.0113	C	ARHGEF28	34785	0.96	0.21	2.66E-06
16	72107660	rs117689220	A	0.0212	G	TXNL4B/HPR	0	0.67	0.15	3.73E-06
5	146442837	rs112998873	A	0.0226	G	PPP2R2B	0	0.65	0.14	3.75E-06
5	146484645	rs115727994	C	0.0177	T	PPP2R2B	23562	0.71	0.15	3.91E-06
16	72059840	rs117350940	C	0.016	T	DHODH	524	0.74	0.16	4.19E-06
5	73280092	rs58647958	G	0.0158	A	ARHGEF28	42274	0.78	0.17	5.45E-06
4	62120962	rs974538	A	0.4192	T	ADGRL3	0	0.24	0.05	5.50E-06
8	144552033	rs61386175	C	0.1534	T	ZC3H3	0	0.30	0.07	6.57E-06
4	62138110	rs28634991	T	0.3754	A	ADGRL3	0	0.24	0.05	6.58E-06

Table 6.6. Top 10 pathways from MAGMA gene-set analysis for the GWAS of Hoehn and Yahr stage 3+.This includes curated gene sets and GO terms from MsigDB. No gene sets/pathways passed Bonferroni correction for the number of tested gene sets (N = 15,496).

FULL_NAME	**NGENES**	**BETA**	**BETA_STD**	**SE**	**P_unadjusted**
Curated_gene_sets:munshi_multiple_myeloma_dn	7	1.10	0.02	0.30	1.51E-04
Curated_gene_sets:martens_bound_by_pml_rara_fusion	417	0.14	0.02	0.04	1.73E-04
GO_bp:go_tetrapyrrole_biosynthetic_process	27	0.57	0.02	0.16	1.86E-04
Curated_gene_sets:hummel_burkitts_lymphoma_up	38	0.45	0.02	0.14	4.87E-04
Curated_gene_sets:sana_tnf_signaling_up	78	0.29	0.02	0.09	5.23E-04
GO_bp:go_regulation_of_systemic_arterial_blood_pressure_by_norepinephrine_epinephrine	8	1.07	0.02	0.33	5.52E-04
Curated_gene_sets:golub_all_vs_aml_up	23	0.51	0.02	0.16	5.63E-04
GO_mf:go_norepinephrine_binding	5	1.27	0.02	0.40	7.74E-04
GO_bp:go_alpha_amino_acid_biosynthetic_process	56	0.33	0.02	0.11	7.84E-04
Curated_gene_sets:kuuselo_pancreatic_cancer_19q13_amplification	30	0.84	0.03	0.26	7.87E-04

Table 6.7. Linkage statistics for common GBA variants with rs72771919. Generated using LDpair in European populations excluding Finnish.

GBA variant	**rsID**	**D’**	**R^2**
p.E326K	rs2230288	1.0	0.0001
p.N370S	rs76763715	1.0	0.0
p.T369M	rs75548401	1.0	0.0001
p.L444P	rs421016	1.0	0.0002

Published summary statistics from the Iwaki survival GWAS [98] to Hoehn and Yahr Stage 3 or greater were downloaded from https://pdgenetics.shinyapps.io/pdprogmetagwasbrowser/ (accessed November 2019). These summary statistics only included variants that passed heterogeneity filters, with MAF > 0.05 and total number of participants > 1000. In total, data was available for 431,602 variants that had rsIDs. No further filtering of these variants was applied prior to meta-analysis.

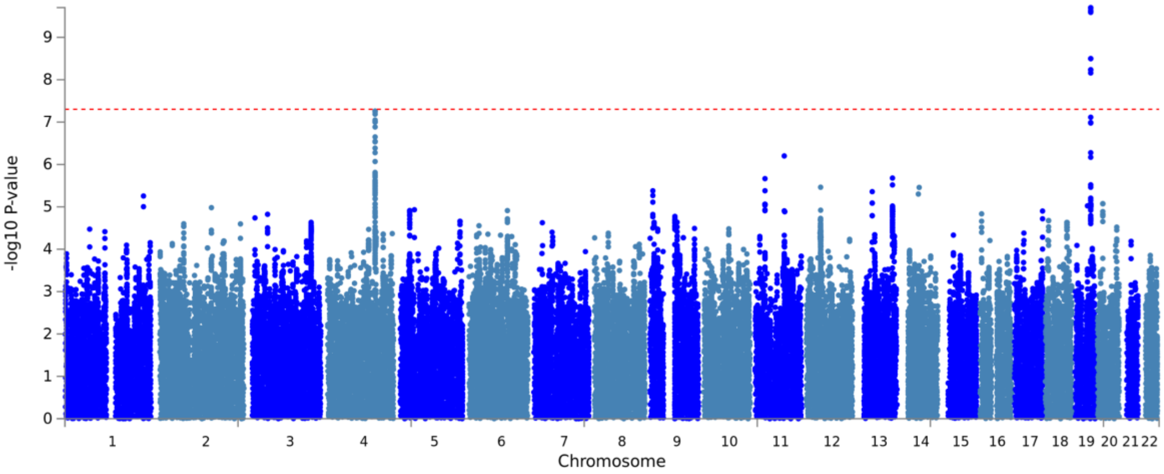
I meta-analysed the current datasets, excluding PPMI as this was included in the Iwaki GWAS. 215,814 variants were present in all 3 datasets and passed heterogeneity filters. The lambda was 0.96. No variants passed genome-wide significance. The top SNP was rs11174375, an intronic variant in *TAFA2* in Chromosome 12 ($p = 1.7 \times 10^{-6}$, N = 3803).

GWAS of dementia

One locus in Chromosome 19 was significantly associated with progression to dementia (Figure 6.10, Table 6.8). The top SNP was the APOE ε4 tagging SNP rs429358 (HR = 1.6, beta = 0.45, p = 2.0 x 10^{-10}).

In the MAGMA gene analysis, *APOC1* was the top gene strongly associated with progression to dementia. *TOMM40* and *APOE* also passed genome-wide significance (p < 2.8 x 10^{-6}, correcting for the number of mapped protein coding genes). There was no enrichment of any gene sets in MAGMA after Bonferroni correction, but the top 10 pathways are shown in Table 6.9.

Figure 6.10. Manhattan plot for the meta-analysis GWAS of progression to dementia (MoCA ≤ 21 or withdrawal due to dementia).The cohorts included in this analysis were Tracking Parkinson's, Oxford Discovery, and PPMI.



No other loci reached genome-wide significance, however there was a nominally significant locus in Chromosome 4. The top SNP in this locus was rs66882945, located at (hg19/GRCh37 position 4:137,129,406, beta = -0.50, p = 5.5 x 10^{-8}). It is an intronic variant in the long intergenic non-coding RNA *RP11-775H9.2* (also known as ENSG00000251567). It is not near *SNCA* (located at 4:90,645,250-90,759,447, hg19/GRCh37 build) in which both rare and common variants have been associated with PD risk and age at onset [96]. The top variant was not in linkage disequilibrium with any previously reported *SNCA* variants: rs356182, rs5019538, or rs2870004 [280].

Neither the Chromosome 10 SNP rs61871952 or the Chromosome 9 SNP rs3808753 were associated with progression to Hoehn and Yahr stage 3 or dementia (all p's > 0.1).

Table 6.8. Top 10 independent SNPs from survival GWAS of dementia.

Chr	Position (GRCh37)	SNP	Effect allele (minor)	Effect allele freq	Ref allele	Nearest gene	Distance to gene (kb)	Beta	SE	p value
19	45411941	rs429358	C	0.1427	T	APOE	0	0.45	0.07	1.99E-10
19	45390333	rs283815	G	0.2103	A	PVRL2	0	0.39	0.06	2.05E-10
4	137129406	rs66882945	C	0.0568	G	RP11-775H9.2	0	0.51	0.09	5.54E-08
11	84641700	rs118188129	T	0.0886	C	DLG2	0	0.40	0.08	6.30E-07
13	101813637	rs594524	C	0.4147	T	NALCN	0	-0.25	0.05	2.09E-06
11	28954555	rs77905407	C	0.077	T	RP11-115J23.1	0	0.43	0.09	2.16E-06
4	137123682	rs2053895	G	0.1134	A	RP11-775H9.2	0	0.34	0.07	2.92E-06
12	41319351	rs2405296	G	0.0444	A	CNTN1	0	0.49	0.10	3.44E-06
19	45404857	rs112019714	C	0.0265	T	TOMM40	0	0.64	0.14	3.45E-06
14	52519137	rs72680322	A	0.0224	G	NID2	0	0.69	0.15	3.48E-06

Table 6.9. Top 10 pathways from MAGMA gene-set analysis for the GWAS of dementia. This includes curated gene sets and GO terms from MsigDB. No gene sets/pathways passed Bonferroni correction for the number of tested gene sets (N = 15,496).

FULL_NAME	NGENES	BETA	BETA_STD	SE	P_unadjusted
GO_mf:go_dolichyl_phosphate_mannose_protein_mannosyltransferase_activity	8	1.14	0.02	0.26	3.87E-06
GO_bp:go_serotonin_transport	12	0.90	0.02	0.24	1.20E-04
GO_bp:go_regulation_of_guanylate_cyclase_activity	11	0.78	0.02	0.22	1.83E-04
GO_bp:go_protein_nitrosylation	12	0.96	0.02	0.28	2.41E-04
Curated_gene_sets:reactome_nephrin_family_interactions	21	0.64	0.02	0.19	3.13E-04
GO_mf:go_interleukin_1_receptor_activity	6	1.33	0.02	0.39	3.52E-04
GO_bp:go_regulation_of_mast_cell_cytokine_production	5	1.10	0.02	0.33	3.69E-04
Curated_gene_sets:mikkelsen_es_hcp_with_h3_unmethylated	55	0.36	0.02	0.11	3.98E-04
GO_bp:go_subpallium_development	22	0.65	0.02	0.19	4.27E-04
GO_bp:go_serotonin_uptake	5	1.16	0.02	0.36	5.51E-04

Targeted assessment of PD risk loci and candidate variants

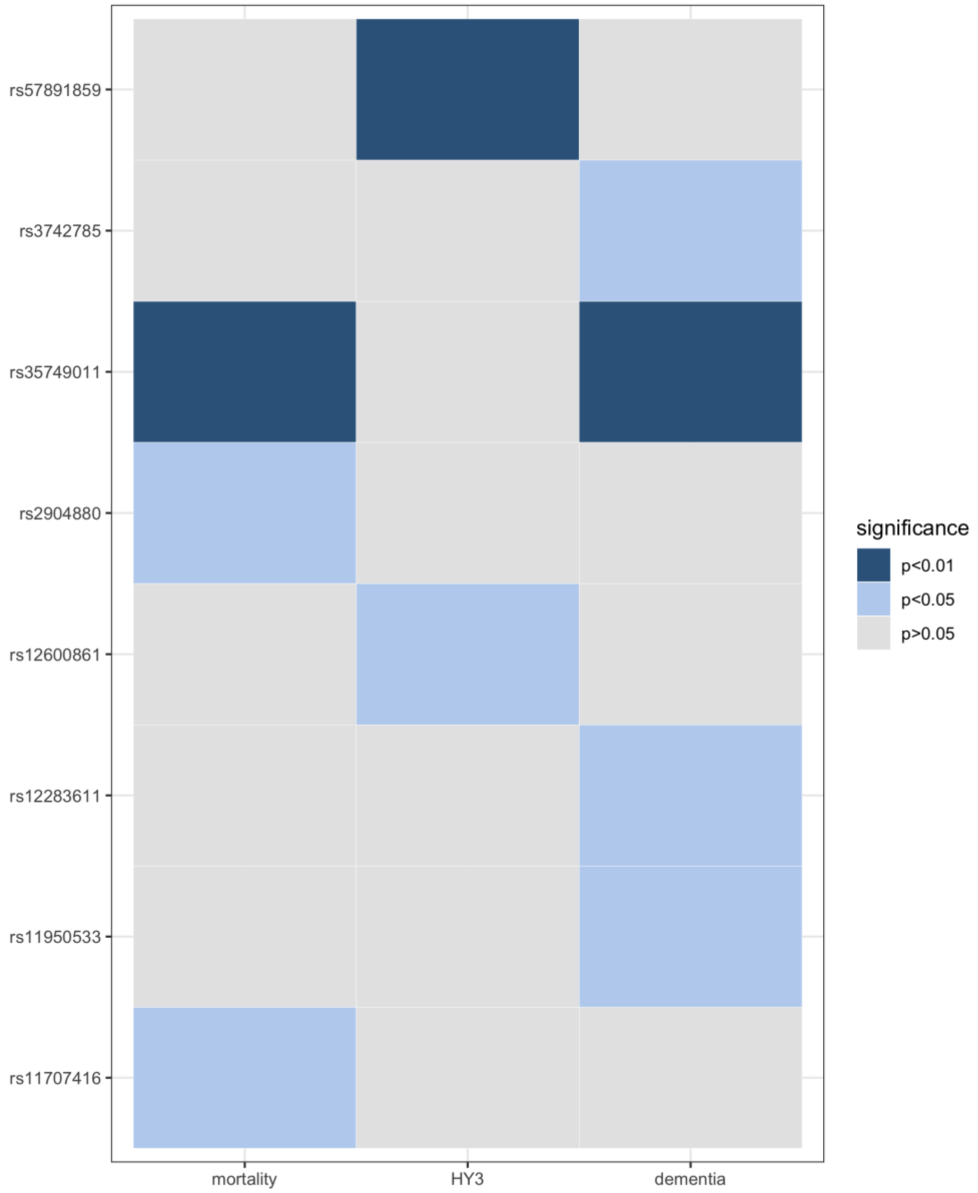
72 out of 90 PD risk variants were present in the final mortality meta-analysis dataset, after filtering, although 2 were not present in the HY3 meta-analysis and 4 were not present in the dementia meta-analysis. No variants passed genome-wide significance or analysis-wide significance (0.05/72). Variants that were nominally associated (p < 0.05) with either progression to mortality, HY3+, or dementia are shown in Figure 6.11.

I found that only a small number of risk variants were associated with progression with p-values < 0.05. rs35749011 was associated with both progression to mortality (beta = 0.40, p = 0.003) and dementia (beta = 0.41, p = 0.008), but not HY3+ (beta = 0.15, p = 0.37). This variant is in linkage disequilibrium with the *GBA* p.E326K variant, D’=0.90, R^2=0.78. rs57891859, an intronic variant in *TMEM163*, was nominally associated with progression to H&Y Stage 3 (beta = -0.20, p = 0.001).

I also examined candidate variants that have previously been associated with progression, including the *LRRK2* SNP rs2242367 associated with PSP mortality [269] (Table 6.10). The *LRRK2* variant rs34637584, as well as other rare *GBA* variants, were not covered in the analysis.

The *ATP8B2* gene identified from the motor progression PCA GWAS (Chapter 5) was not significantly associated with progression to any of the clinical milestones in these studies (all p’s > 0.1).

Figure 6.11. Heatmap of PD risk variants that were associated with progression to mortality, Hoehn and Yahr stage 3 or greater, or dementia.



HY3 = Hoehn and Yahr stage 3 or greater

Table 6.10. Results for candidate variants that have previously been reported for progression.Results were extracted from the meta-analysis results for mortality, Hoehn and Yahr stage 3 or greater, and dementia. Bonferroni correction was applied for the number of variants tested in each outcome (0.05/6 = 0.0083). Variants that passed this threshold are highlighted in bold.

Gene	**rsID**	**Effect**	**Ref**	**Association**	**Mortality**		**HY3+**		**Dementia**	
					beta	**pval**	**beta**	**pval**	**beta**	**pval**
LRRK2	rs2242367	A	G	PSP survival	**0.14**	**0.004**	0.03	0.64	0.01	0.80
LRRK2	rs76904798	T	C	PD risk HY3	-0.10	0.08	-0.03	0.65	-0.05	0.48
GBA	E326K rs2230288	T	C	PD risk PD progression	**0.39**	**0.005**	0.15	0.36	**0.43**	**0.005**
SLC44A1	rs382940	A	T	HY3	0.10	0.20	-0.06	0.55	0.06	0.50
MAPT	H1/H2 rs8070723	G	A	Dementia	0.004	0.94	0.06	0.34	-0.01	0.86
APOE	rs7412	T	C	Dementia (protective)	-0.11	0.17	0.10	0.27	0.01	0.91

HY3+ = Hoehn and Yahr stage 3 or greater

PD Genetic Risk Score

There was no association between the PD GRS and mortality in a meta-analysis across all cohorts, excluding PPMI (HR = 1.001, p = 0.96). In the analysis of Hoehn and Yahr 3 and dementia in Tracking Parkinson's, Oxford Discovery, and PPMI, there was also no association between the GRS and progression (p's > 0.1). However, I was able to replicate the effect of GRS on age at onset [96,259], with higher GRS associated with decreased age at onset (beta = -0.7, p = 0.0007).

AD Polygenic Risk Scores

To determine whether the association results were specific to *APOE* or more general AD risk, I analysed AD PRSs (excluding the *APOE* region) in relation to PD mortality and dementia. Table 6.11 shows the number of SNPs including in the PRSs at each p-value threshold for each cohort. The number of SNPs in the PRS differs between cohort as this is dependent on the overlapping SNPs between the AD GWAS summary statistics and each target genotype dataset. There was no association between the

AD PRS and mortality at any p-value threshold (Table 6.12). However, when the *APOE* region was included, the AD PRS was associated with mortality at p-value threshold 0.0001 (p = 0.003) and 0.001 (p = 3.4 x 10^{-5}).

Table 6.11. Number of SNPs including in each AD PRS (excluding the *APOE* region) in each cohort.

Selection threshold of SNPs in AD GWAS	**Calypso N SNPs**	**Oxford N SNPs**	**Tracking Parkinson's N SNPs**	**QSBB N SNPs**	**UKB N SNPs**	**PPMI N SNPs**
p < 0.0001	368	277	299	340	382	420
p < 0.001	1879	1308	1363	1647	2053	2238
p < 0.01	12468	7769	8095	10568	13857	15150
p < 0.05	45714	25330	26437	37699	51486	54856
p < 0.1	78167	40617	42555	63754	88502	92694
p < 0.2	129818	62252	65611	104883	147231	150608
p < 0.3	171301	77731	82252	137527	194389	196779
p < 0.4	205940	89737	95288	164163	233720	234354
p < 0.5	235172	99349	105600	186805	266550	265855

Table 6.12. Results of random-effects meta-analysis for AD PRS (excluding the *APOE* region) in relation to PD mortality across cohorts, excluding PPMI.

Selection threshold of SNPs in AD GWAS	**HR random effects**	**p-value**
p < 0.0001	1.01	0.62
p < 0.001	1.06	0.03
p < 0.01	1.04	0.14
p < 0.05	1.02	0.51
p < 0.1	1.03	0.28
p < 0.2	1.03	0.30
p < 0.3	1.05	0.18
p < 0.4	1.05	0.21
p < 0.5	1.05	0.22

There was also no association between the AD PRS excluding *APOE* and dementia at any p-value threshold (Table 6.13). When the *APOE* region was included, the AD PRS was still not associated with dementia. There was only a nominal association between the AD PRS including *APOE* at p-value threshold 0.0001 (p = 0.08) and 0.001 (p = 0.07).

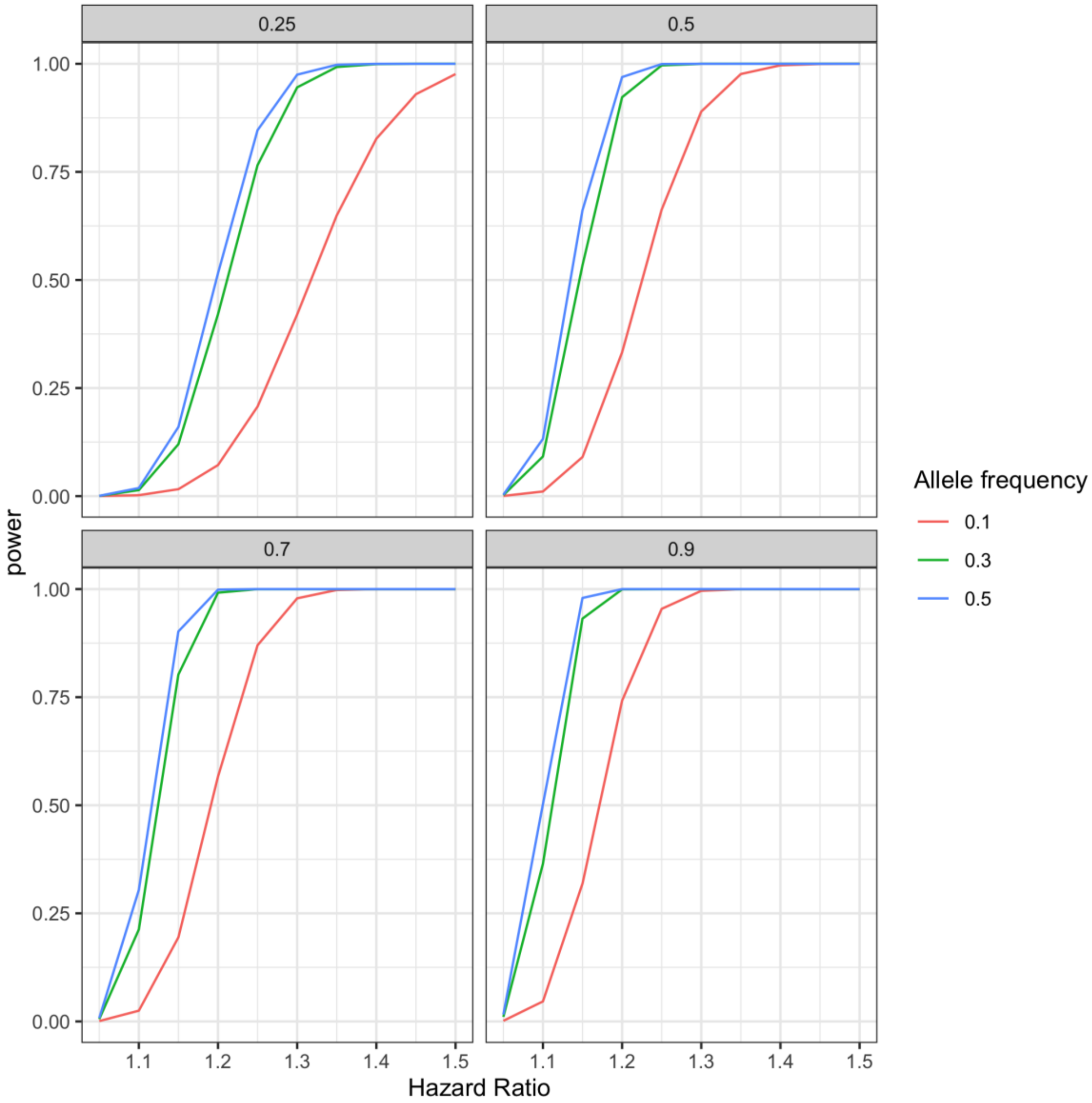
Table 6.13. Results of random-effects meta-analysis for AD PRS (excluding the *APOE* region) in relation to PD dementia in Tracking Parkinson's, Oxford, and PPMI.

Selection threshold of SNPs in AD GWAS	**HR random effects**	**p-value**
p < 0.0001	1.04	0.63
p < 0.001	1.05	0.34
p < 0.01	1.00	0.98
p < 0.05	1.01	0.82
p < 0.1	1.01	0.74
p < 0.2	1.00	0.97
p < 0.3	1.00	0.77
p < 0.4	1.00	0.69
p < 0.5	1.00	0.98

Power

The power to detect a signal in GWAS depends on a number of factors, including effect size, allele frequency of the effect allele, and the proportion of individuals meeting the outcomes. Using the survSNP package, I estimated that this study had 59.4% power to detect a significant effect ($p < 5 \times 10^{-8}$) for the *APOE* SNP rs429358, in the mortality GWAS, based on an allele frequency of 15%, Hazard Ratio of 1.35, and event rate of 25%. Figure 6.12 illustrates how power changes with different effect sizes, effect allele frequencies, and event rates. It is likely that longer follow-up of cohorts is needed to detect significant effects in mortality due to the low event rate in the current GWAS.

Figure 6.12. Power to detect a significant GWAS effect (alpha = 5 x 10^{-8}) at different effect sizes, allele frequencies, and event rates (in the grey headers, plots are faceted by event rate).



Discussion

I conducted GWASs of progression to clinical milestones, including the first large-scale GWAS of mortality in PD, as well as Hoehn and Yahr stage 3 or greater, and dementia.

There are a number of ways of measuring and analysing progression in PD, and there is no clear gold standard for assessing clinical progression. The MDS-UPDRS is commonly used in clinical trials, but this is subject to rater subjectivity and low within-subject reliability over longitudinal assessments [240]. Here, I used clinical milestones as measure of progression, which is a common method and has been used in previous candidate gene studies in PD [17,71,82,263,264], and progression GWAS studies in other diseases [281] and PD [98,259].

This study identified three main findings. Firstly, I found evidence that *APOE* ε4 is associated with progression to mortality and dementia in PD. Analysis of AD PRSs suggest that these results were specific to *APOE*, and not due to more general AD polygenic risk. Previous studies show that *APOE* is associated with AD risk [282], PD age at onset [96], and PD dementia [85,87,88,92], but not PD risk [75]. In the GWAS of PD age at onset, the effect of *APOE* on age/age at onset was similar in cases and controls, unlike other variants (*SNCA*, *TMEM175/GAK*) where the effect was only seen in PD cases [96]. This suggests that the effect of *APOE* is more generally related to aging, and not specific to PD. Indeed, GWASs of longevity have identified *APOE* as an important factor, with the ε4 allele found less frequently in long-living individuals [283,284]. *APOE* also increases risk of ischaemic heart disease/coronary artery disease, and cholesterol levels. It is likely that all these factors are interrelated – health conditions such as heart disease, high cholesterol, and AD increase risk for mortality, and dementia is a strong predictor of mortality in PD [216,285,286], and the general population [287–289].

My results are in line with these previous studies, as I show that *APOE* ε4 is associated with increased risk of mortality in PD. Considering the existing literature on *APOE*, it is likely that this effect is not PD-specific but more general. I found that the mean time from PD onset to death was 10.1 years, and mean follow-up time from PD onset to censoring for cases that did not die was 8.7 years. This shorter follow-up time may explain why I did not identify a genome-wide significant signal for the *APOE* locus, and longer follow-up is needed to confirm this result. The mean time from PD onset to death is longer than estimates from UK-based community samples (6.7 years to 8.3 years) [229,290]. This may be due to bias in clinical studies, including the UK Biobank cohort, which tend to attract generally healthier participants [291] and potentially slower progressing PD patients who are able to attend repeated assessments.

The mean time from PD onset to death in this study was also slightly shorter than that reported in a community-based study in Sweden (12.8 years) [292]. This may be because of longer life expectancy in Sweden compared to the UK, or the shorter follow-up time in this study, Another factor to consider is that the time of PD onset in the UK Biobank patients has been taken as the first record of PD in Hospital Episode Statistics, and this likely does not reflect the actual onset of symptoms.

Secondly, I identified a novel locus in Chromosome 10 in the long non-coding RNA *LOC105378484* which was nominally associated with progression to mortality. This locus is near to the protein-coding gene *ADRA2A*, Adrenoceptor Alpha 2A, although further work is needed to determine if our locus is involved in the regulation or expression of *ADRA2A.* I did not find evidence linking this locus to *ADRA2A* when looking at eQTL databases.

Another locus near this gene has previously been reported to be associated with insomnia at baseline in PD [98]. In that study, the lead SNP rs61863020 was a significant eQTL for *ADRA2A*.

None of the top variants in the Chromosome 10 locus in my GWAS of mortality were covered in the Iwaki meta-analysis of insomnia (either baseline or longitudinal survival analysis). They were also not in linkage disequilibrium with the SNP rs61863020 associated with baseline insomnia in the Iwaki GWAS (D' = 0.06, R^2 = 0.0001 for rs113423519 and rs61863020).

ADRA2A encodes for α2A adrenoceptors (or adrenergic receptors), which are involved in regulating the release of neurotransmitters from sympathetic nerves and from adrenergic neurons in the central nervous system. A study has shown that β2 adrenergic receptor activation reduced α-synuclein in neuronal cells [293]. In a series of experiments in human neuronal cells, rat neurons, and mice, Mittal et al. showed that β2-adrenoreceptor activation modulated endogenous *SNCA* expression and α-synuclein protein levels. β2-adrenoreceptor agonists, meta-proterenol, salbutamol, and clenbuterol, reduced endogenous *SNCA* mRNA. Antagonism of β2-adrenoreceptors using propranolol (a β-blocker) increased *SNCA* mRNA and α-synuclein protein levels. Furthermore, in a longitudinal population study in Norway, they found that salbutamol reduced the risk of developing PD, and this was not explained by smoking and treatment of asthma with salbutamol, while propranolol use increased the risk of developing PD [293].

At present there is limited evidence linking our Chromosome 10 locus to the *ADRA2A* gene so it is not known whether the rs61871952 variant or *LOC105378484* gene may be involved in *ADRA2A* regulation. No previous studies have been published on this. There was no data available for either *LOC105378484* or rs61871952 in GWAS catalogues including the NHGRI-EBI GWAS Catalog [294] (https://www.ebi.ac.uk/gwas/home), the HUNT fast-track GWAS catalog (https://www.ntnu.no/huntgenes/fasttrack), and the Global Biobank Engine. Further work and replication GWAS studies are needed to determine if this locus is important for PD progression and what role it may play.

It is possible that the variants I have identified for mortality are not PD-specific, but more general to neurodegeneration, aging, or even COVID-19. There appeared to be heterogeneity in the effect of the Chromosome 10 SNP, which had a stronger effect in the UK Biobank cohorts for which death data was downloaded very recently in 2020. No GWAS signals have been identified for COVID risk in Chromosome 10 or 19 [295]. There has been one locus identified in Chromosome 9 for COVID, top variant rs657152, but this was not near the Chromosome 9 locus associated with PD mortality (top variant rs3808753, D’ = 0.05, $R^2 = 0$). I also examined publicly available summary statistics for a COVID GWAS in European samples in UK Biobank (https://grasp.nhlbi.nih.gov/Covid19GWASResults.aspx). Interestingly, the *APOE* variant rs429358 was almost genome-wide significant ($p = 9.0 \times 10^{-7}$, beta = 0.27, N = 8486) in the UK Biobank COVID GWAS. The other two top signals in the PD mortality GWAS were not significant ($p > 0.1$). This suggests that *APOE* plays a role in both PD mortality and COVID risk, but this could be due to the role of *APOE* in aging or related diseases, e.g. if patients with AD or other dementias are overrepresented in COVID testing in care homes and hospitals.

The fact that only *APOE* was associated with COVID and not the other GWAS signals indicates that mortality in our PD cases is not due to COVID. If it was, I would expect more overlap between the genetic signals for COVID and mortality.

I did not replicate the finding for *SLC44A1* from the previous progression GWAS of survival to Hoehn and Yahr stage 3 or greater [98]. When I meta-analysed the current data together with summary statistics from the previous GWAS, there were no genome-wide significant signals. This may be due to the inclusion of different covariates in the survival models. Iwaki et al. used a data-driven approach and included quadratic age at diagnosis, quadratic years from diagnosis, education, H&Y 2 or more at baseline, and medication status as covariates, based on a backwards stepwise regression model in each cohort. This may explain the differences in our results. I did not include baseline scores as covariates in the survival models, as baseline performance may be a marker for progression rather than a confounder. Therefore, adjusting for baseline performance may mask true genetic associations with progression.

Replication is key for GWAS studies. It is clear from the GWAS studies in this thesis, and our lack of replication of previous findings, that the phenotypic measure of PD progression is a lot noisier and more variable than that in PD case-control studies. This difficulty means that more large, well-powered studies are needed to robustly replicate results, both the current findings and those from the previous PD progression GWAS.

I also identified a nominal signal in Chromome 9 at rs3808753 for progression to mortality. This is a non-coding transcript variant in *SH3GL2* (SH3 Domain Containing GRB2 Like 2, Endophilin A1). Another locus in this gene is associated with PD risk from case-control GWAS (top SNP rs10756907) [75], however the variants were not in linkage disequilibrium (D’ = 0.14, R^2 = 0.0002). *SH3GL2* is involved in regulation of synaptic vesicle endocytosis, and together with other PD-associated genes, *SYNJ1* and *DNAJC6*, highlights the endocytic membrane-trafficking pathway as important for the pathogenesis of PD [296,297]. The association of this gene with PD progression is interesting and needs to be replicated.

There was minimal overlap between the genetic variants associated with PD risk and PD progression. I showed that only a small number of PD risk loci were nominally associated with either progression to mortality, Hoehn and Yahr stage 3, or dementia. These results are similar to those from the previous PD progression GWAS, which showed minimal overlap between PD risk variants and progression apart from *GBA* variants [98,259]. The GRS was also not associated with any of the outcomes, though I replicated the association between GRS and PD age at onset. This could be due to two reasons: either the effect of the GRS on PD progression is much smaller than on age at onset and larger sample sizes are needed to detect this association, or that there is no association between the PD GRS and progression.

In the targeted candidate variant analysis, I found an association between the *LRRK2* variant rs2242367 and mortality. This variant was associated with PSP progression to mortality and is an eQTL for *LRRK2* [269]. This finding could potential contamination of PSP cases in the current study. Alternatively, it may suggest that the effect of this locus is not specific to PSP but influences progression and survival in other neurodegenerative conditions through the effect of *LRRK2* on mechanisms such as inflammatory response or tau pathology, which is present in some PD cases at postmortem [24].

Limitations

This study has several limitations which need to be recognised. Firstly, phenotyping of PD cases in the UK Biobank is different to other clinical cohorts. Incident PD cases were identified from HES. This data could be from unrelated hospital visits, but could also be linked to PD (e.g. falls) or PD progression (e.g. other comorbidities which contribute to rapid progression). This means the incident PD cohort from UK Biobank may be more rapidly progressing than the general PD population, and this is supported by the very short time to death in patients who died.

Secondly, this was not a population-based study and included several cohorts in which patients may not be representative of the general PD population (e.g. QSBB, UK Biobank). These cohorts may be more rapidly progressing or atypical than other PD patients.

A third limitation is incomplete death data from the clinical cohorts, Tracking Parkinson's and Oxford Discovery. Death was only recorded when provided by the site or the patient's relatives, not from health records, and there were many cases which were lost to follow-up from the study but with no known reason so I may be missing some cases who have died. Efforts to link these clinical cohorts to NHS death records are currently underway.

Fourthly, it is important to recognise that even mortality in PD is not a gold standard of progression. There are some patients who may live for a long time but with substantial impairment, and other PD patients who die from other causes. Mortality in more rapidly progressing diseases, such as PSP [269], are usually clearer as death can be attributed to disease, whereas in PD this phenotype can relate to multiple factors.

To try to clarify this issue, I performed analysis stratified by cause of death in the cohorts that had data available. These results suggest that the Chromosome 10 SNP and the *APOE* SNP were more important for PD deaths, and not interrupted deaths. However, this method of classifying deaths is fairly crude and based on small numbers of patients, and it is difficult to disentangle PD-related or unrelated causes of death. One mortality study found that PD patients had an increased risk of death from ischaemic heart disease, cerebrovascular disease, and other respiratory disease, compared to controls [229], although other studies have not found significant differences between cases and controls [290,292]. If PD patients are more at risk of death from causes which may seem unrelated (e.g. heart attacks), either due to drug treatments, shared causal factors, or immune response, then this stratification of cause of death may not be valid.

Finally, the study is limited by relatively small sample sizes. As shown in the power calculations, the power to detect significant effects is limited by low allele frequencies, small effect sizes, and low event rates (the proportion of individuals meeting the outcome). Particularly for GWASs of mortality, longer follow-up of cohorts is needed to increase power to detect effects. This may explain why I did not identify any significant loci in the mortality GWAS. In addition, it is likely that the heritability of PD progression is low, due to the subjective nature of measuring progression and noise in the phenotypes. This may also explain, in part, why I did not identify more GWAS significant loci. The PD age at onset GWAS only identified 3 genome-wide significant signals with a sample size of 28,586 patients [96]. Despite the limited power, it is still important to conduct these GWASs so that early data can be shared and over time and through collaborations, studies can be meta-analysed to increase sample numbers.

Conclusion

This study is the first GWAS of mortality in PD, and one of the largest GWASs of progression to other clinical milestones. I found that *APOE* is an important determinant of survival/mortality and dementia, and have identified other candidate loci which may be associated with mortality.

Chapter 7 : Conclusions and future directions

This research is crucial to understand the biology of disease progression in PD, which will facilitate the development of new disease modifying treatments. The majority of therapies in the pipeline for PD are for symptomatic treatments, and there is a lack of disease modifying therapies in clinical trials, particularly in Phase 3 studies [298]. Many drugs fail at Phase 3 trials [299,300]. This points to our incomplete understanding of the precise pathological mechanisms that drive disease progression and cell death [300].

The first aim of this PhD was to establish the frequency and clinical characteristics at baseline of PD Mendelian mutation carriers in the Tracking Parkinson's cohort. By using a variety of genetic screening methods, I showed that patients carrying pathogenic mutations are rare (<1%) and most are clinically indistinguishable from idiopathic PD. However, patients carrying *Parkin* or *PINK1* mutations appeared to have earlier onset, longer disease duration at study entry, and better cognition than other early-onset non-carriers, suggestive of slower progression. Further work to analyse the longitudinal progression of these mutation carriers is needed, though was not the focus of this PhD.

The second aim was to understand the clinical predictors of progression (Chapter 4). I showed that age at onset, baseline severity, and disease duration at study entry were associated with progression to clinical milestones, with more moderate evidence for gender. However, some of these may be on the causal pathway between genotype and the progression outcome, so arguably should not be included as covariates in GWAS models. It is likely that baseline severity and disease duration are both surrogate measures of rate of progression and this work implies that "malignant" PD is apparent at presentation, and that these patients should be targeted for disease modifying trials and for intensive multi-disciplinary support. The relationship between age and onset and progression is complex with both disease biology and co-morbidities (e.g. vascular risk factors) playing a role in disease progression, and increasing the risk of "incidental" mortality.

The third aim was to conduct GWASs on scores of motor, cognitive, and composite progression (Chapter 5). By combining multiple clinical scales to improve the phenotypic measure of progression, I showed that the *APOE* e4 allele was strongly associated with cognitive progression. This confirms the results of many previous candidate gene studies, but for the first time on a genome-wide scale. I also identified variation across *ATP8B2* as potentially important for motor progression, although this was not genome-wide significant and further studies are needed.

My fourth aim was to use survival to clinical milestones in PD as markers of progression in GWASs. This included the first large, well-powered GWAS of progression to mortality in PD. I showed that *APOE* was strongly associated with progression to dementia and potentially mortality. I also found sub-genome-wide significant loci in *LOC105378484* and *SH3GL2* which were associated with mortality. Further studies are needed to replicate these results.

Comparison of different GWAS approaches

In this PhD, I have used a range of different measures of PD progression and statistical approaches to analyse them. This speaks to the fact that there is no gold standard of measuring or analysing clinical progression in PD.

The fact that more novel approaches, such as the PCA GWAS, have identified similar findings to previous candidate gene studies, suggests that this approach is valid and can be used to improve the phenotypic measure of progression. This approach could be broadened further to test other scales for motor and cognitive progression. If the results are similar, then this could be a way of combining clinical data across studies which use different assessments.

Overall, I found similar results across the different approaches, e.g. the strong signal for *APOE* in all approaches. It is likely that the different measures of progression (PCA-derived composite scores, clinical milestones, rating scale change) are related and correlated. However each approach may measure slightly different aspects of clinical progression, e.g. the PCA-derived scores include variation in motor features such as tremor which are not included in the clinical milestone measures. Some measures, such as the PCA-derived scores, may be predictive of other markers such as mortality. In other measures, there is direct overlap between the scales that were used, e.g. the PCA-derived motor score and progression to Hoehn and Yahr stage 3.

Each measure likely has its own strengths and weaknesses. Survival analysis of clinical milestones may capture variation in patients who are more rapidly progressing and unable to fully complete detailed scales at follow-up visits. The PCA-derived composite scores may capture more variation in a range of different symptoms.

It would be interesting to look at the genetic correlations using LDSC between the different GWASs, once these progression GWASs access larger sample sizes and identify more genome-wide significant hits. This may give an indication of how much overlap there is between the genetic associations using different measures of progression.

Mechanisms of progression

The mechanisms by which genotypes influence progression remain unknown. It is well known that *APOE* is a strong risk factor for AD, and the likely mechanism is through modulation of amyloid-β accumulation [301]. Comorbid AD pathology is common in PD patients with dementia at postmortem [24]. Thus, *APOE* may influence PD progression solely through AD pathology. However, there may also be interactions between pathological substrates or alternative pathways through which *APOE* drives cognitive impairment.

Key studies in transgenic mice have shown that there are direct interactions between amyloid-β and α-synuclein, and they may have distinct as well as converging pathogenic effects [272]. Masliah et al. showed that mice expressing human α-synuclein in combination with human β-amyloid precursor protein (APP) had both motor and memory deficits, and these deficits seemed to be accelerated/worsened compared to mice expressing just α-synuclein or just amyloid-β [272]. They found that amyloid-β promotes accumulation of α-synuclein, but α-synuclein expression did not affect deposition of amyloid-β or neuritic plaques. Overall, they hypothesised that α-synuclein affected motor function more than cognitive function, whereas the reverse was true for amyloid-β. However other studies, albeit using different models, suggest this may not be the case as mice overexpressing α-synuclein without amyloid-β pathology still have learning and memory deficits [254].

Gallardo et al. found that overexpression of α-synuclein in transgenic mice induced neurodegeneration and motor symptoms, and altered the ubiquitin-proteasome system [302]. This was associated with increased levels of ApoE, insoluble amyloid-β peptides, and increased inflammatory response. Deletion of ApoE alleviated but did not completely abolish α-synuclein neurodegeneration [302]. They suggest that α-synuclein overexpression activates a pathogenic extracellular signalling loop that involves ApoE and amyloid-β and promotes neurodegeneration.

Most recently, Zhao et al. showed that in a mouse model overexpressing α-synuclein without amyloid pathology, human APOE4, but not APOE2 or APOE3, exacerbated α-synuclein pathology in cerebral cortex, hippocampus, amygdala, thalamus, but not the substantia nigra. APOE4 also enhanced motor and cognitive behavioural deficits, neuronal and synaptic loss, and astrogliosis [254]. Though this model does not account for how α-synuclein may interact with amyloid pathology, it suggests that APOE may have a direct effect on α-synuclein, independent of amyloid-β.

It is clear that further work needs to be done to investigate the mechanisms which drive clinical progression. Studying the correlation between pathology, progression, and genetic factors will be helpful. Animal studies have used different methods to model α-synuclein overexpression, and this makes it somewhat difficult to compare results. However, the evidence so far suggests that α-synuclein and amyloid-β can interact, but may also have independent effects on neurodegeneration, and that *APOE* may influence α-synuclein through other mechanisms.

Whether and how α-synuclein is important for clinical progression in PD also needs to be determined. There is substantial evidence showing that α-synuclein is important for the pathogenesis of PD [303]. Overproduction of synuclein through whole gene duplications and triplications, as well as specific mutations, is associated with the development of PD and sometimes a more rapidly progressing phenotype, though there is mixed evidence for patients carrying *SNCA* duplications. However, whether α-synuclein is important for clinical progression in PD needs to be determined, and if so which aspects of α-synuclein – whether overall burden, the spread to different brain regions, or downstream effects (e.g. inflammation, lysosomal dysfunction, dysfunction of other degredation pathways, mitochondrial dysfunction [303]). It is likely that clinical progression relates to a number of these factors rather than a single factor. In addition, different domains of progression may be more affected by particular pathways.

Cognitive impairment and dementia in PD is associated with the cholinergic system [304], in addition to α-synuclein, tau, and amyloid pathologies [24,29,305].

In my GWASs, I did not find *SNCA* variants were associated with PD progression. However, this may be due to a lack of power, especially if common variants are associated with smaller increases in expression. In addition, other genes may modulate α-synuclein through regulatory networks to affect core genes, in an omnigenic model [306]. Furthermore, it may suggest that other factors are important for clinical progression, such as the spread of α-synuclein pathology to other regions of the brain, rather than expression or overall burden.

Recent studies suggest that CSF α-synuclein may track with clinical progression and could be used as a potential biomarker – one study reported a correlation between the ratio of total and oligomeric α-synuclein and UPDRS motor change [307]. Although this research field is fairly new and some studies have reported conflicting results [308], it could suggest that α-synuclein pathology is dynamic during the course of disease and could be important for predicting and tracking clinical progression. In addition, other pathological substrates including tau have been associated with motor progression in small studies [308].

General limitations and considerations

There are a number of limitations to be aware of in these studies. Firstly, GWASs can only detect variation in common SNPs, and there may be rare variants or larger structural variants which also contribute to variation in PD progression – possibly with larger effect sizes. GWASs can identify common variation that tags causal rare variants. Future studies integrating rare variant data from methods such as whole genome sequencing would help to address this issue.

Identifying the causal variant from GWAS loci is another challenge. Only a small number of SNPs are genotyped and imputed, and further fine mapping is needed to cover other variants [120]. In addition, it is helpful to assess evidence to identify causal variants from GWAS loci, including functional, expression, and rare variant burden evidence. These efforts are currently underway for PD through the International PD Genomics Consortium (IPDGC) and online tools such as the GWAS Locus Browser (https://pdgenetics.shinyapps.io/GWASBrowser/).

Another key consideration for these studies is power and sample size. These are some of the largest GWASs for PD progression – the only other well-powered PD progression GWASs have been conducted by Iwaki et al. in 4,093 PD patients [98], although in that study not all patients had data available for all clinical outcomes. The sample sizes in both studies are relatively small for GWASs and are therefore limited to detect variants that have smaller effects or are less common. It has been a challenge to collect and put together cohorts with longitudinal, detailed clinical data in PD but new initiatives such as the Global Parkinson's Genetics Program (GP2) and sharing of publicly available data (AMP-PD, UK Biobank) will make this easier in the future. With these initiatives, it will be possible to conduct larger progression GWASs in PD. I believe there are more loci and genes to discover for PD progression, although likely not to the scale as PD risk, due to the lower heritability and more complex phenotypes. The GWAS of PD age at onset only identified 3 significant signals with over 28,500 PD cases and this is because of the lower heritability of the phenotype [96].

Another consideration is the possibility of collider bias when studying a selected group of individuals, i.e. case-only studies. This has been flagged as an issue in Mendelian Randomisation studies of progression [309]. If there are multiple independent factors that influence disease risk, this can introduce spurious associations when only cases with disease have been selected for progression GWASs. This is also termed as index event bias by Dudbridge et al. [310]. There are methods to adjust for index event bias, and this could be tested in future PD progression studies. However, these methods also assume that genotype effects on prognosis are independent of disease risk [310], and this may not be in the case in PD, for instance with *GBA* variants.

Misdiagnosis of patients is another factor to consider, and may contribute to weaker signals in my GWASs. As discussed earlier, it is likely that with the length of follow-up in the cohort studies, the majority of non-PD patients have been excluded and I performed sensitivity analyses to exclude participants who may have different conditions (Chapter 5). However, this also depends on the frequency of follow-up and whether there is a clear system of notifying study coordinators of change in diagnosis, particularly for patients who are unable to attend study visits. Future cohort studies could include light-touch assessments over the phone or online to collect key data points (including change in diagnosis, clinical milestones, and death) for patients who cannot attend in-person assessments. In addition, it would be useful to link clinical cohorts to death data and encourage registration with brain banks, so that post-mortem confirmation and death cause data can be collected systematically.

A further limitation is that all these studies are done in European and Caucasian populations. Therefore, these results cannot be extrapolated to other populations. There may be different allele frequencies, LD structures, causal variants, and effect sizes in other populations, and this has been shown in PD case-control GWASs in different populations [311]. Studies of PD progression are needed in other populations, and I hope this will be facilitated by the GP2 initiative where a key focus is on underrepresented populations.

Future work

A crucial aspect of future studies is to look at the genetic associations with pathology in addition to clinical progression. As discussed throughout this thesis, a major limitation of this research is the lack of a gold standard/marker for measuring and analysing PD progression. It is possible that biomarkers of pathology are more accurate markers of disease progression. I know that studies are already underway to look at the associations between pathological burden, clinical progression, and genotype in post-mortem brain bank data and this will reveal important insights. Building and analysing longitudinal cohorts with detailed genotyping and sequencing, biomarkers, and post-mortem data (such as PPMI) will be essential to study the relationships between genotype, biomarkers, pathology, and clinical progression.

With the advances in wearable technology, more accurate measures of clinical progression could also be analysed, by incorporating more fine-grained data at more frequent timepoints (i.e. not just at clinic visits which are typically assessed in the ‘on’ medication state). This could also help to include more rapidly progressing patients who are likely to drop out or not join intensive research studies with frequent in-person assessments.

Some of my results suggest that the genetic factors that influence progression may not be PD specific, and could be due to more general neurodegeneration or aging, e.g. *APOE*. Further studies should conduct GWASs with progression across a range of neurodegenerative diseases as well as healthy controls, to determine if the effects of genotype are different across disease groups or if they have the same effect. I would expect that genotype effects on progression are stronger in PD and other neurodegenerative diseases compared to healthy aging, possibly because of interactions between pathologies and comorbidities.

With larger sample sizes in progression GWASs, it would be interesting to look further into sex differences. There are clear differences in clinical progression between men and women, as I showed in Chapter 4. In this PhD, I have conducted preliminary sex-stratified analyses and not identified GWAS significant hits. This is likely to be due to lack of power, and it is too early to conclude that there are no sex differences in the genetics of PD progression. The PD age at onset GWAS also did not identify any significant differences in sex-stratified analysis, but highlighted the *COMT* variant rs4680 which has a different direction of effect in men and women [96]. Early studies in other diseases, such as AD [312], suggest that there may be sex differences in the genetics of progression.

The next steps following large-scale, well-powered PD progression GWASs is to conduct functional studies in cell and animal models to understand the mechanisms of candidate genes and pathways that are important for PD progression. In addition, the results from genetic studies could be used for prediction of progression and better stratification of clinical trials. For instance, individual variants or a cumulative PD progression risk score, similar to the PD GRS, could be used in combination with clinical and demographic variables to predict individual patient trajectories and better detect the effect of therapies on predicted progression [313]. Although the effects of individual genetic variants are small, simulations have shown that even if patients are randomly assigned to trial arms, there can be large differences in the GRS, particularly with small sample sizes (< 1000) [314]. Genetically-mismatched arms in clinical trials can have a large effect, leading to 34% of false negatives in a simulated drug effect [314]. Identification of genetic variants that influence PD progression can be used to balance clinical trial arms and predicted progression.

In addition, if there are variants and genes with large effects on progression, these could be used to design targeted clinical trials. Recent examples of this are the trial of Ambroxol which was targeted towards lysosomal function and *GBA* carriers [315]. The trial of EPI-589 is another example. PD patients carrying *Parkin* and *PINK1* mutations, as well as other genetic and sporadic forms of PD, are known to have mitochondrial dysfunction [316]. This compound has been repurposed from childhood mitochondrial diseases in an attempt to target mitochondrial dysfunction and oxidative stress in PD.

There is clearly a lot of further research that needs to be done in this field. In this PhD, I have conducted some of the earliest large-scale GWASs of PD progression and I hope this work can be used as the starting point for further research, and eventually clinical trials of disease modifying therapies.

Chapter 8 : References

1. Parkinson J. “An essay on the shaking palsy. London: Sherwood, Neely, and Jones; 1817.
2. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson’s disease : a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry. 1992;55:181–4.
3. Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, et al. MDS clinical diagnostic criteria for Parkinson’s disease. Mov Disord. 2015;30(12):1591–601.
4. Spillantini MG, Schmidt ML, Lee VM-Y, Trojanowski JQ. alpha-Synuclein in Lewy bodies. Nature. 1997;388:839–40.
5. Lees AJ, Hardy J, Revesz T. Parkinson’s disease. Lancet [Internet]. 2009 Jun [cited 2011 Jan 13];373(9680):2055–66. Available from: http://www.ncbi.nlm.nih.gov/pubmed/19524782
6. Braak H, Del Tredici K, Rüb U, de Vos RAI, Steur ENHJ, Braak E. Staging of brain pathology related to sporadic Parkinson’s disease. Neurobiol Aging. 2003;24:197–211.
7. Visanji NP, Brooks PL, Hazrati LN, Lang AE. The prion hypothesis in Parkinson’s disease: Braak to the future. Acta Neuropathol Commun. 2014;2(1):1–12.
8. Kalia L V, Lang AE. Parkinson’s disease. Lancet. 2015;386(9996):896–912.
9. Pringsheim T, Jette N, Frolkis A, Steeves TDL. The prevalence of Parkinson’s disease: a systematic review and meta-analysis. Mov Disord [Internet]. 2014;29(13):1583–90. Available from: http://www.ncbi.nlm.nih.gov/pubmed/24976103
10. Schrag A, Ben-Shlomo Y, Quinn NP. Cross sectional prevalence survey of idiopathic Parkinson’s disease and parkinsonism in London. Bmj. 2000;321(July):21–2.
11. Dorsey ER, Bloem BR. The Parkinson pandemic - A call to action. JAMA Neurol. 2018;75(1):9–10.
12. Schrag A, Horsfall L, Walters K, Noyce A, Petersen I. Prediagnostic presentations of Parkinson’s disease in primary care: A case-control study. Lancet Neurol [Internet]. 2015;14(1):57–64. Available from:

http://dx.doi.org/10.1016/S1474-4422(14)70287-X

13. Hely MA, Morris JGL, Reid WGJ, O'Sullivan DJ, Williamson PM, Rail D, et al. The Sydney Multicentre Study of Parkinson's disease: A randomised, prospective five year study comparing low dose bromocriptine with low dose levodopa-carbidopa. J Neurol Neurosurg Psychiatry. 1994;57(8):903–10.

14. Hely MA, Morris JGL, Traficante R, Reid WGJ, O'Sullivan DJ, Williamson PM. The Sydney multicentre study of Parkinson's disease: progression and mortality at 10 years. J Neurol Neurosurg Psychiatry. 1999;67:300–7.

15. Hely M a, Morris JGL, Reid WGJ, Trafficante R. Sydney Multicenter Study of Parkinson's disease: non-L-dopa-responsive problems dominate at 15 years. Mov Disord [Internet]. 2005 Feb [cited 2012 Apr 18];20(2):190–9. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15551331

16. Hely M a, Reid WGJ, Adena M a, Halliday GM, Morris JGL. The Sydney multicenter study of Parkinson's disease: the inevitability of dementia at 20 years. Mov Disord [Internet]. 2008 Apr [cited 2010 Jun 28];23(6):837–44. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18307261

17. Williams-Gray CH, Mason SL, Evans JR, Foltynie T, Brayne C, Robbins TW, et al. The CamPaIGN study of Parkinson's disease: 10-year outlook in an incident population-based cohort. J Neurol Neurosurg Psychiatry [Internet]. 2013;84:1258–64. Available from: http://www.ncbi.nlm.nih.gov/pubmed/23781007

18. Lewis SJG, Foltynie T, Blackwell AD, Robbins TW, Owen AM, Barker RA. Heterogeneity of Parkinson's disease in the early clinical stages using a data driven approach. J Neurol Neurosurg Psychiatry. 2005;76(January 2000):343–8.

19. Fereshtehnejad SM, Zeighami Y, Dagher A, Postuma RB. Clinical criteria for subtyping Parkinson's disease: Biomarkers and longitudinal progression. Brain. 2017;140(7):1959–76.

20. Lawton M, Ben-Shlomo Y, May MT, Baig F, Barber TR, Klein JC, et al. Developing and validating Parkinson's disease subtypes and their motor and cognitive progression. J Neurol Neurosurg Psychiatry [Internet]. 2018;89(12). Available from: http://jnnp.bmj.com/content/89/12/1279.full.pdf

21. Braak H, Rüb U, Jansen Steur ENH, Del Tredici K, De Vos RAI. Cognitive status correlates with neuropathologic stage in Parkinson disease. Neurology.

2005;64(8):1404–10.

22. Apaydin H, Ahlskog JE, Parisi JE, Boeve BF, Dickson DW. Parkinson Disease Neuropathology. Arch Neurol. 2002;59(1):102–12.

23. Hurtig HI, Trojanowski JQ, Galvin J, Ewbank D, Schmidt ML, Lee VM-Y, et al. Alpha-synuclein cortical Lewy bodies correlate with dementia in Parkinson's disease. Neurology. 2000;54:1916–21.

24. Smith C, Malek N, Grosset K, Cullen B, Gentleman S, Grosset DG. Neuropathology of dementia in patients with Parkinson's disease: A systematic review of autopsy studies. J Neurol Neurosurg Psychiatry. 2019;90(11):1234–43.

25. Colosimo C, Hughes AJ, Kilford L, Lees AJ. Lewy body cortical involvement may not always predict dementia in Parkinson's disease. J Neurol Neurosurg Psychiatry. 2003;74(7):852–6.

26. Parkkinen L, Kauppinen T, Pirttilä T, Autere JM, Alafuzoff I. A-Synuclein Pathology Does Not Predict Extrapyramidal Symptoms or Dementia. Ann Neurol. 2005;57(1):82–91.

27. Halliday G, Hely M, Reid W, Morris J. The progression of pathology in longitudinally followed patients with Parkinson's disease. Acta Neuropathol. 2008;115(4):409–15.

28. Kalaitzakis ME, Graeber MB, Gentleman SM, Pearce RK. Striatal beta-Amyloid Deposition in Parkinson Disease With Dementia. J Neuropatholy Exp Neurol. 2008;67(2):155–61.

29. Aarsland D, Creese B, Politis M, Chaudhuri KR, Ffytche DH, Weintraub D, et al. Cognitive decline in Parkinson disease. Nat Rev Neurol. 2017;13(4):217–31.

30. Lashley T, Holton JL, Gray E, Kirkham K, O'Sullivan SS, Hilbig A, et al. Cortical α-synuclein load is associated with amyloid-β plaque burden in a subset of Parkinson's disease patients. Acta Neuropathol. 2008;115(4):417–25.

31. Goetz CG, Fahn S, Martinez-Martin P, Poewe W, Sampaio C, Stebbins GT, et al. Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): Process, format, and clinimetric testing plan. Mov Disord. 2007;22(1):41–7.

32. Schrag A, Spottke A, Quinn NP, Dodel R. Comparative responsiveness of Parkinson's disease scales to change over time. Mov Disord. 2009;24(6):813–8.

33. Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson’s disease. Lancet Neurol [Internet]. 2020;19(2):170–8. Available from: http://dx.doi.org/10.1016/S1474-4422(19)30287-X

34. Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, et al. Alpha-Synuclein Locus Triplication Causes Parkinson’s Disease. Science (80-) [Internet]. 2003;302(5646):841–841. Available from: http://www.sciencemag.org/cgi/doi/10.1126/science.1090278

35. Gloeckner CJ, Kinkl N, Schumacher A, Braun RJ, Neill EO, Meitinger T, et al. The Parkinson disease causing LRRK2 mutation I2020T is associated with increased kinase activity. Hum Mol Genet. 2006;15(2):223–32.

36. West AB, Moore DJ, Biskup S, Bugayenko A, Smith WW, Ross CA, et al. Parkinson ’ s disease-associated mutations in leucine-rich repeat kinase 2 augment kinase activity. Proc Natl Acad Sci U S A. 2005;102(46):16842–7.

37. Taymans J-M, Greggio E. LRRK2 Kinase Inhibition as a Therapeutic Strategy for Parkinson’s Disease, Where Do We Stand? Curr Neuropharmacol. 2016 Apr;14(3):214–25.

38. Fuchs J, Nilsson C, Kachergus J, Munz M, Larsson EM, Schüle B, et al. Phenotypic variation in a large Swedish pedigree due to SNCA duplication and triplication. Neurology. 2007;68(12):916–22.

39. Ahn T-B, Kim SY, Kim JY, Park S-S, Lee DS, Min HJ, et al. Alpha-synuclein gene duplication is present in sporadic Parkinson disease. Neurology [Internet]. 2008;71(16):1294; author reply 1294. Available from: http://www.ncbi.nlm.nih.gov/pubmed/18852448

40. Farrer M, Kachergus J, Forno L, Lincoln S, Wang DS, Hulihan M, et al. Comparison of Kindreds with Parkinsonism and α-Synuclein Genomic Multiplications. Ann Neurol. 2004;55(2):174–9.

41. Muenter MD, Forno LS, Hornykiewicz O, Kish SJ, Maraganore DM, Caselli RJ, et al. Hereditary form of parkinsonism-dementia. Ann Neurol. 1998;43(6):768–81.

42. Alcalay RN, Mirelman A, Saunders-Pullman R, Tang MX, Mejia Santana H, Raymond D, et al. Parkinson disease phenotype in Ashkenazi jews with and without LRRK2 G2019S mutations. Mov Disord. 2013;28(14):1966–71.

43. Healy DG, Falchi M, O’Sullivan SS, Bonifati V, Durr A, Bressman S, et al. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated

Parkinson's disease: a case-control study. Lancet Neurol. 2008;7(7):583–90.

44. Saunders-Pullman R, Mirelman A, Alcalay RN, Wang C, Ortega RA, Raymond D, et al. Progression in the LRRK2-Asssociated Parkinson disease population. JAMA Neurol. 2018;75(3):312–9.

45. Yahalom G, Orlev Y, Cohen OS, Kozlova E, Friedman E, Inzelberg R, et al. Motor progression of Parkinson's disease with the leucine-rich repeat kinase 2 G2019S mutation. Mov Disord. 2014;29(8):1057–60.

46. Dodson MW, Guo M. Pink1 , Parkin , DJ-1 and mitochondrial dysfunction in Parkinson ' s disease. Curr Opin Neuroiology. 2007;17:331–7.

47. Clark IE, Dodson MW, Jiang C, Cao JH, Huh JR, Seol JH, et al. Drosophila pink1 is required for mitochondrial function and interacts genetically with parkin. Nature. 2006;441:1162–1166.

48. Pickrell AM, Youle RJ. The Roles of PINK1 , Parkin , and Mitochondrial Fidelity in Parkinson's Disease. Neuron [Internet]. 2015;85(2):257–73. Available from: http://dx.doi.org/10.1016/j.neuron.2014.12.007

49. Alcalay RN, Caccappolo E, Mejia-Santana H, Tang MX, Rosado L, Orbe Reilly M, et al. Cognitive and motor function in long-duration PARKIN-associated Parkinson disease. JAMA Neurol [Internet]. 2014;71(1):62–7. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3947132&tool=pmcentrez&rendertype=abstract

50. Bonifati V, Rohe CF, Breedveld GJ, Fabrizio E, Mari M De, Tassorelli C, et al. Early-onset parkinsonism associated with PINK1 mutations: frequency, genotypes, and phenotypes. Neurology [Internet]. 2005;65(1):87–95. Available from: http://www.ncbi.nlm.nih.gov/pubmed/16606941

51. Kasten M, Hartmann C, Hampf J, Schaake S, Westenberger A, Vollstedt EJ, et al. Genotype-Phenotype Relations for the Parkinson's Disease Genes Parkin, PINK1, DJ1: MDSGene Systematic Review. Mov Disord. 2018;33(5):730–41.

52. Khan NL, Graham E, Critchley P, Schrag AE, Wood NW, Lees AJ, et al. Parkin disease: A phenotypic study of a large case series. Brain. 2003;126(6):1279–92.

53. Lohmann E, Periquet M, Bonifati V, Wood NW, De Michele G, Bonnet AM, et al. How much phenotypic variation can be attributed to parkin genotype? Ann Neurol. 2003;54(2):176–85.

54. Lohmann E, Dursun B, Lesage S, Hanagasi HA, Sevinc G, Honore A, et al.

Genetic bases and phenotypes of autosomal recessive Parkinson disease in a Turkish population. Eur J Neurol. 2012;19(5):769–75.

55. Lücking CB, Dürr A, Bonifati V, Vaughan J, De Michele G, Gasser T, et al. Association between early-onset Parkinson's disease and Mutations in the Parkin Gene. N Engl J Med. 2000;342:1560–7.

56. Tan EK, Yew K, Chua E, Puvan K, Shen H, Lee E, et al. PINK1 mutations in sporadic early-onset Parkinson's Disease. Mov Disord. 2006;21(6):789–93.

57. Valente EM, Bentivoglio AR, Dixon PH, Ferraris A, Ialongo T, Frontali M, et al. Localization of a Novel Locus for Autosomal Recessive Early-Onset Parkinsonism, PARK6, on Human Chromosome 1p35-p36. Am J Hum Genet. 2001;68(4):895–900.

58. Sidransky E, Nalls MA, Aasly JO, Aharon-Peretz J, Annesi G, Barbosa ER, et al. Multicenter Analysis of Glucocerebrosidase Mutations in Parkinson's Disease. N Engl J Med [Internet]. 2009;361(17):1651–61. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMoa0901281

59. Aharon-Peretz J, Rosenbaum H, Gershoni-Baruch R. The glucocerebrosidase gene and Parkinson's disease in Ashkenazi Jews. N Engl J Med [Internet]. 2004;351(19):1972–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15719452

60. Neumann J, Bras J, Deas E, O'sullivan SS, Parkkinen L, Lachmann RH, et al. Glucocerebrosidase mutations in clinical and pathologically proven Parkinson's disease. Brain. 2009;132(7):1783–94.

61. Toft M, Pielsticker L, Ross OA, Aasly JO, Farrer MJ. Glucocerebrosidase gene mutations and Parkinson disease in the Norwegian. Neurology. 2006;66:415–7.

62. Riboldi GM, Fonzo AB Di. GBA, Gaucher Disease, and Parkinson's Disease: From Genetic to Clinic to New Therapeutic Approaches. Cells. 2019;8(4):364.

63. Alcalay RN, Caccappolo E, Mejia-Santana H, Tang M-X, Rosado L, Orbe Reilly M, et al. Cognitive performance of GBA mutation carriers with early-onset PD: the CORE-PD study. Neurology [Internet]. 2012;78:1434–40. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3345785&tool=pmcentrez&rendertype=abstract

64. Brockmann K, Srulijes K, Pflederer S, Hauser AK, Schulte C, Maetzler W, et al. GBA-associated Parkinson's disease: Reduced survival and more rapid progression in a prospective longitudinal study. Mov Disord. 2015;30(3):407–

11.

65. Brockmann K, Srulijes K, Hauser a. K, Schulte C, Csoti I, Gasser T, et al. GBA-associated PD presents with nonmotor characteristics. Neurology. 2011;77:276–80.

66. Cilia R, Tunesi S, Marotta G, Cereda E, Tesei S, Zecchinelli A, et al. Survival and dementia in GBA -associated Parkinson Disease : the mutation matters . Ann Neurol. 2016;80:662–73.

67. Clark LN, Ross BM, Wang Y, Mejia-Santana H, Harris J, Louis ED, et al. Mutations in the glucocerebrosidase gene are associated with early-onset Parkinson disease. Neurology. 2007;69(12):1270–7.

68. Crosiers D, Verstraeten A, Wauters E, Engelborghs S, Peeters K, Mattheijssens M, et al. Mutations in glucocerebrosidase are a major genetic risk factor for Parkinson's disease and increase susceptibility to dementia in a Flanders-Belgian cohort. Neurosci Lett [Internet]. 2016;629:160–4. Available from: http://dx.doi.org/10.1016/j.neulet.2016.07.008

69. Davis AA, Andruska KM, Benitez B a., Racette B a., Perlmutter JS, Cruchaga C. Variants in GBA, SNCA, and MAPT Influence Parkinson Disease Risk, Age at Onset, and Progression. Neurobiol Aging [Internet]. 2016;37:209.e1-209.e7. Available from: http://linkinghub.elsevier.com/retrieve/pii/S0197458015004716

70. Gan-Or Z, Giladi N, Orr-Urtreger A. Differential phenotype in Parkinson's disease patients with severe versus mild GBA mutations. Brain. 2009;132(10):2009.

71. Winder-Rhodes SE, Evans JR, Ban M, Mason SL, Williams-Gray CH, Foltynie T, et al. Glucocerebrosidase mutations influence the natural history of Parkinson's disease in a community-based incident cohort. Brain. 2013;136(2):392–9.

72. Di Maio R, Hoffman EK, Rocha EM, Keeney MT, Sanders LH, De Miranda BR, et al. LRRK2 activation in idiopathic Parkinson's disease. Sci Transl Med. 2018;10(451):1–13.

73. Reed X, BandréS-Ciga S, Blauwendraat C, Cookson MR. The role of monogenic genes in idiopathic Parkinson's disease. Neurobiol Dis. 2018;

74. Schork NJ, Murray SS, Frazer KA, Topol EJ. Common vs. Rare Allele Hypotheses for Complex Diseases. Curr Opin Genet Dev. 2009;19(3):212–9.

75. Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D,

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(12):1091–102.

76. Chang D, Nalls MA, Hallgrímsdóttir IB, Hunkapiller J, van der Brug M, Cai F, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson’s disease risk loci. Nat Genet [Internet]. 2017;49(10):1511–6. Available from: http://www.nature.com/doifinder/10.1038/ng.3955

77. Edwards TL, Scott WK, Almonte C, Burt A, Powell EH, Beecham GW, et al. Genome-Wide association study confirms SNPs in SNCA and the MAPT region as common risk factors for parkinson disease. Ann Hum Genet. 2010;74(2):97–109.

78. International Parkinson’s Disease Genomics Consortium. Imputation of sequence variants for identification of genetic risks for Parkinson’s disease: A meta-analysis of genome-wide association studies. Lancet [Internet]. 2011;377(9766):641–9. Available from: http://dx.doi.org/10.1016/S0140-6736(10)62345-8

79. Simón-Sánchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, et al. Genome-wide association study reveals genetic risk underlying Parkinson’s disease. Nat Genet [Internet]. 2009;41(12):1308–12. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2787725&tool=pmcentrez&rendertype=abstract

80. The UK Parkinson’s Disease Consortium, The Wellcome Trust Case Control Consortium 2. Dissection of the genetics of Parkinson’s disease identifies an additional association 5’ of SNCA and multiple associated haplotypes at 17q21. Hum Mol Genet. 2011;20(2):345–53.

81. Singleton AB, Farrer MJ, Bonifati V. The Genetics of Parkinson’s Disease: Progress and Therapeutic Implications Monogenic Loci. Mov Disord. 2013;28(1):14–23.

82. Evans JR, Mason SL, Williams-Gray CH, Foltynie T, Brayne C, Robbins TW, et al. The natural history of treated Parkinson’s disease in an incident, community based cohort. J Neurol Neurosurg Psychiatry. 2011;82(10):1112–8.

83. Goris A, Williams-Gray CH, Clark GR, Foltynie T, Lewis SJG, Brown J, et al. Tau and alpha-synuclein in susceptibility to, and dementia in, Parkinson’s disease. Ann Neurol [Internet]. 2007;62(2):145–53. Available from:

http://www.ncbi.nlm.nih.gov/pubmed/17683088

84. Williams-Gray CH, Evans JR, Goris A, Foltynie T, Ban M, Robbins TW, et al. The distinct cognitive syndromes of Parkinson's disease: 5 year follow-up of the CamPaIGN cohort. Brain. 2009;132(11):2958–69.

85. Nombela C, Rowe JB, Winder-Rhodes SE, Hampshire A, Owen AM, Breen DP, et al. Genetic impact on cognition and brain function in newly diagnosed Parkinson's disease: ICICLE-PD study. Brain. 2014;137(10):2743–58.

86. Setó-Salvia N, Clarimón J, Pagonabarraga J, Pascual-Sedano B, Campolongo A, Combarros O, et al. Dementia Risk in Parkinson Disease: Disentangling the Role of MAPT Haplotypes. Arch Neurol [Internet]. 2011;68(3):359–64. Available from: http://search.proquest.com/docview/858144943/

87. Mata IF, Leverenz JB, Weintraub D, Trojanowski JQ, Hurtig HI, Van Deerlin VM, et al. APOE, MAPT, and SNCA genes and cognitive performance in Parkinson disease. JAMA Neurol. 2014;71(11).

88. Morley JF, Xie SX, Hurtig HI, Stern MB, Colcher A, Horn S, et al. Genetic influences on cognitive decline in Parkinson's disease. Mov Disord. 2012;27(4):512–8.

89. Paul KC, Rausch R, Creek MM, Sinsheimer JS, Bronstein JM, Bordelon Y, et al. APOE, MAPT, and COMT and Disease Susceptibility and Cognitive Symptom Progression. J Parkinsons Dis. 2016;6:349–59.

90. Huang X, Chen P, Kaufer DI, Troster AI, Poole C. Apolipoprotein E and Dementia in Parkinson Disease. Arch Neurol. 2006;63(2):189–93.

91. Kurz MW, Dekomien G, Nilsen OB, Larsen JP, Aarsland D, Alves G. APOE Alleles in Parkinson Disease and Their Relationship to Cognitive Decline: A Population-based, Longitudinal Study. J Geriatr Psychiatry Neurol. 2009;22(3):166–70.

92. Williams-Gray CH, Goris A, Saiki M, Foltynie T, Compston DAS, Sawcer SJ, et al. Apolipoprotein e genotype as a risk factor for susceptibility to and dementia in Parkinson's Disease. J Neurol. 2009;256(3):493–8.

93. Latourelle JC, Beste MT, Hadzi TC, Miller RE, Oppenheim JN, Valko MP, et al. Large-scale identification of clinical and genetic predictors of motor progression in patients with newly diagnosed Parkinson's disease: a longitudinal cohort study and validation. Lancet Neurol [Internet]. 2017;16(11):908–16. Available from: http://dx.doi.org/10.1016/S1474-4422(17)30328-9

94. Paul KC, Schulz J, Bronstein JM, Lill CM, Ritz BR. Association of Polygenic Risk Score With Cognitive Decline and Motor Progression in Parkinson Disease. JAMA Neurol [Internet]. 2018; Available from: http://archneur.jamanetwork.com/article.aspx?doi=10.1001/jamaneurol.2017.4206

95. Pihlstrøm L, Morset KR, Grimstad E, Vitelli V. A cumulative genetic risk score predicts motor progression in Parkinson’s disease. Mov Disord. 2016;31(4):487–90.

96. Blauwendraat C, Heilbron K, Vallerga CL, Bandres-Ciga S, von Coelln R, Pihlstrøm L, et al. Parkinson’s disease age at onset genome-wide association study: Defining heritability, genetic loci, and α-synuclein mechanisms. Mov Disord [Internet]. 2019;34(6):866–75. Available from: https://onlinelibrary.wiley.com/doi/abs/10.1002/mds.27659

97. Hensman-Moss DJ, Pardiñas AF, Langbehn D, Lo K, Leavitt BR, Roos R, et al. Identification of genetic variants associated with Huntington’s disease progression. Lancet Neurol. 2017;16(9):701–11.

98. Iwaki H, Blauwendraat C, Leonard HL, Kim JJ, Liu G, Maple-Grødem J, et al. Genomewide association study of Parkinson’s disease clinical biomarkers in 12 longitudinal patients’ cohorts. Mov Disord. 2019;(July):1–12.

99. Malek N, Swallow DMA, Grosset KA, Lawton MA, Marrinan SL, Lehn AC, et al. Tracking Parkinson’s: Study Design and Baseline Patient Data. J Parkinsons Dis. 2015;5:947–59.

100. Szewczyk-Krolikowski K, Tomlinson P, Nithi K, Wade-Martins R, Talbot K, Ben-Shlomo Y, et al. The influence of age and gender on motor and non-motor features of early Parkinson’s disease: Initial findings from the Oxford Parkinson Disease Center (OPDC) discovery cohort. Park Relat Disord [Internet]. 2014;20(1):99–105. Available from: http://dx.doi.org/10.1016/j.parkreldis.2013.09.025

101. Marek K, Jennings D, Lasch S, Siderowf A, Tanner C, Simuni T, et al. The Parkinson Progression Marker Initiative (PPMI). Prog Neurobiol. 2011;95(4):629–35.

102. Knipe MDW, Wickremaratchi MM, Wyatt-Haines E, Morris HR, Ben-Shlomo Y. Quality of life in young- compared with late-onset Parkinson’s disease. Mov

Disord. 2011;26(11):2011–8.

103. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature. 2018;562(7726):203–9.

104. Sudlow C, Gallacher J, Allen N, Beral V, Burton P, Danesh J, et al. UK Biobank: An Open Access Resource for Identifying the Causes of a Wide Range of Complex Diseases of Middle and Old Age. PLoS Med. 2015;12(3):1–10.

105. Farlow JL, Robak LA, Hetrick K, Bowling K, Boerwinkle E, Coban-Akdemir ZH, et al. Whole-Exome Sequencing in Familial Parkinson Disease. JAMA Neurol [Internet]. 2016;73(1):68–75. Available from: http://www.ncbi.nlm.nih.gov/pubmed/26595808

106. Wang K, Li M, Hakonarson H. ANNOVAR : functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.

107. Blauwendraat C, Faghri F, Pihlstrom L, Geiger JT, Elbaz A, Lesage S, et al. NeuroChip, an updated version of the NeuroX genotyping platform to rapidly screen for variants associated with neurological diseases. Neurobiol Aging [Internet]. 2017;57:247.e9-247.e13. Available from: http://dx.doi.org/10.1016/j.neurobiolaging.2017.05.009

108. IPDGC, WTCC2. A Two-Stage Meta-Analysis Identifies Several New Loci for Parkinson’s Disease. PLOS Genet. 2011 Jun;7(6):e1002142.

109. Altman DG, Bland JM. Time to event (survival) data. BMJ. 1998;317:468–9.

110. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). BMJ. 1998;317(7172):1572.

111. Spruance SL, Reid JE, Grace M, Samore M. Hazard Ratio in Clinical Trials. Antimicrob Agents Chemother. 2004;48(8):2787–92.

112. Stare J, Maucort-Boulch D. Odds ratio, hazard ratio and relative risk. Metod Zv. 2016;13(1):59–67.

113. Walker DA, Smith TJ. Nine pseudo R2 indices for binary logistic regression models. J Mod Appl Stat Methods. 2016;15(1):848–54.

114. Willer CJ, Li Y, Abecasis GR. METAL: Fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010;26(17):2190–1.

115. Zeggini E, Ioannidis JPA. Meta-analysis in genome-wide association studies. Pharmacogenomics. 2009;10(2):191–201.

116. Borenstein M, Hedges L V, Higgins JP, Rothstein HR. Introduction to meta-analysis. John Wiley & Sons, Ltd; 2009.

117. West S, Gartlehner G, Mansfield A, Poole C, Tant E, Lenfestey N, et al. Comparative Effectiveness Review Methods : Clinical Heterogeneity. Rockville; 2010.

118. Singleton A, Hardy J. A generalizable hypothesis for the genetic architecture of disease: Pleomorphic risk loci. Hum Mol Genet. 2011;20(R2):158–62.

119. Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: An efficient and comprehensive tool for rare variant association analysis using sequence data. Bioinformatics. 2016;32(9):1423–6.

120. McCarthy MI, Abecasis GR, Cardon LR, Goldstein DB, Little J, Ioannidis JPA, et al. Genome-wide association studies for complex traits: Consensus, uncertainty and challenges. Nat Rev Genet. 2008;9(5):356–69.

121. Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. Nat Commun [Internet]. 2017;8(1):1–10. Available from: http://dx.doi.org/10.1038/s41467-017-01261-5

122. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. PLoS Comput Biol. 2015;11(4):1–19.

123. Wickremaratchi MM, Perera D, O'Loghlen C, Sastry D, Morgan E, Jones A, et al. Prevalence and age of onset of Parkinson's disease in Cardiff: a community based cross sectional study and meta-analysis. J Neurol Neurosurg Psychiatry. 2009;80:805–7.

124. Lesage S, Brice A. Role of mendelian genes in "sporadic" Parkinson's disease. Parkinsonism Relat Disord [Internet]. 2012;18 Suppl 1:S66-70. Available from: http://www.ncbi.nlm.nih.gov/pubmed/22166458

125. Puschmann A. Monogenic Parkinson's disease and parkinsonism: Clinical phenotypes and frequencies of known mutations. Parkinsonism Relat Disord [Internet]. 2013;19(4):407–15. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1353802013000552

126. Lubbe S, Morris HR. Recent advances in Parkinson's disease genetics. J Neurol [Internet]. 2014;261:259–66. Available from: http://link.springer.com/10.1007/s00415-013-7003-2

127. Golub Y, Berg D, Calne DB, Pfeiffer RF, Uitti RJ, Stoessl a J, et al. Genetic factors influencing age at onset in LRRK2-linked Parkinson disease.

Parkinsonism Relat Disord [Internet]. 2009;15(7):539–41. Available from: http://dx.doi.org/10.1016/j.parkreldis.2008.10.008

128. Klebe S, Golmard J-L, Nalls M a., Saad M, Singleton a. B, Bras JM, et al. The Val158Met COMT polymorphism is a modifier of the age at onset in Parkinson's disease with a sexual dimorphism. J Neurol Neurosurg Psychiatry [Internet]. 2013;84:666–73. Available from: http://jnnp.bmj.com/cgi/doi/10.1136/jnnp-2012-304475

129. Mata IF, Leverenz JB, Weintraub D, Trojanowski JQ, Chen-Plotkin A, Van Deerlin VM, et al. GBA Variants are associated with a distinct pattern of cognitive deficits in Parkinson's disease. Mov Disord [Internet]. 2015;31(1):95–102. Available from: http://doi.wiley.com/10.1002/mds.26359

130. Alcalay RN, Caccappolo E, Mejia-Santana H, Tang MX, Rosado L, Ross BM, et al. Frequency of Known Mutations in Early-Onset Parkinson Disease. Arch Neurol [Internet]. 2010;67(9):1116–22. Available from: http://archpsyc.jamanetwork.com/pdfaccess.ashx?ResourceID=1412925&PDFSource=13

131. Kilarski LL, Pearson JP, Newsway V, Majounie E, Knipe MDW, Misbahuddin A, et al. Systematic review and UK-based study of PARK2 (parkin), PINK1, PARK7 (DJ-1) and LRRK2 in early-onset Parkinson's disease. Mov Disord. 2012 Oct;27(12):1522–9.

132. Clark LN, Wang Y, Karlins E, Saito L, Mejia-Santana H, Harris J, et al. Frequency of LRRK2 mutations in early- and late-onset Parkinson disease. Neurology. 2006;67(10):1786–91.

133. Hughes AJ, Daniel SE, Lees AJ. Improved accuracy of clinical diagnosis of Lewy body Parkinson's disease. Neurology. 2001;57:1497–1499.

134. Lill CM, Mashychev A, Hartmann C, Lohmann K, Marras C, Lang AE, et al. Launching the movement disorders society genetic mutation database (MDSGene). Mov Disord. 2016;31(5):607–9.

135. Malek N, Weil RS, Bresner C, Lawton MA, Grosset KA, Tan M, et al. Features of GBA -associated Parkinson's disease at presentation in the UK Tracking Parkinson's study. J Neurol Neurosurg Psychiatry [Internet]. 2018;89:702–9. Available from: http://jnnp.bmj.com/lookup/doi/10.1136/jnnp-2017-317348

136. The 1000 Genomes Project Consortium. A global reference for human genetic variation. Nature. 2015;526(7571):68–74.

137. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. Nat Genet [Internet]. 2016;48(10):1284–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27571263%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5157836

138. Office for National Statistics. Estimates of the population for the UK, England and Wales, Scotland and Northern Ireland [Internet]. 2017. Available from: https://www.ons.gov.uk/peoplepopulationandcommunity/populationandmigration/populationestimates/datasets/populationestimatesforukenglandandwalesscotlandandnorthernireland

139. Sanchez-Contreras M, Heckman MG, Tacik P, Diehl N, Brown PH, Soto-Ortolaza AI, et al. Study of LRRK2 variation in tauopathy: Progressive supranuclear palsy and corticobasal degeneration. Mov Disord. 2017;32(1):115–23.

140. Vilas D, Sharp M, Gelpi E, Genís D, Marder KS, Cortes E, et al. Clinical and neuropathological features of progressive supranuclear palsy in Leucine rich repeat kinase (LRRK2) G2019S mutation carriers. Mov Disord [Internet]. 2018;33(2):335–8. Available from: http://doi.wiley.com/10.1002/mds.27225

141. Malek N, Swallow DM a., Grosset K a., Lawton M a., Smith CR, Bajaj NP, et al. Olfaction in Parkin single and compound heterozygotes in a cohort of young onset Parkinson’s disease patients. Acta Neurol Scand [Internet]. 2015;134(4):271–6. Available from: http://doi.wiley.com/10.1111/ane.12538

142. Trinh J, Zeldenrust FMJ, Huang J, Kasten M, Schaake S, Petkovic S, et al. Genotype-phenotype relations for the Parkinson’s disease genes SNCA, LRRK2, VPS35: MDSGene systematic review. Mov Disord. 2018;33(12):1857–70.

143. Parkinson’s UK. The incidence and prevalence of Parkinson’s in the UK. Results from the Clinical Practice Research Datalink Reference Report. 2017; Available from: https://www.parkinsons.org.uk/sites/default/files/2018-01/Prevalence Incidence Report Latest_Public_2.pdf

144. Atashrazm F, Dzamko N. LRRK2 inhibitors and their potential in the treatment of Parkinson’s disease: current perspectives. Clin Pharmacol. 2016;8:177–89.

145. Alessi DR, Sammler E. LRRK2 kinase in Parkinson’s disease. Science. 2018;360(6384):36–7.

146. Paisán-Ruíz C, Jain S, Evans EW, Gilks WP, Simón J, Van Der Brug M, et al. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron. 2004;44(4):595–600.

147. Zimprich A, Biskup S, Leitner P, Lichtner P, Farrer M, Lincoln S, et al. Mutations in LRRK2 cause autosomal-dominant parkinsonism with pleomorphic pathology. Neuron. 2004;44(4):601–7.

148. Funayama M, Hasegawa K, Kowa H, Saito M, Tsuji S, Obata F. A new locus for Parkinson's Disease (PARK8) maps to chromosome 12p11.2-q13.1. Ann Neurol. 2002;51(3):296–301.

149. Di Fonzo A, Rohé CF, Ferreira J, Chien HF, Vacca L, Stocchi F, et al. A frequent LRRK2 gene mutation associated with autosomal dominant Parkinson's disease. Lancet. 2005;365(9457):412–5.

150. Nichols WC, Pankratz N, Hernandez D, Paisán-Ruíz C, Jain S, Halter CA, et al. Genetic screening for a single common LRRK2 mutation in familial Parkinson's disease. Lancet. 2005;365(9457):410–2.

151. Gilks WP, Abou-Sleiman PM, Gandhi S, Jain S, Singleton A, Lees AJ, et al. A common LRRK2 mutation in idiopathic Parkinson's disease. Lancet. 2005;365(9457):415–6.

152. Hernandez D, Paisan Ruiz C, Crawley A, Malkani R, Werner J, Gwinn-Hardy K, et al. The dardarin G2019S mutation is a common cause of Parkinson's disease but not other neurodegenerative diseases. Neurosci Lett. 2005;389(3):137–9.

153. Lesage S, Dürr A, Tazir M, Lohmann E, Leutenegger A-L, Janin S, et al. *LRRK2* G2019S as a Cause of Parkinson's Disease in North African Arabs. N Engl J Med [Internet]. 2006;354(4):422–3. Available from: http://www.nejm.org/doi/abs/10.1056/NEJMc055540

154. Williams-Gray CH, Goris A, Foltynie T, Brown J, Maranian M, Walton A, et al. Prevalence of the LRRK2 G2019S mutation in a UK community based idiopathic Parkinson's disease cohort. J Neurol Neurosurg Psychiatry [Internet]. 2006;77(5):665–7. Available from: http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=pubmed&cmd=Retrieve&dopt=AbstractPlus&list_uids=16614029

155. Lesage S, Ibanez P, Lohmann E, Pollak P, Tison F, Tazir M, et al. G2019S LRRK2 mutation in French and North African families with Parkinson's disease. Ann Neurol. 2005;58(5):784–7.

156. Ozelius LJ, Senthil G, Saunders-Pullman R, Ohmann E, Deligtisch A, Tagliati M, et al. LRRK2 G2019S as a cause of Parkinson's disease in Ashkenazi Jews. N Engl J Med. 2006;354(4):424–5.

157. Farrer M, Stone J, Mata IF, Lincoln S, Kachergus J, Hulihan M, et al. LRRK2 mutations in Parkinson disease. Neurology. 2005;65(5):738–40.

158. Zabetian CP, Samii A, Mosley AD, Roberts JW, Leis BC, Yearout D, et al. A clinic-based study of the LRRK2 gene in Parkinson disease yields new mutations. Neurology. 2005;65(5):741–4.

159. Deng H, Le W, Guo Y, Hunter CB, Xie W, Jankovic J. Genetic and clinical identification of Parkinson's disease patients with LRRK2 G2019S mutation. Ann Neurol. 2005;57(6):933–4.

160. Pankratz N, Pauciulo MW, Elsaesser VE, Marek DK, Halter CA, Rudolph A, et al. Mutations in LRRK2 other than G2019S are rare in a north-American based sample of familial Parkinson's didease. Mov Disord. 2006;21(12):2257–60.

161. Möller JC, Rissling I, Mylius V, Höft C, Eggert KM, Oertel WH. The prevalence of the G2019S and R1441C/G/H mutations in LRRK2 in German patients with Parkinson's disease. Eur J Neurol. 2008;15(7):743–5.

162. Lee AJ, Wang Y, Alcalay RN, Mejia-Santana H, Saunders-Pullman R, Bressman S, et al. Penetrance estimate of LRRK2 p.G2019S mutation in individuals of non-Ashkenazi Jewish ancestry. Mov Disord. 2017;32(10):1432–8.

163. Srivatsal S, Cholerton B, Leverenz JB, Wszolek ZK, Uitti RJ, Dickson DW, et al. Cognitive profile of LRRK2-related Parkinson's disease. Mov Disord. 2015;30(5):728–33.

164. Kasten M, Marras C, Klein C. Nonmotor Signs in Genetic Forms of Parkinson's Disease [Internet]. 1st ed. Vol. 133, International Review of Neurobiology. Elsevier Inc.; 2017. 129–178 p. Available from: http://dx.doi.org/10.1016/bs.irn.2017.05.030

165. Iwaki H, Blauwendraat C, Makarious MB, Bandrés-Ciga S, Leonard HL, Gibbs JR, et al. Penetrance of Parkinson's Disease in LRRK2 p.G2019S Carriers Is Modified by a Polygenic Risk Score. Mov Disord. 2020;35(5):774–80.

166. Trinh J, Gustavsson EK, Vilariño-güell C, Bortnick S, Latourelle J, Mckenzie MB, et al. DNM3 and genetic modifiers of age of onset in LRRK2 Gly2019Ser parkinsonism : a genome-wide linkage and association study. 2016;4422(16):1–9.

167. Fernández-Santiago R, Garrido A, Infante J, González-Aramburu I, Sierra M, Fernández M, et al. α-synuclein (SNCA) but not dynamin 3 (DNM3) influences age at onset of leucine-rich repeat kinase 2 (LRRK2) Parkinson's disease in Spain. Mov Disord. 2018;33(4):637–41.

168. Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, et al. Mutation in the Alpha-Synuclein Gene Identified in Families with Parkinson's Disease. Science (80-). 1997;276(June):2045–7.

169. Scott WK, Yamaoka LH, Stajich JM, Scott BL, Vance JM, Roses AD, et al. The alpha-synuclein gene is not a major risk factor in familial Parkinson disease. Neurogenetics. 1999;2(3):191–2.

170. Nuytemans K, Meeus B, Crosiers D, Brouwers N, Goossens D, Engelborghs S, et al. Relative contribution of simple mutations vs. copy number variations in five Parkinson disease genes in the Belgian population. Hum Mutat. 2009;30(7):1054–61.

171. Berg D, Niwar M, Maass S, Zimprich A, Möller JC, Wuellner U, et al. Alpha-synuclein and Parkinson's disease: Implications from the screening of more than 1,900 patients. Mov Disord. 2005;20(9):1191–4.

172. Bozi M, Papadimitriou D, Antonellou R, Moraitou M, Maniati M, Vassilatis DK, et al. Genetic assessment of familial and early-onset Parkinson's disease in a Greek population. Eur J Neurol. 2014;21(7):963–8.

173. Ibáñez P, Bonnet A-M, Débarges B, Lohmann E, Tison F, Pollak P, et al. Causal relation between alpha-synuclein gene duplication and familial Parkinson's disease. Lancet. 2004;364(9440):1169–71.

174. Nishioka K, Ross OA, Ishii K, Kachergus JM, Ishiwata K, Kitagawa M, et al. Expanding the clinical phenotype of SNCA duplication carriers. Mov Disord. 2009;24(12):1811–9.

175. Bonifati V. Genetics of Parkinson's disease – state of the art, 2013. Parkinsonism Relat Disord [Internet]. 2014;20S1:S23–8. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1353802013700099

176. Schneider SA, Alcalay RN. Neuropathology of genetic synucleinopathies with parkinsonism: Review of the literature. Mov Disord. 2017;32(11):1504–23.

177. Chartier-Harlin M-CJK, Roumier C, Mouroux V, Douay X, Lincoln S, Levecque C, et al. α-synuclein locus duplication as a cause of familial Parkinson's disease. Lancet [Internet]. 2004;364(9440):1169–71. Available from:

http://www.sciencedirect.com/science/article/pii/S0140673604171043

178. Hernandez DG, Reed X, Singleton AB. Genetics in Parkinson disease: Mendelian versus non-Mendelian inheritance. J Neurochem. 2016;139:59–74.

179. Kim CY, Alcalay RN. Genetic Forms of Parkinson ' s Disease. Semin Neurol. 2017;37(2):135–46.

180. Marder K, M-x T, Mejia-Santana H, Rosada L, Louis E. Predictors of parkin mutations in early onset parkinson disease: the CORE-PD study. Arch Neurol. 2010;67(6):731–8.

181. Periquet M, Latouche M, Lohmann E, Rawal N, De Michele G, Ricard S, et al. Parkin mutations are frequent in patients with isolated early-onset parkinsonism. Brain. 2003;126(6):1271–8.

182. Abbas N, Lücking CB, Ricard S, Dürr A, Bonifati V, De Michele G, et al. A wide variety of mutations in the parkin gene are responsible for autosomal recessive parkinsonism in Europe. French Parkinson's Disease Genetics Study Group and the European Consortium on Genetic Susceptibility in Parkinson's Disease. Hum Mol Genet [Internet]. 1999;8(4):567–74. Available from: http://www.hmg.oxfordjournals.org/cgi/doi/10.1093/hmg/8.4.567%5Cnhttp://www.ncbi.nlm.nih.gov/pubmed/10072423

183. Hatano Y, Li Y, Sato K, Asakawa S, Yamamura Y, Tomiyama H, et al. Novel PINK1 mutations in early-onset parkinsonism. Ann Neurol. 2004;56(3):424–7.

184. Li Y, Tomiyama H, Sato K, Hatano Y, Yoshino H, Atsumi M, et al. Clinicogenetic study of PINK1 mutations in autosomal recessive early-onset parkinsonism. Neurology. 2005;64(11):1955–7.

185. Valente EM, Salvi S, Ialongo T, Marongiu R, Elia AE, Caputo V, et al. PINK1 mutations are associated with sporadic early-onset Parkinsonism. Ann Neurol. 2004;56(3):336–41.

186. Healy DG, Abou-Sleiman PM, Gibson JM, Ross OA, Jain S, Gandhi S, et al. PINK1 (PARK6) associated Parkinson disease in Ireland. Neurology. 2004;63:1486–8.

187. Rogaeva E, Johnson J, Lang AE, Gulick C, Gwinn-Hardy K, Kawarai T, et al. Analysis of the PINK1 Gene in a Large Cohort of Cases With Parkinson Disease. Arch Neurol [Internet]. 2004;61(12). Available from: http://archneur.jamanetwork.com/article.aspx?doi=10.1001/archneur.61.12.1898

188. Koros C, Simitsi A, Stefanis L. Genetics of Parkinson's Disease: Genotype–Phenotype Correlations [Internet]. 1st ed. Vol. 132, International Review of Neurobiology. Elsevier Inc.; 2017. 197–231 p. Available from: http://dx.doi.org/10.1016/bs.irn.2017.01.009

189. Lohmann E, Thobois S, Lesage S, Broussolle E, Du Montcel ST, Ribeiro MJ, et al. A multidisciplinary study of patients with early-onset PD with and without parkin mutations. Neurology. 2009;72(2):110–6.

190. Takahashi H, Ohama E, Suzuki S, Horikawa Y, Ishikawa A, Morita T, et al. Familial juvenile parkinsonism: Clinical and pathologic study in a family. Neurology. 1994;44(March):437–41.

191. Mori H, Kondo T, Yokochi M, Matsumine H, Nakagawa-Hattori Y, Miyake T, et al. Pathologic and biochemical studies of juvenile parkinsonism linked to chromosome 6q. Neurology. 1998;51(3):890–2.

192. Farrer M, Chan P, Chen R, Tan L, Lincoln S, Hernandez D, et al. Lewy bodies and parkinsonism in families with parkin mutations. Ann Neurol. 2001;50(3):293–300.

193. Liu G, Boot B, Locascio JJ, Jansen IE, Winder-Rhodes S, Eberly S, et al. Specifically neuropathic Gaucher's mutations accelerate cognitive decline in Parkinson's. Ann Neurol. 2016;80(5):674–85.

194. Davis MY, Johnson CO, Leverenz JB, Weintraub D, Trojanowski JQ, Chen-Plotkin A, et al. Association of GBA mutations and the E326K polymorphism with motor and cognitive progression in parkinson disease. JAMA Neurol. 2016;73(10):1217–24.

195. Gan-Or Z, Giladi N, Rozovski U, Shifrin C, Rosner S, Gurevich T, et al. Genotype-phenotype correlations between GBA mutations and Parkinson disease risk and onset. Neurology [Internet]. 2008;70(24):2277–83. Available from: http://www.neurology.org/cgi/doi/10.1212/01.wnl.0000304039.11891.29

196. Foroud T, Uniacke SK, Liu L, Pankratz N, Rudolph A, Halter C, et al. Heterozygosity for a mutation in the parkin gene leads to later onset Parkinson disease. Neurology. 2003;60(5):796–801.

197. Klein C, Hedrich K, Wellenbrock C, Kann M, Harris J, Marder K, et al. Frequency of parkin mutations in late-onset Parkinson's disease. Ann Neurol. 2003;54(3):415–6.

198. Schrag A, Ben-Shlomo Y, Quinn N. How valid is the clinical diagnosis of

Parkinson’s disease in the community? J Neurol Neurosurg Psychiatry. 2002;73:529–534.

199. Tobin MD, Sheehan NA, Scurrah KJ, Burton PR. Adjusting for treatment effects in studies of quantitative traits: Antihypertensive therapy and systolic blood pressure. Stat Med. 2005;24(19):2911–35.

200. Vendette M, Gagnon J-F, Décary A, Massicotte-Marquez J, Postuma RB, Doyon J, et al. REM sleep behavior disorder predicts cognitive impairment in Parkinson disease without dementia. Neurology. 2007;69(19):1843–9.

201. Marion MH, Qurashi M, Marshall G, Foster O. Is REM sleep Behaviour Disorder (RBD) a risk factor of dementia in idiopathic Parkinson’s disease? J Neurol. 2008;255(2):192–6.

202. Schrag A, Siddiqui UF, Anastasiou Z, Weintraub D, Schott JM. Clinical variables and biomarkers in prediction of cognitive impairment in patients with newly diagnosed Parkinson’s disease: a cohort study. Lancet Neurol [Internet]. 2016;16(1):66–75. Available from: http://linkinghub.elsevier.com/retrieve/pii/S1474442216303283

203. Marras C, Mcdermott MP, Rochon PA, Tanner CM, Naglie G, Rudolph A, et al. Survival in Parkinson disease. Neurology. 2005;64:87–93.

204. Marras C, Rochon P, Lang AE. Predicting motor decline and disability in Parkinson Disease: A systematic review. Arch Neurol. 2011;59(7):1724–8.

205. Suchowersky O, Reich S, Perlmutter J, Zesiewicz T, Gronseth G, Weiner WJ. Appendix A: Practice parameter: Diagnosis and prognosis of new onset Parkinson disease (an evidence-based review): Report of the Quality Standards Subcommitte of the American Academy Neurology. Contin Lifelong Learn Neurol. 2007;13(1):158–65.

206. Post B, Merkus MP, De Haan RJ, Speelman JD. Prognostic factors for the progression of Parkinson’s disease: A systematic review. Mov Disord. 2007;22(13):1839–51.

207. Martino R, Candundo H, Lieshout P van, Shin S, Crispo JAG, Barakat-Haddad C. Onset and progression factors in Parkinson’s disease: A systematic review. Neurotoxicology [Internet]. 2017;61:132–41. Available from: http://dx.doi.org/10.1016/j.neuro.2016.04.003

208. Willis AW, Schootman M, Kung N, Evanoff BA, Perlmutter JS, Racette BA. Predictors of survival in patients with Parkinson disease. Arch Neurol.

2012;69(5):601–7.

209. Auyeung M, Tsoi TH, Mok V, Cheung CM, Lee CN, Li R, et al. Ten year survival and outcomes in a prospective cohort of new onset Chinese Parkinson's disease patients. J Neurol Neurosurg Psychiatry. 2012;83(6):607–11.

210. Alves G, Wentzel-Larsen T, Aarsland D, JP L. Progression of motor impairment and disability in Parkinson disease: a population-based study. Neurology [Internet]. 2005;65(9):1436-1441 6p. Available from: http://search.ebscohost.com/login.aspx?direct=true&db=ccm&AN=106150175&site=ehost-live

211. Keener AM, Paul KC, Folle A, Bronstein JM, Ritz B. Cognitive impairment and mortality in a population-based Parkinson's disease cohort. J Parkinsons Dis. 2018;8(2):353–62.

212. Zhao YJ, Wee HL, Chan YH, Seah SH, Au WL, Lau PN, et al. Progression of Parkinson's disease as evaluated by Hoehn and Yahr stage transition times. Mov Disord. 2010;25(6):710–6.

213. Vu TC, Nutt JG, Holford NHG. Progression of motor and nonmotor features of Parkinson's disease and their response to treatment. Br J Clin Pharmacol. 2012;74(2):267–83.

214. Burn DJ, Rowan EN, Allan LM, Molloy S, O'Brien JT, McKeith IG. Motor subtype and cognitive decline in Parkinson's disease, Parkinson's disease with dementia, and dementia with Lewy bodies. J Neurol Neurosurg Psychiatry. 2006;77(5):585–9.

215. Levy G, Tang MX, Cote LJ, Louis ED, Alfaro B, Mejia H, et al. Motor impairment in PD: Relationship to incident dementia and age. Neurology. 2000;55(4):539–44.

216. Levy G, Tang M-X, Louis ED, Côté LJ, Alfaro B, Mejia H, et al. The association of incident dementia with mortality in PD. Neurology. 2002;59:1708–13.

217. Cereda E, Cilia R, Klersy C, Siri C, Pozzi B, Reali E, et al. Dementia in Parkinson's disease: Is male gender a risk factor? Park Relat Disord [Internet]. 2016;26:67–72. Available from: http://dx.doi.org/10.1016/j.parkreldis.2016.02.024

218. Pigott K, Rick J, Xie SX, Hurtig H, Chen-Plotkin A, Duda JE, et al. Longitudinal study of normal cognition in Parkinson disease. Neurology [Internet]. 2015;85(15):1276–82. Available from:

http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=psyc12&NEWS=N&AN=2015-47278-004%0Ahttp://www.neurology.org%0Ahttp://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed17&NEWS=N&AN=606479610

219. Pedersen KF, Larsen JP, Tysnes O-B, Alves G. Prognosis of Mild Cognitive Impairment in Early Parkinson Disease The Norwegian ParkWest Study. JAMA Neurol. 2013;70(510):580–6.

220. Domellöf ME, Ekman U, Forsgren L, Elgh E. Cognitive function in the early phase of Parkinson's disease, a five-year follow-up. Acta Neurol Scand. 2015;132(2):79–88.

221. Liu G, Locascio JJ, Corvol JC, Boot B, Liao Z, Page K, et al. Prediction of cognition in Parkinson's disease with a clinical–genetic score: a longitudinal analysis of nine cohorts. Lancet Neurol. 2017;16(8):620–9.

222. Malek N, Lawton MA, Swallow DMA, Grosset KA, Marrinan SL, Bajaj N, et al. Vascular disease and vascular risk factors in relation to motor features and cognition in early Parkinson's disease. Mov Disord. 2016;31(10):1518–26.

223. Hely MA, Morris JGL, Reid WGJ, O'Sullivan DJ, Williamson PM, Broe GA, et al. Age at onset: the major determinant of outcome in Parkinson's disease. Acta Neurol Scand. 1995;92(6):455–63.

224. Louis ED, Tang MX, Cote L, Alfaro B, Mejia H, Marder K. Progression of parkinsonian signs in Parkinson disease. Arch Neurol [Internet]. 1999;56(3):334–7. Available from: http://www.ncbi.nlm.nih.gov/pubmed/10190824

225. Hoehn MM, Yahr MD, Hoehn MM, Yahr MD. Parkinsonism : onset , progression , and mortality. Neurology. 1967;17(5).

226. Schwab RS. Progression and prognosis in parkinson's disease. J Nerv Ment Dis. 1960;130(6):556–66.

227. Morgante L, Salemi G, Meneghini F, Di Rosa AE, Epifanio A, Grigoletto F, et al. Parkinson Disease Survival. Arch Neurol. 2000;57(4):507.

228. Uitti RJ, Ahlskog JE, Maraganore DM, Muenter MD, Atkinson EJ, Cha RH, et al. Levodopa therapy and survival in idiopathic parkinson's disease: Olmsted county project. Neurology. 1993;43(10):1918–26.

229. Ben-Shlomo Y, Marmot MG. Survival and cause of death in a cohort of patients with parkinsonism: Possible clues to aetiology? J Neurol Neurosurg Psychiatry.

1995;58(3):293–9.

230. Ishihara LS, Cheesbrough A, Brayne C, Schrag A. Estimated life expectancy of Parkinson’s patients compared with the UK population. J Neurol Neurosurg Psychiatry. 2007;78(12):1304–9.

231. Diamond SG, Markham CH, Hoehn MM, McDowell FH, Muenter MD. An examination of male-female differences in progression and mortality of Parkinson’s disease. Neurology [Internet]. 1990 May 1;40(5):763 LP – 763. Available from: http://n.neurology.org/content/40/5/763.abstract

232. Jankovic J, Kapadia AS. Functional decline in Parkinson disease. Arch Neurol. 2001;58(10):1611–5.

233. Anang JBM, Gagnon J, Bertrand J, Romentes SR, Latreille V, Passinet M, et al. Predictors of dementia in Parkinson disease. Neurology. 2014;83:1253–60.

234. De Lau LML, Verbaan D, Marinus J, van Hilten JJ. Survival in Parkinson’s disease. Relation with motor and non-motor features. Park Relat Disord [Internet]. 2014;20(6):613–6. Available from: http://dx.doi.org/10.1016/j.parkreldis.2014.02.030

235. Alves G, Larsen JP, Emre M, Wentzel-Larsen T, Aarsland D. Changes in motor subtype and risk for incident dementia in Parkinson’s disease. Mov Disord. 2006;21(8):1123–30.

236. Louis ED, Marder K, Cote L, Tang M, Mayeux R. Mortality from Parkinson’s Disease. Arch Neurol. 1997;54(3):260–4.

237. Hoops S, Nazem S, Siderowf AD, Duda JE, Xie SX, Stern MB, et al. Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. Neurology. 2009;73(21):1738–45.

238. Chahine LM, Siderowf A, Barnes J, Seedorff N, Caspell-Garcia C, Simuni T, et al. Predicting Progression in Parkinson’s Disease Using Baseline and 1-Year Change Measures. J Parkinsons Dis. 2019;9(4):665–79.

239. Maetzler W, Liepelt I, Berg D. Progression of Parkinson’s disease in the clinical phase: potential markers. Lancet Neurol [Internet]. 2009;8(12):1158–71. Available from: http://dx.doi.org/10.1016/S1474-4422(09)70291-1

240. Evers LJW, Krijthe JH, Meinders MJ, Bloem BR, Heskes TM. Measuring Parkinson’s disease over time: The real-world within-subject reliability of the MDS-UPDRS. Mov Disord. 2019;34(10):1480–7.

241. Kerr GK, Worringham CJ, Cole MH, Lacherez PF, Wood JM, Silburn PA.

Predictors of future falls in Parkinson disease. Neurology. 2010;75(2):116–24.

242. Lawton M, Baig F, Toulson G, Morovat A, Evetts SG, Ben-Shlomo Y, et al. Blood biomarkers with Parkinson’s disease clusters and prognosis: the Oxford Discovery cohort. Mov Disord. 2019;1:1–9.

243. R Core Team (R Foundation for Statistical Computing). R: A language and environment for statistical computing. [Internet]. Vienna, Austria; 2017. Available from: http://www.r-project.org/

244. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA : A Tool for Genome-wide Complex Trait Analysis. Am J Hum Genet [Internet]. 2011;88(1):76–82. Available from: http://dx.doi.org/10.1016/j.ajhg.2010.11.011

245. Yang J, Ferreira T, Morris AP, Medland SE, Madden PAF, Heath AC, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. Nat Genet [Internet]. 2012;44(4):369–75. Available from: http://dx.doi.org/10.1038/ng.2213

246. Bulik-Sullivan B, Loh PR, Finucane HK, Ripke S, Yang J, Patterson N, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat Genet. 2015;47(3):291–5.

247. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. Nat Genet. 2015;47(11):1236–41.

248. den Heijer JM, Cullen VC, Quadri M, Schmitz A, Hilt DC, Lansbury P, et al. A Large-Scale Full GBA1 Gene Screening in Parkinson’s Disease in the Netherlands. Mov Disord. 2020;

249. Fahn S, Oakes D, Shoulson I, Kieburtz K, Rudolph A, Lang A, et al. Levodopa and the progression of Parkinson’s disease. N Engl J Med [Internet]. 2004;351(24):2498–508. Available from: http://www.ncbi.nlm.nih.gov/pubmed/15590952

250. Goetz CG, Stebbins GT, Tilley BC. Calibration of unified Parkinson’s disease rating scale scores to Movement Disorder Society-unified Parkinson’s disease rating scale scores. Mov Disord. 2012;27(10):1239–42.

251. Gan-Or Z, Liong C, Alcalay RN. GBA-Associated Parkinson’s Disease and Other Synucleinopathies. Curr Neurol Neurosci Rep. 2017;18(8).

252. Nalls MA, McLean CY, Rick J, Eberly S, Hutten SJ, Gwinn K, et al. Diagnosis of Parkinson’s disease on the basis of clinical and genetic classification: A

population-based modelling study. Lancet Neurol [Internet]. 2015;14(10):1002–9. Available from: http://dx.doi.org/10.1016/S1474-4422(15)00178-7

253. O'Donoghue MC, Murphy SE, Zamboni G, Nobre AC, Mackay CE. APOE genotype and cognition in healthy individuals at risk of Alzheimer's disease: A review. Cortex [Internet]. 2018;104:103–23. Available from: https://doi.org/10.1016/j.cortex.2018.03.025

254. Zhao N, Attrebi ON, Ren Y, Qiao W, Sonustun B, Martens YA, et al. APOE4 exacerbates alpha-synuclein pathology and related toxicity independent of amyloid. Sci Transl Med. 2020;12:1809.

255. Tsuang D, Leverenz JB, Lopez OL, Hamilton RL, Bennett DA, Schneider JA, et al. APOE ε4 increases risk for dementia in pure synucleinopathies. JAMA Neurol. 2013;70(2):223–8.

256. Moreno-Grau S, Hernández I, Heilmann-Heimbach S, Ruiz S, Rosende-Roca M, Mauleón A, et al. Genome-wide significant risk factors on chromosome 19 and the APOE locus. Oncotarget. 2018;9(37):24590–600.

257. Paulusma CC, Oude Elferink RPJ. The type 4 subfamily of P-type ATPases, putative aminophospholipid translocases with a role in human disease. Biochim Biophys Acta - Mol Basis Dis. 2005;1741(1–2):11–24.

258. Verschuur CVM, Suwijn SR, Boel JA, Post B, Bloem BR, Van Hilten JJ, et al. Randomized delayed-start trial of levodopa in Parkinson's disease. N Engl J Med. 2019;380(4):315–24.

259. Iwaki H, Blauwendraat C, Leonard HL, Liu G, Maple-Grødem J, Corvol JC, et al. Genetic risk of Parkinson disease and progression: An analysis of 13 longitudinal cohorts. Neurol Genet. 2019;5(4).

260. Maple-Grødem J, Chung J, Aaser K, Tzoulis C, Tysnes O, Freddy K, et al. Alzheimer disease associated variants in SORL1 accelerate dementia development in Parkinson disease. Neurosci Lett [Internet]. 2018;674:123–6. Available from: https://doi.org/10.1016/j.neulet.2018.03.036

261. Fagan ES, Pihlstrøm L. Genetic risk factors for cognitive decline in Parkinson's disease: a review of the literature. Eur J Neurol. 2017;24(4):561-e20.

262. Ritz B, Rhodes SL, Bordelon Y, Bronstein J. Alpha-Synuclein genetic variants predict faster motor symptom progression in idiopathic Parkinson disease. PLoS One. 2012;7(5).

263. Wang G, Huang Y, Wei Chen, Chen S, Wang Y, Xiao Q, et al. Variants in the

SNCA gene associate with motor progression while variants in the MAPT gene associate with the severity of Parkinson’s disease. Park Relat Disord [Internet]. 2016;24:89–94. Available from: http://dx.doi.org/10.1016/j.parkreldis.2015.12.018

264. Markopoulou K, Biernacka JM, Armasu SM, Anderson KJ, Ahlskog JE, Chase BA, et al. Does α-synuclein have a dual and opposing effect in preclinical vs. clinical Parkinson’s disease? Park Relat Disord [Internet]. 2014;20(6):584–9. Available from: http://dx.doi.org/10.1016/j.parkreldis.2014.02.021

265. Huang Y, Rowe DB, Halliday GM. Interaction between α-synuclein and tau genotypes and the progression of Parkinson’s disease. J Parkinsons Dis. 2011;1(3):271–6.

266. Chung SJ, Armasu SM, Biernacka JM, Anderson KJ, Lesnick TG, Rider DN, et al. Genomic determinants of motor and cognitive outcomes in Parkinson’s disease. Park Relat Disord. 2012;18(7):881–6.

267. Hauser RA, Grosset DG. [123I]FP-CIT (DaTscan) SPECT brain imaging in patients with suspected parkinsonian syndromes. J Neuroimaging. 2012;22(3):225–30.

268. Sanchez-Contreras MY, Kouri N, Cook CN, Serie DJ, Heckman MG, Finch NA, et al. Replication of progressive supranuclear palsy genome-wide association study identifies SLCO1A2 and DUSP10 as new susceptibility loci. Mol Neurodegener. 2018;13(1):1–10.

269. Jabbari E, Tan MMX, Reynolds RH, Mok KY, Ferrari R, Murphy DP, et al. Common variation at the LRRK2 locus is associated with survival in the primary tauopathy progressive supranuclear palsy. bioRxiv. 2020;

270. Adler CH, Beach TG, Hentz JG, Shill HA, Caviness JN, Driver-Dunckley E, et al. Low clinical diagnostic accuracy of early vs advanced Parkinson disease: Clinicopathologic study. Neurology. 2014;83(5):406–12.

271. Marsh SE, Blurton-Jones M. Examining the mechanisms that link β-amyloid and α-synuclein pathologies. Alzheimer’s Res Ther. 2012;4(2):1–8.

272. Masliah E, Rockenstein E, Veinbergs I, Sagara Y, Mallory M, Hashimoto M, et al. β-Amyloid peptides enhance α-synuclein accumulation and neuronal deficits in a transgenic mouse model linking Alzheimer’s disease and Parkinson’s disease. Proc Natl Acad Sci U S A. 2001;98(21):12245–50.

273. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al.

Genome-wide meta-analysis identifies new loci and functional pathways influencing Alzheimer’s disease risk. Nat Genet. 2019;51(3):404–13.

274. Escott-Price V, Nalls M, Morris H, Lubbe S, Brice A, Gasser T, et al. Polygenic risk to Parkinson’s Disease is correlated with disease age at onset. Ann Neurol. 2015;77:582–91.

275. Escott-Price V, Sims R, Bannister C, Harold D, Vronskaya M, Majounie E, et al. Common polygenic variation enhances risk prediction for Alzheimer’s disease. Brain. 2015;138(12):3673–84.

276. Hannon E, Shireby GL, Brookes K, Attems J, Sims R, Cairns NJ, et al. Genetic risk for Alzheimer’s disease influences neuropathology via multiple biological pathways. Brain Commun. 2020;2(2):1–13.

277. Choi SW, O’Reilly PF. PRSice-2: Polygenic Risk Score software for biobank-scale data. Gigascience. 2019;8(7):1–6.

278. Choi SW, Mak TSH, O’Reilly PF. Tutorial: a guide to performing polygenic risk score analyses. Nat Protoc [Internet]. 2020;15(9):2759–72. Available from: http://dx.doi.org/10.1038/s41596-020-0353-1

279. Owzar K, Li Z, Cox N, Jung SH. Power and Sample Size Calculations for SNP Association Studies With Censored Time-to-Event Outcomes. Genet Epidemiol. 2012;36(6):538–48.

280. Pihlstrøm L, Blauwendraat C, Cappelletti C, Berge-Seidl V, Langmyhr M, Henriksen SP, et al. A comprehensive analysis of SNCA-related genetic risk in sporadic parkinson disease. Ann Neurol. 2018;84(1):117–29.

281. Parsa A, Kanetsky PA, Xiao R, Gupta J, Mitra N, Limou S, et al. Genome-Wide Association of CKD Progression: The Chronic Renal Insufficiency Cohort Study. J Am Soc Nephrol [Internet]. 2016;1–12. Available from: http://www.jasn.org/cgi/doi/10.1681/ASN.2015101152

282. Lambert JC, Ibrahim-Verbaas CA, Harold D, Naj AC, Sims R, Bellenguez C, et al. Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer’s disease. Nat Genet. 2013;45(12):1452–8.

283. Broer L, Buchman AS, Deelen J, Evans DS, Faul JD, Lunetta KL, et al. GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. Journals Gerontol - Ser A Biol Sci Med Sci. 2015;70(1):110–8.

284. Bathum L, Christiansen L, Jeune B, Vaupel J, McGue M, Christensen K. Apolipoprotein E genotypes: Relationship to cognitive functioning, cognitive

decline, and survival in nonagenarians. J Am Geriatr Soc. 2006;54(4):654–8.

285. Hughes TA, Ross HF, Mindham RHS, Spokes EGS. Mortality in Parkinson's disease and its association with dementia and depression. Acta Neurol Scand. 2004;110(2):118–23.

286. de Lau LML, Schipper CMA, Hofman A, Koudstaal PJ, Breteler MMB. Prognosis of Parkinson Disease. Arch Neurol. 2005;62(8):1265.

287. Todd S, Barr S, Roberts M, Passmore AP. Survival in dementia and predictors of mortality: A review. Int J Geriatr Psychiatry. 2013;28(11):1109–24.

288. Garcia-Ptacek S, Farahmand B, Kareholt I, Religa D, Cuadrado ML, Eriksdotter M. Mortality risk after dementia diagnosis by dementia type and underlying factors: A cohort of 15,209 patients based on the swedish dementia registry. J Alzheimer's Dis. 2014;41(2):467–77.

289. Koller D, Kaduszkiewicz H, Van Den Bussche H, Eisele M, Wiese B, Glaeske G, et al. Survival in patients with incident dementia compared with a control group: A five-year follow-up. Int Psychogeriatrics. 2012;24(9):1522–30.

290. Pennington S, Snell K, Lee M, Walker R. The cause of death in idiopathic Parkinson's disease. Park Relat Disord [Internet]. 2010;16(7):434–7. Available from: http://dx.doi.org/10.1016/j.parkreldis.2010.04.010

291. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants with Those of the General Population. Am J Epidemiol. 2017;186(9):1026–34.

292. Fall PA, Saleh A, Fredrickson M, Olsson JE, Granérus AK. Survival time, mortality, and cause of death in elderly patients with Parkinson's disease: A 9-year follow-up. Mov Disord. 2003;18(11):1312–6.

293. Mittal S, Bjørnevik K, Im DS, Flierl A, Dong X, Locascio JJ, et al. β2-Adrenoreceptor is a regulator of the α-synuclein gene driving risk of Parkinson's disease. Science (80-). 2017;357(6354):891–8.

294. Buniello A, Macarthur JAL, Cerezo M, Harris LW, Hayhurst J, Malangone C, et al. The NHGRI-EBI GWAS Catalog of published genome-wide association studies, targeted arrays and summary statistics 2019. Nucleic Acids Res. 2019;47(D1):D1005–12.

295. Ellinghaus D, Degenhardt F, Bujanda L, Buti M, Albillos A, Invernizzi P, et al. Genomewide Association Study of Severe Covid-19 with Respiratory Failure. N

Engl J Med. 2020;

296. Nguyen M, Wong YC, Ysselstein D, Severino A, Krainc D. Synaptic, Mitochondrial, and Lysosomal Dysfunction in Parkinson’s Disease. Trends Neurosci [Internet]. 2019;42(2):140–9. Available from: https://doi.org/10.1016/j.tins.2018.11.001

297. Bandres-Ciga S, Saez-Atienzar S, Bonet-Ponce L, Billingsley K, Vitale D, Blauwendraat C, et al. The endocytic membrane trafficking pathway plays a major role in the risk of Parkinson’s disease. Mov Disord. 2019;34(4):460–8.

298. Mcfarthing K, Buff S, Rafaloff G, Dominey T, Wyse RK. Parkinson’s Disease Drug Therapies in the Clinical Trial Pipeline : 2020. J Parkinsons Dis. 2020;10:757–74.

299. Stocchi F, Rascol O, Hauser RA, Huyck S, Tzontcheva A, Capece R, et al. Randomized trial of preladenant, given as monotherapy, in patients with early Parkinson disease. Neurology. 2017;88(23):2198–206.

300. Olanow CW, Kieburtz K, Schapira AHV. Why have we failed to achieve neuroprotection in Parkinson’s disease? Ann Neurol. 2008;64(SUPPL. 2):101–10.

301. Verghese PB, Castellano JM, Holtzman DM. Apolipoprotein E in Alzheimer’s disease and other neurological disorders. Lancet Neurol [Internet]. 2011;10(3):241–52. Available from: http://dx.doi.org/10.1016/S1474-4422(10)70325-2

302. Gallardo G, Schlüter OM, Südhof TC. A molecular pathway of neurodegeneration linking α-synuclein to ApoE and Aβ peptides. Nat Neurosci. 2008;11(3):301–8.

303. Rocha EM, Miranda B De, Sanders LH. Alpha-synuclein : Pathology , mitochondrial dysfunction and neuroinflammation in Parkinson’ s disease. Neurobiol Dis. 2018;109:249–57.

304. Shimada H, Hirano S, Shinotoh H, Aotsuka A, Sato K, Tanaka N, et al. Mapping of brain acetylcholinesterase alterations in Lewy body disease by PET. Neurology. 2009;73(4):273–8.

305. Halliday GM, Leverenz JB, Schneider JS, Adler CH. The neurobiological basis of cognitive impairment in Parkinson’s disease. Mov Disord. 2014;29(5):634–50.

306. Boyle EA, Li YI, Pritchard JK. An Expanded View of Complex Traits: From

Polygenic to Omnigenic. Cell [Internet]. 2017;169(7):1177–86. Available from: http://dx.doi.org/10.1016/j.cell.2017.05.038

307. Majbour NK, Vaikath NN, Eusebi P, Chiasserini D, Ardah M, Varghese S, et al. Longitudinal changes in CSF alpha-synuclein species reflect Parkinson's disease progression. Mov Disord. 2016;31(10):1535–42.

308. Hall S, Surova Y, Öhrfelt A, Zetterberg H, Lindqvist D, Hansson O. CSF biomarkers and clinical progression of Parkinson disease. Neurology. 2015;84(1):57–63.

309. Paternoster L, Tilling KM, Davey Smith G, Smith GD. Genetic epidemiology and Mendelian randomization for informing disease therapeutics : Conceptual and methodological challenges. PLoS Genet [Internet]. 2017;13(10):1–9. Available from: http://dx.doi.org/10.1371/journal.pgen.1006944

310. Dudbridge F, Allen RJ, Sheehan NA, Schmidt AF, Lee JC, Jenkins RG, et al. Adjustment for index event bias in genome-wide association studies of subsequent events. Nat Commun [Internet]. 2019;10(1561). Available from: http://dx.doi.org/10.1038/s41467-019-09381-w

311. Foo JN, Chew EGY, Chung SJ, Peng R, Blauwendraat C, Nalls MA, et al. Identification of Risk Loci for Parkinson Disease in Asians and Comparison of Risk between Asians and Europeans: A Genome-Wide Association Study. JAMA Neurol. 2020;77(6):746–54.

312. Fan CC, Banks SJ, Thompson WK, Chen CH, McEvoy LK, Tan CH, et al. Sex-dependent autosomal effects on clinical progression of Alzheimer's disease. Brain. 2020;143(7):2272–80.

313. Kent DM, Hayward R a. Limitations of applying summary results of clinical trials to individual patients. Jama. 2007;298:1209–12.

314. Leonard H, Blauwendraat C, Krohn L, Faghri F, Iwaki H, Ferguson G, et al. Genetic variability and potential effects on clinical trial outcomes: Perspectives in Parkinson's disease. J Med Genet. 2020;57(5):331–8.

315. Mullin S, Smith L, Lee K, D'Souza G, Woodgate P, Elflein J, et al. Ambroxol for the Treatment of Patients with Parkinson Disease with and Without Glucocerebrosidase Gene Mutations: A Nonrandomized, Noncontrolled Trial. JAMA Neurol. 2020;77(4):427–34.

316. Bose A, Beal MF. Mitochondrial dysfunction in Parkinson's disease. J Neurochem. 2016;216–31.