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. The objective of the current study was to identify genetic variants associated with PD progression.

Methods: We analyzed 3 large longitudinal cohorts: Tracking Parkinson’s, Oxford Discovery, and the Parkinson’s Progression Markers Initiative. We included clinical data for 3364 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, in which we combined multiple assessments using a principal components analysis to derive scores for composite, motor, and cognitive progression. These scores were analyzed in linear regression in genome-wide association studies. We also performed a targeted analysis of the 90 PD risk loci from the latest 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 ε4 region.

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 ε4 region.

Relevant conflicts of interest/financial disclosures: None to report.

Funding agencies: Parkinson’s UK.

Received: 29 April 2020; Revised: 10 September 2020; Accepted: 5 October 2020

Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.28342
Progression in Parkinson’s disease (PD) is heterogeneous, with some patients progressing rapidly, whereas others remain relatively stable over time. 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 the 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. However, the majority of PD GWASs have compared cases with healthy controls to identify variants linked to disease status. 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, and there is no gold standard measure of progression, although the Movement Disorder Society Unified Parkinson’s Disease Rating Scale (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, including rater subjectivity and practice effects in cognitive assessments. Therefore, combining multiple measures may improve the accuracy of measuring progression, as shown in the Huntington’s disease (HD) progression GWAS. In this study, we analyzed data from 3 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 analyzed 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 3 data sets were merged after imputation, with only shared variants retained. Genetic principal components were generated and outliers removed (see Supplementary Methods and Figs. 1 and 2).

Clinical Outcome Measures

Individual-level data from the cohorts were merged. To increase the power and the accuracy of the final progression scores, we performed all transformations and created progression scores from the merged data set as follows (Fig. 1).

Motor progression was assessed using 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). In PPMI, we used motor assessments conducted in the “off” medication state.

Cognitive progression was assessed using the Montreal Cognitive Assessment, 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 standardized 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 as covariates. We included terms for subject random effects to account for individual heterogeneity in the intercept (baseline value) and slope (rate of progression).

We used random-effect slope values as the measure of “residual” progression not predicted by age at onset, cohort, sex, and education, for each individual. We performed PCA on these values after zero centering and scaling to have unit variance. The final progression scores from the PCA relate to the variability explained,
and therefore the direction cannot be strictly interpreted. Patients who were missing clinical data (e.g. MDS-UPDRS part III total) at all visits were not included in the PCA and subsequent GWAS analysis.

**Removal of Non-PD Cases**

Any patients who were diagnosed with a different condition during follow-up were removed from analyses. We also conducted sensitivity analyses to remove any cases that may have non-PD conditions but an alternative diagnosis had not yet been confirmed. First, we removed patients in Tracking Parkinson’s and Oxford Discovery who had a clinician-rated diagnostic certainty of PD < 90%. Second, 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 genotyping data (to adjust for population substructure). GWASs were conducted in rvtests using the single-variant Wald test. Genome-wide complex trait analysis conditional and joint analysis (GCTA-COJO) was used to identify independent signals. 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, and we analyzed the association with each progression score using linear regression.

**GBA**

We analyzed GBA rare variant carriers compared with noncarriers 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. We analyzed GBA status in relation to the progression scores using linear regression, adjusting for cohort and the first 5 genetic principal components.

**Levodopa-Equivalent Daily Dose-Adjusted Sensitivity Analyses**

Medication may affect MDS-UPDRS part III scores, in particular in Tracking Parkinson’s and Oxford
Discovery, in which patients were assessed in the “on” state. To address this, we performed a sensitivity analysis adjusting for levodopa-equivalent daily dose (LEDD), as described in a previous study, in which we estimated the effect of levodopa on MDS-UPDRS part III scores (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 and most likely with higher doses.

Results

We included clinical data for 3364 PD patients with 12,144 observations (Table 1). Mean follow-up time ± SD was 4.2 ± 1.5 years, and mean disease duration at study entry was 2.9 ± 2.6 years. A total of 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 (Figs. S3–S6).

We found that the first principal components for motor and cognitive progression were moderately correlated ($r = -0.35$, $P < 2.2 \times 10^{-16}$; Table S1). 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 (Figs. S7 and S8). Tables S2–S6 show how the raw scales and the motor, cognitive, and composite principal components are correlated. None of the principal components were associated with cohort (all $P > 0.9$).

GWAS of Motor Progression

A total of 2848 PD patients had motor progression scores and genotype data. No variants passed genome-wide significance (Fig. 3, Table S8). However, in the gene-based test, $ATP8B2$ in chromosome 1 was associated with motor progression ($P = 5.3 \times 10^{-6}$; Figs. S12 and S13), although this did not reach significance correcting for the number of mapped genes ($P = 2.81 \times 10^{-6}$).

We conducted follow-up GWASs in each cohort separately (Table S9) 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 (Table S10).

Our top variant in chromosome 1, rs35950207, was associated with motor progression, $P = 5.0 \times 10^{-6}$. We examined the associations for this SNP in the previous progression GWAS (https://pdgenetics.shinyapps.io/pdprogmetagwasbrowser/); rs35950207 was not significantly associated with binomial analysis of Hoehn and Yahr stage 3 or more at baseline ($\beta = 0.27$, $P = 0.03$).

The variant rs35950207 is 2 kb upstream of $AQP10$. It is an expression quantitative trait loci (eQTL) for $AQP10$ in whole blood (GTEx, $P = 1.7 \times 10^{-6}$; eQTLGen, $P = 3.62 \times 10^{-139}$) and other tissues (subcutaneous adipose, skin, esophagus, testis, and heart). It is also an eQTL for $ATP8B2$ in blood (GTEx, $P = 1.5 \times 10^{-5}$; eQTLGen, $P = 7.84 \times 10^{-42}$) and in the cerebellum (GTEx, $P = 7.8 \times 10^{-5}$). $GBA$ is also located in chromosome 1, and $GBA$ variants are associated with both PD risk and progression. 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).

In chromosome 5, the top SNP in the variant-based analysis was rs17367669, 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

A total of 2788 patients had cognitive progression scores and genotype data. The top variant was rs429358, which tags the $APOE$ e4 allele ($P = 2.53 \times 10^{-13}$; Fig. 4, Table S11, Figs. S14 and S15). Figure S16 shows that e4 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 (Tables S12 and S13).
When we performed conditional analysis on the top SNP, rs429358, a group of SNPs still passed genome-wide significance, indicating independent signals (Fig. S17). 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 harbor the C alleles, then this codes the e4 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.

### TABLE 1. Cohort demographics at baseline

<table>
<thead>
<tr>
<th>Demographics at baseline</th>
<th>Tracking Parkinson’s</th>
<th>Oxford Discovery</th>
<th>PPMI</th>
<th>ALL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of PD patients</td>
<td>1966</td>
<td>985</td>
<td>413</td>
<td>3364</td>
</tr>
<tr>
<td>Total number of visits analyzed</td>
<td>5936</td>
<td>3142</td>
<td>3066</td>
<td>12,144</td>
</tr>
<tr>
<td>Mean length of follow-up (years)</td>
<td>3.8 (1.4)</td>
<td>4.3 (1.7)</td>
<td>5.4 (1.2)</td>
<td>4.2 (1.5)</td>
</tr>
<tr>
<td>Male (%)</td>
<td>65.2%</td>
<td>64.2%</td>
<td>65.4%</td>
<td>65.0%</td>
</tr>
<tr>
<td>Age at onset (years)</td>
<td>64.4 (9.8)</td>
<td>64.5 (9.8)</td>
<td>59.5 (10.0)</td>
<td>63.9 (10.0)</td>
</tr>
<tr>
<td>Age at diagnosis (years)</td>
<td>66.3 (9.3)</td>
<td>66.1 (9.6)</td>
<td>61.0 (9.7)</td>
<td>65.6 (9.6)</td>
</tr>
<tr>
<td>Age at study entry (years)</td>
<td>67.6 (9.3)</td>
<td>67.4 (9.6)</td>
<td>61.5 (9.8)</td>
<td>66.8 (9.7)</td>
</tr>
<tr>
<td>Disease duration — time from symptom onset to assessment (years)</td>
<td>3.2 (3.0)</td>
<td>2.9 (1.9)</td>
<td>2.0 (2.0)</td>
<td>2.9 (2.6)</td>
</tr>
<tr>
<td>Time from diagnosis to assessment (years)</td>
<td>1.3 (0.9)</td>
<td>1.3 (0.9)</td>
<td>0.5 (0.5)</td>
<td>1.2 (0.9)</td>
</tr>
<tr>
<td>MDS-UPDRS part III</td>
<td>22.9 (12.3)</td>
<td>26.8 (11.1)</td>
<td>20.7 (8.8)</td>
<td>23.8 (11.7)</td>
</tr>
<tr>
<td>MDS-UPDRS part III annual change$^a$</td>
<td>1.9 (3.7)</td>
<td>2.1 (3.5)</td>
<td>1.8 (2.2)</td>
<td>2.1 (6.2)</td>
</tr>
<tr>
<td>MDS-UPDRS part II</td>
<td>9.9 (6.6)</td>
<td>8.9 (6.2)</td>
<td>5.8 (4.1)</td>
<td>9.0 (6.3)</td>
</tr>
<tr>
<td>MDS-UPDRS part II annual change$^a$</td>
<td>1.3 (1.6)</td>
<td>1.3 (1.6)</td>
<td>0.9 (1.1)</td>
<td>1.3 (2.8)</td>
</tr>
<tr>
<td>Hoehn and Yahr stage mean$^b$</td>
<td>1.8 (0.6)</td>
<td>1.9 (0.6)</td>
<td>1.6 (0.5)</td>
<td>1.8 (0.6)</td>
</tr>
<tr>
<td>Hoehn and Yahr stage annual change</td>
<td>0.1 (0.2)</td>
<td>0.06 (0.1)</td>
<td>0.08 (0.1)</td>
<td>0.06 (0.3)</td>
</tr>
<tr>
<td>Hoehn and Yahr stage 0 to 1.5 (%)</td>
<td>48.1%</td>
<td>23.2%</td>
<td>44.8%</td>
<td>40.4%</td>
</tr>
<tr>
<td>Hoehn and Yahr stage 2 to 2.5 (%)</td>
<td>45.1%</td>
<td>68.8%</td>
<td>54.7%</td>
<td>53.2%</td>
</tr>
<tr>
<td>Hoehn and Yahr stage 3 (%)</td>
<td>6.8%</td>
<td>8.1%</td>
<td>0.5%</td>
<td>6.4%</td>
</tr>
<tr>
<td>MoCA total (adjusted for education)</td>
<td>24.9 (3.6)</td>
<td>24.5 (3.5)</td>
<td>27.1 (2.3)</td>
<td>25.0 (3.6)</td>
</tr>
<tr>
<td>MoCA total annual change $^c$</td>
<td>$-0.1$ (0.9)</td>
<td>$-0.1$ (0.8)</td>
<td>$-0.2$ (0.6)</td>
<td>$-0.1$ (1.5)</td>
</tr>
<tr>
<td>Semantic fluency$^c$</td>
<td>21.8 (6.9)</td>
<td>34.7 (9.0)</td>
<td>21.0 (5.4)</td>
<td>25.5 (9.5)</td>
</tr>
<tr>
<td>Semantic fluency annual change $^c$</td>
<td>$-0.2$ (1.5)</td>
<td>$-0.5$ (2.0)</td>
<td>$-0.1$ (0.9)</td>
<td>$-0.5$ (3.0)</td>
</tr>
<tr>
<td>MDS-UPDRS part I.1</td>
<td>0.5 (0.7)</td>
<td>0.5 (0.6)</td>
<td>0.3 (0.5)</td>
<td>0.5 (0.7)</td>
</tr>
<tr>
<td>MDS-UPDRS part I.1 annual change</td>
<td>0.07 (0.2)</td>
<td>0.05 (0.2)</td>
<td>0.07 (0.1)</td>
<td>0.05 (0.3)</td>
</tr>
</tbody>
</table>

SD, standard deviation; PPMI, Parkinson’s Progression Markers Initiative; PD, Parkinson’s disease; MDS-UPDRS, Movement Disorder Society Unified Parkinson’s Disease Rating Scale; MoCA, Montreal Cognitive Assessment. Mean (SD) shown unless otherwise indicated.

$^a$Annual change score derived from a mixed-effects model of the raw scores as a function of years from onset, with subject random effects to account for individual heterogeneity in the intercept (baseline values) and slope (rate of progression). No other covariates were included in the model within each cohort. For the overall value, we adjusted for cohort and the interaction between cohort and years from onset.

$^b$Tracking Parkinson’s used the modified Hoehn and Yahr stage scale, whereas Oxford Discovery and PPMI used the original scale. Hoehn and Yahr stage proportions are shown as a total of the number of people with nonmissing Hoehn and Yahr ratings at baseline.

$^c$Instructions and timing for the semantic fluency task were slightly different between cohorts (completed within 60 or 90 seconds). To account for these differences, we standardized all scales within each cohort separately (see Methods section).

**Fig. 2.** Manhattan plot for GWAS of composite progression. The red dashed line indicates the genome-wide significance threshold, $P = 5 \times 10^{-8}$. The top genes from the MAGMA gene-based analysis and $P$ values are shown on the right. [Color figure can be viewed at wileyonlinelibrary.com]
We found frequencies of APOE genotypes similar to those of previous studies\(^2\) (Table S14).

**LEDG-Adjusted Analyses**

When we performed GWASs of composite progression and motor progression after adjusting for LEDG, 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 (beta = 0.33, \(P = 8.8 \times 10^{-8}\)). For motor progression, the top SNP was also the same as in the main analysis and \(A\)T\(P\)8\(B\)2 and \(A\)QP10 still the top genes in the MAGMA gene analysis, although not genome-wide significant.

**Sex-Stratified Analyses**

The APOE locus passed genome-wide significance only in men for composite progression and cognitive progression (\(P < 5 \times 10^{-8}\)). 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\(^2\) \(73\) were present in our final data set, including the SNCA and TMEM175/GAK variants associated with PD age at onset\(^3\). No variants passed analysis-wide significance (\(P = 0.05/73\)). Variants with at least 1 association, \(P < 0.05\), are shown in Figure S18.

We found that only a small number of risk variants were associated with progression, with \(P < 0.05\). The variant 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), \(D^\prime = 0.90, R^2 = 0.78\).

We also extracted results for other candidate variants that have been implicated in PD progression (Fig. S19). We did not find that the top variant, rs382940, in \(S\)LC44A1 that was associated in progression to Hoehn and Yahr stage 3 from the Iwaki GWAS\(^1\) was

---

**FIG. 3.** Manhattan plot for the GWAS of motor progression. Genome-wide significance is the standard \(P = 5 \times 10^{-8}\) (not indicated in the figure). The top genes from the MAGMA gene-based analysis and \(P\) values are shown on the right. [Color figure can be viewed at wileyonlinelibrary.com]

**FIG. 4.** Manhattan plot for the variant-based GWAS of cognitive progression. The red dashed line indicates the genome-wide significance threshold, \(P = 5 \times 10^{-8}\). The top genes from the MAGMA gene-based analysis and \(P\) values are shown on the right. [Color figure can be viewed at wileyonlinelibrary.com]
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 Linkage Disequilibrium Score Regression (LDSC) results also suggested very little overlap between each of the progression GWASs and PD case-control GWAS (all Ps > 0.5).

**PD Genetic Risk Score**

A total of 73 PD risk SNPs were present in our genotype data, and 2 proxies were identified for missing variants (Table S15). 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 2020 patients from Tracking Parkinson’s and PPMI. 194 (9.6%) carried a pathogenic variant in GBA (Table S16). 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 the larger P values, although the direction of effects were the same (Tables S17 and S18).

**Discussion**

We used a new method of analyzing 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 with 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. 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. In PD, there are many different scales for assessing symptoms. Each scale has a degree of measurement error and different sensitivity to progression of underlying symptoms. 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 2 main findings. First, we replicated previous findings for APOE ɛ4. Many studies have shown that the ɛ4 allele is associated with dementia in PD, and potentially separately from the risk of Alzheimer’s disease (AD). 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. Alternatively, APOE may drive cognitive decline independently of amyloid/AD pathology. Recent animal model work has shown that the ɛ4 allele is independently associated with α-synuclein pathology and toxicity. In addition, the ɛ4 allele is overrepresented in dementia with Lewy body cases with “pure” Lewy body pathology, compared with PDD cases. A systematic review showed that limbic and neocortical α-synuclein pathology had the strongest association with PD dementia. 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, in which multiple risk loci have been 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 genotype, as has been done in AD. 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. 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-UPDRS part III but does not influence the rate of progression, and this was shown in a previous study. We also found that the mean rate of change in MDS-UPDRS part III was comparable in 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; 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 at an earlier stage of the disease. 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. 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. 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. We have used different methods and a different phenotype to analyze PD progression. Further progression GWASs are needed to replicate both sets of results, and other metrics for PD progression could be analyzed, such as mortality.

Although 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, COMT, BDNF, MTHFR, and SORL1 have been reported to influence cognitive decline (reviewed in Fagan and Pihlstrom). 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. A small GWAS of motor and cognitive progression identified suggestive loci in C8orf4 and CLRN3, although these have not been replicated. A novel machine-learning approach found that variation in LINGO2 was associated with change in the MDS-UPDRS, although again this finding needs independent replication. We did not replicate these findings, possibly because we were underpowered as a GWAS to detect variants with smaller effects or because we have analyzed progression using different methods. However, many of these previous 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 that 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, analysis of PD progression is more complex because of 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 were not available for PPMI, and abnormal dopamine transporter scans cannot differentiate between PD and other degenerative parkinsonian conditions. Despite this, our sensitivity analysis suggests 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, or rs2242367 near LRRK2 associated with PSP progression.

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-year duration of disease; 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 because of comorbidities such as vascular burden or interactions between synuclein and copathologies (such as amyloid, and tau) in the rapid progressors that 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 that may be important. These advances are essential to understanding the biology of disease progression and nominating therapeutic targets to stop or slow PD progression.
Acknowledgments: 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) Dementia 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 the University of Oxford. This research was supported by the National Institute for Health Research University College London Hospitals Biomedical Research Centre. This research was also supported in part by the Intramural Research Program of the NIH, National Institute on Aging. This research was supported in part by the RCUK/UKRI Research Innovation Fellowship (Medical Research Council) and the NIHR Cambridge Biomedical Research Centre Dementia and Neurodegeneration theme (C. H.W-G). Work in Cambridge was funded in part through the NIHR Bio-Medical Research Centre as well as funding from Parkinson’s UK and Cure Parkinson’s Trust. 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/).

Data Availability Statement
Anonymized 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). The PPMI data are publicly available on application (https://www.ppmi-info.org/access-data-specimens/download-data/). Code is available at https://github.com/huw-morris-lab/PCD-PCA-progression-GWAS.

References

Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher’s web-site.
Author Contributions


M.M.X.T.: 1A, 1B, 1C, 2A, 2B, 2C, 3A, 3B.
M.A.L.: 1B, 2C, 3B.
E.J.: 2C, 3B.
R.H.R.: 2C, 3B.
H.L.: 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.

Financial Disclosures for the Preceding 12 Months

M.M.X.T. is supported by 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. is 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, and Boehringer and honoraria from UCB Pharma, GE Healthcare, Lily Pharma, and 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 the Michael J. Fox Foundation. T.F. has received grants from the Michael J. Fox Foundation, Cure Parkinson’s Trust, Brain Research Trust, John Black Charitable Foundation, and 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 an 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, and has received grants from the Michael J Fox Foundation, Parkinson’s UK, the Evelyn Trust, Cure Parkinson’s Trust, and the Cambridge Centre for Parkinson’s Plus. C.H.W-G. has also received consultancy payments from Modus Outcomes and Evidera. 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.