nature genetics

LETTERS

https://doi.org/10.1038/s41588-021-00847-6

Check for updates

Genome-wide survival study identifies a novel synaptic locus and polygenic score for cognitive progression in Parkinson's disease

Ganqiang Liu[1,2,3], Jiajie Peng[1,2,4], Zhixiang Liao[1,2], Joseph J. Locascio[1,2,5], Jean-Christophe Corvol[6], Frank Zhu[1,2], Xianjun Dong[1,2], Jodi Maple-Grødem[7,8], Meghan C. Campbell[9], Alexis Elbaz[10], Suzanne Lesage[6], Alexis Brice[6], Graziella Mangone[6], John H. Growdon[5], Albert Y. Hung[5], Michael A. Schwarzschild[5], Michael T. Hayes[1,11], Anne-Marie Wills[5], Todd M. Herrington[5], Bernard Ravina[12], Ira Shoulson[13], Pille Taba[14], Sulev Kõks[15,16], Thomas G. Beach[17], Florence Cormier-Dequaire[6], Guido Alves[7,8,18], Ole-Bjørn Tysnes[19,20], Joel S. Perlmutter[9,21,22], Peter Heutink[23], Sami S. Amr[24], Jacobus J. van Hilten[25], Meike Kasten[26,27], Brit Mollenhauer[28,29], Claudia Trenkwalder[29,30], Christine Klein[31], Roger A. Barker[32,33], Caroline H. Williams-Gray[32], Johan Marinus[25], International Genetics of Parkinson Disease Progression (IGPP) Consortium* and Clemens R. Scherzer[1,2,5,11] ✉

A key driver of patients' well-being and clinical trials for Parkinson's disease (PD) is the course that the disease takes over time (progression and prognosis). To assess how genetic variation influences the progression of PD over time to dementia, a major determinant for quality of life, we performed a longitudinal genome-wide survival study of 11.2 million variants in 3,821 patients with PD over 31,053 visits. We discover *RIMS2* as a progression locus and confirm this in a replicate population (hazard ratio (HR) = 4.77, $P = 2.78 \times 10^{-11}$), identify suggestive evidence for *TMEM108* (HR = 2.86, $P = 2.09 \times 10^{-8}$) and *WWOX* (HR = 2.12, $P = 2.37 \times 10^{-8}$) as progression loci, and confirm associations for *GBA* (HR = 1.93, $P = 0.0002$) and *APOE* (HR = 1.48, $P = 0.001$). Polygenic progression scores exhibit a substantial aggregate association with dementia risk, while polygenic susceptibility scores are not predictive. This study identifies a novel synaptic locus and polygenic score for cognitive disease progression in PD and proposes diverging genetic architectures of progression and susceptibility.

The past decade has seen success in identifying genetic variants linked to susceptibility for common disease from genome-wide association studies (GWAS) through time-static, two-group comparisons of unaffected controls and cases captured in one single snapshot of time[1,2]. The genetic architecture of progression and prognosis, which are fundamental for patients, has not been established. Which genes determine whether a patient will have an aggressive or benign course, and which variants influence who will develop dementia? To shift from the genetics of susceptibility to precision medicine, longitudinal designs[3] are needed that examine the critical time dimension and provide information about individual change[4].

The number of patients with PD is projected to double to 14 million worldwide by 2040 (ref. [5]). The pace of progression varies considerably between patients[6–8]. Parkinson's disease dementia (PDD) is one of the most debilitating manifestations of disease progression in PD[9] with the greatest influence on quality of life[9], caregivers and health costs[10]. In clinical trials, the heterogeneity of progression rates obfuscates drug effects. None of the existing PD therapies slow the underlying neuropathology, which relentlessly advances from brainstem to cortex[11] and clinically correlates with progression from motor to cognitive symptoms[12].

Limited evidence exists on the genetic architecture of cognitive decline in PD beyond the *GBA* (β-glucocerebrosidase) locus established by us[6,7,13] and others[14]. *APOE* (apolipoprotein E) is implicated chiefly based on cross-sectional studies[15]. Evidence for other candidate genes and GWAS-derived susceptibility variants is controversial (for example, *LRRK2*[16], *SNCA*[17,18], *MAPT*[13,19] and others[20,21]).

We determined the effects of 11.2 million deeply imputed variants on cognitive decline in 4,872 patients with PD in 15 cohorts[13,22–30] from North America and Europe between 1986 and 2017, who were prospectively assessed with 36,123 study visits (Supplementary Fig. 1 and Supplementary Table 1). We evaluated thousands more patients, tens of thousands more follow-up visits and millions more SNPs than previous longitudinal explorations[7,20,21], and confirmed associations in an independent replicate population. We performed whole-genome genotyping on our cohorts with the new-generation, high-density Illumina Infinium Multi-Ethnic Global Array that harnesses content from Phase 3 of the 1000 Genomes Project[31] and transancestry tagging strategies to maximize imputation accuracy for low-frequency variants. Imputation[32–35] provides power of detection comparable to whole-genome sequencing (WGS) for low frequency (minor allele frequency (MAF) of ≥1% but <5%)[35] and common variants[35]. We genotyped 1.8 million variants and imputed 11.2 million variants (Methods and Supplementary Fig. 2). Concordance of imputation compared to WGS was 99.4% based on 562 samples probed with both methods (Supplementary Fig. 3).

A full list of affiliations appears at the end of the paper.

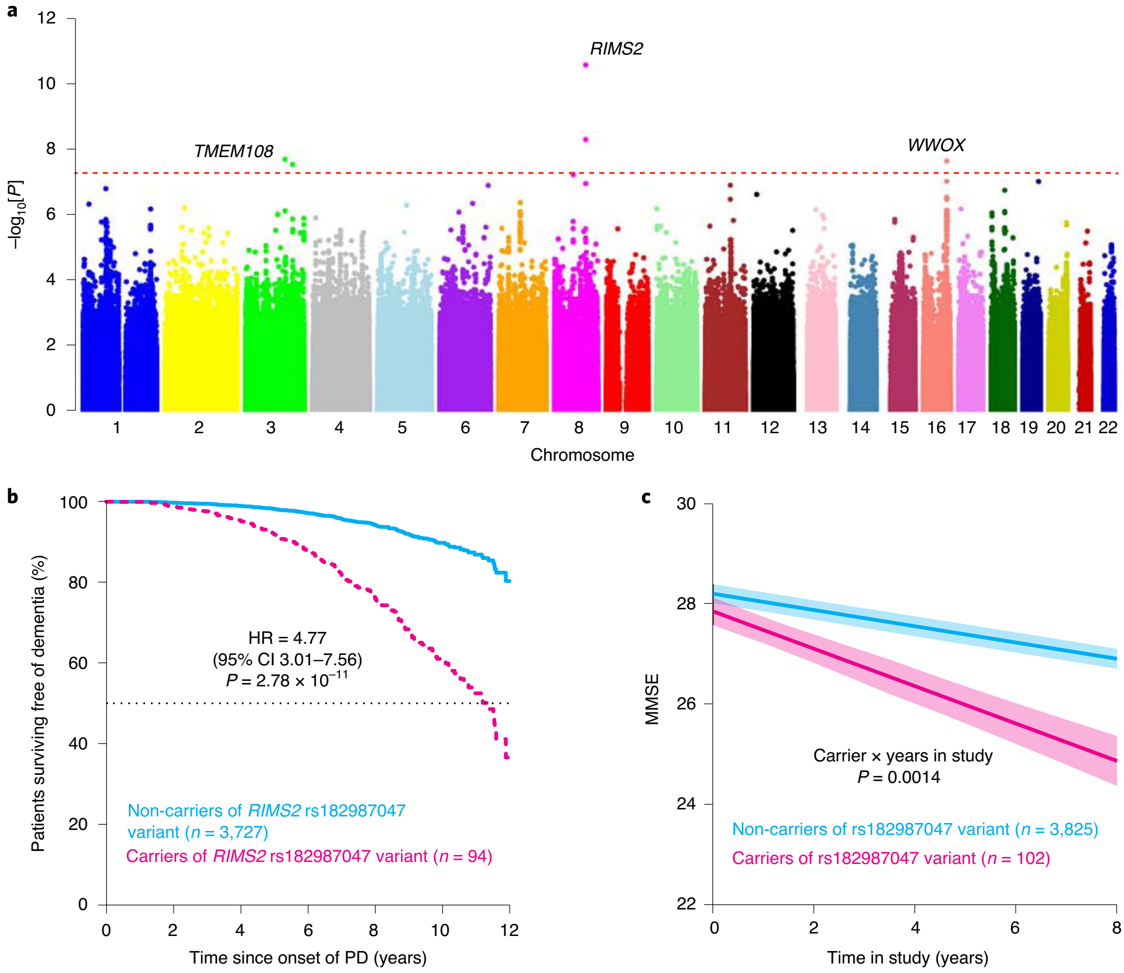


Fig. 1 | Within-cases longitudinal GWSS identifies three loci associated with progression to PDD. a, Manhattan plot of the GWSS. $-\log_{10}[P$ value$]$ from the Cox proportional hazards (Cox PH) model with two-sided Wald test for 12-year survival free of dementia are plotted against chromosomal position for the combined population ($n=3{,}821$ cases with PD tracked in 31,053 longitudinal visits for up to 12 years). Each point represents a SNP. The dashed red line corresponds to the genome-wide significance threshold. **b**, Covariate-adjusted survival curves for patients with PD without the *RIMS2* rs182987047 variant (light blue line) and for those carrying the variant (dashed magenta line). Cox PH model with two-sided Wald test. **c**, Adjusted mean MMSE scores across time predicted from the estimated fixed-effect parameters in the LMM analysis are shown for cases carrying the *RIMS2* rs182987047 variant (magenta) and cases without the variant (non-carriers; light blue) adjusting for covariates. Shaded ribbons indicate ± standard error of the mean (s.e.m.) across time. *P* values from LMM.

A total of 4,491 samples passed genotyping quality control. Patients were left-censored, and those with missing or non-quality clinical data were excluded ($n=670$, Extended Data Fig. 1). To identify genetic variants associated with progression from PD to PDD (Supplementary Table 2), we performed a longitudinal genome-wide survival study (GWSS) (Fig. 1 and Methods) on the remaining 3,821 patients. We assigned 2,650 patients and 11,744 visits to the discovery population. 1,171 patients and 19,309 visits comprised the replicate population. We used a GWSS to estimate the influence of common and low-frequency genetic variants on time from the onset of PD to progression to the endpoint of PDD. We performed Cox proportional hazards analyses adjusting for age at onset, sex, years of education at enrollment, ten principal components of genetic population substructure, and a 'cohort' term as a random effect (frailty model[36]). Physicians recruited and longitudinally assessed the participants without knowledge of their genotypes.

An association signal in the *RIMS2* locus reached genome-wide significance in the discovery population and was confirmed in the replicate population (Fig. 1 and Table 1). The genomic control inflation factor (λ_{GC}) was 1.067 in the combined analysis (Supplementary Fig. 4), and the linkage disequilibrium (LD) score regression intercept was lower (1.057)[37], consistent with a contribution of polygenicity to inflation[37,38]. The lead variant rs182987047 in the *RIMS2* locus (NC_000008.10:g.105249272A>T; Extended Data Fig. 2) was associated with progression to PDD with HR = 4.74 (95% confidence interval (95% CI) 2.87–7.83) and $P=1.16\times10^{-9}$ in the discovery cohort. This was confirmed in the replicate population with a HR = 6.2 (95% CI 1.78–21.29) with $P=0.004$). In the combined analysis, the lead *RIMS2* variant showed HR = 4.77 (95% CI 3.01–7.56) and $P=2.78\times10^{-11}$ (Fig. 1a,b). Another linked variant in this locus (rs116918991; NC_000008.10:g.105158401G>A; correlated with $r^2=0.49$) also achieved genome-wide significance in the combined analysis, with $P=5.21\times10^{-9}$ (Extended Data Fig. 3). We next investigated whether a different measure of longitudinal cognitive function would confirm this association. Generalized linear mixed model (LMM) meta-analysis of serial Mini Mental State Exam (MMSE) scores, a measure of global cognitive function in PD[39], in carriers compared to non-carriers confirmed these results. Serial MMSE scores in patients carrying the lead *RIMS2* variant declined more rapidly over time compared to

Table 1 | Variants linked to progression from PD to PDD

Chr.	Position (Mb)	SNP	Risk allele	RAF	HR	95% CI	*P* discovery	*P* replication	*P* combined	Nearest gene
8	**105.25**	**rs182987047**	**T**	**0.013**	**4.77**	**3.01-7.56**	**1.16×10^{-9}**	**4.14×10^{-3}**	**2.78×10^{-11}**	***RIMS2***
3	132.99	rs138073281	C	0.017	2.86	1.98-4.13	3.43×10^{-5}	4.23×10^{-5}	2.09×10^{-8}	*TMEM108*
16	78.28	rs8050111	G	0.066	2.12	1.63-2.75	1.08×10^{-6}	0.01	2.37×10^{-8}	*WWOX*

Hazard ratio for developing PDD in patients with PD carrying a risk allele. Three variants were imputed, and imputation accuracy was confirmed by WGS. Bold font, replicated association; regular font, suggestive associations. Chr., chromosome; RAF, risk allele frequency; HR, hazard ratio from the combined analysis. *P* values from Cox proportional hazards models with two-sided Wald test.

patients who were non-carriers, with $P = 0.0014$ (Fig. 1c) in the LMM adjusting for fixed covariates of age, sex, disease duration upon enrollment, years of education, ten principal components, and random effects (Methods). The *RIMS2* variant was not associated with motor progression (Supplementary Fig. 5), possibly due to power and design limitations (confounding from PD medications, which treat motor symptoms[40] but not dementia).

RIMS2 (chromosome 8) encodes the regulating synaptic membrane exocytosis 2 protein, a RIM family member, which is involved in docking and priming of presynaptic vesicles[41,42]. Mutations in *RIMS2* cause cone-rod synaptic disorder syndrome (MIM 618970)[43]. In mice, knockout of the *RIMS2* ortholog leads to critical defects in memory[44]. The paralog *RIMS1* (chromosome 6) is a PD susceptibility locus[45] that was not associated with progression. Human *RIMS2* showed preferential expression in brain compared to 53 tissues (GTEx[46] v7; Extended Data Fig. 4) with high expression in dopamine and pyramidal neurons laser-captured from 86 and 13 human brains, respectively (BRAINcode[47]; Extended Data Fig. 5).

Two suggestive association signals were located in transmembrane protein 108 (*TMEM108*; NC_000003.11:g.132985956A>C) and WW domain containing oxidoreductase (*WWOX;* NC_000016.9:g.78281160A>G) loci, respectively (Fig. 1a). These loci achieved genome-wide significance ($P < 5 \times 10^{-8}$) in the combined analysis of discovery and replicate populations with suggestive $P < 5 \times 10^{-5}$ in the discovery and $P < 0.05$ in the replication cohort (Table 1). These two loci can now be prioritized for further evaluation. Six additional loci reached genome-wide significance in the discovery cohort but were not replicated (Supplementary Table 3 and Supplementary Fig. 4). The rs138073281 variant in the *TMEM108* locus, which is implicated in synaptic spine formation[48] and cognition[48], was associated with progression to PDD, with HR = 2.86 (95% CI 1.98–4.13) and $P = 2.09 \times 10^{-8}$ in the combined Cox analysis (Table 1). LMM meta-analysis confirmed that patients carrying the *TMEM108* variant had a more rapid decline in serial MMSE scores compared to non-carriers, with $P = 0.0019$ (Extended Data Fig. 3). The *WWOX* locus had HR = 2.12 (95% CI 1.63–2.75) and $P = 2.37 \times 10^{-8}$. *WWOX* is mutated in autosomal recessive ataxia with mental retardation and epilepsy[49], and has been associated with Alzheimer's disease[50] while this manuscript was in preparation. Patients carrying the *WWOX* variant had a more rapid longitudinal cognitive decline in MMSE scores compared to non-carriers, with $P = 0.009$ in the LMM analysis (Extended Data Fig. 3). *TMEM108* and *WWOX* are both expressed in human brain (Extended Data Fig. 4) and specifically in dopamine and pyramidal neurons[47] (Extended Data Fig. 5).

The *RIMS2* locus and the suggestive prognosis-associated loci have not been associated with PD susceptibility in any case-control GWAS, including large meta-analyses, which reported non-significant *P* values for the three variants[45]. Therefore, if they were to modulate disease susceptibility, their effect sizes would likely be very modest. As sub-threshold variants may contribute to genetic architecture[51], we examined 505 sub-threshold progression variants ($P < 10^{-5}$ and $> 5 \times 10^{-8}$ in the combined analysis) for overlap with susceptibility variants[45]. None of the sub-threshold progression variants was significantly associated with susceptibility, adjusting for multiple testing (for example, 497 had *P* values > 0.05, 8 had nominal *P* values between 0.01 and 0.05). Thus, lead variants associated with cognitive progression differed from susceptibility variants.

We next evaluated the effects of two previously nominated candidate prognostic genes, *GBA*[6,7,13,14] and *APOE*[15], on risk of dementia in patients with PD (Supplementary Table 4) in the combined population. Patients carrying a pathogenic mutation for Gaucher's disease or protein-coding variants associated with PD (as defined previously[6]) in *GBA* had a HR of 1.93 (95% CI 1.36–2.73) for dementia, with $P = 0.0002$ (Fig. 2a) in the Cox analysis, extending previous results[6,7,14]. They had a more rapid longitudinal decline in MMSE scores compared to non-carriers in LMM analysis ($\beta = -0.087$, $P = 0.011$, Fig. 2b). Patients carrying the *APOE* ε4 allele had HR 1.48 (95% CI 1.17–1.87) for PDD and $P = 0.001$ (Fig. 2c), and a more rapid decline in MMSE scores ($\beta = -0.078$, $P = 0.0003$) compared to non-carriers (Fig. 2d). The *RIMS2* variant was a stronger predictor of PD dementia than *GBA* and *APOE* (by approximately 2.5 and 3 times, respectively).

It has been assumed that GWAS-derived susceptibility variants constitute progression drivers with limited evidence (for example, ref. [20]). The aggregate effect of 90 GWAS-derived susceptibility loci[45] can be captured in a polygenic risk score (PRS) (Methods) that estimates the cumulatively genetic susceptibility for PD[52]. We tested the PRS for association with dementia prognosis in our longitudinal PD cohorts. Contrary to expectation, no statistically significant association between PRS and progression to PDD was found in the Cox analysis (HR = 0.95, 95% CI, 0.80–1.13, $P = 0.57$). The area under the curve (AUC) for 10-year prediction of PDD was 0.496 (95% CI 0.444–0.548; Table 2 and Fig. 3a), which was not different from chance. Furthermore, we compared patients in the highest PRS quartile to those in the lowest PRS quartile using survival curves (Fig. 3b, $P = 0.91$) and LMM (Extended Data Fig. 6) and detected no appreciable differences. Individually, none of the 90 susceptibility variants achieved multiple-testing-corrected significance thresholds for predicting PDD (Supplementary Table 5). They were also not significantly linked to motor progression in PD as measured by transition to HY (Hoehn and Yahr) stage 3 using Cox model analysis and change in the MDS-UPDRS (Movement Disorder Society–sponsored revision of the unified Parkinson's disease rating scale) part III subscale score by LMM model analysis, respectively, adjusting for covariates (Supplementary Data 1). There was no correlation between the statistical power to detect effects at these SNPs and the observed *P* values (Pearson correlation $r^2 = 0.016$, $P = 0.88$). This suggests that genetic variants and scores linked to susceptibility are not significantly associated with cognitive progression.

We then used the lead variant from each of the three prognosis loci to develop an innovative cognitive polygenic hazard score (PHS) to predict PD dementia (Methods). The HR was 2.54 (95% CI 2.10–3.08) with $P = 4.51 \times 10^{-20}$ for a one-unit value increase in PHS. The PHS was associated with prediction of PDD with a 10-year cumulative AUC of 0.589 (95% CI 0.552–0.626; Fig. 3a). Out of 3,821 cases with PD, 688 (18%) carried at least one of the three novel progression alleles (rs182987047, rs138073281, rs8050111), of which 639 cases carried only one progression allele, 47 cases carried two risk alleles, and two cases carried three risk alleles. Cox proportional hazards analysis stratified for carriers of 1, 2 (either homozygous

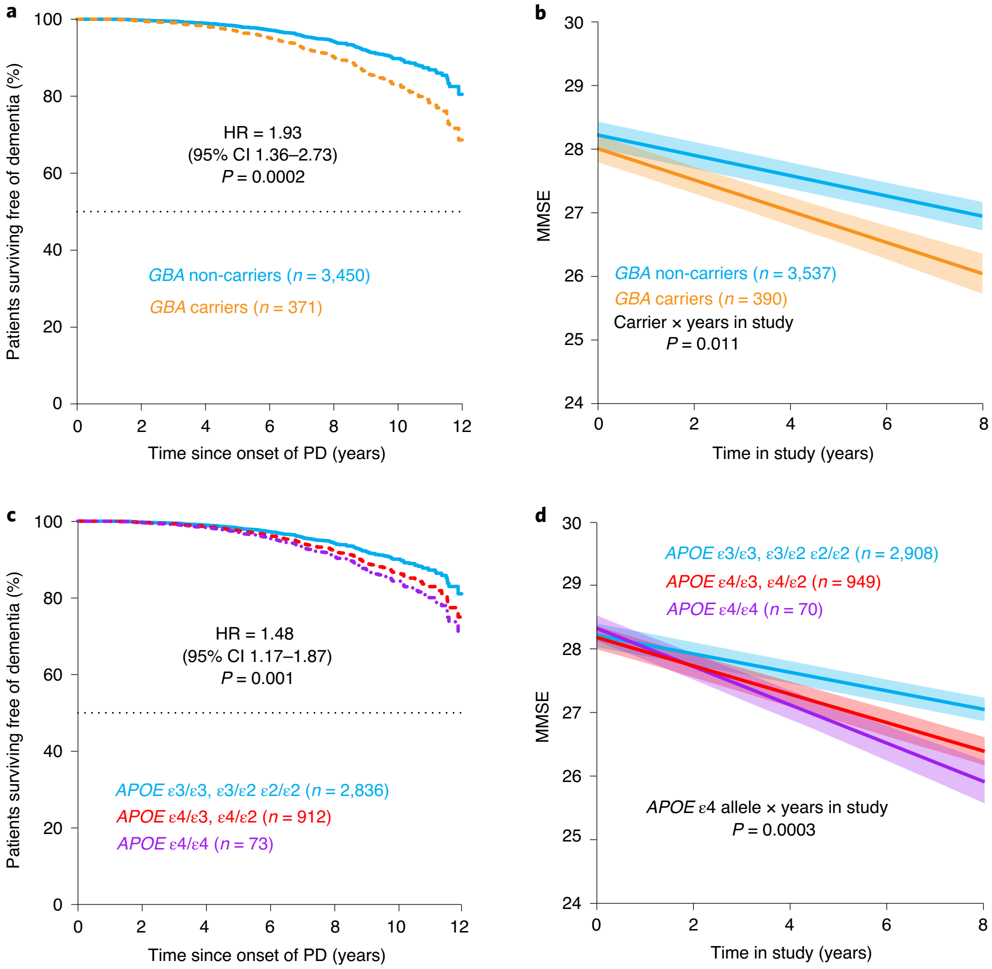


Fig. 2 | *GBA* and *APOE* ε4 accelerate cognitive decline in individuals with PD. a, Covariate-adjusted survival curves for patients with PD without *GBA* mutation (light blue line) and those carrying *GBA* mutation (orange dashed line). **b**, Adjusted mean MMSE scores across time predicted from the estimated fixed-effect parameters in the LMM for carriers (orange) and non-carriers (light blue) of a *GBA* variant. **c**, Covariate-adjusted survival curves for patients with PD without an *APOE* ε4 allele (light blue line), carriers of one *APOE* ε4 allele (red line) and carriers of two *APOE* ε4 alleles (purple line). Cox PH model with two-sided Wald test. **d**, Adjusted mean MMSE scores across time predicted from the estimated fixed-effect for non-carriers (light blue line), *APOE* ε4 heterozygous (red line) and *APOE* ε4 homozygous (purple line) carriers. Cox PH model with two-sided Wald test (**a**,**c**); shaded ribbons indicate ± s.e.m. across time (**b**–**d**); *P* values from LMM.

or heterozygous for two loci) and 3 risk alleles compared to non-carrier cases indicated an additive effect with HRs of 2.79 (95% CI 2.12–3.67) with P of 2.70×10^{-13}, 5.65 (95% CI 3.27–9.74) with $P = 4.81 \times 10^{-10}$, and 30.4 (95% CI 3.77–245.4), respectively.

We evaluated different genetic Cox proportional hazards models for prediction of PDD in the combined population (Table 2, Fig. 3a and Methods). The most robust genetic hazard model included the three new prognosis loci plus *GBA* and *APOE* (model concordance = 0.618). This PHS was a significant predictor of PDD with an AUC of 0.623 (95% CI 0.576–0.670). It was significantly more accurate in estimating whether a patient will develop dementia within 10 years from disease onset than chance alone ($P = 2.68 \times 10^{-22}$) or compared to the PRS ($P = 0.0009$). The Cox-adjusted survival curves of patients (Fig. 3c) showed that 89.6% of patients with a low (zero) PHS survived for 10 years after onset of PD without dementia. By contrast, only 73.3% of patients in the highest quartile of the PHS remained free of dementia for 10 years after onset of PD.

To further test the performance of the PHS in patients whose data were not previously used to discover and replicate the progression variants, or to build and optimize the PHS, we analyzed three new independent cohorts (EPIPARK, DeNoPa (De Novo Parkinson Cohort), HBS2). The association of the PHS with PDD prediction was replicated, with $P = 0.01$ and HR = 2.05 (1.16–3.61; Table 2) across the three new cohorts. AUCs in the independent development and validation stages were consistent; for example, 0.623 (0.576–0.670) and 0.668 (0.519–0.817), respectively (Table 2). Similarly, stratified covariate-adjusted survival analysis comparing cases scoring in the highest quartile of PHS with cases scoring zero on the PHS were consistent in development and validation stages, with HR = 2.82 (2.14–3.72) and HR = 3.20 (1.26–8.11), respectively (Fig. 3c,d; longitudinal follow-up period was considerably shorter in the new cohorts). The PRS was again not predictive of PDD in the three new cohorts ($P = 0.25$).

This study uncovered genetic variation linked to cognitive progression in PD with substantial effect sizes. These progression variants were not associated with susceptibility. Susceptibility variants and scores did not appear to predict progression. This is consistent with the hypothesis that disease initiation and progression may, in part, be governed by diverging genetics and mechanisms[8,53,54]. Cognitive progression in PD strongly correlates with cortical spread of Lewy bodies and neurites[11,55]. Furthermore, amyloid plaques and tangles present in up to one-third of patients[55,56]. Our study indicates

Table 2 | Performance of different genetic Cox proportional hazards models for predicting PDD

Genetic factor[a]	PHS stage[b]	HR	95% CI	P	Concordance of model	AUC[c]	95% CI
PRS: PD GWAS (90 SNPs)	Development	0.95	0.80–1.13	0.57	0.510	0.496	0.444–0.548
PHS: novel 3 variants	Development	2.54	2.10–3.08	4.51×10^{-20}	0.597	0.589	0.552–0.626
PHS: novel 3 variants + *GBA*	Development	2.49	2.08–2.98	1.30×10^{-21}	0.601	0.611	0.569–0.652
PHS: novel 3 variants + *APOE* ε4	Development	2.47	2.06–2.96	5.12×10^{-21}	0.616	0.604	0.559–0.649
PHS: novel 3 variants + *GBA* + *APOE* ε4	Development	2.42	2.04–2.88	2.68×10^{-22}	0.618	0.623	0.576–0.670
PRS: PD GWAS (90 SNPs)	Validation	0.60	0.25–1.42	0.25	0.551	0.588	0.399–0.778
PHS: novel 3 variants + *GBA* + *APOE* ε4	Validation	2.05	1.16–3.61	0.01	0.617	0.668	0.519–0.817

[a] Exclusively genetic factors were used in these Cox proportional hazards models. [b] The PHS development stage comprises the discovery and replication cohorts; the PHS validation stage comprises three additional, independent cohorts that were not available and not used during the variant discovery, variant replication and PHS development stages. [c] Area under the 10-year cumulative or dynamic ROC curves for PHS development cohorts (for example, combined discovery and replication cohorts), and the area under the 6-year ROC curves for the independent PHS validation cohorts; 95% CI was estimated using a simulation method. P values from Cox proportional hazards models with two-sided Wald test.

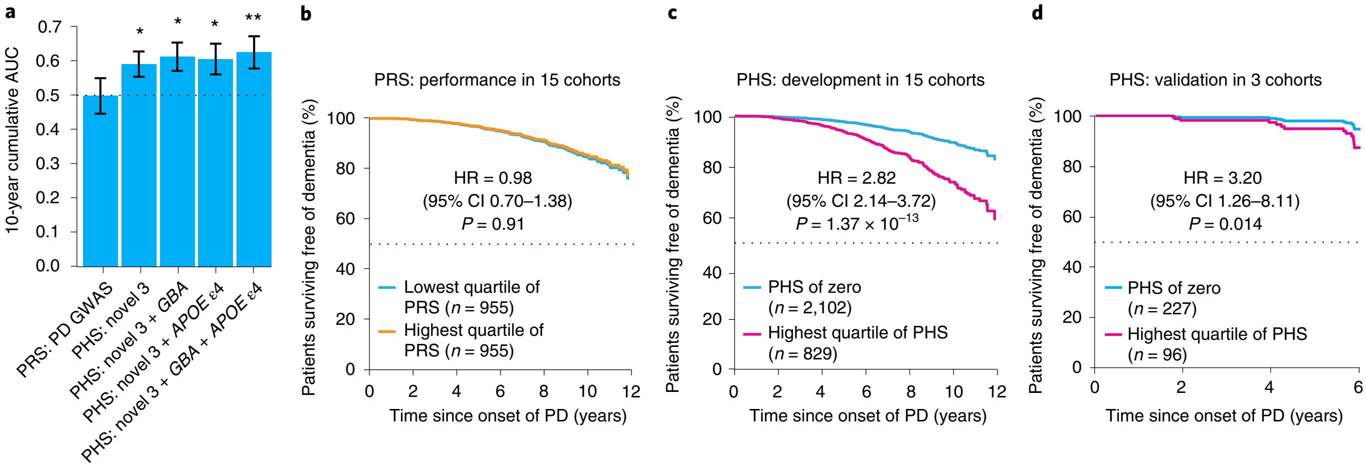


Fig. 3 | A polygenic hazard score outperforms polygenic risk scores in dementia prediction. a, Comparison of polygenic Cox PH models for predicting progression to PDD in patients with PD ($n = 3{,}821$ with 31,053 longitudinal visits). Data are visualized as the 10-year cumulative AUC (bars) and the 95% CI (error bars), estimated as described in ref. [58], implemented in the timeROC package. P values shown are the AUC of individual PHS models (based on prognosis variants) compared to the AUC of a PRS model (based on 90 susceptibility variants[45]). $*P < 0.05$ (exact values, from left to right, are $P = 0.006$, $P = 0.002$ and $P = 0.003$), $**P = 0.0009$, two-sided z-tests. **b**, Cox-adjusted survival curves for survival free of PDD for cases scoring in the highest quartile of PRS (orange) compared to cases scoring in the lowest quartile of PRS (light blue). **c**, Cox-adjusted survival curves for survival free of PDD for cases scoring in the highest quartile of PHS (magenta) compared to cases scoring zero on the PHS (light blue) for the combined dataset. **d**, Cox-adjusted survival curves for survival free of PDD for cases scoring in the highest quartile of PHS (magenta) compared to cases scoring zero on the PHS (light blue) for the new PHS validation dataset. Cox PH model with two-sided Wald test (**b**–**d**), results of stratified analyses are shown (HR, 95% CI, P values); results of non-stratified analyses are given in Table 2. Patients assigned to the highest quartile of PRS or PHS were those with a score greater than the score separating the fourth (highest) and third quartile of values.

that genetic drivers of PD progression may comprise PD-specific loci (for example, *RIMS2* and potentially *TMEM108*), loci shared with dementia with Lewy bodies (for example, *APOE* and *GBA*[57]), and possibly loci shared with Alzheimer's disease (for example, *APOE* and *WWOX*[50]). Analyses of larger longitudinal populations will be required to detect variants with small effect sizes, to increase statistical power for motor phenotypes confounded by PD medications, and to systematically decode the divergent and convergent features of the genetic architecture underlying susceptibility, progression and dementias.

These results suggest a new paradigm for drug development. Disease-modifying drugs that target the genetic drivers of disease progression could potentially turn fast progressors into slow progressors and substantially improve quality of life. Clinically, this study provides a polygenic score that could be used to enrich trials with patients who have a more aggressive disease course and are therefore likely to show the greatest benefits from interventions. This may be useful because ascertaining therapeutic efficacy in patients who naturally progress slowly is exceedingly difficult.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41588-021-00847-6.

Received: 3 April 2019; Accepted: 16 March 2021;
Published online: 6 May 2021

References

1. Nalls, M. A. et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat. Genet.* **46**, 989–993 (2014).

2. Chang, D. et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat. Genet.* **49**, 1511–1516 (2017).
3. Wijmenga, C. & Zhernakova, A. The importance of cohort studies in the post-GWAS era. *Nat. Genet.* **50**, 322–328 (2018).
4. Locascio, J. J. & Atri, A. An overview of longitudinal data analysis methods for neurological research. *Dement. Geriatr. Cogn. Dis. Extra* **1**, 330–357 (2011).
5. Dorsey, E. R. & Bloem, B. R. The Parkinson pandemic—a call to action. *JAMA Neurol.* **75**, 9–10 (2018).
6. Liu, G. et al. Specifically neuropathic Gaucher's mutations accelerate cognitive decline in Parkinson's. *Ann. Neurol.* **80**, 674–685 (2016).
7. Liu, G. et al. Prediction of cognition in Parkinson's disease with a clinical-genetic score: a longitudinal analysis of nine cohorts. *Lancet Neurol.* **16**, 620–629 (2017).
8. Aarsland, D. et al. Cognitive decline in Parkinson disease. *Nat. Rev. Neurol.* **13**, 217–231 (2017).
9. Schrag, A., Jahanshahi, M. & Quinn, N. What contributes to quality of life in patients with Parkinson's disease? *J. Neurol. Neurosurg. Psychiatry* **69**, 308–312 (2000).
10. Svenningsson, P., Westman, E., Ballard, C. & Aarsland, D. Cognitive impairment in patients with Parkinson's disease: diagnosis, biomarkers, and treatment. *Lancet Neurol.* **11**, 697–707 (2012).
11. Braak, H. et al. Staging of brain pathology related to sporadic Parkinson's disease. *Neurobiol. Aging* **24**, 197–211 (2003).
12. Langston, J. W. The Parkinson's complex: parkinsonism is just the tip of the iceberg. *Ann. Neurol.* **59**, 591–596 (2006).
13. Williams-Gray, C. H. et al. The CamPaIGN study of Parkinson's disease: 10-year outlook in an incident population-based cohort. *J. Neurol. Neurosurg. Psychiatry* **84**, 1258–1264 (2013).
14. Cilia, R. et al. Survival and dementia in GBA-associated Parkinson's disease: the mutation matters. *Ann. Neurol.* **80**, 662–673 (2016).
15. Pang, S., Li, J., Zhang, Y. & Chen, J. Meta-analysis of the relationship between the *APOE* gene and the onset of Parkinson's disease dementia. *Parkinsons Dis.* **2018**, 9497147 (2018).
16. Healy, D. G. et al. Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. *Lancet Neurol.* **7**, 583–590 (2008).
17. Guella, I. et al. alpha-synuclein genetic variability: a biomarker for dementia in Parkinson disease. *Ann. Neurol.* **79**, 991–999 (2016).
18. Markopoulou, K. et al. Does alpha-synuclein have a dual and opposing effect in preclinical vs. clinical Parkinson's disease? *Parkinsonism Relat. Disord.* **20**, 584–589 (2014).
19. Mata, I. F. et al. APOE, MAPT, and SNCA genes and cognitive performance in Parkinson disease. *JAMA Neurol.* **71**, 1405–1412 (2014).
20. Paul, K. C., Schulz, J., Bronstein, J. M., Lill, C. M. & Ritz, B. R. Association of polygenic risk score with cognitive decline and motor progression in Parkinson disease. *JAMA Neurol.* **75**, 360–366 (2018).
21. Mata, I. F. et al. Large-scale exploratory genetic analysis of cognitive impairment in Parkinson's disease. *Neurobiol. Aging* **56**, 211 e1–211 e7 (2017).
22. Locascio, J. J. et al. Association between alpha-synuclein blood transcripts and early, neuroimaging-supported Parkinson's disease. *Brain* **138**, 2659–2671 (2015).
23. Pankratz, N. et al. Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2. *Ann. Neurol.* **71**, 370–384 (2012).
24. Jankovic, J. et al. Variable expression of Parkinson's disease: a base-line analysis of the DATATOP cohort. The Parkinson Study Group. *Neurology* **40**, 1529–1534 (1990).
25. Ravina, B. et al. A longitudinal program for biomarker development in Parkinson's disease: a feasibility study. *Mov. Disord.* **24**, 2081–2090 (2009).
26. Winder-Rhodes, S. E. et al. Glucocerebrosidase mutations influence the natural history of Parkinson's disease in a community-based incident cohort. *Brain* **136**, 392–399 (2013).
27. Marinus, J. et al. A short scale for the assessment of motor impairments and disabilities in Parkinson's disease: the SPES/SCOPA. *J. Neurol. Neurosurg. Psychiatry* **75**, 388–395 (2004).
28. Breen, D. P., Evans, J. R., Farrell, K., Brayne, C. & Barker, R. A. Determinants of delayed diagnosis in Parkinson's disease. *J. Neurol.* **260**, 1978–1981 (2013).
29. Rosenthal, L. S. et al. The NINDS Parkinson's disease biomarkers program. *Mov. Disord.* **30**, 915–923 (2016).
30. Writing Group for the NINDS Exploratory Trials in Parkinson Disease (NET-PD) Investigators *et al*. Effect of creatine monohydrate on clinical progression in patients with Parkinson disease: a randomized clinical trial. *JAMA* **313**, 584–593 (2015).
31. 1000 Genomes Project Consortium *et al*. A global reference for human genetic variation. *Nature* **526**, 68–74 (2015).
32. Browning, S. R. & Browning, B. L. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am. J. Hum. Genet.* **81**, 1084–1097 (2007).
33. Li, Y., Willer, C. J., Ding, J., Scheet, P. & Abecasis, G. R. MaCH: using sequence and genotype data to estimate haplotypes and unobserved genotypes. *Genet. Epidemiol.* **34**, 816–834 (2010).
34. Marchini, J., Howie, B., Myers, S., McVean, G. & Donnelly, P. A new multipoint method for genome-wide association studies by imputation of genotypes. *Nat. Genet.* **39**, 906–913 (2007).
35. Visscher, P. M. et al. 10 years of GWAS discovery: biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
36. Ripatti, S. & Palmgren, J. Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics* **56**, 1016–1022 (2000).
37. Bulik-Sullivan, B. K. et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).
38. Yang, J. et al. Genomic inflation factors under polygenic inheritance. *Eur. J. Hum. Genet.* **19**, 807–812 (2011).
39. Dubois, B. et al. Diagnostic procedures for Parkinson's disease dementia: recommendations from the movement disorder society task force. *Mov. Disord.* **22**, 2314–2324 (2007).
40. Armstrong, M. J. & Okun, M. S. Diagnosis and treatment of Parkinson disease: a review. *JAMA* **323**, 548–560 (2020).
41. Kaeser, P. S. et al. RIM proteins tether Ca2+ channels to presynaptic active zones via a direct PDZ-domain interaction. *Cell* **144**, 282–295 (2011).
42. Liu, C., Kershberg, L., Wang, J., Schneeberger, S. & Kaeser, P. S. Dopamine secretion is mediated by sparse active zone-like release sites. *Cell* **172**, 706–718 e15 (2018).
43. Mechaussier, S. et al. Loss of function of RIMS2 causes a syndromic congenital cone-rod synaptic disease with neurodevelopmental and pancreatic involvement. *Am. J. Hum. Genet.* **106**, 859–871 (2020).
44. Powell, C. M. et al. The presynaptic active zone protein RIM1alpha is critical for normal learning and memory. *Neuron* **42**, 143–153 (2004).
45. Nalls, M. A. 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.* **18**, 1091–1102 (2019).
46. GTEx, Consortium et al. Genetic effects on gene expression across human tissues. *Nature* **550**, 204–213 (2017).
47. Dong, X. et al. Enhancers active in dopamine neurons are a primary link between genetic variation and neuropsychiatric disease. *Nat. Neurosci.* **21**, 1482–1492 (2018).
48. Jiao, H. F. et al. Transmembrane protein 108 is required for glutamatergic transmission in dentate gyrus. *Proc. Natl Acad. Sci. USA.* **114**, 1177–1182 (2017).
49. Mallaret, M. et al. The tumour suppressor gene *WWOX* is mutated in autosomal recessive cerebellar ataxia with epilepsy and mental retardation. *Brain* **137**, 411–419 (2014).
50. Kunkle, B. W. et al. Genetic meta-analysis of diagnosed Alzheimer's disease identifies new risk loci and implicates Aβ, tau, immunity and lipid processing. *Nat. Genet.* **51**, 414–430 (2019).
51. Khera, A. V. et al. Genome-wide polygenic scores for common diseases identify individuals with risk equivalent to monogenic mutations. *Nat. Genet.* **50**, 1219–1224 (2018).
52. Nalls, M. A. et al. Diagnosis of Parkinson's disease on the basis of clinical and genetic classification: a population-based modelling study. *Lancet Neurol.* **14**, 1002–1009 (2015).
53. Lee, J. K., Tran, T. & Tansey, M. G. Neuroinflammation in Parkinson's disease. *J. Neuroimmune Pharmacol.* **4**, 419–429 (2009).
54. Johnson, M. E., Stecher, B., Labrie, V., Brundin, L. & Brundin, P. Triggers, facilitators, and aggravators: redefining Parkinson's disease pathogenesis. *Trends Neurosci.* **42**, 4–13 (2019).
55. Irwin, D. J. et al. Neuropathologic substrates of Parkinson disease dementia. *Ann. Neurol.* **72**, 587–598 (2012).
56. Irwin, D. J. et al. Neuropathological and genetic correlates of survival and dementia onset in synucleinopathies: a retrospective analysis. *Lancet Neurol.* **16**, 55–65 (2017).
57. Guerreiro, R. et al. Investigating the genetic architecture of dementia with Lewy bodies: a two-stage genome-wide association study. *Lancet Neurol.* **17**, 64–74 (2018).
58. Blanche, P., Dartigues, J. F. & Jacqmin-Gadda, H. Estimating and comparing time-dependent areas under receiver operating characteristic curves for censored event times with competing risks. *Stat. Med.* **32**, 5381–5397 (2013).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

[1]Center for Advanced Parkinson Research, Harvard Medical School, Brigham and Women's Hospital, Boston, MA, USA. [2]Precision Neurology Program of Brigham and Women's Hospital, Harvard Medical School, Boston, MA, USA. [3]School of Medicine, Sun Yat-sen University, Shenzhen, Guangdong, China. [4]School of Computer Science, Northwestern Polytechnical University, Xi'an, Shaanxi, China. [5]Department of Neurology, Massachusetts General Hospital and Harvard Medical School, Boston, MA, USA. [6]Sorbonne Université, Paris Brain Institute – Institut du Cerveau – ICM, Institut National de Santé et en Recherche Médicale, Centre National de Recherche Scientifique, Assistance Publique Hôpitaux de Paris, Département de Neurologie et de Génétique, Centre d'Investigation Clinique Neurosciences, Hôpital Pitié-Salpêtrière, Paris, France. [7]The Norwegian Centre for Movement Disorders, Stavanger University Hospital, Stavanger, Norway. [8]Department of Chemistry, Bioscience and Environmental Engineering, University of Stavanger, Stavanger, Norway. [9]Departments of Neurology and Radiology, Washington University School of Medicine, St. Louis, MO, USA. [10]Paris-Saclay University, Université de Versailles Saint-Quentin-en-Yvelines (UVSQ), Inserm, Gustave Roussy, 'Exposome and heredity' team, Centre de researche en épidémiologie et santé des populations (CESP), Villejuif, France. [11]Department of Neurology, Brigham and Women's Hospital, Boston, MA, USA. [12]Praxis Precision Medicines, Cambridge, MA, USA. [13]Department of Neurology, Center for Health + Technology, University of Rochester, Rochester, NY, USA. [14]Department of Neurology and Neurosurgery, University of Tartu, Tartu, Estonia. [15]Centre for Molecular Medicine and Innovative Therapeutics, Murdoch University, Perth, Western Australia, Australia. [16]Perron Institute for Neurological and Translational Science, Perth, Western Australia, Australia. [17]Banner Sun Health Research Institute, Sun City, AZ, USA. [18]Department of Neurology, Stavanger University Hospital, Stavanger, Norway. [19]Department of Neurology, Haukeland University Hospital, Bergen, Norway. [20]Department of Clinical Medicine, University of Bergen, Bergen, Norway. [21]Department of Neuroscience, Washington University School of Medicine, St. Louis, MO, USA. [22]Program of Physical Therapy and Program of Occupational Therapy, Washington University School of Medicine, St. Louis, MO, USA. [23]German Center for Neurodegenerative diseases (DZNE), Tübingen, Germany. [24]Translational Genomics Core of Partners HealthCare Personalized Medicine, Cambridge, MA, USA. [25]Department of Neurology, Leiden University Medical Center, Leiden, the Netherlands. [26]Institute of Neurogenetics, University of Lübeck, University Hospital of Schleswig-Holstein, Lübeck, Germany. [27]Department of Psychiatry and Psychotherapy, University of Lübeck, Lübeck, Germany. [28]Department of Neurology, University Medical Center Göttingen, Göttingen, Germany. [29]Paracelsus-Elena-Klinik, Kassel, Germany. [30]Department of Neurosurgery, University Medical Center Göttingen, Göttingen, Germany. [31]Institute of Neurogenetics, University of Lübeck, University Hospital of Schleswig-Holstein, Lübeck, Germany. [32]John Van Geest Centre for Brain Repair, Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK. [33]Wellcome - MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: cscherzer@rics.bwh.harvard.edu

International Genetics of Parkinson Disease Progression (IGPP) Consortium

Ganqiang Liu[1,2,3], Jiajie Peng[1,2,4], Zhixiang Liao[1,2], Joseph J. Locascio[1,2,5], Jean-Christophe Corvol[6], Xianjun Dong[1,2], Jodi Maple-Grødem[7,8], Meghan C. Campbell[9], Alexis Elbaz[10], Suzanne Lesage[6], Alexis Brice[6], Graziella Mangone[6], Bernard Ravina[12], Ira Shoulson[13], Pille Taba[14], Sulev Kõks[15,16], Thomas G. Beach[17], Florence Cormier-Dequaire[6], Guido Alves[7,8,18], Ole-Bjørn Tysnes[19,20], Joel S. Perlmutter[9,21,22], Peter Heutink[23], Jacobus J. van Hilten[25], Meike Kasten[26,27], Brit Mollenhauer[28,29], Claudia Trenkwalder[29,30], Christine Klein[31], Roger A. Barker[32,33], Caroline H. Williams-Gray[32], Johan Marinus[25] and Clemens R. Scherzer[1,2,5,11]

Methods

Study participants. Supplementary Table 1 describes the cohorts included in this work.

Discovery, replication and PHS development stages. We used 15 cohorts[13,22–30,59–64] from North America and Europe to discover and replicate progression variants and to build the PHS. The 15 cohorts comprised a total of 4,872 patients with PD (with available genotyping data), who were longitudinally assessed with 36,123 study visits between 1986 and 2017 (Supplementary Fig. 1).

Written informed consent for DNA collection and phenotypic data collection for secondary research use for each cohort was obtained from the participants, with approval from the local ethics committees. The institutional review board of Partners HealthCare approved the current genotyping and analyses. For the Parkinson's Progression Markers Initiative (PPMI), approval was obtained to download and analyze the publicly accessible WGS and clinical data. Patients with a diagnosis of PD established according to modified UK PD Society Brain Bank diagnostic criteria, as previously reported[1,22,25,28,60,61,63–66], were recruited to 13 cohorts. In DATATOP (Deprenyl and Tocopherol Antioxidative Therapy of Parkinsonism), the eligibility criteria required a clinical diagnosis of early, idiopathic PD (HY stages 1 or 2) with patients not on anti-parkinsonian medications[65]. For the Arizona Study of Aging/Brain and Body Donation Program, all subjects had come to autopsy and had full neuropathological examinations with neuropathological diagnosis[62]. Diagnostic certainty was increased by confirming the clinical diagnosis of PD during longitudinal follow-up visits[67] in all cohorts. Patients whose longitudinal follow-up evaluations were not consistent with a diagnosis of PD were excluded. Cohorts were a priori assigned to discovery or replication cohorts to achieve an approximately two-thirds to one-third split among the two stages (while considering fixed cohort sizes) and to achieve an balanced distribution of the distinct types of cohorts (for example, purpose-designed biomarkers studies, phase 3 clinical trials, population-based cohorts) across the two stages.

Serial MMSE scores[68] were longitudinally collected in ten cohorts. Montreal Cognitive Assessment (MoCA)[69] scores were collected in PDBP (Parkinson's Disease Biomarkers Program)[29] and PPMI[60] cohorts and converted to MMSE scores according to a published formula[70]. SCOPA-COG (scales for outcomes in Parkinson's disease-cognition) were collected in PROPARK (Profiling Parkinson's disease)[61], PROPARK-C (PROPARK-cross sectional cohort) and NET-PD Long term Study-1 (LS1)[30] cohorts and converted to MMSE scores. Cohort-specific definitions of PDD were used (Supplementary Table 2). For seven cohorts, operationalized level 1 diagnostic criteria for PDD according to the Movement Disorders Society Task Force[39] were available; PreCEPT (Parkinson Research Examination of CEP-1347 Trial) and DATATOP used distinct definitions. PreCEPT defined PDD as a score of 4 on the UPDRS subscale 1 item 1 defined as 'cognitive dysfunction [that] precludes the patient's ability to carry out normal activities and social interactions'. For DATATOP published criteria for cognitive impairment leading to functional impairment were used[71]. Depression status was defined according to cohort-specific assessments. Ancestry was self-reported. For several cohorts in this analysis, we evaluated previously collected longitudinal phenotypic data; for the active HBS, PDBP and DIGPD (Drug Interaction with Genes in Parkinson's Disease) cohorts, both retrospectively and prospectively collected longitudinal data elements were included. HBS, Arizona Study of Aging/Brain and Body Donation Program, NET-PD LS1, CamPaIGN (Cambridgeshire Parkinson's Incidence from GP to Neurologist), PICNICS (Parkinsonism: Incidence, Cognitive and Non-motor heterogeneity In Cambridgeshire), DIGPD, PDBP, PIB, ParkWest, PROPARK and PROPARK-C cohorts comprised the discovery population and DATATOP, PPMI, PreCEPT and Tartu the replicate population.

PHS validation stage. To avoid overfitting, we tested the performance of the pre-specified PHS in 520 patients from three independent cohorts with detailed longitudinal clinical phenotyping; DeNoPa[72], EPIPARK[73] and HBS2 (Supplementary Table 1), from Germany and the United States. These three longitudinal PD cohorts were not used to discover and replicate progression variants, or to build the PHS. PD was diagnosed in these cohorts according to modified UK PD Society Brain Bank diagnostic criteria. Cohort-specific definitions of PDD are listed in Supplementary Table 2.

Genotyping and data quality control and processing. Quality control steps are shown in Extended Data Fig. 1. In brief, the DNA of patients with PD was quality controlled on an Agilent 2100 Bioanalyzer. DNA was quantified against an eight-point standard curve using the Quant-iT Picogreen dsDNA Assay Kit (Life Technologies, P7589) with a SpectraMax Gemini plate reader from Molecular Devices. Sample were genotyped at the Translational Genomics Core of Partners HealthCare using the Illumina Multi-Ethnic Genotyping Array (MEGA A1)[74], which includes 1,779,819 markers (MEG array kit, Illumina, WG-316–1001). DNA was amplified using a whole-genome amplification process. After fragmentation of the DNA, the sample was hybridized to 50-mer probes attached to the BeadChips, stopping one base before the interrogated base. Single base extension was then carried out to incorporate a labeled nucleotide. Dual color (Cy3 and Cy5) staining allowed the nucleotide to be detected by the iSCAN reader. Data from the iSCAN were collected in the Illumina LIMS and automated conversion to genotype occured using Autocall v2.0.1. In total, 4,510 PD samples were genotyped with the MEGA array; 512 PD samples from the PPMI had whole-genome sequences available.

We used PLINK[75] (v1.90 beta) and in-house scripts to conduct genotyping data processing and perform rigorous subject and SNP quality control (Extended Data Fig. 1). SNPs with overall missingness > 0.05 were excluded. Samples with mismatched sex were excluded. Samples with a genotype missingness > 0.05 or heterozygosity rate > 4 s.d. from the mean were also excluded. To check relatedness among samples, 279,933 LD-independent SNPs were selected and pairwise identity by descent was estimated using PLINK routine '--indep 50 5 2'. For any related sample (pi-hat between 0.1875 to 0.9)[76], one case with higher genotyping call rate was selected and kept, and the others were excluded. For those sample pairs with pi-hat > 0.9, both cases were excluded. To identify geographical outliers, a pruned data set containing 86,998 LD-independent SNPs were merged with the 1000 Genomes Project data set[77]. Principal component analysis (smartpca)[78] was used to identify and exclude the geographical outliers.

For 4,491 patients with PD, 31,885 (95.4%) of visits occurred within 12 years of longitudinal follow-up from disease onset, with a median follow-up time of 6.7 years (interquartile range, 4.2 years). We therefore focused our survival analyses on the 12-year time frame from disease onset. In total, 3,821 samples passed genotyping and clinical data quality control (Extended Data Fig. 1). Patients were left-censored and those with missing or non-quality clinical data were excluded ($n = 670$; Extended Data Fig. 1). Specifically, 24 were excluded for whom clinical data were not available. Another 646 patients were excuded owing to missing critical individual data points or left-censoring (Extended Data Fig. 1) (for example, 138 participants already had PDD at the baseline visit and were left-censored; 39 subjects had missing data for age at onset or age at the baseline visit; 238 subjects had a first study visit that occurred more than 12 years from disease onset; and 231 were missing dementia ascertainment data). To identify genetic variants associated with progression from PD to PDD, we performed a longitudinal GWSS on these 3,821 patients, of whom 2,650 (and 11,744 visits) were assigned to the discovery population, and 1,171 (and 19,309 visits) were assigned to the replicate population.

For 520 independent patients from DeNoPa, EPIPARK and HBS2, the same genotyping quality control was performed, and 425 samples passed quality control and 21 patients were excluded due to left-censoring. Thus, a total of 404 patients with 1,028 visits were used in the PHS validation stage.

Genotype imputation. Genotype imputation was performed using Minimac3 (v2.0.1) on the Michigan online imputation server[79]. The haplotype reference consortium (HRC version r1.1)[80] was selected as the reference panel. This consists of 64,940 haplotypes of predominantly European ancestry with ~39.2 million SNPs, all with an estimated minor allele count of ≥ 5. Eagle2 (v2.3)[81] with 20-Mb chunk size was used to estimate haplotype phasing; pipeline details, including quality check, phasing and imputation, are available at https://imputationserver.sph.umich.edu. Samples from all discovery and replication cohorts were prephased and imputed in a single batch to avoid batch effects attributable to the imputation process: Multi-Ethnic Genotyping Array (MEGA) data of 4,020 subjects with PD with 1,635,580 SNPs at autosomes were used as input for the online server. To estimate imputation accuracy, imputed genotype calls for 1,052,012 SNPs were compared with directly genotyped data using EmpR to calculate the correlation between the true genotyped values and the imputed values from the output of Minimac3. Mean R^2 was 0.996 and EmpR was 0.979 for variants with MAF ≥ 0.1% (Supplementary Fig. 2). Imputed variants with MAF < 0.1% and/or $R^2 < 0.3$ were excluded. In total, 11,220,132 imputed SNPs remained for further analysis (Supplementary Fig. 2). In addition, we removed 26,785 variants with discordant MAF (with Fisher's exact test false discovery rate < 0.05) observed with MEGA array plus imputation (14 cohorts) compared to WGS (PPMI cohort). In total, 7,741,751 variants with MAF ≥ 1% remained for further analysis. Imputation for the PHS validation cohorts was performed separately.

PPMI and HBS whole-genome sequencing datasets. The PPMI and HBS data consist of 512 and 699 individuals with PD, respectively. WGS was performed by Macrogen under the direction of A. Singleton (National Institute on Ageing (NIA)). Samples were prepared according to the Illumina TruSeq PCR Free DNA sample Preparation Guide. The libraries were sequenced using a Illumina HiSeq X Ten Sequencer. Detailed methods are available at https://ida.loni.usc.edu/pages/access/geneticData.jsp.

Evaluating the concordance between imputed genotypes and sequencing. We used the SnpSift tool (http://snpeff.sourceforge.net/SnpSift.version_4_0.html) to evaluate the concordance between imputed SNPs (based on the MEGA array) and SNPs directly called from WGS in 562 individuals from HBS for whom both assays were available (Supplementary Fig. 3). The percentage concordance between 10,421,270 imputed SNPs and WGS was 99.4% (standard error, 0.0006%). For the three SNPs associated with PDD (whose genotypes came from imputation), we observed high concordance rates of 99.5%, 99.6% and 98.9% for rs182987047, rs138073281 and rs8050111, respectively. Imputation average call rates (AvgCall) and imputation R^2 values were 0.998 and 0.98 for rs182987047, 0.996 and 0.888 for rs138073281, and 0.997 and 0.961 for rs8050111, respectively.

Candidate loci *GBA* and *APOE*. *GBA* gene variants were defined as described[6], and included pathogenic mutations associated with Gaucher's disease as well as the PD-associated coding risk variants (E326K, T369M and E388K). We previously reported[6] *GBA* genotypes (largely based on targeted or Sanger sequencing of the locus) for 2,625 of the 4,491 patients with PD included here. For the remaining 1,866 patients with PD, *GBA* variants and mutations were identified based on the MEGA array. Participants were classified as carriers (with one or more *GBA* mutations) or non-carriers (no *GBA* mutation) as reported[6].

APOE alleles ε2 ε3 and ε4 were identified based on rs7412 and rs429358 from MEGA chip plus imputation data (14 cohorts) or WGS (PPMI cohort). We compared imputed *APOE* alleles of 531 HBS patients with PD to the results of a TaqMan SNP genotyping assay for the two SNPs. The concordance rate was 98.7%. We classified the 4,491 patients with PD into three groups for downstream analysis: 81 homozygous ε4 carriers (ε4/ε4), 1,068 heterozygous ε4 carriers (ε2/ε4, ε3/ε4), and 3,342 non-ε4 carriers (ε2/ε2, ε2/ε3, ε3/ε3).

Statistical analysis. The Cox proportional hazards statistic was used to estimate the influence of each genotype on time (years from onset of PD) to reaching the endpoint of PDD. Age at onset of PD, sex, years of education, and the top ten principal components of population substructure were included as covariates in the Cox analyses. For the meta-analyses across cohorts, a 'cohort' term was included as a random effect (a random effects Cox model is often termed a 'frailty' model). Regarding 'cohort' as a random term will permit inferences about study level variance among a hypothetical universe of studies in the reference population. For 4,491 patients with PD, 31,885 (95.4%) of visits occurred within 12 years of longitudinal follow-up from disease onset with a median follow-up time of 6.7 years (interquartile range, 4.2 years). We therefore focused our survival analyses on the 12-year time frame from disease onset. Cox proportional hazards analyses were performed using the coxph function in the Survival package (v2.38–1) in R, and the 'Breslow' method was used for handling observations that have tied survival times. *P* values of less than or equal to 5×10^{-8} were considered indicative of genome-wide significance.

Generalized longitudinal mixed fixed and random effects analysis (LMM)[4] of cognitive decline was performed using serial MMSE scores longitudinally assessed (enrollment visit and multiple longitudinal follow-up visits) in the combined data set. The PROPARK-C and Tartu cohorts were excluded from the LMM because no longitudinal MMSE scores were available. The MMSE score was the dependent variable and the primary predictors were group status (for example, genotype carrier status or alleles), time in the study (years), and their interaction. An intercept term and linear rate of change across time per subject were the random terms (permitted to be correlated). Subject-level fixed covariates were age at baseline, sex, years of education, duration of PD illness at baseline, as well as ten principal components. A study term was included as a random effect. The significance, direction and effect size of the group × time terms answers the question of differential progression for the carriers, compared to the non-carrier group. To avoid problems with somewhat non-normal residuals for MMSE, *P* values were obtained by penalized quasi-likelihood ratio tests of the full model with the effect in question contrasted with the model without the effect in question. This analysis was performed using the glmmPQL function in the MASS package (v7.3–37). All analyses were conducted in the R statistical environment, v3.3.1. Nominal *P* values (not adjusted for multiple testing) were shown except where indicated otherwise. Evidence for genome-wide significance in the discovery population was defined as $P \leq 5 \times 10^{-8}$; *P* values ≤ 0.05 were considered evidence of significance in the replicate population and in the PHS validation population. Associations for previously established candidate loci were considered significant if they met Bonferroni-adjusted significance thresholds (for example, 0.05/number of established candidates evaluated).

Polygenic risk score. A PRS was calculated as the weighted sum of the number of risk alleles possessed by an individual, in which the weight was taken as the natural log of the odds ratio associated with each individual SNP. We used 90 lead GWAS variants associated with susceptibility for PD and the odds ratios from a recent meta-analysis study[45] to calculate the PRS (Supplementary Table 5).

Polygenic hazard score. For each patient in this study, we calculated a PHS using a similar method to that described in ref. [82]. We used the hazard ratios of the lead associated SNPs (from the combined data set) in each of the three prognosis loci to calculate the PHS. In brief, we added the number of risk alleles (0, 1 or 2) for a lead variant multiplied by the effect size (natural log of hazard ratio from combined dataset) for that variant. In other versions of the PHS, we additionally included one or both of the candidate cognitive prognosis genes (*GBA* mutation status and *APOE* ε4 allele haplotype). To evaluate the performance of the PHS models, the cumulative or dynamic receiver operating characteristic (ROC), AUCs, confidence intervals of the AUC (simulation method), and comparisons between two AUCs were calculated using the timeROC package (v0.2)[58] in R with the inverse probability of censoring weights method used to compute the weights.

Characterization of genomic risk loci. We used FUMA (http://fuma.ctglab.nl) to characterize the cognitive prognosis loci. Tag SNPs with suggestive $P < 1 \times 10^{-5}$ were input; additional SNPs in high LD with a tag SNP (with $r^2 > 0.6$ and independent from each other with $r^2 < 0.6$) were identified using the 1000 Genomes Phase 3 reference panel for Europeans. If LD blocks of independent significant SNPs were closely located to each other (<250 kb based on the most right and left SNPs from each LD block), they were merged into one genomic locus.

Gene expression analysis. Gene expression profiles of the three significant loci in human tissues was downloaded directly from GTEx portal v7 (https://gtexportal.org/). Downloaded gene expression profiles were normalized. Detailed processing methods can be found in the GTEx portal v7. Human brain cell type-specifc expression of the three cognitive prognosis loci was evaluated using the BRAINcode dataset[47] and portal (http://www.humanbraincode.org).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

A Nature Research Reporting Summary is available for this paper. Human brain cell type-specific expression data from BRAINcode[47] RNA-seq data are accessible through a user-friendly webportal at http://www.humanbraincode.org and individual-level data through dbGAP (accession no. phs001556.v1.p1). The gene expression profiles of human tissues used in this study can be downloaded from the GTEx Portal v7 (https://gtexportal.org/). The GWSS summary statistics for the combined analysis of discovery and replicate populations are publicly accessible through the EGA database at https://ega-archive.org/ (accession no. EGAS00001005110). Individual-level genetic data for the NIH-funded Illumina Multi-Ethnic Genotyping Array analyses of the HBS2 and EPIPARK cohorts are accessible in dbGAP with accession no. phs002328.v1.p1 in accordance with NIH Genomic Data Sharing Policy. The WGS and clinical data for PPMI included in this study are publicly available upon request to ppmi@loni.usc.edu through a PPMI Whole Genome Sequencing Data Agreement. Clinical data for PDBP included in this study are publicly available through https://pdbp.ninds.nih.gov. Clinical longitudinal data for the other cohorts included are accessible through appropriate data sharing agreements that protect patient privacy with the institutions that conducted or are conducting study consents and clinical assessments under local institutional review board approvals.

Code availability

Analysis code is available at https://github.com/sixguns1984/GWSS.PDD.

References

59. Alves, G. et al. Incidence of Parkinson's disease in Norway: the Norwegian ParkWest study. *J. Neurol. Neurosurg. Psychiatry* **80**, 851–857 (2009).
60. Parkinson Progression Marker, I. The Parkinson Progression Marker Initiative (PPMI). *Prog. Neurobiol.* **95**, 629–635 (2011).
61. Verbaan, D. et al. Patient-reported autonomic symptoms in Parkinson disease. *Neurology* **69**, 333–341 (2007).
62. Beach, T. G. et al. Arizona Study of Aging and Neurodegenerative Disorders and Brain and Body Donation Program. *Neuropathology* **35**, 354–389 (2015).
63. Lucero, C. et al. Cognitive reserve and β-amyloid pathology in Parkinson disease. *Parkinsonism Relat. Disord.* **21**, 899–904 (2015).
64. Corvol, J. C. et al. Longitudinal analysis of impulse control disorders in Parkinson disease. *Neurology* **91**, e189–e201 (2018).
65. Parkinson Study Group. DATATOP: a multicenter controlled clinical trial in early Parkinson's disease. *Arch. Neurol.* **46**, 1052–1060 (1989).
66. Williams-Gray, C. H. et al. The distinct cognitive syndromes of Parkinson's disease: 5 year follow-up of the CamPaIGN cohort. *Brain* **132**, 2958–2969 (2009).
67. Hughes, A. J., Daniel, S. E., Ben-Shlomo, Y. & Lees, A. J. The accuracy of diagnosis of parkinsonian syndromes in a specialist movement disorder service. *Brain* **125**, 861–870 (2002).
68. Goetz, C. G. et al. Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations. *Mov. Disord.* **19**, 1020–1028 (2004).
69. Hoops, S. et al. Validity of the MoCA and MMSE in the detection of MCI and dementia in Parkinson disease. *Neurology* **73**, 1738–1745 (2009).
70. van Steenoven, I. et al. Conversion between mini-mental state examination, montreal cognitive assessment, and dementia rating scale-2 scores in Parkinson's disease. *Mov. Disord.* **29**, 1809–1815 (2014).
71. Uc, E. Y. et al. Incidence of and risk factors for cognitive impairment in an early Parkinson disease clinical trial cohort. *Neurology* **73**, 1469–1477 (2009).
72. Mollenhauer, B. et al. Baseline predictors for progression 4 years after Parkinson's disease diagnosis in the De Novo Parkinson Cohort (DeNoPa). *Mov. Disord.* **34**, 67–77 (2019).
73. Kasten, M. et al. Cohort Profile: a population-based cohort to study non-motor symptoms in parkinsonism (EPIPARK). *Int. J. Epidemiol.* **42**, 128–128k (2013).

74. Bien, S. A. et al. Strategies for enriching variant coverage in candidate disease loci on a multiethnic genotyping array. *PLoS ONE* **11**, e0167758 (2016).
75. Purcell, S. et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
76. Anderson, C. A. et al. Data quality control in genetic case-control association studies. *Nat. Protoc.* **5**, 1564–1573 (2010).
77. 1000 Genomes Project Consortium *et al.* A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073 (2010).
78. Price, A. L. et al. Principal components analysis corrects for stratification in genome-wide association studies. *Nat. Genet.* **38**, 904–909 (2006).
79. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
80. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
81. Loh, P. R. et al. Reference-based phasing using the Haplotype Reference Consortium panel. *Nat. Genet.* **48**, 1443–1448 (2016).
82. Desikan, R. S. et al. Genetic assessment of age-associated Alzheimer disease risk: development and validation of a polygenic hazard score. *PLoS Med.* **14**, e1002258 (2017).
83. Cuellar-Partida, G., Renteria, M. E. & MacGregor, S. LocusTrack: integrated visualization of GWAS results and genomic annotation. *Source Code Biol. Med.* **10**, 1 (2015).

Acknowledgements

We are grateful to O. Nemirovsky for his philanthropic support, encouragement and insights. We thank all study participants, their families and friends for their support and participation, and our study coordinators for making this work possible. We thank A. Brown at Partners HealthCare Personalized Medicine and Y. Kuras at Brigham and Women's Hospital for technical assistance. The study was funded in part by philanthropic support (to Brigham & Women's Hospital and C.R.S.) for Illumina MEGA chip genotyping; and in part by NINDS and NIA R01NS115144 (to C.R.S.), which funded genotyping for the PHS validation cohorts EPIPARK and HBS2. The work of C.R.S. was supported by NIH grants NINDS/NIA R01NS115144, U01NS095736, U01NS100603, and the American Parkinson Disease Association Center for Advanced Parkinson Research. T.M.H. is funded by NIH grant K23NS099380. While this manuscript was in revision, G.L. transferred to a new position at Sun Yat-sen University, where he received support from the National Natural Science Foundation of China (project no. 31900475), the Fundamental Research Funds for the Central Universities (project no. 19ykpy146), Young Talent Recruitment Project of Guangdong (project no. 2019QN01Y139) and Shenzhen Basic Research Project (project no. JCYJ20190807161601692). For each individual cohort, acknowledgements are listed in the Supplementary Note.

Author contributions

C.R.S. conceived and designed the study. G.L. contributed to the study design, and carried out the statistical and bioinformatics analyses with J.P., J.J.L. and C.R.S. X.D. contributed to the analysis. Z.L. and S.S.A. performed genotyping. Patient samples and phenotypic data were collected by J.-C.C., F.Z., J.M.-G., M.C.C., A.E., S.L., A.B., G.M., J.H.G., A.Y.H., M.A.S., M.T.H., A.-M.W., T.M.H., B.R., I.S., S.K., P.T., T.G.B., F.C.-D., G.A., O.-B.T., J.S.P., P.H., J.J.v.H., R.A.B., C.H.W.-G., J.M., M.K., C.K., C.T., B.M. and C.R.S. C.R.S. and G.L. drafted the manuscript. All authors reviewed, edited and approved the manuscript prior to submission.

Competing interests

Brigham and Women's Hospital holds a US provisional patent application on the PHS for predicting PD progression, on which C.R.S. is named as inventor. Outside this work, C.R.S. has served as consultant, scientific collaborator or on scientific advisory boards for Sanofi, Berg Health, Pfizer and Biogen, and has received grants from the NIH, US Department of Defense, American Parkinson Disease Association, and the Michael J Fox Foundation (MJFF). G.L., J.J.L., J.M., A.E., J.H.G., A.Y.H., S.K., P.T., S.S.A., J.S.P. and M.C.C. report no relevant financial or other conflicts of interest in relation to this study. M.A.S. has no conflict of interest related to this work. Outside this work, M.A.S. has received grants from NINDS, DoD, MJFF, Farmer Family Foundation, and has served as a consultant to commercial programs for Eli Lilly & Co (data monitoring committee), Prevail Therapeutics (scientific advisory board), Denali Therapeutics (scientific advisory board), nQ Medical (scientific advisory board), Chase Therapeutics (scientific advisory board) and Partner Therapeutics (scientific advisory board). A.-M.W. has received research funding from the ALS Association, the Parkinson's Foundation, has participated in clinical trials funded by Acorda, Biogen, Bristol-Myers Squibb, Sanofi/Genzyme, Pfizer and Abbvie, and received consultant payments from Mitsubishi Tanabe and from Accordant. T.M.H. has no conflict of interest related to this work. Outside this work, he has received honoraria for consulting in advisory boards for Boston Scientific and Medtronic. J.-C.C. has no conflict of interest related to this work. Outside this work, J.C.C. has received honoraria for consulting in advisory boards for Abbvie, Actelion, Air Liquide, Biogen, BMS, BrainEver, Clevexel, Denali, Pfizer, Theranexus and Zambon. B.R. is an employee of and holds equity in Praxis Precision Medicines and is advisor for Caraway Therapeutics and Brain Neurotherapy Bio. I.S. is Principal Investigator of a MJFF Computational Science Grant (2017–19). S.K. is supported by Multiple Sclerosis of Western-Australia (MSWA) and the Perron Institute. P.H. is a Scientific Advisor of Neuron23. J.J.v.H. has no conflict of interest related to this work. Outside this work, J.J.v.H. has received grants from the Alkemade-Keuls Foundation, Stichting Parkinson Fonds, Parkinson Vereniging, the Netherlands Organisation for Health Research and Development, the Netherlands Organisation for Scientific Research, Hersenstichting, AbbVie, MJFF, and research support from Hoffmann-La-Roche, Lundbeck and the Centre of Human Drug Research. B.M. has no conflict related to this work. Outside this work, B.M. has received honoraria for consultancy from Roche, Biogen, AbbVie, Servier and Amprion. B.M. is member of the executive steering committee of the Parkinson Progression Marker Initiative and Principal Investigator of the Systemic Synuclein Sampling Study of the MJFF for Parkinson's Research and has received research funding from the Deutsche Forschungsgemeinschaft (DFG), EU (Horizon2020), Parkinson Fonds Deutschland, Deutsche Parkinson Vereinigung, Parkinson's Foundation and MJFF. R.A.B. has no conflict of interest related to this work. Outside this work, R.A.B. has received consultancy monies from LCT, FCDI, Novo Nordisk, Cellino, Sana, UC, has received royalties from Wiley and Springer Nature, grant funding from CPT, NIHR Cambridge Biomedical Research Centre (146281), the UK Medical Research Council (MRC), Wellcome Trust (203151/Z/16/Z) and Rosetrees Trust (A1519 M654). C.H.W.-G. has no conflict of interest related to this work. C.H.W.-G. is supported by a RCUK UKRI Research Innovation Fellowship awarded by the MRC (MR/R007446/1) and the NIHR Cambridge Biomedical Research Centre, and receives grant support from MJFF, the Evelyn Trust, the Cure Parkinson's Trust, Parkinson's UK, the Rosetrees Trust and the Cambridge Centre for Parkinson-Plus. C.H.W.-G. has received honoraria from Lundbeck and consultancy payments from Modus Outcomes and Evidera. C.T. is supported by EU Grant Horizon 2020/propag-ageing and the MJFF. C.K. serves as a medical advisor to Centogene for genetic testing reports in the fields of movement disorders and dementia, excluding Parkinson's disease.

Additional information

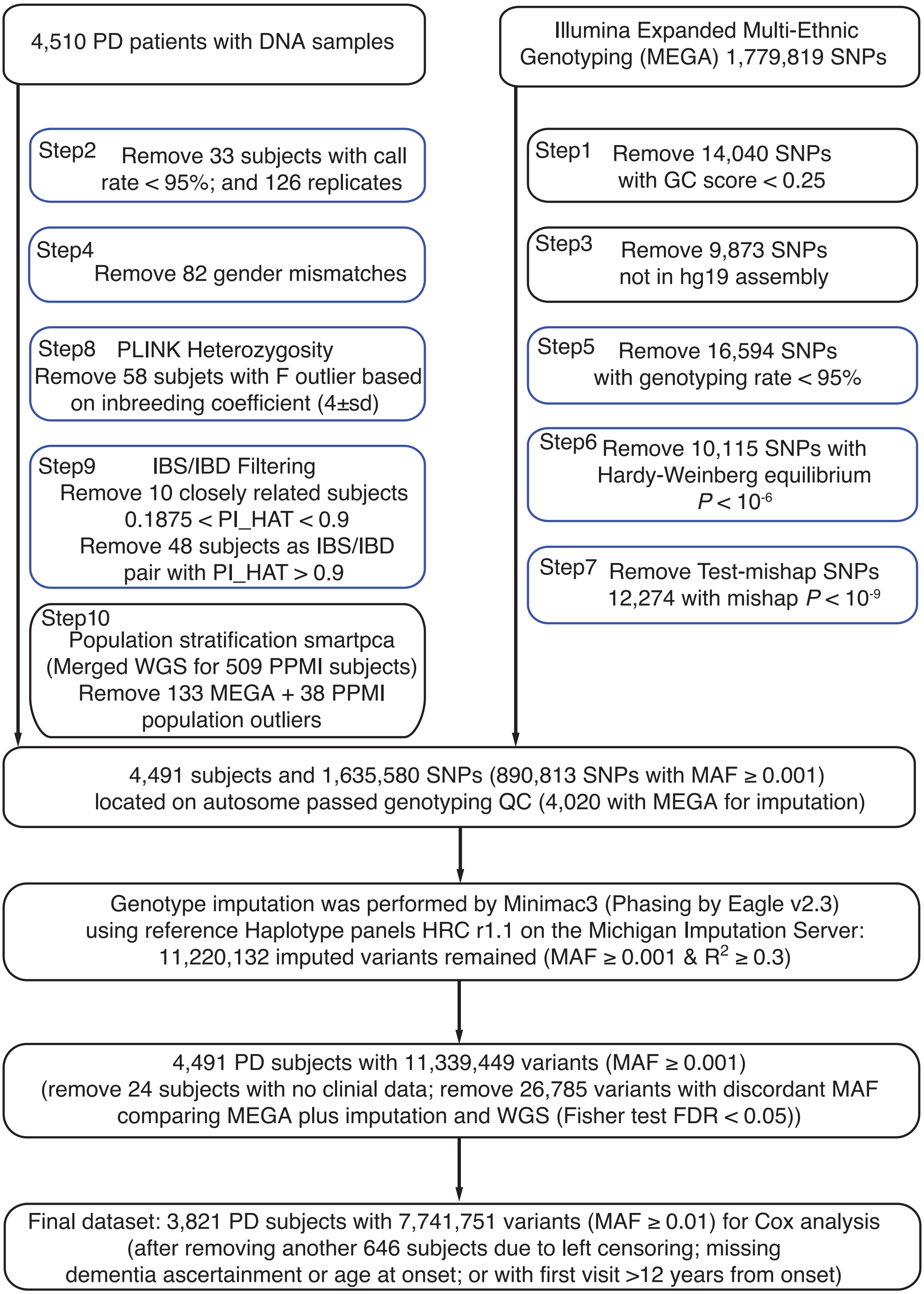
Extended data is available for this paper at https://doi.org/10.1038/s41588-021-00847-6.

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41588-021-00847-6.

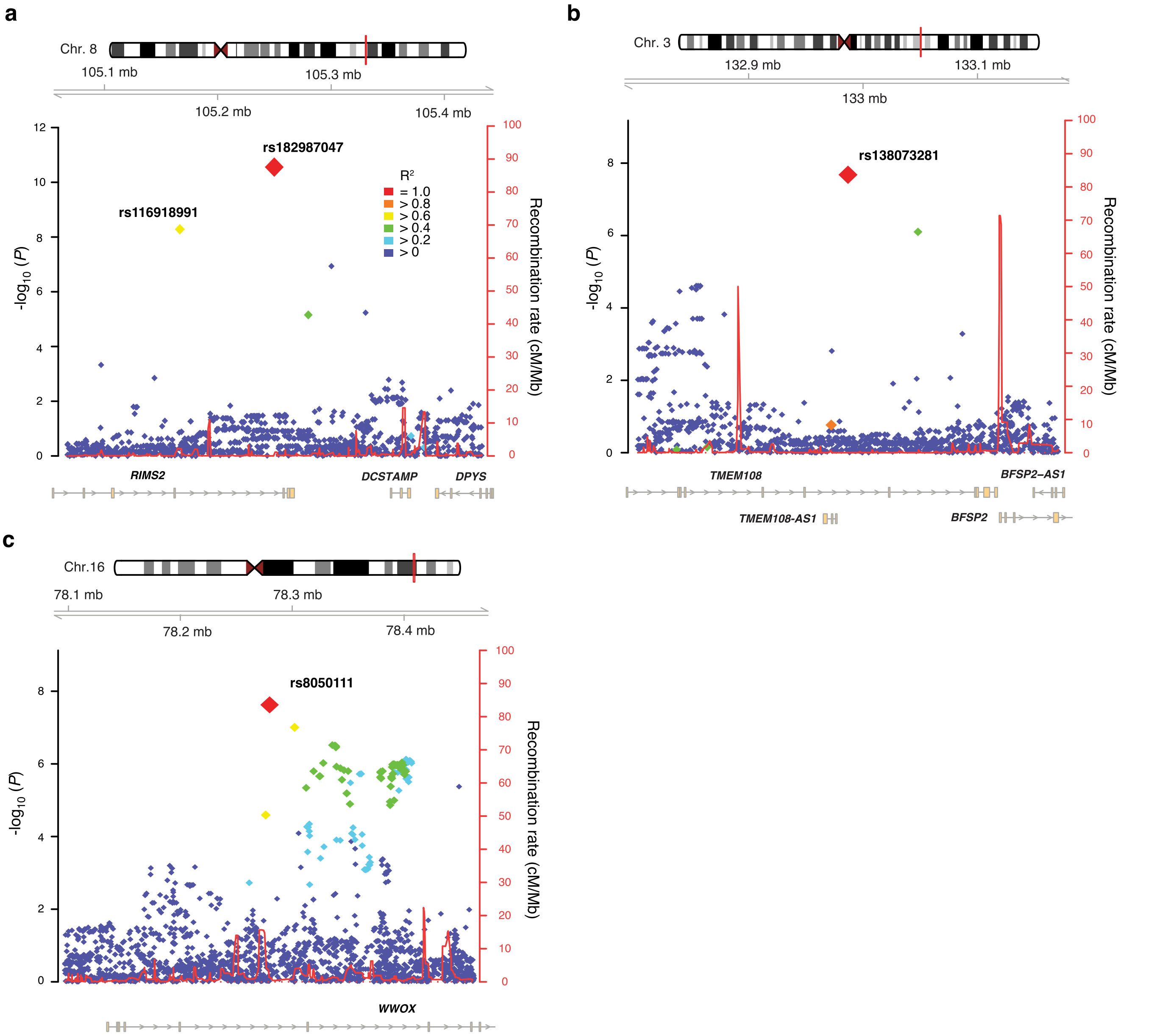
Correspondence and requests for materials should be addressed to C.R.S.

Peer review information *Nature Genetics* thanks Robert Graham and the other, anonymous, reviewer for their contribution to the peer review of this work.

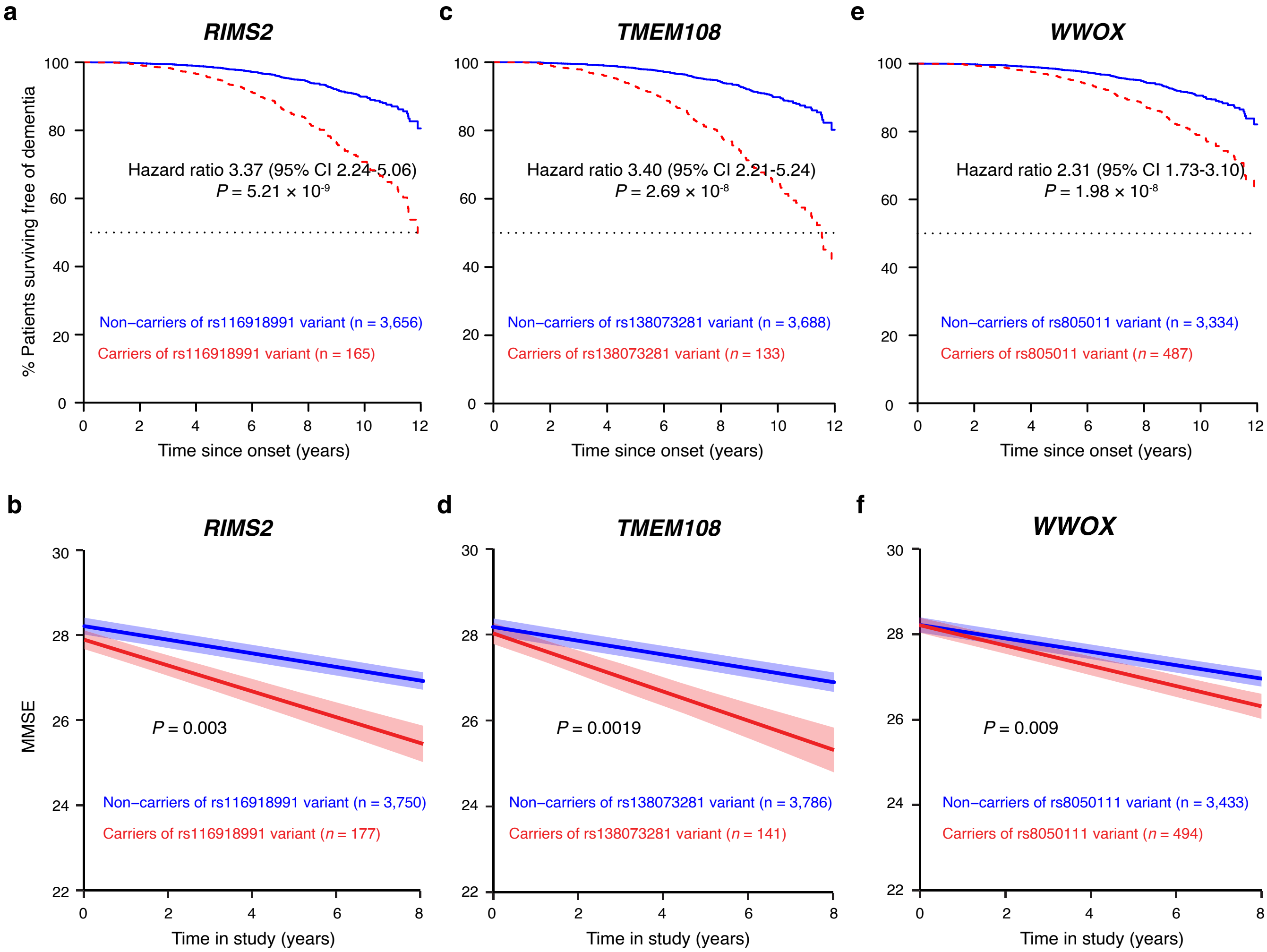
Reprints and permissions information is available at www.nature.com/reprints.


4,510 PD patients with DNA samples
Illumina Expanded Multi-Ethnic Genotyping (MEGA) 1,779,819 SNPs
Step2 Remove 33 subjects with call rate < 95%; and 126 replicates
Step1 Remove 14,040 SNPs with GC score < 0.25
Step4 Remove 82 gender mismatches
Step3 Remove 9,873 SNPs not in hg19 assembly
Step8 PLINK Heterozygosity Remove 58 subjets with F outlier based on inbreeding coefficient (4±sd)
Step5 Remove 16,594 SNPs with genotyping rate < 95%
Step9 IBS/IBD Filtering Remove 10 closely related subjects 0.1875 < PI_HAT < 0.9 Remove 48 subjects as IBS/IBD pair with PI_HAT > 0.9
Step6 Remove 10,115 SNPs with Hardy-Weinberg equilibrium $P < 10^{-6}$
Step7 Remove Test-mishap SNPs 12,274 with mishap $P < 10^{-9}$
Step10 Population stratification smartpca (Merged WGS for 509 PPMI subjects) Remove 133 MEGA + 38 PPMI population outliers
4,491 subjects and 1,635,580 SNPs (890,813 SNPs with MAF ≥ 0.001) located on autosome passed genotyping QC (4,020 with MEGA for imputation)
Genotype imputation was performed by Minimac3 (Phasing by Eagle v2.3) using reference Haplotype panels HRC r1.1 on the Michigan Imputation Server: 11,220,132 imputed variants remained (MAF ≥ 0.001 & $R^2 \geq 0.3$)
4,491 PD subjects with 11,339,449 variants (MAF ≥ 0.001) (remove 24 subjects with no clinial data; remove 26,785 variants with discordant MAF comparing MEGA plus imputation and WGS (Fisher test FDR < 0.05))
Final dataset: 3,821 PD subjects with 7,741,751 variants (MAF ≥ 0.01) for Cox analysis (after removing another 646 subjects due to left censoring; missing dementia ascertainment or age at onset; or with first visit >12 years from onset)

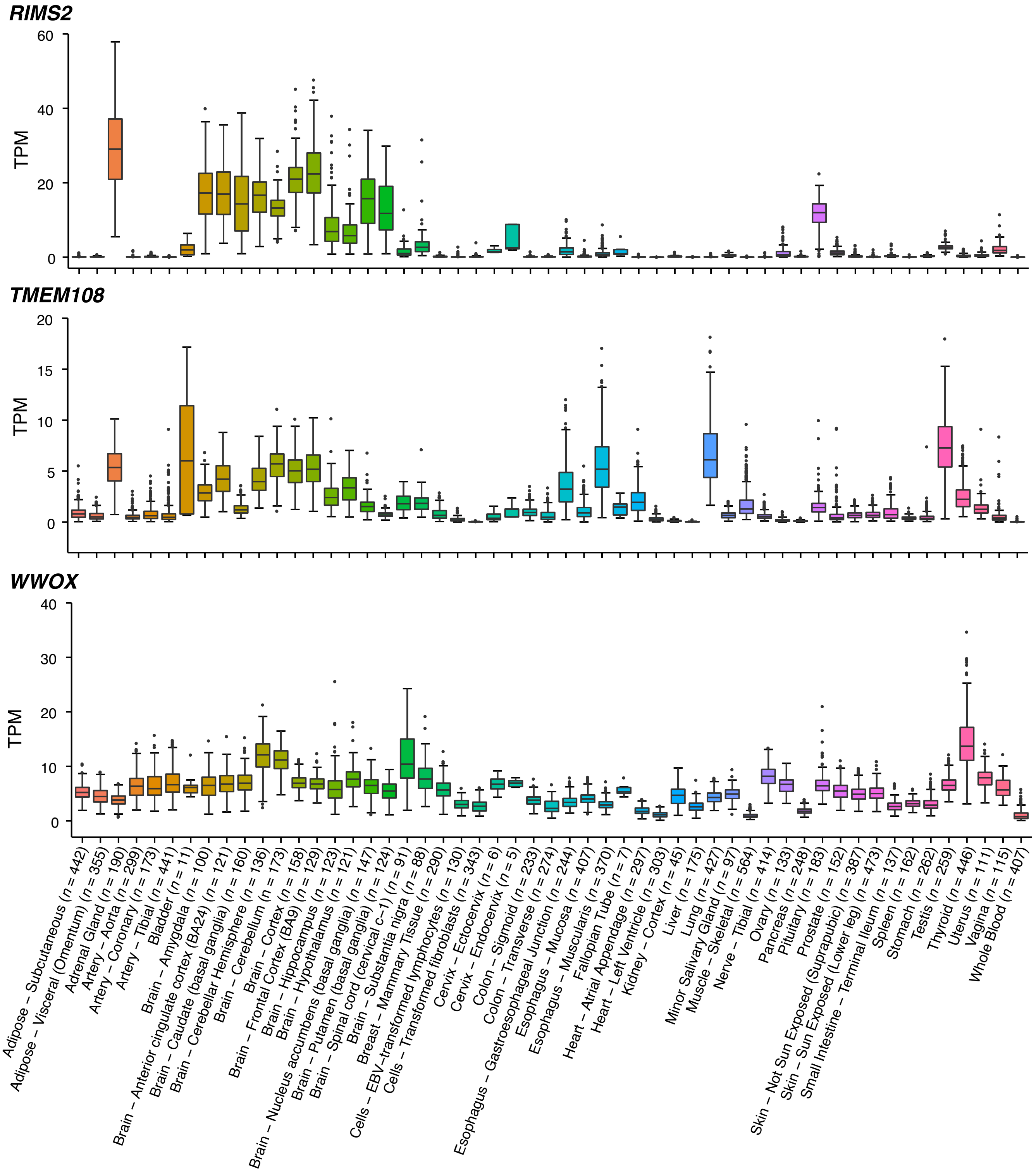
Extended Data Fig. 1 | Genotyping pipeline for discovery and replication cohorts. Quality control (QC) steps outlined in blue were performed using PLINK v1.90beta[75]. Note that 509 samples with WGS from the PPMI cohort (after removing three with gender mismatches) were added in Step 10.



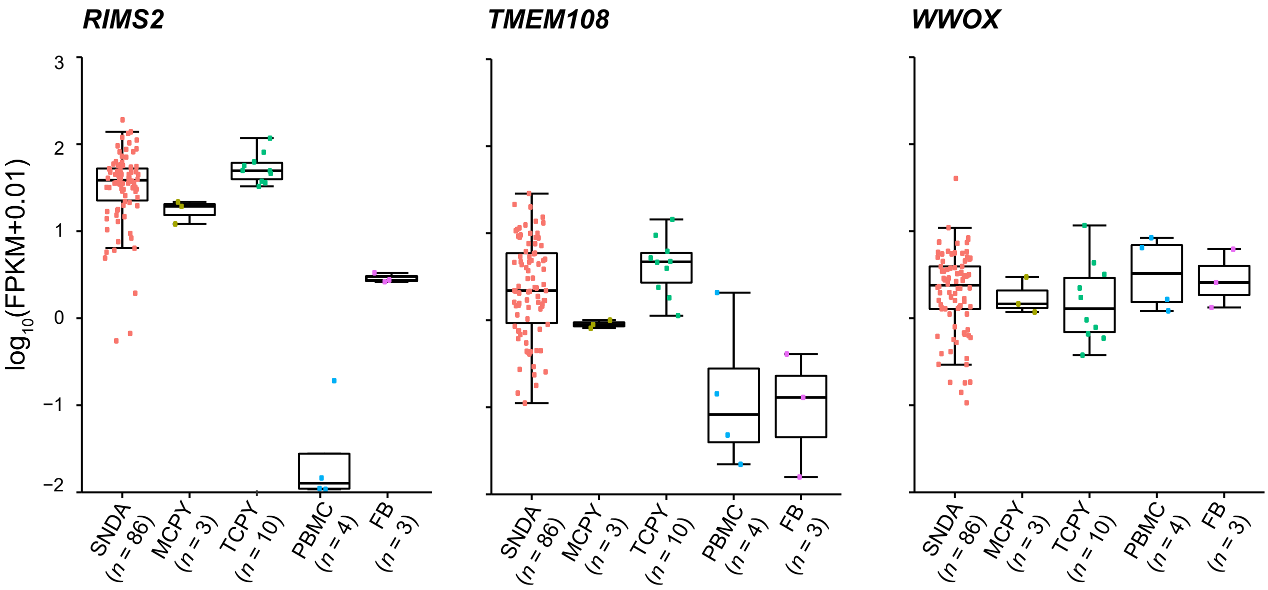
Extended Data Fig. 2 | Characteristics of loci associated with cognitive progression in PD. a, *RIMS2* locus. **b**, *TMEM108* locus. **c**, *WWOX* locus. Top, chromosomal position; middle, -log$_{10}$(*P* values) for individual SNPs at each locus (left *y*-axis) with the rate of recombination indicated by the red line (right *y*-axis); bottom, gene positions with the locus. Each point represents a SNP colored according to LD with the lead associated variant. Figure panels were generated with LocusTrack[83] and *r*^{2} values were calculated based on CEU population in the 1000 Genomes Project data set[77].



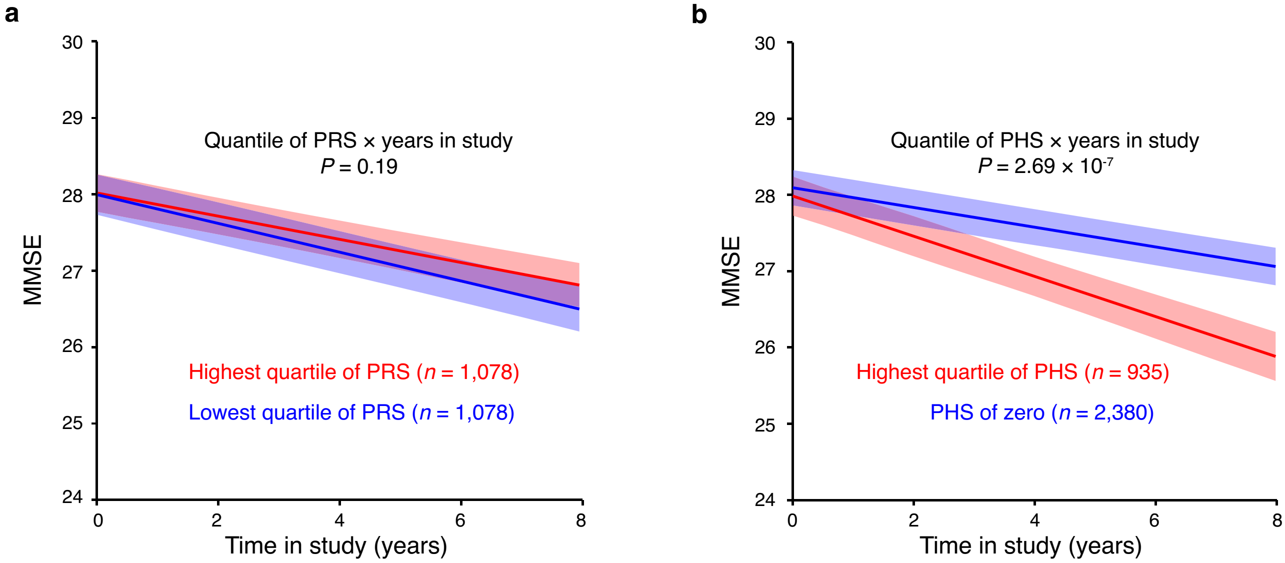
Extended Data Fig. 3 | Associations between a second *RIMS2* variant rs116918991, *TMEM108* rs138073281, and *WWOX* rs8050111 with cognitive PD progression. a,**c**,**e**, Covariate-adjusted survival curves for PD patients without the indicated variant (blue line) and for those carrying the indicated variant (heterozygotes and homozygotes; red dashed line) are shown. *P* values Cox PH models with two-sided Wald test. **b**,**d**,**f**, Adjusted mean MMSE scores across time predicted from the estimated fixed-effect parameters of the LMM analysis are shown for cases carrying the variant (heterozygotes and homozygotes; red) and cases without the variant (non-carriers; blue) adjusting for covariates. Shaded ribbons indicate ± s.e.m. around predicted MMSE scores across time. Note that a second *RIMS2* variant rs116918991 (correlated with $r^2 = 0.49$ with the lead variant rs182987047; Fig. 1) is shown in **a** and **b**, and that the HR and *P* values shown here for *TMEM108* rs138073281 and *WWOX* rs8050111 are different from the HR and *P* values from the main analysis (Table 1), where variant alleles were coded as 0, 1, 2. *P* values from LMM analysis with two-sided *t*-test.



Extended Data Fig. 4 | *RIMS2*, *TMEM108*, and *WWOX* are expressed in human brain. Gene expression profiles were downloaded directly from the GTEx Portal V7[46]. Expression values are shown in Transcript per Million (TPM), calculated from a gene model with isoforms collapsed to a single gene. Box plots visualize first, third quartiles and medians; the ends of the whiskers represent the lowest (or highest) value still within 1.5-times the interquartile range. Outliers are displayed as dots, if they are above or below 1.5-times the interquartile range. *n* indicates number of individuals for each tissue analyzed in GTEx V7.



Extended Data Fig. 5 | Cell-type specific expression of *RIMS2*, *TMEM108*, and *WWOX* in human brain. Cell type-specific transcriptomes were assayed using laser-capture RNA sequencing (lcRNAseq) as we reported[47]. Gene expression (FPKM) profiles of *RIMS2*, *TMEM108*, and *WWOX* are from BRAINcode consortium (http://www.humanbraincode.org). *n* indicates the number of individuals assayed for each cell type. SNDA, indicates dopamine neurons laser-captured from human substantial nigra pars compacta; MCPY, pyramidal neurons from human motor cortex; TCPY, pyramidal neurons from human temporal cortex; PBMC, human peripheral blood mononuclear white cells; FB, primary human fibroblasts. Box plots visualize first, third quartiles, and medians; the ends of the whiskers represent the lowest (or highest) value still within 1.5-times the interquartile range. Each dot represents a sample.



Extended Data Fig. 6 | The polygenic hazard score (PHS) is associated with decline in serial MMSE scores. a, PD cases scoring in the highest quartile (red) of a polygenic risk score (PRS based on 90 susceptibility variants[45]) compared to PD cases scoring in the lowest quartile of the PRS (blue) are shown. **b**, PD cases scoring in the highest quartile (red) of the PHS (comprising *GBA* + *APOE* ε4 + the 3 novel progression variants) compared to PD cases scoring zero on the PHS (blue) are shown. For **a** and **b**, adjusted mean MMSE scores across time predicted from the estimated fixed-effect parameters in the LMM analysis for the combined data set comprising discovery and replication populations are shown. The shaded ribbons indicate ± s.e.m. around predicted MMSE scores across time. The *P* values from LMM analysis with two-sided *t*-tests.

nature research

Corresponding author(s): Clemens R. Scherzer

Last updated by author(s): 2021/2/15

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a | Confirmed

☐ ☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement

☐ ☒ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly

☐ ☒ The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.

☐ ☒ A description of all covariates tested

☐ ☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons

☐ ☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)

☐ ☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted
Give P values as exact values whenever suitable.

☒ ☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings

☒ ☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes

☐ ☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection.
Data analysis	Genotype imputation was performed using Minimac3 (v2.0.1) on the Michigan online imputation server. The SnpSift tool (v4.0) was used to evaluate the concordance between imputed SNPs (based on the MEGA array) and SNPs directly called from whole genome sequencing in 562 individuals from HBS for which both assays were available. Cox proportional hazards analyses were performed using the coxph function in the Survival package (version 2.38-1). Generalized longitudinal mixed fixed and random effects analysis was performed using the glmmPQL function in the MASS package (version 7.3-37) in R (version 3.3.1). R version 3.3.1 was also used to compute polygenic hazard and polygenic risk scores as described in the methods. Analysis code is made available in https://github.com/sixguns1984/GWSS.PDD. To evaluate the performance of the PHS models, the cumulative/dynamic receiver operating characteristic (ROC) curves, area under curves (AUC), confidence intervals of the AUC (simulation method), and comparisons between two AUCs were calculated using the timeROC package (Version 0.2) in R (version 3.3.1). FUMA (v1.3.3. http://fuma.ctglab.nl) was used to evaluate prognosis loci.

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

A Life Sciences Reporting Summary for this paper is available. Human brain cell type-specific expression data from BRAINcode. RNAseq data are accessible through a user-friendly webportal at www.humanbraincode.org and individual-level data through dbGAP (acc. number phs001556.v1.p1). The gene expression profiles of human tissues used in this study can be downloaded from the GTEx Portal V7 (https://gtexportal.org/). The full GWSS summary statistics for the combined analysis of discovery and replication populations are publicly accessible through the EGA database https://www.ebi.ac.uk/ega/ (acc. Number EGAS00001005110). Individual-level genetic data for the NIH-funded Illumina Multi-Ethnic Genotyping Array analyses of the HBS2 and EPIPARK cohorts are accessible in dbGAP with accession number phs002328.v1.p1 in accordance with NIH Genomic Data Sharing Policy. The whole genome sequencing and clinical data for PPMI included in this study are publicly available through a PPMI Whole Genome Sequencing Data Agreement at ppmi@loni.usc.edu. Clinical data for PDBP included in this study are publicly available through https://pdbp.ninds.nih.gov. Clinical longitudinal data for the other cohorts included are accessible through appropriate data sharing agreements that protect patient privacy with the institutions which conducted or are conducting study consents and clinical assessments under local IRB approvals.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   □ Behavioural & social sciences   □ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size

No sample size calculations were performed. Sample sizes were chosen based on availability of cohorts.

Data exclusions

Sample and variant quality control was performed as outlined in the methods. Samples not meeting Q/C criteria were excluded as shown in Extended Data Figure 1. The exclusion criteria of genotyping Q/C were pre-established. In genome-wide survival analysis, total 670 patients were left-censored and those with missing or non-quality clinical data were excluded.

Replication

Progression variants meeting genome-wide significance in the discovery phase were evaluated in an independent replication population once. Moreover, an exploratory joint-phase (combined discovery and replication populations) meta-analysis was conducted. The PHS score was developed in the PHS Development Stage population (combined discovery and replication populations) and evaluated in an independent PHS Validation Population once. All attempts at replication were successful.

Randomization

In this within-cases longitudinal cohort study, PD patients were enrolled and longitudinally assessed; neither patients nor investigators were aware of an individual's whole genome genotyping status; after completion of longitudinal follow-up assessments, PD patients were assigned to carrier and non-carrier groups based on genotypes.

Blinding

Patients and investigators were blinded to the participants' whole genome genotyping data during enrollment and longitudinal follow-up.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study	
☒	□	Antibodies
☒	□	Eukaryotic cell lines
☒	□	Palaeontology
☒	□	Animals and other organisms
□	☒	Human research participants
☒	□	Clinical data

Methods

n/a	Involved in the study	
☒	□	ChIP-seq
☒	□	Flow cytometry
☒	□	MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics

Characteristics of the cohorts are described in Supplementary Table 1 and Supplementary Fig. 1 .

Recruitment

All participants were recruited into individual cohorts as part of previous studies. The cohorts include population-based, incident cohort studies; purpose-built biomarkers studies with carefully standardized, highly compatible clinical and biospecimens collection methods; and carefully phenotyped, failed Phase III clinical trials. Genetic analyses were performed at the end of the clinical longitudinal follow-up period. Physicians therefore recruited and longitudinally assessed the participants without knowledge of their genotypes. This cohort study design is thought be less vulnerable to recruitment and ascertainment bias than case-control studies. The population-based cohorts included in the analysis were designed to guard against self-selection bias, which may affect clinical trial cohorts.

Ethics oversight	Written informed consent for DNA collection and phenotypic data collection for secondary research use for each cohort was obtained from the participants with approval from the local ethics committees. The Institutional Review Board of Partners HealthCare approved the current genotyping and analyses. For PPMI, approval was obtained to download and analyze the publicly available WGS and clinical data.

Note that full information on the approval of the study protocol must also be provided in the manuscript.