Genome-wide association studies of cognitive and motor progression in Parkinson's disease

Authors: Manuela MX TAN, BPsych,[1,2] Michael A LAWTON, PhD,[3] Edwin JABBARI, MRCP,[1,2] Regina H REYNOLDS, MSc,[4] Hirotaka IWAKI, MD, PhD,[5,6] Cornelis BLAUWENDRAAT, PhD,[5] Sofia KANAVOU, MSc,[3] Miriam I POLLARD, BSc,[1] Leon HUBBARD, PhD,[7] Naveed MALEK, MRCP,[8] Katherine A GROSSET, MD,[8] Sarah L MARRINAN, MD,[9] Nin BAJAJ, PhD,[10] Roger A BARKER, PhD,[11,12] David J BURN, MD,[13] Catherine BRESNER, BSc,[7] Thomas FOLTYNIE, PhD,[1,2] Nicholas W WOOD, PhD,[1,2] Caroline H WILLIAMS-GRAY, MRCP, PhD,[11] John HARDY, PhD,[2,4,14,15,16,17] Michael A NALLS, PhD,[5,6] Andrew B SINGLETON, PhD,[5] Nigel M WILLIAMS, PhD,[7] Yoav BEN-SHLOMO, MD, PhD,[3] Michele TM HU, PhD,[18,19,20] Donald G GROSSET, MD,[8] Maryam SHOAI, PhD,[4,15] Huw R MORRIS, PhD, FRCP[1,2]

Affiliations

[1]Department of Clinical and Movement Neurosciences, Queen Square Institute of Neurology, University College London, London, UK

[2]UCL Movement Disorders Centre, University College London, London, UK

[3]Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK

[4]Department of Neurodegenerative Diseases, Queen Square Institute of Neurology, University College London, London, UK

[5]Molecular Genetics Section, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, USA

[6]Data Tecnica International, Glen Echo, Maryland, USA

[7]Institute of Psychological Medicine and Clinical Neurosciences, MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University, Cardiff, UK

[8]Department of Neurology, Institute of Neurological Sciences, Queen Elizabeth University Hospital, Glasgow, UK

[9]Institute for Ageing and Health, Newcastle University, Newcastle upon Tyne, UK

[10]Department of Neurology, Queen's Medical Centre, Nottingham, UK

[11]Department of Clinical Neurosciences, John van Geest Centre for Brain Repair, University of Cambridge, Cambridge, UK

[12]Wellcome-MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK

[13]Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK

[14]Reta Lila Weston Institute, UCL Queen Square Institute of Neurology, London, UK

[15]UK Dementia Research Institute, University College London, London, UK

[16]National Institute for Health Research (NIHR) University College London Hospitals Biomedical Research Centre, London, UK

[17]Institute for Advanced Study, The Hong Kong University of Science and Technology, Hong Kong SAR, China

[18]Nuffield Department of Clinical Neurosciences, Division of Clinical Neurology, University of Oxford, Oxford, UK

[19]Oxford Parkinson's Disease Centre, University of Oxford, Oxford, UK

[20]Department of Clinical Neurology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK

Corresponding authors: Miss Manuela Tan and Professor Huw Morris, Department of Clinical and Movement Neurosciences, Queen Square Institute of Neurology, Royal Free Hospital, Rowland Hill Street, London NW3 2PF. Email: manuela.tan@ucl.ac.uk; h.morris@ucl.ac.uk. Telephone: +44 7448 290037

Word count (excluding abstract, tables and table/figure titles and legends): 3681

Running title: PD progression GWAS

Key Words: Parkinson's disease, genetics, progression, genome-wide association study

Financial Disclosure/Conflict of Interest: None to report

Funding sources: Parkinson's UK

Abbreviations: FUMA = Functional Mapping and Annotation of GWAS ; HD = Huntington's Disease; GWAS = Genome-wide Association Study; LD = Linkage Disequilibrium; MDS-UPDRS = Movement Disorder Society Unified Parkinson's Disease Rating Scale; MoCA = Montreal Cognitive Assessment; PCA = Principal Components Analysis; PD = Parkinson's Disease; PPMI = Parkinson's Progression Markers Initiative; SNP = Single Nucleotide Polymorphism; SD = Standard Deviation.

ABSTRACT

Background: There are currently no treatments that stop or slow the progression of Parkinson's disease (PD). Case-control genome-wide association studies have identified variants associated with disease risk, but not progression.

Objective: To identify genetic variants associated with PD progression.

Methods: We analysed three large, longitudinal cohorts: Tracking Parkinson's, Oxford Discovery, and the Parkinson's Progression Markers Initiative. We included clinical data for 3,364 patients with 12,144 observations (mean follow-up 4.2 years). We used a new method in PD, following a similar approach in Huntington's disease, where we combined multiple assessments using a principal components analysis to derive scores for composite, motor, and cognitive progression. These scores were analysed in linear regressions in genome-wide association studies. We also performed a targeted analysis of the 90 PD risk loci from the latest case-control meta-analysis.

Results: There was no overlap between variants associated with PD risk, from case-control studies, and PD age at onset versus PD progression. The *APOE* ε4 tagging variant, rs429358, was significantly associated with composite and cognitive progression in PD. Conditional analysis revealed several independent signals in the *APOE* locus for cognitive progression. No single variants were associated with motor progression. However in gene-based analysis, *ATP8B2,* a phospholipid transporter related to vesicle formation, was nominally associated with motor progression ($p=5.3x10^{-6}$).

Conclusions: We provide early evidence that this new method in PD improves measurement of symptom progression. We show that the *APOE* ε4 allele drives progressive cognitive impairment in PD. Replication of this method and results in independent cohorts is needed.

INTRODUCTION

Progression in Parkinson's disease (PD) is heterogeneous, with some patients progressing rapidly while others remain relatively stable over time[1]. There is a clear need to identify genetic variants that affect symptom progression in PD. These genes and pathways could be targeted to develop therapies to stop or slow progression of PD. Genetic factors could also help to stratify patients and predict progression more accurately in clinical trials.

Genome-wide association studies (GWASs) in PD have identified 90 independent loci associated with disease risk[2]. 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[3] 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[4], including rater subjectivity and practice effects in cognitive assessments. Therefore, combining multiple measures may improve the accuracy of measuring progression[5,6], as shown in the Huntington's disease (HD) progression GWAS[7]. In this study, we analysed data from three large, prospective, longitudinal studies: Tracking Parkinson's, Oxford Parkinson's Disease Centre Discovery, and Parkinson's Progression Markers Initiative (PPMI). We combined multiple measures of motor and cognitive progression using Principal Components Analysis (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

Standard quality control procedures were performed in PLINK v1.9. The cohorts were genotyped, filtered and imputed separately, but following the same quality control steps. Only variants with minor allele frequency >1% were included. The three datasets were merged after imputation, with only shared variants retained. Genetic

principal components were generated and outliers removed (see Supplementary Methods and Supplementary Figures 1-2).

Clinical outcome measures

Individual-level data from the cohorts was merged. In order to increase power and the accuracy of the final progression scores, we performed all transformations and created progression scores from the merged dataset as follows (Figure 1).

Motor progression was assessed using the MDS-UPDRS Part III (clinician-assessed movement examination), MDS-UPDRS Part II (patient-reported experiences of daily living), and Hoehn and Yahr stage (clinician-assessed rating of impairment and disability)[8,9]. In PPMI, we used motor assessments conducted in the 'off' medication state.

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).

Raw scores were transformed into percentages and standardised to the population baseline mean and standard deviation within each cohort (Supplementary Methods).

Analysis

Progression scores

We 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. PD onset was based on participants' self-reported symptom onset. For the cognitive measures, we included the number of years of education before higher education, and whether higher education was undertaken. We included terms for subject random effects to account for individual heterogeneity in the intercept (baseline values) and slope (rate of progression).

We used the random effect slope values as the measure of 'residual' progression not predicted by age at onset, cohort, gender, and education, for each individual. We performed PCA on these values after zero centring and scaling to have unit variance.

The final progression scores from the PCAs relate to the variability explained, and therefore the direction cannot be strictly interpreted. Patients who were missing clinical data (e.g. MDS-UPDRSIII total) at all visits were not included in the PCA and subsequent GWAS analysis.

Removal of non-PD cases

Any patients that were diagnosed with a different condition during follow-up were removed from analyses. We also conducted sensitivity analyses to remove any cases which may have non-PD conditions but an alternative diagnosis had not yet been confirmed. Firstly, we removed patients in Tracking Parkinson's and Oxford Discovery who had a clinician-rated diagnostic certainty of PD <90%[10,11]. Secondly, we 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. multiple system atrophy) conditions.

GWAS

For each GWAS, we included the following covariates: cohort (to adjust for differences in genotyping data and measurement error) and the first 5 genetic principal components from the merged genotype data (to adjust for population substructure). GWASs were conducted in rvtests[12] using the single variant Wald test. Genome-wide Complex Trait Analysis conditional and joint analysis (GCTA-COJO) was used to identify independent signals[13,14]. Individuals carrying rare variants in *GBA*, *LRRK2* or other PD genes were not excluded from the GWASs. We also performed sex-stratified analysis to identify if there are different genetic associations in men and women.

Genetic risk scores were calculated from the 90 loci from the PD case-control GWAS[2] and we analysed the association with each progression score using linear regression.

GBA

We 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. We classified patients as carrying a pathogenic

GBA variant, including Gaucher's Disease variants and variants associated with PD but excluding novel variants, using previous studies[15,16]. We 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, we performed a sensitivity analysis adjusting for LEDD, as described in a previous study, where we estimated the effect of levodopa on the MDS-UPDRSIII[11] (Supplementary Methods). 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[17], and most likely with higher doses.

RESULTS

We included clinical data for 3,364 PD patients with 12,144 observations (Table 1). The mean follow-up time was 4.2 years (SD=1.5 years), and mean disease duration at study entry 2.9 years (SD=2.6 years). 79.7% of patients had completed the 72-month follow-up visit.

Within the motor progression PCA, the first principal component explained 61.0% of the total variance. Within the cognitive domain PCA, the first principal component explained 59.8% of the total variance (Supplementary Figures 3-6).

We found that the first principal components for motor and cognitive progression were moderately correlated ($r=-0.35$, $p < 2.2 \times 10^{-16}$; Supplementary Table 1). We 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 (Supplementary Figures 7-8). Supplementary Tables 2-6 show the how the raw scales and the motor, cognitive, and

composite principal components are correlated. None of the principal components were associated with cohort (all p-values >0.9).

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. All GWAS lambdas were <1.05. One variant rs429358 in Chromosome 19 passed genome-wide significance ($p=1.2x10^{-8}$, Figure 2, Supplementary Table 7, Supplementary Figures 9-10). This variant tags the *APOE* ε4 allele. In the gene-based test, *APOE, TOMM40* and *APOC1* reached significance ($p<2.8x10^{-6}$, correcting for the number of mapped protein coding genes). When we performed conditional analysis on the top SNP rs429358, there were no other SNPs that passed significance in this region (Supplementary Figure 11). The Reactome pathway cytosolic sulfonation of small molecules pathway was significantly enriched ($p=6.9x10^{-6}$).

GWAS of motor progression

2,848 PD patients had motor progression scores and genotype data. No variants passed genome-wide significance (Figure 3, Supplementary Table 8). However, in the gene-based test, *ATP8B2* in Chromosome 1 was associated with motor progression ($p=5.3x10^{-6}$, Supplementary Figures 12-13), although this did not reach significance correcting for the number of mapped genes ($p=2.81x10^{-6}$).

We conducted follow-up GWASs in each cohort separately (Supplementary Table 9) and each motor scale separately (without combining in PCA) to confirm that the results were not driven by a single cohort, or a single scale. These results show that associations are strengthened with the PCA approach (Supplementary Table 10).

Our top variant in Chromosome 1, rs35950207, was associated with motor progression, $p=5.0x10^{-6}$. We examined the associations for this SNP in the previous progression GWAS[18] (https://pdgenetics.shinyapps.io/pdprogmetagwasbrowser/); rs35950207 was not significantly associated with binomial analysis of Hoehn and Yahr stage 3 or more at baseline (beta=0.27, p=0.03).

rs35950207 is a variant 2kb upstream of *AQP10*. It is an expression quantitative trait loci (eQTL) for *AQP10* in whole blood (GTEx $p=1.7x10^{-6}$, eQTLGen $p=3.62x10^{-139}$) and other tissues (subcutaneous adispose, skin, esophagus, testis, and heart). It is also an eQTL for *ATP8B2* in blood (GTEx $p=1.5x10^{-5}$, eQTLGen $p=7.84x10^{-42}$) and in the cerebellum (GTEx $p=7.8x10^{-5}$). *GBA* is also located in Chromosome 1 and *GBA* variants are associated with both PD risk and progression[19]. 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 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 top variant was rs429358, which tags the *APOE* ε4 allele ($p=2.53x10^{-13}$, Figure 4, Supplementary Table 11, Supplementary Figure 14-15). Supplementary Figure 16 shows that ε4 carriers had more severe cognitive progression. *APOE,* was also significantly associated with cognitive progression in the gene-based analysis, in addition to *APOC1* and *TOMM40*. Follow-up analyses showed that the effects for the top 5 independent SNPs were consistent in each cohort and each scale (Supplementary Tables 12-13).

When we performed conditional analysis on the top SNP rs429358, a group of SNPs still passed genome-wide significance, indicating independent signals (Supplementary Figure 17). The top SNP was rs6857 (beta=-0.33, $p=4.4 \times 10^{-11}$). This is a 3' UTR Variant in *NECTIN2*. We 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*.

We found similar frequencies of *APOE* genotypes to previous studies[20] (Supplementary Table 14).

LEDD-adjusted analyses

When we performed GWASs of composite progression and motor progression after adjusting for LEDD, we 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 (β=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* locus passed genome-wide significance only in men for composite progression and cognitive progression ($p<5x10^{-8}$). 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[2], 73 were present in our final dataset, including the *SNCA* and *TMEM175/GAK* variants associated with PD age at onset[21]. No variants passed analysis-wide significance (p=0.05/73). Variants with at least one association p<0.05 are shown in Supplementary Figure 18.

We 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.

We also extracted results for other candidate variants that have been implicated in PD progression (Supplementary Figure 19). We did not find that the top variant rs382940 in *SLC44A1* that was associated in progression to H&Y stage 3 from the Iwaki GWAS[18] was associated with either composite, motor or cognitive progression in our GWASs.

Overall, we did not find any overlap between the variants associated with PD risk, age at onset, and progression. Our 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 (Supplementary Table 15). The risk score was nominally associated with cognitive progression (beta=-0.098, p=0.04) but not composite (beta=0.09, p=0.12), or motor progression (beta=0.02, p=0.69).

GBA

GBA data was available for 2,020 patients from Tracking Parkinson's and PPMI. 194 (9.6%) carried a pathogenic variant in *GBA* (Supplementary Table 16). *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).

Removal of potential non-PD cases

Removing patients with <90% diagnostic certainty did not substantially affect our results; the top signals had slightly weaker associations in these sensitivity analyses. When we 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 (Supplementary Tables 17-18).

DISCUSSION

We 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 in large clinical cohorts.

Our study contributes to evidence that improving the phenotypic measure can increase power in genetic studies. We 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[7]. Other studies show prediction accuracy of PD status or progression (such as development of cognitive impairment) is improved by combining multiple clinical, genetic, and biomarker factors[6,22].

In PD, there are many different scales for assessing symptoms. Each scale has a degree of measurement error[4] and different sensitivity to progression of underlying symptoms[23]. PCA is a data-driven approach that combines multiple measures to identify latent components that explain the most variability in the data, and these may more accurately reflect disease progression.

Our progression GWASs have identified two main findings. Firstly, we replicated previous findings for *APOE* ε4. Many studies have shown that the ε4 allele is associated with dementia in PD[20,24–26], and potentially separately from the risk of Alzheimer's disease (AD)[27]. 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[28]. 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[29]. In addition, the ε4 allele is overrepresented in Dementia with Lewy Body cases with 'pure' Lewy body pathology, compared to PDD cases[30]. A systematic review showed that limbic and neocortical α-synuclein pathology had the strongest association with PD dementia[28]. 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[31]. Further work is needed to fine-map this region and determine if there are other genes that contribute to cognitive progression.

We 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[32]. This gene has not been reported in PD or other diseases, and needs to be tested in other 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[33]. We also found that the mean rate of change in the MDS-UPDRSIII was comparable between Tracking Parkinson's/Oxford Discovery and PPMI (Table 1), despite the different medication states. Together, these suggest that medication has not influenced our results for motor progression.

We have shown that the genetics of PD risk and progression are largely separate. In our targeted analysis of PD risk variants, *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[34–39], however many of these studies have longer follow-up, or patients with longer disease duration. This may explain why we did not find a strong effect for motor progression, and is supported by analysis of *GBA* in patients earlier in disease stage[15]. In addition, previous studies have used different methods to measure progression. Our unbiased genome-wide search suggests that, in addition to *GBA*, there are potentially other genes that are important for PD progression.

Our targeted analysis showed that only a few PD risk variants were nominally associated with progression, similar to the previous PD progression GWAS[18,40]. This suggests that there is minimal overlap in the genetic architecture of PD risk and PD progression. Similarly, the age at onset GWAS showed only a partial overlap with the genetics of PD risk[21]. We now have the ability 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.

We 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[18]. We

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*[1,41–43], *COMT*[24,42]*, BDNF, MTHFR, and SORL1*[44] have been reported to influence cognitive decline (reviewed in Fagan & Pihlstrom[45]) . 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[26,46–49]. A small GWAS of motor and cognitive progression identified suggestive loci in *C8orf4* and *CLRN3*[50], although these have not been replicated. A novel machine learning approach found that variation in *LINGO2* was associated with change in the MDS-UPDRS[51], although again this finding needs independent replication. We did not replicate these findings, possibly because we are underpowered as a GWAS to detect variants with smaller effects, or because we have analysed progression using different methods. However, many of these candidate gene studies are small and some associations have not been convincingly replicated.

Our 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.*

We may also be underpowered to detect variants with smaller effects on progression. Although the HD GWAS identified significant signals in smaller samples[7], analysis of PD progression is more complex due to slower 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. We did not find that our results changed substantially when we 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[52]. Despite this, our sensitivity analysis suggest that our results are not being driven by non-PD conditions. Our GWASs also did not identify loci that are associated with PSP risk, including *MAPT, MOBP*[53], or rs2242367 near *LRRK2* associated with PSP progression[54].

Many of our top variants had weaker signals when we excluded the fastest and slowest progressing patients. With our duration of follow-up, we should have excluded the majority of non-PD patients as diagnostic accuracy improves after 5 years of disease duration[1,55] 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[56], or interactions between synuclein and co-pathologies (such as amyloid, and tau)[57,58] in the rapid progressors which exacerbates clinical progression.

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

Data Availability

Anonymised data from Tracking Parkinson's and Oxford Discovery are available to researchers on application. Please apply via the project coordinators (tracking-parkinsons@glasgow.ac.uk and parkinsons.discovery@nhs.net respectively). The PPMI data is publicly available on application (https://www.ppmi-info.org/access-data-specimens/download-data/).

Code is available at https://github.com/huw-morris-lab/PD-PCA-progression-GWAS.

Author Contributions

1. Research project: A. Conception, B. Organization, C. Execution;

2. Statistical Analysis: A. Design, B. Execution, C. Review and Critique;

3. Manuscript Preparation: A. Writing of the first draft, B. Review and Critique.

M.M.X.T.: 2A, 2B, 2C, 3A, 3B

M.A.L.: 1B, 2C, 3B

E.J.: 2C, 3B

R.H.R.: 2C, 3B

H.I.: 2C, 3B

C.B.: 2C, 3B

S.K.: 1B, 3B

M.P.: 2C, 3B

L.H.: 1C

N.M.: 1C, 3B

K.A.G.: 1A, 1B, 1C, 3B

S.L.M.: 1C

N.B.: 1A, 1B, 1C

R.A.B.: 1A, 1B, 3B

D.J.B.: 1A, 1B, 3B

C.B.: 1C

T.F.: 1A, 1B, 3B

J.H.: 1A, 3B

N.W.: 1A, 1B

C.H.W-G.: 1B, 1C, 2C, 3B

M.A.N.: 2C, 3B

A.B.S.: 3B

N.W.W.: 1A, 1B, 2C, 3B

Y.B-S.: 1A, 1B, 2C, 3B

M.T.M.H.: 1A, 1B, 1C, 3B

D.G.G.: 1A, 1B, 1C, 3B

M.S.: 2A, 2C, 3B

H.R.M.: 1A, 1B, 2A, 2C, 3B

Acknowledgements

Both Tracking Parkinson's and Oxford Discovery are primarily funded and supported by Parkinson's UK. Both studies are supported by the National Institute for Health Research (NIHR) Dementias and Neurodegenerative Diseases Research Network (DeNDRoN). Oxford Discovery is also supported by the NIHR Oxford Biomedical Research Centre based at Oxford University Hospitals NHS Trust, and University of Oxford.

This research was supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre.

This research was supported in part by the Intramural research Program of the NIH, National institute on Aging.

The UCL Movement Disorders Centre is supported by the Edmond J. Safra Philanthropic Foundation.

Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database. For up-to-date information on the study, visit www.ppmi-info.org.

PPMI – a public-private partnership – is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners (listed in https://www.ppmi-info.org/about-ppmi/who-we-are/study-sponsors/).

Financial Disclosures for the preceding 12 months

M.M.X.T. is supported Parkinson's UK. M.A.L. is supported by Parkinson's UK. E.J. is supported by the Medical Research Council UK. R.H.R. supported through the award of a Leonard Wolfson Doctoral Training Fellowship. N.B. has received payment for advisory board attendance from UCB, Teva Lundbeck, Britannia, GSK, Boehringer and honoraria from UCB Pharma, GE Healthcare, Lily Pharma, Medtronic. He has received research grant support from GE Healthcare, Wellcome Trust, Medical Research Council, Parkinson's UK and National Institute for Health Research. R.A.B. has received grants from Parkinson's UK, NIHR, Cure Parkinson's Trust, Evelyn Trust, Rosetrees Trust, MRC, Wellcome Trust, and EU along with payment for advisory board work from Oxford Biomedica, Living Cell Technologies, Fujifilm Cellular Dynamics Inc, Nova Nordisk, BlueRock Therapeutics, Sana Biotherapeutics, Aspen Neuroscience and UCB along with honoraria from Wiley and Springer for books and editorial work.D.J.B. has received grants from NIHR, Wellcome Trust, GlaxoSmithKline Ltd, Parkinson's UK and Michael J Fox Foundation. T.F. has received grants from Michael J Fox Foundation, Cure Parkinson's Trust, Brain Research trust, John Black Charitable Foundation, Rosetrees trust and honoraria for speaking at meetings from Bial, Profile Pharma and Medtronic. N.W.W. is supported by the MRC and NIHR UCLH Biomedical research centre. C.H.W-G is supported by a RCUK/UKRI Research Innovation Fellowship awarded by the Medical Research Council (MR/R007446/1), by the NIHR Cambridge Biomedical Research Centre Dementia and Neurodegeneration Theme (Grant Reference Number 146281), and by the Cambridge Centre for Parkinson-Plus. M.A.N. reports that this work was done under a consulting contract with National Institutes of Health, he also consults for Lysosomal Therapeutics Inc, Neuron23, and Illumina. J.H. is supported by the UK Dementia Research Institute which receives its funding from DRI Ltd, funded by the UK Medical Research Council, Alzheimer's Society and Alzheimer's Research UK. He is also supported by the MRC, Wellcome Trust, Dolby Family Fund, National Institute for Health Research University College London Hospitals Biomedical Research Centre. Y.B-S. has received grant funding from the MRC, NIHR, Parkinson's UK, NIH and ESRC. N.M.W. is supported by Parkinson's UK. M.T.H. receives grants from Parkinson's UK, Oxford NIHR Biomedical Research Centre, and MJFF and is an adviser to the Roche Prodromal Advisory and Biogen Digital advisory boards. D.G.G. has received grants from Michael's Movers, The Neurosciences Foundation, and Parkinson's UK, and honoraria from UCB Pharma and GE Healthcare, and consultancy fees from Acorda Therapeutics. H.R.M. is supported by the PSP Association, CBD Solutions, Drake Foundation, the Medical Research Council UK, Parkinson's UK, and Cure Parkinson's Trust. All other authors did not declare any funding sources that directly contributed to this study. M.M.X.T. takes responsibility for the integrity of the data and the accuracy of the data analysis.

References

1. Williams-Gray CH, Mason SL, Evans JR, et al. The CamPaIGN study of Parkinson's disease: 10-year outlook in an incident population-based cohort. *J Neurol Neurosurg Psychiatry*. 2013;84:1258-1264. doi:10.1136/jnnp-2013-305277

2. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol*. 2019;18(12):1091-1102. doi:10.1016/S1474-4422(19)30320-5

3. Maetzler W, Liepelt I, Berg D. Progression of Parkinson's disease in the clinical phase: potential markers. *Lancet Neurol*. 2009;8(12):1158-1171. doi:10.1016/S1474-4422(09)70291-1

4. 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):1480-1487. doi:10.1002/mds.27790

5. Kerr GK, Worringham CJ, Cole MH, Lacherez PF, Wood JM, Silburn PA. Predictors of future falls in Parkinson disease. *Neurology*. 2010;75(2):116-124. doi:10.1212/WNL.0b013e3181e7b688

6. 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*. 2016;16(1):66-75. doi:10.1016/S1474-4422(16)30328-3

7. Hensman-Moss DJ, Pardiñas AF, Langbehn D, et al. Identification of genetic variants associated with Huntington's disease progression. *Lancet Neurol*. 2017;16(9):701-711. doi:10.1016/S1474-4422(17)30161-8

8. Hoehn MM, Yahr MD, Hoehn MM, Yahr MD. Parkinsonism : onset , progression , and mortality. *Neurology*. 1967;17(5). doi:10.1212/WNL.17.5.427

9. Goetz CG, Poewe W, Rascol O, et al. Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: Status and recommendations. *Mov Disord*. 2004;19(9):1020-1028. doi:10.1002/mds.20213

10. Lawton M, Ben-Shlomo Y, May MT, et al. Developing and validating Parkinson's disease subtypes and their motor and cognitive progression. *J Neurol Neurosurg Psychiatry*. 2018;89(12). http://jnnp.bmj.com/content/89/12/1279.full.pdf.

11. Lawton M, Baig F, Toulson G, et al. Blood biomarkers with Parkinson's disease clusters and prognosis: the Oxford Discovery cohort. *Mov Disord*. 2019;1:1-9. doi:10.1002/mds.27888

12. 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):1423-1426. doi:10.1093/bioinformatics/btw079

13. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA : A Tool for Genome-wide Complex Trait Analysis. *Am J Hum Genet*. 2011;88(1):76-82. doi:10.1016/j.ajhg.2010.11.011

14. Yang J, Ferreira T, Morris AP, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet*. 2012;44(4):369-375. doi:10.1038/ng.2213

15. Malek N, Weil RS, Bresner C, et al. Features of GBA -associated Parkinson's disease at presentation in the UK Tracking Parkinson's study. *J Neurol Neurosurg Psychiatry*. 2018;89:702-709. doi:10.1136/jnnp-2017-317348

16. den Heijer JM, Cullen VC, Quadri M, et al. A Large-Scale Full GBA1 Gene Screening in Parkinson's Disease in the Netherlands. *Mov Disord*. 2020. doi:10.1002/mds.28112

17. 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):2911-2935. doi:10.1002/sim.2165

18. Iwaki H, Blauwendraat C, Leonard HL, et al. Genomewide association study of Parkinson's disease clinical biomarkers in 12 longitudinal patients' cohorts. *Mov Disord*. 2019;(July):1-12. doi:10.1002/mds.27845

19. Gan-Or Z, Liong C, Alcalay RN. GBA-Associated Parkinson's Disease and Other Synucleinopathies. *Curr Neurol Neurosci Rep*. 2017;18(8).

592 doi:10.1007/s11910-018-0860-4

593 20. Williams-Gray CH, Goris A, Saiki M, et al. Apolipoprotein e genotype as a risk
594 factor for susceptibility to and dementia in Parkinson's Disease. *J Neurol*.
595 2009;256(3):493-498. doi:10.1007/s00415-009-0119-8

596 21. Blauwendraat C, Heilbron K, Vallerga CL, et al. Parkinson's disease age at
597 onset genome-wide association study: Defining heritability, genetic loci, and α-
598 synuclein mechanisms. *Mov Disord*. 2019;34(6):866-875.
599 doi:10.1002/mds.27659

600 22. Nalls MA, McLean CY, Rick J, et al. Diagnosis of Parkinson's disease on the
601 basis of clinical and genetic classification: A population-based modelling study.
602 *Lancet Neurol*. 2015;14(10):1002-1009. doi:10.1016/S1474-4422(15)00178-7

603 23. Schrag A, Spottke A, Quinn NP, Dodel R. Comparative responsiveness of
604 Parkinson's disease scales to change over time. *Mov Disord*. 2009;24(6):813-
605 818. doi:10.1002/mds.22438

606 24. Nombela C, Rowe JB, Winder-Rhodes SE, et al. Genetic impact on cognition
607 and brain function in newly diagnosed Parkinson's disease: ICICLE-PD study.
608 *Brain*. 2014;137(10):2743-2758.

609 25. Morley JF, Xie SX, Hurtig HI, et al. Genetic influences on cognitive decline in
610 Parkinson's disease. *Mov Disord*. 2012;27(4):512-518.

611 26. Mata IF, Leverenz JB, Weintraub D, et al. APOE, MAPT, and SNCA genes
612 and cognitive performance in Parkinson disease. *JAMA Neurol*. 2014;71(11).

613 27. O'Donoghue MC, Murphy SE, Zamboni G, Nobre AC, Mackay CE. APOE
614 genotype and cognition in healthy individuals at risk of Alzheimer's disease: A
615 review. *Cortex*. 2018;104:103-123. doi:10.1016/j.cortex.2018.03.025

616 28. Smith C, Malek N, Grosset K, Cullen B, Gentleman S, Grosset DG.
617 Neuropathology of dementia in patients with Parkinson's disease: A systematic
618 review of autopsy studies. *J Neurol Neurosurg Psychiatry*. 2019;90(11):1234-
619 1243. doi:10.1136/jnnp-2019-321111

620 29. Zhao N, Attrebi ON, Ren Y, et al. APOE4 exacerbates alpha-synuclein
621 pathology and related toxicity independent of amyloid. *Sci Transl Med*.

2020;12:1809.

30. Tsuang D, Leverenz JB, Lopez OL, et al. APOE ε4 increases risk for dementia in pure synucleinopathies. *JAMA Neurol*. 2013;70(2):223-228. doi:10.1001/jamaneurol.2013.600

31. Moreno-Grau S, Hernández I, Heilmann-Heimbach S, et al. Genome-wide significant risk factors on chromosome 19 and the APOE locus. *Oncotarget*. 2018;9(37):24590-24600. doi:10.18632/oncotarget.25083

32. 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. doi:10.1016/j.bbadis.2005.04.006

33. Verschuur CVM, Suwijn SR, Boel JA, et al. Randomized delayed-start trial of levodopa in Parkinson's disease. *N Engl J Med*. 2019;380(4):315-324. doi:10.1056/NEJMoa1809983

34. Brockmann K, Srulijes K, Pflederer S, et al. GBA-associated Parkinson's disease: Reduced survival and more rapid progression in a prospective longitudinal study. *Mov Disord*. 2015;30(3):407-411. doi:10.1002/mds.26071

35. Winder-Rhodes SE, Evans JR, Ban M, et al. Glucocerebrosidase mutations influence the natural history of Parkinson's disease in a community-based incident cohort. *Brain*. 2013;136(2):392-399. doi:10.1093/brain/aws318

36. Crosiers D, Verstraeten A, Wauters E, 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*. 2016;629:160-164. doi:10.1016/j.neulet.2016.07.008

37. Davis MY, Johnson CO, Leverenz JB, et al. Association of GBA mutations and the E326K polymorphism with motor and cognitive progression in parkinson disease. *JAMA Neurol*. 2016;73(10):1217-1224. doi:10.1001/jamaneurol.2016.2245

38. Cilia R, Tunesi S, Marotta G, et al. Survival and dementia in GBA -associated Parkinson Disease : the mutation matters . *Ann Neurol*. 2016;80:662-673.

652 39. Alcalay RN, Caccappolo E, Mejia-Santana H, et al. Cognitive performance of
653 GBA mutation carriers with early-onset PD: the CORE-PD study. *Neurology*.
654 2012;78:1434-1440. doi:10.1212/WNL.0b013e318253d54b

655 40. Iwaki H, Blauwendraat C, Leonard HL, et al. Genetic risk of Parkinson disease
656 and progression: An analysis of 13 longitudinal cohorts. *Neurol Genet*.
657 2019;5(4). doi:10.1212/NXG.0000000000000348

658 41. Evans JR, Mason SL, Williams-Gray CH, et al. The natural history of treated
659 Parkinson's disease in an incident, community based cohort. *J Neurol*
660 *Neurosurg Psychiatry*. 2011;82(10):1112-1118. doi:10.1136/jnnp.2011.240366

661 42. Williams-Gray CH, Evans JR, Goris A, et al. The distinct cognitive syndromes
662 of Parkinson's disease: 5 year follow-up of the CamPaIGN cohort. *Brain*.
663 2009;132(11):2958-2969. doi:10.1093/brain/awp245

664 43. Goris A, Williams-Gray CH, Clark GR, et al. Tau and alpha-synuclein in
665 susceptibility to, and dementia in, Parkinson's disease. *Ann Neurol*.
666 2007;62(2):145-153. doi:10.1002/ana.21192

667 44. Maple-Grødem J, Chung J, Aaser K, et al. Alzheimer disease associated
668 variants in SORL1 accelerate dementia development in Parkinson disease.
669 *Neurosci Lett*. 2018;674:123-126. doi:10.1016/j.neulet.2018.03.036

670 45. Fagan ES, Pihlstrøm L. Genetic risk factors for cognitive decline in Parkinson's
671 disease: a review of the literature. *Eur J Neurol*. 2017;24(4):561-e20.
672 doi:10.1111/ene.13258

673 46. Ritz B, Rhodes SL, Bordelon Y, Bronstein J. Alpha-Synuclein genetic variants
674 predict faster motor symptom progression in idiopathic Parkinson disease.
675 *PLoS One*. 2012;7(5). doi:10.1371/journal.pone.0036199

676 47. Wang G, Huang Y, Wei Chen, et al. Variants in the SNCA gene associate with
677 motor progression while variants in the MAPT gene associate with the severity
678 of Parkinson's disease. *Park Relat Disord*. 2016;24:89-94.
679 doi:10.1016/j.parkreldis.2015.12.018

680 48. Markopoulou K, Biernacka JM, Armasu SM, et al. Does α-synuclein have a
681 dual and opposing effect in preclinical vs. clinical Parkinson's disease? *Park*

Relat Disord. 2014;20(6):584-589. doi:10.1016/j.parkreldis.2014.02.021

49. 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):271-276. doi:10.3233/JPD-2011-11027

50. Chung SJ, Armasu SM, Biernacka JM, et al. Genomic determinants of motor and cognitive outcomes in Parkinson's disease. *Park Relat Disord*. 2012;18(7):881-886.

51. Latourelle JC, Beste MT, Hadzi TC, 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*. 2017;16(11):908-916. doi:10.1016/S1474-4422(17)30328-9

52. Hauser RA, Grosset DG. [123I]FP-CIT (DaTscan) SPECT brain imaging in patients with suspected parkinsonian syndromes. *J Neuroimaging*. 2012;22(3):225-230. doi:10.1111/j.1552-6569.2011.00583.x

53. Sanchez-Contreras MY, Kouri N, Cook CN, 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. doi:10.1186/s13024-018-0267-3

54. Jabbari E, Tan MMX, Reynolds RH, et al. Common variation at the LRRK2 locus is associated with survival in the primary tauopathy progressive supranuclear palsy. *bioRxiv*. 2020. doi:https://doi.org/10.1101/2020.02.04.932335

55. Adler CH, Beach TG, Hentz JG, et al. Low clinical diagnostic accuracy of early vs advanced Parkinson disease: Clinicopathologic study. *Neurology*. 2014;83(5):406-412. doi:10.1212/WNL.0000000000000641

56. Malek N, Lawton MA, Swallow DMA, 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):1518-1526.

57. 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.

doi:10.1186/alzrt109

58. Masliah E, Rockenstein E, Veinbergs I, 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):12245-12250. doi:10.1073/pnas.211412398

Figure legends

Figure 1. Steps to create composite, motor, and cognitive progression scores.

Figure 2. Manhattan plot for GWAS of composite progression. The red dashed line indicates the genome-wide significance threshold p-value 5×10^{-8}. The top genes from the MAGMA gene-based analysis and p values are shown on the right.

Figure 3. Manhattan plot for the GWAS of motor progression. Genome-wide significance is the standard p-value 5×10^{-8} (not indicated in the figure). The top genes from the MAGMA gene-based analysis and p values are shown on the right.

Figure 4. 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}. The top genes from the MAGMA gene-based analysis and p values are shown on the right.