Genome-wide determinants of mortality and clinical progression in Parkinson's disease

Authors: Manuela MX TAN, PhD[1,2,3], Michael A LAWTON, PhD[4], Miriam I POLLARD, BSc[2], Emmeline BROWN, BSc[2], Samir BEKADAR, PhD[5], Edwin JABBARI, MRCP, PhD[2,3], Regina H REYNOLDS, PhD[6,31], Hirotaka IWAKI, MD, PhD[7,8], Cornelis BLAUWENDRAAT, PhD[7,9], Sofia KANAVOU, MSc[4], Leon HUBBARD, PhD[10], Naveed MALEK, MRCP[11], Katherine A GROSSET, MD[11], Nin BAJAJ, PhD[12], Roger A BARKER, PhD[13,14,31], David J BURN, MD[15], Catherine BRESNER, BSc[10], Thomas FOLTYNIE, PhD[2,3], Nicholas W WOOD, PhD[2,3,31], Caroline H WILLIAMS-GRAY, MRCP, PhD[13], Ole A ANDREASSEN, MD, PhD[16,17], Mathias TOFT, MD, PhD[1,17], Alexis ELBAZ, MD, PhD[18], Fanny ARTAUD, PhD[18], Alexis BRICE, MD[5], Jean-Christophe CORVOL, MD, PhD[5], Jan AASLY, MD, PhD[19,20], Matthew J FARRER, PhD[21], Michael A NALLS, PhD[7,8], Andrew B SINGLETON, PhD[7,9], Nigel M WILLIAMS, PhD[10], Yoav BEN-SHLOMO, MB BS, PhD[4], John HARDY, PhD[3,22,23,24,25,26,31], Michele TM HU, PhD[27,28,29], Donald G GROSSET, MD[30], Maryam SHOAI, PhD[22,23,31], Lasse PIHLSTRØM, MD, PhD[1], Huw R MORRIS, PhD, FRCP[2,3,31].

Affiliations

[1] Department of Neurology, Oslo University Hospital, Oslo, Norway

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

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

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

[5] Sorbonne University, Paris Brain Institute - ICM, Inserm, CNRS, Assistance Publique Hôpitaux de Paris, Departement of Neurology, Hôpital Pitié-Salpêtrière, Paris, France

[6] Genetics and Genomic Medicine, Great Ormond Street Institute of Child Health, University College London, London, UK

[7] Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, USA

[8] Data Tecnica International, Washington, DC, USA

[9] Center for Alzheimer's and Related Dementias, National Institutes of Health, Bethesda, MD, USA

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

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

[12] Clinical Neurosciences, University of Nottingham, Nottingham, UK

[13] John Van Geest Centre for Brain Repair, Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK
[14] Wellcome-MRC Cambridge Stem Cell Institute, University of Cambridge, Cambridge, UK
[15] Faculty of Medical Sciences, Newcastle University, Newcastle upon Tyne, UK
[16] NORMENT, Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway
[17] Institute of Clinical Medicine, Faculty of Medicine, University of Oslo, Oslo, Norway
[18] Paris-Saclay University, UVSQ, Inserm, Gustave Roussy, "Exposome and Heredity" team, CESP, 94807, Villejuif, France
[19] Department of Neurology, St. Olavs Hospital, Trondheim, Norway
[20] Department of Neuromedicine and Movement Science (INB), Faculty of Medicine and Health Sciences, Norwegian University of Science and Technology (NTNU), Trondheim, Norway
[21] Department of Neurology, University of Florida, Gainesville, FL, United States
[22] Department of Neurodegenerative Diseases, Queen Square Institute of Neurology, University College London, London, UK
[23] Reta Lila Weston Institute, UCL Queen Square Institute of Neurology, London, UK
[24] UK Dementia Research Institute, University College London, London, UK
[25] National Institute for Health Research (NIHR) University College London Hospitals Biomedical Research Centre, London, UK
[26] Institute for Advanced Study, The Hong Kong University of Science and Technology, Hong Kong SAR, China
[27] Nuffield Department of Clinical Neurosciences, Division of Clinical Neurology, University of Oxford, Oxford, UK
[28] Oxford Parkinson's Disease Centre, University of Oxford, Oxford, UK
[29] Department of Clinical Neurology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK
[30] School of Neuroscience and Psychology, University of Glasgow, Glasgow, UK
[31]Aligning Science Across Parkinson's (ASAP) Collaborative Research Network, Chevy Chase, MD, 20815.

Corresponding authors:
Dr Manuela MX Tan, Department of Neurology, Oslo University Hospital, Oslo, Norway, manuela.tan@medisin.uio.no.

Professor Huw R Morris, Department of Clinical and Movement Neurosciences, Queen Square Institute of Neurology, Royal Free Hospital, Rowland Hill Street, London NW3 2PF, h.morris@ucl.ac.uk.

Running title: PD progression GWAS
Key Words: Parkinson's disease, genetics, progression, genome-wide association study, mortality, survival, clinical milestones, cognitive impairment, dementia
Funding sources: Parkinson's UK, Aligning Science Across Parkinson's through the Michael J Fox Foundation for Parkinson's Research, Southern and Eastern Norway Regional Health Authority

Abbreviations: FUMA = Functional Mapping and Annotation of GWAS; GWAS = Genome-Wide Association Study; H&Y = Hoehn and Yahr; LD = Linkage Disequilibrium; MDS-UPDRS = Movement Disorder Society Unified Parkinson's Disease Rating Scale; MoCA = Montreal Cognitive Assessment; MMSE = Mini Mental State Examination; PD = Parkinson's Disease; PPMI = Parkinson's Progression Markers Initiative; SNP = Single Nucleotide Polymorphism; SD = Standard Deviation; SE = Standard Error; UKB = UK Biobank.

ABSTRACT

Background: There are 90 genetic risk variants for Parkinson's disease (PD) but currently only five nominated loci for PD progression. The biology of PD progression is likely to be of central importance in defining mechanisms that can be used to develop new treatments.

Methods: We studied 6,766 PD patients, over 15,340 visits with a mean follow-up of between 4.2 and 15.7 years and carried out a genome wide survival study for time to motor progression, defined by reaching Hoehn and Yahr stage 3 or greater, cognitive impairment as defined by serial cognitive examination, and death (mortality).

Findings: There was a robust effect of the ***APOE*** ε4 allele on mortality and cognitive impairment in PD. We identified three novel loci for mortality and motor progression, and nominated genes based on physical proximity or expression quantitative trait loci data. One locus within the **TBXAS1** gene encoding thromboxane A synthase 1 was associated with mortality in PD (HR = 2.04 [95% CI 1.63 to 2.56], p-value = 7.71×10^{-10}). Another locus near the ***SYT10*** gene encoding synaptotagmin 10 was associated with mortality just above genome-wide significance (HR=1.36 [95% CI 1.21 to 1.51], p-value=5.31×10^{-8}). A genomic variant associated with the expression of ***ADORA2A*****,** encoding the A2A adenosine receptor, was associated with motor progression (HR=4.83 [95% CI 2.89 to 8.08], p-value=1.94×10^{-9}). Only the non-Gaucher disease causing ***GBA*** PD risk variant E326K, of the known PD risk variants, was associated with progression in PD.

Interpretation: We report three novel loci associated with PD progression or mortality. Further work is needed to understand the links between these genomic variants and the underlying disease biology. However, thromboxane synthesis, vesicular peptidergic neurotransmitter release and the A2A adenosine receptor may represent new candidates for disease modification in PD.

RESEARCH IN CONTEXT

Evidence before this study: We searched PubMed for articles on Parkinson's disease (PD) with no language restrictions from database inception up to February 9, 2022. We used the search terms "Parkinson disease AND genetics" and "disease progression OR survival OR mortality OR prognosis OR longitudinal studies". We also conducted this search with the addition of "genome-wide association study" (GWAS) to focus on these genome-wide analyses. There are now three published large-scale GWASs

investigating PD progression, and many candidate variant studies. However, no genome-wide studies have reported on survival/mortality in PD.

Added value of this study: To our knowledge, this is the first GWAS of survival in PD. Our study highlights new loci influencing survival in PD, including *TBXAS1* and *SYT10*. We also conducted GWASs of progression to other clinical milestones, Hoehn and Yahr stage 3 or greater, and cognitive impairment. We show that *APOE* influences both mortality and cognitive progression in PD, and report an additional locus influencing expression of *ADORA2A* which affects the rate of motor progression.

Implications of all the available evidence: With the exception of *APOE*, we report new loci which have not been previously associated with PD progression or for mortality and ageing in the general population. These loci could be investigated in functional studies as potential drug targets to stop or slow progression of PD. In addition, new genetic loci can help to improve our understanding of the biology of PD progression and prediction of progression. Further replication of these loci is also needed in independent, longitudinal PD cohorts.

INTRODUCTION

Parkinson's disease (PD) is a progressive neurodegenerative condition for which there are no treatments to stop or slow disease progression. Large-scale genome-wide case-control association studies (GWASs) of PD have identified 90 independent variants associated with disease risk[1]. However, it is also important to study the genetics and biology of disease progression. This will facilitate the development of potential disease-modifying treatments. There have now been a handful of GWAS which aim to identify genetic variants associated with progression in PD. These have nominated loci in *SLC44A1* (encoding choline transporter like protein -1, involved in membrane synthesis) for progression to Hoehn and Yahr (H&Y) stage 3 or greater,[2] *APOE* for cognitive progression,[3] and *RIMS2* (encoding the RAB3 interacting RIMS2 protein, involved in neurotransmitter release) for progression to PD dementia.[4] In addition, many candidate gene studies have shown that variants in *GBA*, *APOE,* and potentially *MAPT*, are associated with the rate of PD motor and cognitive progression.[5]

PD progression may be determined by differential cellular susceptibility, related to mitochondrial function or proteostasis, differential cell to cell spread of pathology, or novel pathways and mechanisms. Risk factors determined from case control studies indicate aetiological pathways and guide future preventive trials, these may or may not differ from risk factors that determine disease progression. Currently disease modifying treatment trials focus on intervention in recently diagnosed patients, related to disease progression after diagnosis. Work on large scale longitudinal cohorts over the last ten years has now enabled the collaborative study of large clinico-genetic datasets. Here we have carried out GWAS of progression to three key clinical milestones in PD: mortality, H&Y 3 or greater, and cognitive impairment. We have analysed data from 6,766 PD patients with over 15,340 visits and mean follow-up ranging between 4.2 to 15.7 years.

METHODS

We studied 11 cohorts from Europe and America, and included cohorts in each analysis who had sufficient data on the outcomes of interest (see Supplementary Materials). Genotyping, quality control, and imputation was performed in each cohort separately but following the same steps. Only variants with high imputation quality scores (INFO/R2 > 0.8) and minor allele frequency > 1% were retained for analysis.

We assessed the following clinical outcomes: mortality, H&Y stage 3 or greater, and cognitive impairment. Cognitive impairment was defined as a Montreal Cognitive Assessment (MoCA) score of ≤ 21, a Mini Mental State Examination (MMSE) score of ≤

26, dementia using the DSM classification, or dementia using the MDS criteria, or study withdrawal due to reported dementia or cognitive impairment. These criteria were based on previous studies in these cohorts[2,6,7]. Cohorts were excluded if less than 20 individuals met the outcome of interest during the follow-up period, or < 5% of the total cohort size. Progression to each clinical milestone from the starting time point of PD symptom onset was assessed using Cox proportional hazard models. We adjusted for age at onset, gender, and the first five genetic principal components to adjust for population stratification. Meta-analysis was performed in METAL (version 2011-03-25)[8], using an inverse variance weighted fixed effects model. GWASs with a genomic inflation factor above 1.2 were excluded from the meta-analysis. Only SNPs that were present in > 1,000 individuals were included in the final results. SNPs with heterogeneous effects across cohorts were also excluded (p-value < 0.05 for Cochran's Q-test for heterogeneity, and/or I^2 > 80). The null hypothesis was tested with the standard GWAS significance level of 5×10^{-8}. Results were uploaded to Functional Mapping and Annotation of GWAS (FUMA; https://fuma.ctglab.nl/)[9] to annotate, prioritise, and visualize GWAS results and perform gene-set analysis with MAGMA. Forest plots for the top SNPs were generated in R v3.6 using the *forestplot* package. Further details are provided in the Supplementary Materials.

We also performed candidate variant analysis of the 90 PD risk variants from the most recent PD case-control GWAS,[1] and the cumulative PD genetic risk score (GRS). We also examined associations for other candidate variants that have been reported in PD and Progressive Supranuclear Palsy (PSP) progression: *SLC44A1, RIMS2*, *WWOX, TMEM108*, *APOE* ε2 allele, *MAPT* H1 haplotype, and rs2242367 adjacent to the *LRRK2* locus. We analysed the Alzheimer's disease (AD) GRS in relation to PD progression. 38 loci passing genome-wide significance from the latest AD GWAS were used to create the AD GRS.[10] The *APOE* region was excluded from the GRS (19:40,000,000-50,000,000).[10] To clarify whether our genetic results for mortality were related to PD or non-PD causes (e.g. general immunity, cardiovascular disease, or COVID), we performed a competing risk analysis using the Fine-Gray method for each of the top SNPs to evaluate whether the SNPs were related to PD death or non-PD death (Supplementary Materials).

RESULTS

Overall 6,766 participants with PD were analysed with mean follow-up between 4.2 and 15.7 years (Table 1). We did not have data from regular follow-up visits for all studies as some studies only contributed mortality data. However in the studies that had regular follow-up visit data available, over 15,340 visits were analysed (Table 1).

GWAS of mortality

One study (Cambridge PD Research Clinic) was excluded from the meta-analysis of mortality because the study-specific genomic inflation factor was above 1.2 (Supplementary Table 1). The PPMI study was also excluded because less than 20 individuals reached this endpoint.

5,744 patients were included in the meta-analysis of mortality. Of these, 1,846 (32.1%) individuals had died with a median time to death of 10.6 years from PD onset. 7,696,389 SNPs were present in at least 1,000 individuals and 7,313,918 SNPs passed meta-analysis filtering for heterogeneity and MAF variability. The genomic inflation value of the meta-analysis after filtering was 1.04.

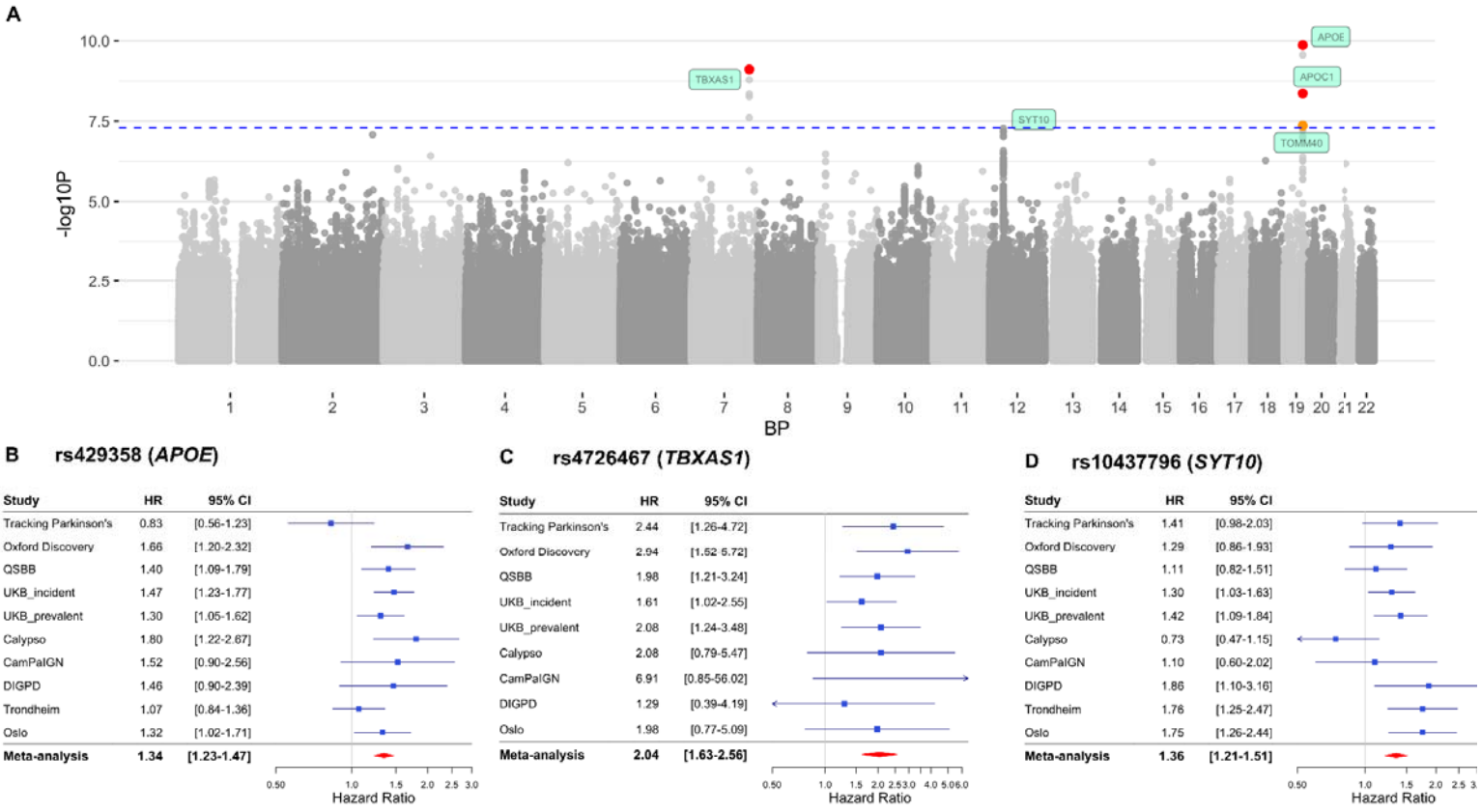


Figure 1. GWAS meta-analysis of mortality. **(A)** The Manhattan plot showing two GWAS significant loci after meta-analysis. The blue dashed line indicates the threshold for genome-wide significance, $p=5x10^{-8}$. SNPs highlighted in red have p-value$<5x10^{-9}$.

SNPs highlighted in orange have p-value<$5x10^{-8}$. One nominal association in Chromosome 12 is also annotated with the nearest gene, *SYT10* (p=$5.3x10^{-8}$). **(B)** Forest plot for the top SNP rs429358 in Chromosome 19, in *APOE*. **(C)** Forest plot for the top SNP rs4726467 in Chromosome 7, in *TBXAS1*. **(D)** Forest plot for the top SNP rs10437796 in Chromosome 12, near *SYT10.* BP = Base Pair, CI = Confidence Interval, HR = Hazard Ratio.

Two loci passed genome-wide significance and were identified to determine mortality in PD (Figure 1). The top SNP was rs429358 in Chromosome 19 (p=$4.0x10^{-10}$), which tags the *APOE* ε4 allele (Table 2). One other loci in Chromosome 7 in *TBXAS1* also reached significance (p<5 x 10^{-8}), and another loci in Chromosome 12 near *SYT10* was nominally associated (p=$5.3x10^{-8}$). Regional association plots are shown in Figure 1 and Supplementary Figures 1-3. The top ten independent SNPs identified from FUMA and the nearest genes are reported in Table 2.

In the MAGMA gene-based test, *APOE* was significantly associated with mortality (p=$1.9x10^{-10}$), and *SYT10* was associated just below genome-wide significance (p=$3.6x10^{-6}$). There was no significant association of any gene-sets or tissues in the MAGMA gene-set analysis and gene property analysis for tissue specificity.

In *TBXAS1*, there were two SNPs associated with mortality, rs4726467 and rs144889025, identified as independent in FUMA with r^2<0.6 (Table 2). However in LDpair these two SNPs are correlated with D'=0.82 and r^2=0.57, in non-Finnish European populations (https://ldlink.nci.nih.gov/). To determine whether these were independently associated with mortality, we performed conditional analysis by including both SNPs in the model. When we did this, the SNPs were no longer significantly associated (both p-values > 0.05), indicating they are not independent signals.

Both SNPs are significant expression quantitative trait loci (eQTLs), with the effect (minor) allele decreasing expression of *TBXAS1* in blood (eQTLGen; https://www.eqtlgen.org/)[11] but not other tissues as reported in GTEx (https://gtexportal.org/). There was no evidence on GTEx that the two SNPs were splicing Quantitative Trait Loci (sQTLs). Brain eQTL data at MetaBrain (https://www.metabrain.nl/)[12] did not indicate the two SNPs were significant cis-eQTLs in any brain regions from European samples. In LDproxy (https://ldlink.nci.nih.gov/), we identified two coding variants in linkage disequilibrium (LD) near the top SNPs, one

synonymous variant in *HIPK2* and one missense variant in *PARP12* (Supplementary Materials).

The top SNP in Chromosome 12, rs10437796, is not directly within *SYT10* but increases *SYT10* expression in the testis and decreases expression in the tibial nerve. The SNP also increases expression of the long noncoding RNA (lncRNA) RP11-438D14.2 (ENSG00000259937) in the brain. This is a 'sense intronic' transcript to *SYT10*, a long non-coding transcript which is within an intron of a coding gene and does not overlap any exons. Brain cis-eQTL data from MetaBrain showed that the effect allele A of rs10437796 significantly increased expression of *SYT10* in cortex but not other brain regions. This SNP was not an eQTL or sQTL for any genes in blood in eQTLGen. There were no coding variants in LD with this SNP in LDproxy within a 500kb window.

We also performed colocalization analysis to determine whether the association signals for PD mortality and gene expression are driven by a shared causal variant (see Supplementary Materials; Supplementary Table 2). We used cis-eQTL data from PsychENCODE and eQTLGen to examine gene expression in whole brain or blood, respectively. However, no PD mortality loci showed evidence of colocalization with eQTLs (PP.H4<0.75).

Cause of death analysis

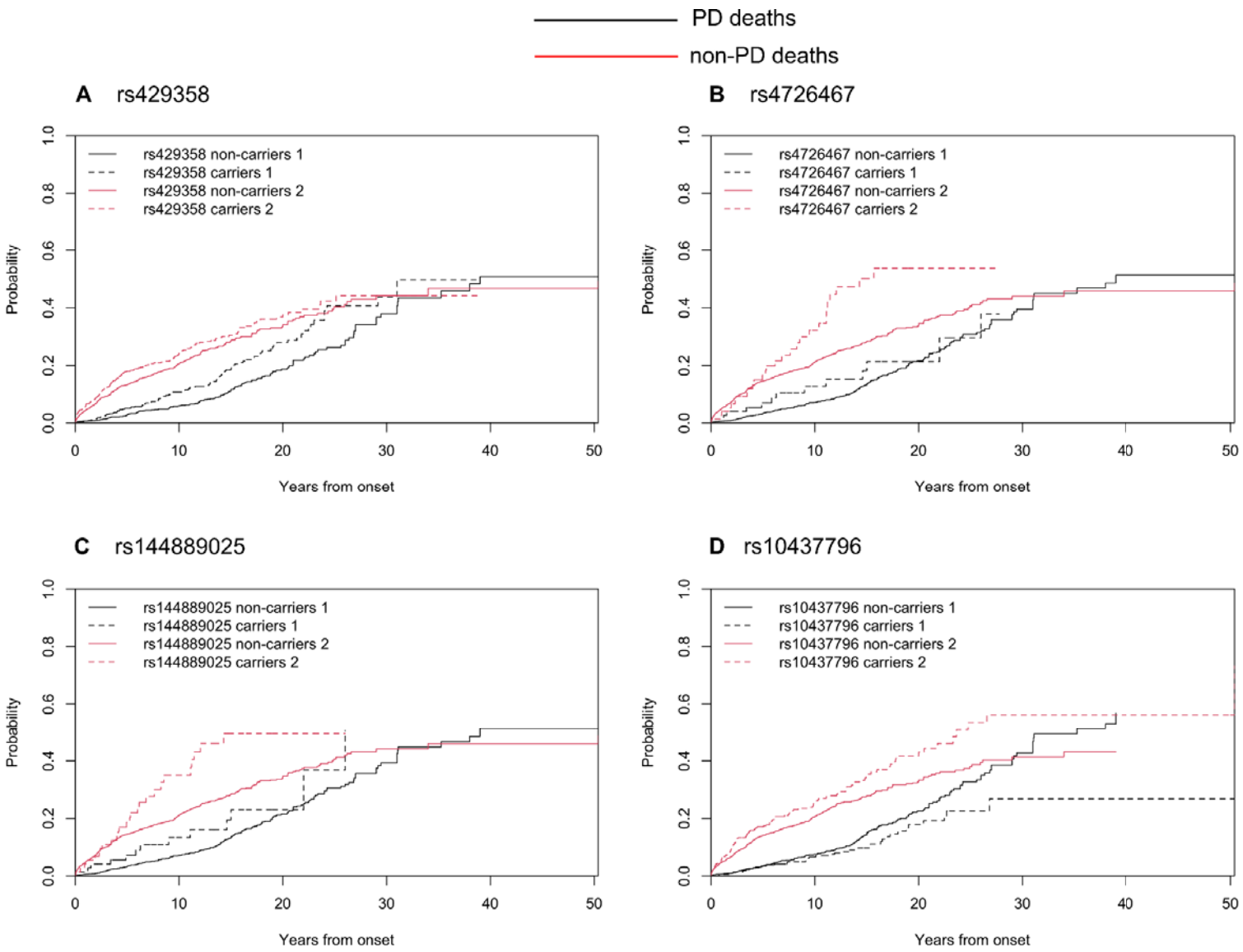
To clarify whether our genetic results for all-cause mortality were related to PD or non-PD causes (e.g. general immunity, cardiovascular disease), we performed a competing risk analysis for PD death and non-PD death. In the UKB and QSBB cohorts which collected cause of death data, we classified the primary cause of death as either 1) related to PD and end of life (e.g. pneumonia, aspiration pneumonia, bronchopneumonia), or 2) non-PD, or 'interrupted', likely unrelated to PD. Non-PD death causes included: cardiac arrest/ heart failure/ myocardial infarct/ heart disease, carcinoma, glioblastoma, gastric intestinal bleed or perforation, head injury, pulmonary embolism, sudden death, or other accidental death causes (Supplementary Materials).

754/1772 patients who had died had cause of death and time to death data available. The median survival time was 6.1 years in the patients who died. Of these, 258 (34.2%) had a cause of death related to PD.

1,950 individuals were included for the competing risk analysis: patients who were still surviving (N=1,196), PD deaths (N=258) and non-PD deaths (N=496). The median time to death was 11.7 years in the PD death group and 4.0 years in the non-PD death

group. For the top *APOE* SNP rs429358, increased mortality appeared to be driven by PD deaths (HR=1.47, p-value=2.6×10^{-3}) (Figure 2A, Table 3). However for the *TBXAS1* SNPs, rs4726467 and rs144889025, there was no significant effect of the SNPs on PD deaths (p-values>0.05) (Figure 2B, 2C). These two *TBXAS1* SNPs were nominally associated with increased non-PD deaths (HR=1.63, p-value=0.01, and HR=1.54, p-value=0.03 respectively). The *SYT10* SNP, rs10437796, was nominally protective against PD death (HR=0.71, p=0.05) and increased non-PD deaths (HR=1.50, p=1.8×10^{-4}) (Figure 2D). This suggests that *APOE* ε4 is having a specific effect on PD mortality but the other SNPs may not be PD specific.

Figure 2. Competing risk cumulative incidence function plots for the top 4 mortality SNPs.

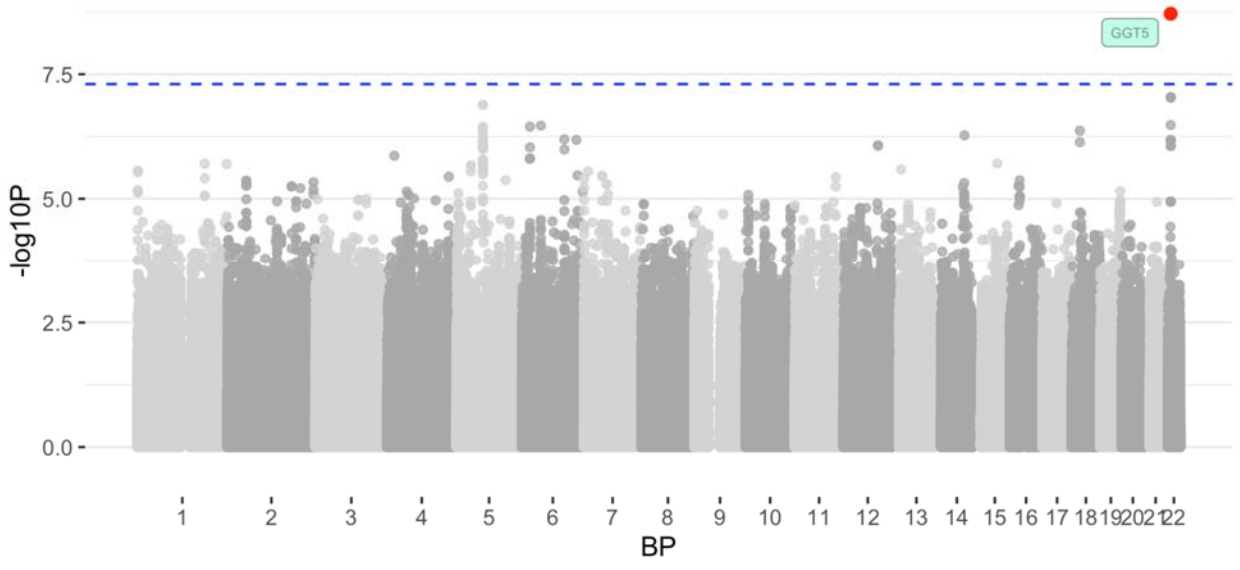


GWAS of H&Y stage 3 or greater

3,299 individuals were analysed for progression to H&Y stage 3 or greater. The Oslo cohort was excluded as the genomic inflation factor was greater than 1.2. After exclusion of the Oslo dataset, 872 individuals (26.4%) met the outcome of H&Y stage 3 or greater, with a median time to H&Y 3 of 5.5 years. 6,553,504 SNPs passed filtering for heterogeneity and MAF variability. The genomic inflation factor of the meta-analysis was 1.01. The top ten independent SNPs from FUMA are reported in Table 4. One locus in Chromosome 22 was significantly associated with progression to H&Y stage 3 or greater, with the lead SNP rs112809886 ($p=1.9x10^{-9}$), close to *GGT5* (Figure 3). The regional association plot is shown in Supplementary Figure 4.

The top SNP, rs112809886, is an eQTL for *ADORA2A,* with the alternate allele increasing gene expression in the tibial nerve and cerebellar hemisphere of the brain. It also decreases gene expression of *UPB1* and *ADORA2A* in whole blood (GTEx). In the eQTLGen data for blood eQTLs, it is a significant cis-eQTL for *UPB1, SUSD2, GSTT1,* and AP000351.10. Brain QTL data from MetaBrain showed that the effect allele A of rs112809886 increased expression of *ADORA2A* in the but not any other brain regions.

Figure 3. GWAS meta-analysis of progression to H&Y stage 3+.

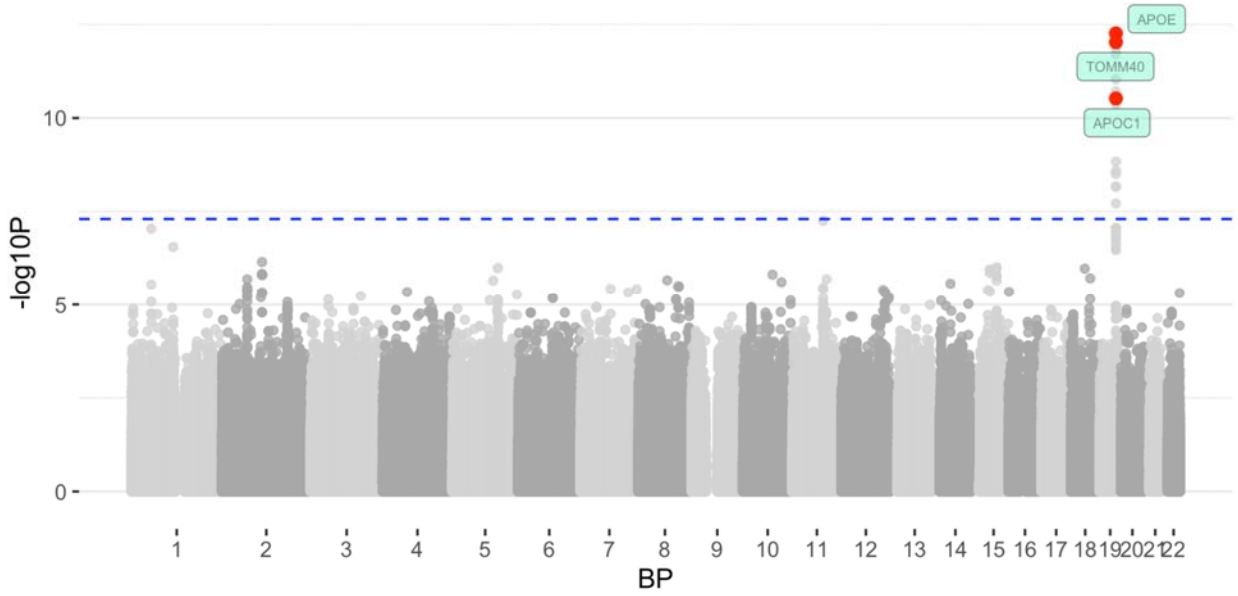


GWAS of cognitive impairment

3,565 individuals were analysed for cognitive impairment. Of those, 1,081 (30.3%) met the outcome of cognitive impairment with a median time of 5.0 years. 6,713,896 SNPs passed filtering for heterogeneity and MAF variability. The genomic inflation factor was 1.02. The top ten independent SNPs from FUMA are reported in Table 5.

The top SNP was rs429358 in Chromosome 19, which is the same *APOE* ε4 tagging SNP as identified in the mortality GWAS (Figure 4). The minor allele C was associated with more rapid progression to cognitive impairment (HR=1.54, $p=5.5x10^{-13}$).

Figure 4. GWAS meta-analysis of progression to cognitive impairment.



Candidate variant analysis

We did not find that any of the 90 PD risk SNPs were associated with PD progression at genome-wide significance (Supplementary Materials; Supplementary Table 4). Only one variant, rs35749011, near *KRTCAP2* but tagging the *GBA* p.E326K variant, was associated with mortality ($p=3.6x10^{-4}$) below the analysis-wide significance threshold (p-value threshold 0.05/88=0.00057) and nominally associated with cognitive impairment (p=0.02). There was also no association between the PD GRS and any of the

progression outcomes (Supplementary Materials). In the candidate variant analysis, only the PSP survival SNP rs2242367 was associated with more rapid progression to mortality in PD (HR=1.13 [95% CI 1.04 to 1.21], p=0.002) (Supplementary Table 5).

AD Genetic Risk Score (GRS)

In the random-effects meta-analysis, the AD GRS without *APOE* was nominally associated with mortality (HR=1.06 [95% CI 1.01 to 1.11], p=0.03) but not cognitive impairment (HR=1.05 [95% CI 0.99 to 1.12], p=0.09).

Power calculations

The power to detect a signal in a survival GWAS depends on a number of factors, including effect size, allele frequency of the effect allele, and the proportion of individuals meeting the outcome of interest. Using the 'survSNP' package[13], we estimate that this study had 92% power to detect a significant effect ($p<5x10^{-8}$) for our top *APOE* SNP rs429358 in the mortality GWAS, given an allele frequency of 16%, Hazard Ratio of 1.34, event/death rate of 32.1% and median time to death of 10.6 years. Supplementary Figure 7 shows how power changes with different event rates and allele frequencies. Clearly power for progression studies will increase with longer follow-up as more individuals meet the outcomes.

DISCUSSION

We have conducted a large meta-analysis GWAS of progression to clinical milestones in PD. We have identified the *APOE, TBXAS1, SYT10*, and *ADORA2A* loci as relevant to survival, motor, and cognitive progression in PD.

***APOE**:* The *APOE* SNP rs429358 was associated with both mortality and cognitive impairment. *APOE* is the strongest genetic risk factor for AD,[10,14] and is also associated with cardiovascular disease (including coronary heart disease / coronary artery disease), and cholesterol levels.[15]

In PD, *APOE* has been associated with age at onset,[16] cognition and dementia, and potentially motor progression,[17] but not PD risk.[1,18] In the GWAS of PD age at onset, the effect of *APOE* was shown to be similar between age at onset in cases and age of entry of controls.[16] This suggests that the effect of *APOE* on PD age at onset is more generally related to ageing, and not specific to PD age at onset. Indeed, GWASs of longevity and survival in the general population have identified *APOE* as the strongest

genetic factor, with the same ε4 (rs429358) allele associated with increased mortality[19] and found less frequently in long-living individuals.[20] However, our analysis of competing cause of death showed that the *APOE* SNP significantly increased PD-related deaths and not non-PD deaths. This suggests that although *APOE* influences survival in general, it may also have a specific effect in PD. Another possibility is that *APOE* increases risk of frailty, both physical and cognitive, leading to death but this is attributed to PD as it would not be possible to disentangle these two influences on mortality.

We also analysed AD GRSs excluding *APOE.* This showed that non-*APOE* AD genetic risk influences mortality but not cognitive impairment in PD, providing some support for a specific non-AD pathology related role of ApoE in PD cognitive impairment. There is some preclinical support for this in that transgenic mice carrying the ApoE e4 allele have increased alpha-synuclein pathology both in a double transgenic (ApoEe4 x SNCA A53T) and a striatal preformed fibril injection model.[21]

TBXAS1: *TBXAS1* encodes Thromboxane A Synthase 1, which catalyses the conversion of prostaglandin H2 to thromboxane A2, which acts as a platelet aggregator and vasoconstrictor. Variants in *TBXAS1* have previously been associated with coronary artery disease in a number of GWASs.[22–25] We showed that a locus in *TBXAS1* was associated with all-cause mortality in PD cohorts, but our competing risk analysis in a subset of cases suggests this may be driven by non-PD deaths rather than being specific to PD progression. Thus, it is possible that the association between *TBXAS1* and PD mortality may be due to increased coronary artery disease or other non-PD causes.

However, there are a few important factors to consider. Firstly, the locus we identified for mortality in PD cohorts is different from those associated with coronary artery disease, so it may be that different SNPs within this gene have different effects. Secondly, the competing risk analysis was performed in a limited sample that had cause of death data available, and UKB and QSBB patients may not be representative of the wider PD population. There was a more rapid time to death in these cohorts compared to our other observational cohorts. This may be due to recruitment of more clinically atypical PD cases in the QSBB,[26,27] and late 'detection' of PD in the UKB through hospital records (patients may experience onset of PD symptoms several years beforehand but are only identified with PD in the UKB when they present in hospital). Thus further analyses of cause of death in larger and more representative cohorts are needed. Thirdly, it is also possible that patients with PD are more susceptible to adverse effects of a SNP than healthy controls, so PD increases vulnerability to the SNP effect on coronary artery disease but the cause of death is not able to be classified as 'PD-

related'. This has also been seen in previous non-genetic studies of PD mortality.[28] Finally, it is also important to consider that PD individuals are less likely to be smokers and hence any effect of the SNP on mortality through coronary artery disease may be less pronounced than in smokers. Thus we need to further clarify the mechanism through which *TBXAS1* variants influence mortality in PD patients, and whether this is specific to PD or more general.

Our search of the most recent longevity GWASs[29] suggests that variation at *TBXAS1* does not influence mortality and longevity in the general population (Supplementary Materials).

SYT10: We also found evidence that a locus near *SYT10* (synaptotagmin 10) in Chromosome 12 was associated with mortality. This protein is a calcium sensor involved in regulation of calcium-dependent exocytosis, specifically related to peptidergic vesicles.[30] Variants in this gene have been linked to heart rate and cardiac cycle phenotypes.[31–33] Our competing risk analysis suggested that *SYT10* largely influenced non-PD deaths, and appeared to be nominally protective against PD deaths. However, this locus does not have a significant effect on longevity in the general population, although there was nominal association in the most recent longevity GWAS,[29] (beta=-0.008, p=0.003) (Supplementary Materials). This suggests that the effect of the SNP is not specific to PD mortality, however our competing risk analysis is underpowered as discussed above. Further cause of death analysis in both PD patients and healthy controls are needed to clarify the effect of this locus.

GGT5 / UPB1 / ADORA2A: A locus near *GGT5* (gamma-glutamyltransferase 5) was associated with progression to H&Y stage 3 or greater. The gene encodes a protein that cleaves the gamma-glutamyl peptide bond in the process of metabolising gamma-glutamyl compounds such as antioxidants, and inflammatory molecules.[34] The top SNP rs112809886 is an eQTL for *ADORA2A* in the brain and tibial nerve, and this variant increases the expression of *ADORA2A* in human cerebellum. *ADORA2A* encodes the adenosine A2A receptor, which is highly expressed in the basal ganglia in enkephalin expressing GABA-ergic striatal-pallidal neurons, so there is preliminary evidence that increased adenosine neurotransmission may increase motor progression in PD. The A2A receptor antagonist istradefylline has been licensed as a treatment of PD in Japan since 2013 and in the US since 2019, and has been shown to decrease daily OFF time.[35,36] This is a promising locus but further work is required to test for replication, as this SNP showed some heterogeneity of effects across cohorts.

PD risk variants and candidate variants: In line with previous PD progression GWASs, the majority of the 90 PD risk SNPs were not associated with PD

progression.[2–4] We showed that one variant, rs35749011, linked to *GBA* p.E326K (also known as p.E365K) was associated with mortality and cognitive impairment. Interestingly this variant does not cause Gaucher disease or have a major effect on glucosylceramide levels suggesting a dissociation between glucosylceramide and the role of *GBA* in PD progression. This is consistent with the recently reported trial data reporting a lack of effect of the glucosylceramide inhibitor venglustat in modifying PD progression.

We were not able to replicate findings for other candidate variants nominated from previous PD progression GWASs. We examined results for rs382940 in *SLC44A1* for progression to H&Y stage 3 or greater,[2] however this was not associated with progression in any of our results.

We also did not replicate findings for variants in *RIMS2*, *WWOX,* and *TMEM108* which have been reported for PD dementia.[4] The p-values for these variants were all > 0.3 in our GWASs (Supplementary Table 4). This may be due to different criteria for defining the cognition outcomes, as the Liu study used a more stringent MDS criteria to define PD dementia whereas we have used more liberal criteria thus capturing cognitive impairment less severe than dementia. Further work is needed to replicate nominated variants both from the current study and previous studies. We did not find evidence to support *APOE* ε2 and *MAPT* H1 haplotype as factors for mortality or cognitive progression. We found some evidence suggesting the PSP mortality SNP, rs2242367,[37] was also associated with more rapid mortality in PD. This SNP was shown to increase expression of *LRRK2*, although there was no colocalisation with *LRRK2*.[37] This finding could indicate that there is some 'contamination' of PSP cases in our PD cohorts, as PSP can be frequently misdiagnosed as PD and we did not have pathological diagnosis data on the majority of cases.

Limitations: This study has some limitations. This study is one of the largest GWASs of PD progression and the first large-scale GWAS of PD mortality. However larger sample sizes and longer follow-up are needed to detect variants with smaller effects or lower allele frequencies. Secondly, more data is needed on postmortem pathological diagnosis to improve power for cause of death analyses. Another limitation is the heterogeneity between cohorts and PD case selection. Our cohorts tend to be recruited from specialist clinics and groups of patients, and this may lead to a tendency to recruit more atypical patients, or rapidly progressing patients. More population-based studies are needed to improve generalisability of these results. Several of our cohorts are also non-incident, with a delay between symptom onset and study entry, and this mean that we are not able to capture the most rapidly progressing patients.

In addition, we nominated genes from the top SNPs based on physical proximity and eQTL databases, however additional fine-mapping and annotation is needed to prioritise causal variants and genes for each locus. Finally, the interpretation of GWAS for neurological disease remains limited by the resolution of the effects of genomic variants on gene expression from bulk RNA sequencing studies. Rapidly increasing sample sizes, and the development of single cell resources will enable a more direct interpretation of the relationship between genomic variants and disease biology.

Conclusions: We conducted three large-scale GWASs of PD progression, including the first GWAS of mortality in PD. We identified three genome-wide significant signals, including *TBXAS1*. We also showed that the genetic factors influencing progression in PD are largely different to those influencing PD risk, emphasising the need for further studies of progression. This work will help us to better understand the biology of PD progression and develop new disease-modifying treatments.

Data availability

GWAS summary statistics are available for download at: https://tinyurl.com/PDprogressionSumstats.

All code for our analyses are publicly available at https://github.com/huw-morris-lab/PD-survival-GWAS.

Tracking Parkinson's data is available through the Tracking Parkinson's portal: https://www.trackingparkinsons.org.uk/about-1/data/. The Oxford Parkinson's Disease Centre Discovery Cohort data (https://doi.org/10.1016/j.parkreldis.2013.09.025) was provided on request (Michele Hu, michele.hu@ndcn.ox.ac.uk). The Cambridgeshire Parkinson's Incidence from GP to Neurologist (CamPaIGN) (https://www.thebarkerwilliamsgraylab.co.uk/parkinsons-disease/current-studies-pd/) data and the Cambridge clinic data were provided on request (Caroline Williams-Gray/ Roger Barker; chm27@cam.ac.uk). Parkinson's Progression Markers Initiative (PPMI) data was accessed from the PPMI platform: https://www.ppmi-info.org/access-data-specimens/download-data. Queen Square Brain Bank for Neurological Disorders (QSBB) data was provided upon request (qsbbmtas@ucl.ac.uk). UK Biobank data were accessed through the UK Biobank: https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access. Drug Interaction With Genes in Parkinson's Disease (DIGPD) data (https://clinicaltrials.gov/ct2/show/NCT01564992) was provided upon request (Jean-Christophe Corvol, jean-christophe.corvol@aphp.fr). Calypso data was provided on request to the study team (Huw Morris, h.morris@ucl.ac.uk). The

Trondheim Parkinson's Disease Study (Trondheim) data was provided on request (https://doi.org/10.14802/jmd.21029). The Oslo Parkinson's Disease data was provided on request (Lasse Pihlstrom/ Mathias Toft, lasse.pihlstrom@medisin.uio.no).

Acknowledgements

This research was funded in whole or in part by Aligning Science Across Parkinson's [Grant number: ASAP-000478] through the Michael J. Fox Foundation for Parkinson's Research (MJFF). For the purpose of open access, the author has applied a CC BY public copyright license to all Author Accepted Manuscripts arising from this submission. This research has been conducted using the UK Biobank Resource under Application Number 46450.

Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database (https://www.ppmi-info.org/accessdata-specimens/download-data). 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, including [list the full names of all of the PPMI funding partners found at https://www.ppmi-info.org/about-ppmi/who-we-are/studysponsors].

Declaration of interests

MMXT is employed by Oslo University Hospital. She has received grant support from Parkinson's UK, the Michael J Fox Foundation, and South-Eastern Norway Regional Health Authority (Helse Sør-Øst). MAL received fees for advising on a secondary analysis of an RCT sponsored by North Bristol NHS trust. KG is an independent medical consultant and receives payments from the University of Glasgow. She has no other fees or honoraria to report.J-CC has served on advisory boards for Biogen, Denali, Idorsia, Prevail Therapeutic, Servier, Theranexus, UCB ; and received grants from Sanofi and the Michael J Fox Foundation outside of this work. MAN's participation in this project was part of a competitive contract awarded to Data Tecnica International LLC by the National Institutes of Health to support open science research, he also currently serves on the scientific advisory board for Clover Therapeutics and is an advisor to Neuron23 Inc as a data science fellow. OAA is a consultant to HealthLytix. CWG is employed by the University of Cambridge and holds a RCUK/UKRI Research Innovation Fellowship awarded by the Medical Research Council (MR/R007446/1). In the last 36 months, she has received research funding from the Cambridge Centre for Parkinson-Plus, the NIHR Cambridge Biomedical Research Centre, Cure Parkinson's, Parkinson's UK, The Evelyn Trust, and the Michael J Fox Foundation; speaker payments from GSK and World Parkinson's Coalition; and consulting fees from Evidera

Inc. RAB has served as an advisor to UCB; BlueRock Therapeutics; Novo Nordisk; Aspen Neuroscience, UCB and Transine Therapeutics. He has received lecture fees from Novo Nordisk. He has received grant support from the MRC, Wellcome, Parkinson's UK, Cure Parkinson's Trust, EU, Novo Nordisk, DRI and ASAP. 'This research was supported by the NIHR Cambridge Biomedical Research Centre (BRC-1215-20014). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care'. He has received royalties from Wiley and Springer-Nature. DGG is an employee of the University of Glasgow. In the past 12 months he reports consultancy fees from the Glasgow Memory Clinic; honoraria for chairing or attending meetings from AbbVie and BIAL Pharma. HRM is employed by UCL. In the last 12 months he reports paid consultancy from Roche and Amylyx ; lecture fees/honoraria - BMJ, Kyowa Kirin, Movement Disorders Society. Research Grants from Parkinson's UK, Cure Parkinson's Trust, PSP Association, Medical Research Council, Michael J Fox Foundation. HRM is a co-applicant on a patent application related to C9ORF72 - Method for diagnosing a neurodegenerative disease (PCT/GB2012/052140).

Ethics

Tracking Parkinson's: West of Scotland Research Ethics Service (WoSRES) Research Ethics Committee gave ethical approval for this study (ref 11/AL/0163). Oxford Discovery: NRES Committee, South Central Oxford A Research Ethics Committee gave ethical approval for this study (ref 16/SC/0108). CamPaIGN: Cambridge Research Ethics Committee gave ethical approval for this study. Cambridge PD Research Clinic: Cambridge Research Ethics Committee gave ethical approval for this study. PPMI: The Research Subjects Review Board at the University of Rochester approved the PPMI study protocol. UK Biobank: UK Biobank has approval from the North West Multi-centre Research Ethics Committee (MREC) as a Research Tissue Bank (RTB). QSBB: London Central Research Ethics Committee gave ethical approval for this research tissue bank (ref 18/LO/0721). DIGPD: the Ethical committee Ile-De-France VI gave ethical approval for this study (ID project: 2009-A00109-48). Calypso: Wales Research Ethics Committee 3 gave ethical approval for this study. Trondheim: the Ethics Committee of Central Norway gave ethical approval for this study (ref 2011/1137). Oslo: the Regional Committee for Medical Research Ethics in South-East Norway gave ethical approval for this study.

References

1 Nalls MA, Blauwendraat C, Vallerga CL, *et al.* Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol* 2019; **18**: 1091–102.

2 Iwaki H, Blauwendraat C, Leonard HL, *et al.* Genomewide association study of Parkinson's disease clinical biomarkers in 12 longitudinal patients' cohorts. *Mov Disord* 2019; : 1–12.

3 Tan MMX, Lawton MA, Jabbari E, *et al.* Genome-Wide Association Studies of Cognitive and Motor Progression in Parkinson's Disease. *Mov Disord* 2020. DOI:10.1002/mds.28342.

4 Liu G, Peng J, Liao Z, *et al.* Genome-wide survival study identifies a novel synaptic locus and polygenic score for cognitive progression in Parkinson ' s disease. *Nat Genet* 2021. DOI:10.1038/s41588-021-00847-6.

5 Fagan ES, Pihlstrøm L. Genetic risk factors for cognitive decline in Parkinson's disease: a review of the literature. *Eur J Neurol* 2017; **24**: 561-e20.

6 Malek N, Weil RS, Bresner C, *et al.* Features of GBA -associated Parkinson's disease at presentation in the UK Tracking Parkinson's study. *J Neurol Neurosurg Psychiatry* 2018; **89**: 702–9.

7 Dalrymple-Alford JC, Macaskill MR, Nakas CT, *et al.* The MoCA: Well-suited screen for cognitive impairment in Parkinson disease. *Neurology* 2010; **75**: 1717–25.

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

9 Watanabe K, Taskesen E, Van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun* 2017; **8**: 1–10.

10 Wightman DP, Jansen IE, Savage JE, *et al.* A genome-wide association study with 1,126,563 individuals identifies new risk loci for Alzheimer's disease. *Nat Genet* 2021; **53**: 1276–82.

11 Võsa U, Claringbould A, Westra HJ, *et al.* Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet* 2021; **53**: 1300–10.

12 Klein N de, Tsai EA, Vochteloo M, *et al.* Brain expression quantitative trait locus and network analysis reveals downstream effects and putative drivers for brain-related diseases. *bioRxiv* 2021; : 2021.03.01.433439.

13 Owzar K, Li Z, Cox N, Jung SH. Power and Sample Size Calculations for SNP Association Studies With Censored Time-to-Event Outcomes. *Genet Epidemiol* 2012; **36**: 538–48.

14 Schwartzentruber J, Cooper S, Liu JZ, *et al.* Genome-wide meta-analysis, fine-mapping and integrative prioritization implicate new Alzheimer's disease risk

genes. *Nat Genet* 2021; **53**: 392–402.

15 Lahoz C, Schaefer EJ, Cupples LA, *et al.* Apolipoprotein E genotype and cardiovascular disease in the Framingham Heart Study. *Atherosclerosis* 2001; **154**: 529–37.

16 Blauwendraat C, Heilbron K, Vallerga CL, *et al.* Parkinson's disease age at onset genome-wide association study: Defining heritability, genetic loci, and α-synuclein mechanisms. *Mov Disord* 2019; **34**: 866–75.

17 Pu JL, Jin CY, Wang ZX, *et al.* Apolipoprotein E Genotype Contributes to Motor Progression in Parkinson's Disease. *Mov Disord* 2021; : 1–6.

18 Williams-Gray CH, Goris A, Saiki M, *et al.* Apolipoprotein e genotype as a risk factor for susceptibility to and dementia in Parkinson's Disease. *J Neurol* 2009; **256**: 493–8.

19 Rosvall L, Rizzuto D, Wang HX, Winblad B, Graff C, Fratiglioni L. APOE-related mortality: Effect of dementia, cardiovascular disease and gender. *Neurobiol Aging* 2009; **30**: 1545–51.

20 Deelen J, Evans DS, Arking DE, *et al.* A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat Commun* 2019; **10**. DOI:10.1038/s41467-019-11558-2.

21 Davis AA, Inman CE, Wargel ZM, *et al.* APOE genotype regulates pathology and disease progression in synucleinopathy. *Sci Transl Med* 2020; **12**: 1–15.

22 Van Der Harst P, Verweij N. Identification of 64 novel genetic loci provides an expanded view on the genetic architecture of coronary artery disease. *Circ Res* 2018; **122**: 433–43.

23 Matsunaga H, Ito K, Akiyama M, *et al.* Transethnic Meta-Analysis of Genome-Wide Association Studies Identifies Three New Loci and Characterizes Population-Specific Differences for Coronary Artery Disease. *Circ Genomic Precis Med* 2020; : 128–38.

24 Koyama S, Ito K, Terao C, *et al.* Population-specific and trans-ancestry genome-wide analyses identify distinct and shared genetic risk loci for coronary artery disease. *Nat Genet* 2020; **52**: 1169–77.

25 Shadrina AS, Shashkova TI, Torgasheva AA, *et al.* Prioritization of causal genes for coronary artery disease based on cumulative evidence from experimental and in silico studies. *Sci Rep* 2020; **10**: 1–15.

26 Parkkinen L, Soininen H, Laakso M, Alafuzoff I. Alpha-Synuclein Pathology Is Highly Dependent on the Case Selection. *Neuropathol Appl Neurobiol* 2001; **27**: 314–25.

27 Ince PG, McArthur FK, Bjertness E, Torvik A, Candy JM, Edwardson JA. Neuropathological diagnoses in elderly patients in Oslo: Alzheimer's disease, Lewy body disease, vascular lesions. *Dementia* 1995; **6**: 162–8.

28 Ben-Shlomo Y, Marmot MG. Survival and cause of death in a cohort of patients

with parkinsonism: Possible clues to aetiology? *J Neurol Neurosurg Psychiatry* 1995; **58**: 293–9.

29 Timmers PRHJ, Wilson JF, Joshi PK, Deelen J. Multivariate genomic scan implicates novel loci and haem metabolism in human ageing. *Nat Commun* 2020; **11**: 1–10.

30 Cao P, Maximov A, Südhof TC. Activity-dependent IGF-1 exocytosis is controlled by the Ca 2+-sensor synaptotagmin-10. *Cell* 2011; **145**: 300–11.

31 Verweij N, Van De Vegte YJ, Van Der Harst P. Genetic study links components of the autonomous nervous system to heart-rate profile during exercise. *Nat Commun* 2018; **9**: 1–9.

32 Verweij N, Benjamins JW, Morley MP, *et al.* The Genetic Makeup of the Electrocardiogram. *Cell Syst* 2020; **11**: 229-238.e5.

33 Nolte IM, Munoz ML, Tragante V, *et al.* Genetic loci associated with heart rate variability and their effects on cardiac disease risk. *Nat Commun* 2017; **8**. DOI:10.1038/ncomms15805.

34 Wickham S, West MB, Cook PF, Hanigan MH. Gamma-glutamyl compounds: Substrate specificity of gamma-glutamyl transpeptidase enzymes. *Anal Biochem* 2011; **414**: 208–14.

35 Torti M, Vacca L, Stocchi F. Istradefylline for the treatment of Parkinson's disease: is it a promising strategy? *Expert Opin Pharmacother* 2018; **19**: 1821–8.

36 Chen JF, Cunha RA. The belated US FDA approval of the adenosine A2A receptor antagonist istradefylline for treatment of Parkinson's disease. *Purinergic Signal* 2020; **16**: 167–74.

37 Jabbari E, Koga S, Valentino RR, *et al.* Genetic determinants of survival in progressive supranuclear palsy: a genome-wide association study. *Lancet Neurol* 2021; **20**: 107–16.

Table 1. Cohort demographics.

Cohort	N PD patients	N male (%)	Age onset, years	Age diagnosis, years	Age entry, years	Disease duration at baseline (time from symptom onset to baseline), years	N observations	Follow-up, years	Array
Tracking Parkinson's	1782	1154 (64·8%)	64·4 (9·8)	66·2 (9·3)	67·5 (9·3)	3·2 (3·0)	7629	4·2 (1·7)	Illumina HumanCore Exome array with custom content (over 27,000 custom variants implicated in neurological and psychiatric disorders)
Oxford Discovery	842	544 (64·6%)	64·5 (9·6)	66·2 (9·5)	67·4 (9·4)	2·9 (1·8)	3181	4·7 (2·8)	Illumina HumanCore Exome-12 v1.1 or Illumina InfiniumCoreExome-24 v1.1
PPMI	403	263 (65·3%)	59·5 (10·0)	61·0 (9·7)	61·5 (9·7)	2·0 (1·9)	2200	6·0 (1·8)	Whole genome sequencing (see https://www.ppmi-info.org/). Only variants passing quality filters (PASS)
QSBB	308	187 (60·7%)	61·8 (10·1)	NA	NA	NA	NA	15·7 (7·7)	Illumina NeuroChip (Illumina Infinium HumanCore-24 array with approximately 180,000 custom variants implicated in neurological diseases
UKB - incident	1174	728 (62·0%)	NA	69·6 (5·5)	64·1 (5·2)	NA	NA	10·1 (2·1)	Affymetrix Applied Biosystems BiLEVE Axium Array and UK Biobank Axium Array
UKB - prevalent	839	526 (62·7%)	NA	57·4 (7·1)	62·8 (5·5)	5·5 (4·8)	NA	9·8 (2·5)	Affymetrix Applied Biosystems BiLEVE Axium Array and UK Biobank Axium Array
Calypso	186	125 (67·2%)	60·0 (10·0)	61·6 (9·8)	67·7 (9·5)	7·7 (5·3)	NA	9·3 (3·9)	Illumina BeadArray Human660-Quad
CamPaIGN	90	49 (54·4%)	67·9 (10·4)	70·0 (9·7)	70·3 (9·7)	2·5 (2·9)	NA	8·2 (4·3)	Illumina BeadArray Human660-Quad
Cambridge PD Research clinic	239	152 (63·6%)	58·6 (12·0)	60·4 (11·4)	65·1 (10·5)	6·8 (6·1)	NA	9·0 (5·0)	Illumina BeadArray Human660-Quad
DIGPD	376	227 (60·4%)	58·7 (10·0)	59·6 (9·9)	62·2 (9·9)	3·5 (1·6)	1940	9·0 (2·1)	Illumina Multi-Ethnic Genotyping Array (MEGA) array
Trondheim	192	109 (56·8%)	62·9 (9·7)	NA	NA	NA	390	15·7 (7·1)	Illumina NeuroChip (Illumina Infinium HumanCore-24 array with approximately 180,000 custom variants implicated in neurological diseases)
Oslo	335	226 (67·5%)	52·2 (10·2)	54·2 (10·2)	64·0 (9·2)	11·8 (6·3)	NA	9·2 (3·6)	Illumina Infinium OmniExpress array

Means (SD) are shown unless otherwise indicated. Data shown are only in individuals who had both clinical and genetic data, after quality control filters have been applied within each cohort.

The number of individuals in each cohort with complete data for each outcome of interest (and covariates) may be less than the total number.

Follow-up time is calculated as the time from study entry to last visit, death, or last known status date (censoring), whichever is the latest. For QSBB and Trondheim, we did not have data on age at study entry so follow-up is calculated as the time from onset to death.

Abbreviations: CamPaIGN = Cambridgeshire Parkinson's Incidence from GP to Neurologist; DIGPD = Drug Interaction With Genes in Parkinson's Disease; HY3 = Hoehn and Yahr stage 3 or greater; PPMI = Parkinson's Progression Markers Initiative; QSBB = Queen Square Brain Bank pathologically-confirmed PD cases; UKB = UK Biobank.

Table 2. Top SNPs from meta-analysis of progression to mortality.

chr	bp	rsID	effect allele	non-effect allele	effect allele freq	nearest gene	distance to gene (BP)	beta	SE	Hazard Ratio	95% CI	p-value
19	45411941	rs429358	C	T	0·160	*APOE*	0	0·295	0·046	1·342	1·23-1·47	1·35E-10
7	139637422	rs4726467	T	C	0·021	*TBXAS1*	0	0·713	0·116	2·039	1·63-2·56	7·71E-10
7	139664899	rs144889025	T	C	0·020	*TBXAS1*	0	0·717	0·119	2·048	1·62-2·59	1·62E-09
12	33635494	rs10437796	A	C	0·098	*SYT10*	42740	0·304	0·056	1·355	1·21-1·51	5·31E-08
19	45390333	rs283815	G	A	0·218	*TOMM40*	4144	0·221	0·041	1·247	1·15-1·35	5·98E-08
2	217782469	rs151251359	G	A	0·012	*TNP1*	57687	1·249	0·233	3·487	2·21-5·51	8·34E-08
9	17616880	rs3808753	G	A	0·035	*SH3GL2*	0	0·448	0·088	1·564	1·32-1·86	3·34E-07
3	113979619	rs142285045	A	C	0·013	*ZNF80*	23194	0·836	0·165	2·308	1·67-3·19	3·80E-07
19	45394336	rs71352238	C	T	0·147	*TOMM40*	141	0·240	0·048	1·272	1·16-1·40	4·16E-07
18	33965217	rs76125680	C	G	0·012	*FHOD3*	0	0·874	0·174	2·397	1·70-3·37	5·32E-07

Independent lead SNPs identified by FUMA using standard settings (r2 threshold 0.6)

Genome coordinates are in build hg19/GRCh37

Abbreviations: BP = base pair, chr = chromosome, CI = Confidence Interval, freq = frequency, SE = Standard Error, SNP = Single Nucleotide Polymorphism

Table 3. Competing risk analysis for cause of death in the UKB and QSBB cohorts, using the Fine-Gray subdistribution hazard model.

				PD death (N = 258)		non-PD death (N = 496)	
chr	**bp**	**rsID**	**main mortality GWAS p-value**	**HR (95% CI)**	**p-value**	**HR (95% CI)**	**p-value**
19	45411941	rs429358	3·97E-10	1·47 (1·14 - 1·89)	2·60E-03	1·12 (0·92 - 1·36)	0·25
7	139637422	rs4726467	1·95E-09	1·28 (0·71 - 2·29)	0·41	1·63 (1·11 - 2·39)	0·012
7	139664899	rs144889025	4·13E-09	1·42 (0·80 - 2·53)	0·24	1·54 (1·03 - 2·30)	0·034
12	33635494	rs10437796	4·44E-08	0·71 (0·51 - 1·00)	0·05	1·50 (1·21 - 1·85)	1·80E-04

Each SNP was coded in a dominant model (0 vs. 1 or 2 allele carriers)

N = 1950 individuals included in competing risk analysis (N = 1196 surviving, N = 258 PD death, N = 496 non-PD death)

Abbreviations: BP = base pair, chr = chromosome, CI = Confidence Interval, freq = frequency, HR = Hazard Ratio, SNP = Single Nucleotide Polymorphism

Table 4. Top SNPs from meta-analysis of progression to Hoehn and Yahr stage 3 or greater.

chr	bp	rsID	effect allele	non-effect allele	effect allele freq	nearest gene	distance to gene (BP)	beta	SE	Hazard Ratio	95% CI	p-value
22	24641838	rs112809886	A	G	0·013	*GGT5*	521	1·575	0·262	4·833	2·89-8·08	1·94E-09
22	24900359	rs116812922	G	A	0·012	*UPB1*	0	1·476	0·276	4·376	2·55-7·52	9·26E-08
5	73225229	rs6870211	G	A	0·013	*ARHGEF28*	0	1·017	0·193	2·765	1·90-4·03	1·30E-07
6	52338210	rs139137718	C	T	0·021	*EFHC1*	0	0·735	0·144	2·085	1·57-2·76	3·41E-07
6	22270017	rs2066265	G	A	0·051	*PRL*	17456	0·491	0·096	1·633	1·35-1·97	3·58E-07
18	22226901	rs56204509	G	A	0·026	*HRH4*	166980	0·682	0·135	1·978	1·52-2·58	4·31E-07
14	84795014	rs189121749	A	G	0·012	*LOC100506718*	1220969	0·905	0·180	2·471	1·73-3·52	5·35E-07
6	117400944	rs79251781	G	A	0·024	*RFX6*	147618	0·671	0·135	1·955	1·50-2·55	6·42E-07
6	150585939	rs146959904	A	G	0·015	*PPP1R14C*	14411	0·777	0·156	2·174	1·60-2·95	6·57E-07
12	95077908	rs117824989	A	C	0·017	*TMCC3*	33584	0·770	0·156	2·159	1·59-2·93	8·55E-07

Independent lead SNPs identified by FUMA using standard settings (r2 threshold 0.6)

Genome coordinates are in build hg19/GRCh37

Abbreviations: BP = base pair, chr = chromosome, CI = Confidence Interval, freq = frequency, SE = Standard Error, SNP = Single Nucleotide Polymorphism

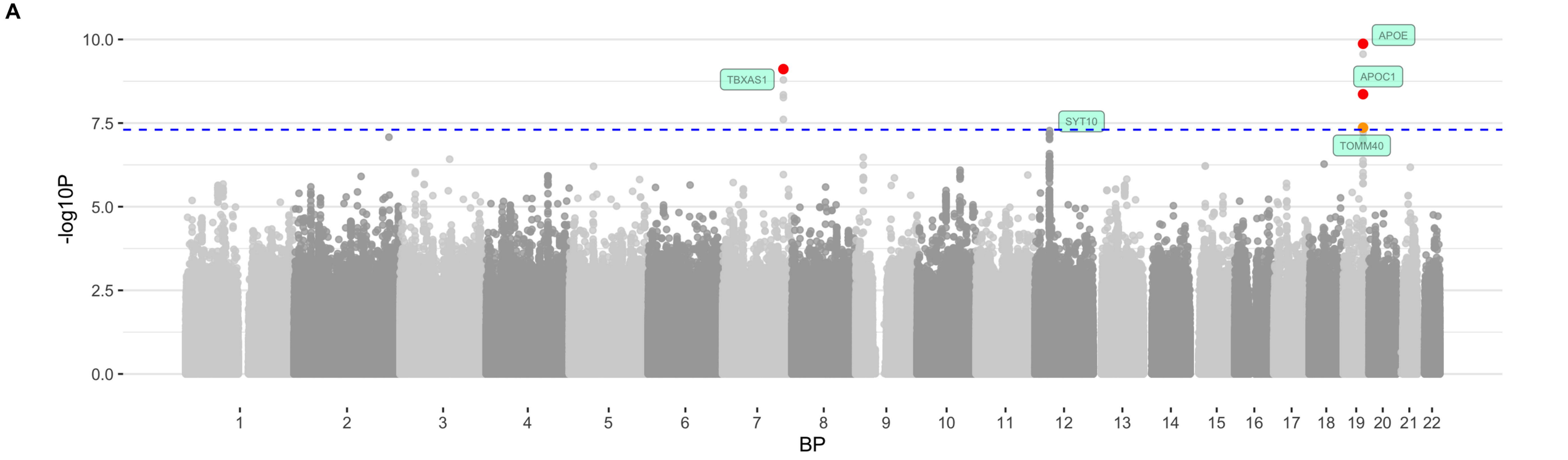
Table 5. Top SNPs from meta-analysis of progression to cognitive impairment.

chr	bp	rsID	effect allele	non-effect allele	effect allele freq	nearest gene	distance to gene (BP)	beta	SE	Hazard Ratio	95% CI	p-value
19	45411941	rs429358	C	T	0·149	*APOE*	0	0·432	0·060	1·541	1·37-1·73	5·45E-13
19	45390333	rs283815	G	A	0·210	*TOMM40*	4144	0·381	0·053	1·463	1·32-1·62	9·39E-13
11	84641700	rs118188129	T	C	0·090	*DLG2*	0	0·387	0·071	1·472	1·28-1·69	5·63E-08
19	45395844	rs34095326	A	G	0·106	*TOMM40*	0	0·373	0·070	1·452	1·27-1·66	8·67E-08
1	55020927	rs1084150	T	C	0·371	*ACOT11*	0	-0·263	0·049	0·769	0·70-0·85	9·09E-08
19	45428234	rs66626994	A	G	0·156	*APOC1*	5628	0·315	0·059	1·370	1·22-1·54	1·02E-07
1	115054084	rs35423939	T	C	0·041	*TRIM33*	303	0·528	0·103	1·695	1·39-2·07	2·79E-07
2	109790451	rs189033073	T	C	0·012	*EDAR*	184623	1·241	0·250	3·457	2·12-5·64	7·07E-07
15	69481630	rs78912421	A	G	0·018	*GLCE*	0	0·795	0·162	2·215	1·61-3·05	9·73E-07
5	125376369	rs13357370	A	C	0·029	*GRAMD3*	319419	0·571	0·117	1·769	1·41-2·22	1·03E-06

Independent lead SNPs identified by FUMA using standard settings (r2 threshold 0.6)

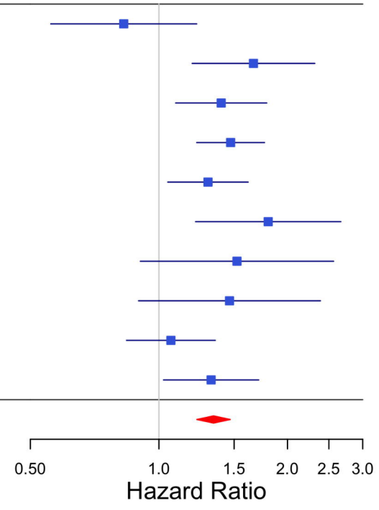
Genome coordinates are in build hg19/GRCh37

Abbreviations: BP = base pair, chr = chromosome, CI = Confidence Interval, freq = frequency, SE = Standard Error, SNP = Single Nucleotide Polymorphism



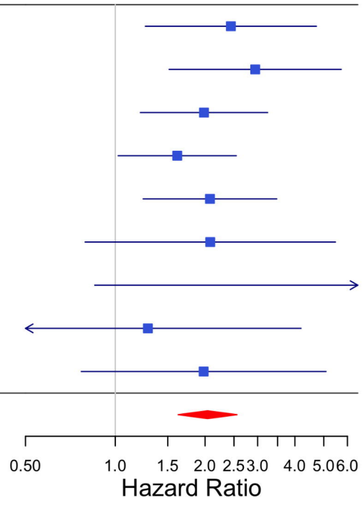
B **rs429358 (*APOE*)**

Study	HR	95% CI
Tracking Parkinson's	0.83	[0.56-1.23]
Oxford Discovery	1.66	[1.20-2.32]
QSBB	1.40	[1.09-1.79]
UKB_incident	1.47	[1.23-1.77]
UKB_prevalent	1.30	[1.05-1.62]
Calypso	1.80	[1.22-2.67]
CamPaIGN	1.52	[0.90-2.56]
DIGPD	1.46	[0.90-2.39]
Trondheim	1.07	[0.84-1.36]
Oslo	1.32	[1.02-1.71]
Meta-analysis	**1.34**	**[1.23-1.47]**



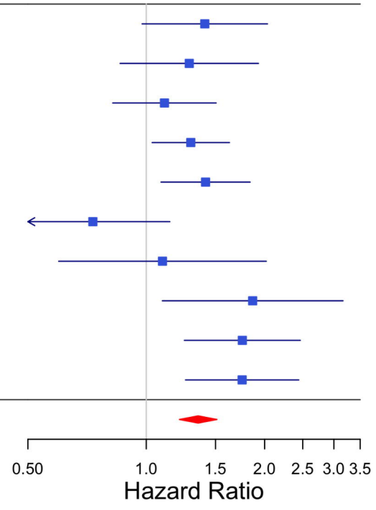
C **rs4726467 (*TBXAS1*)**

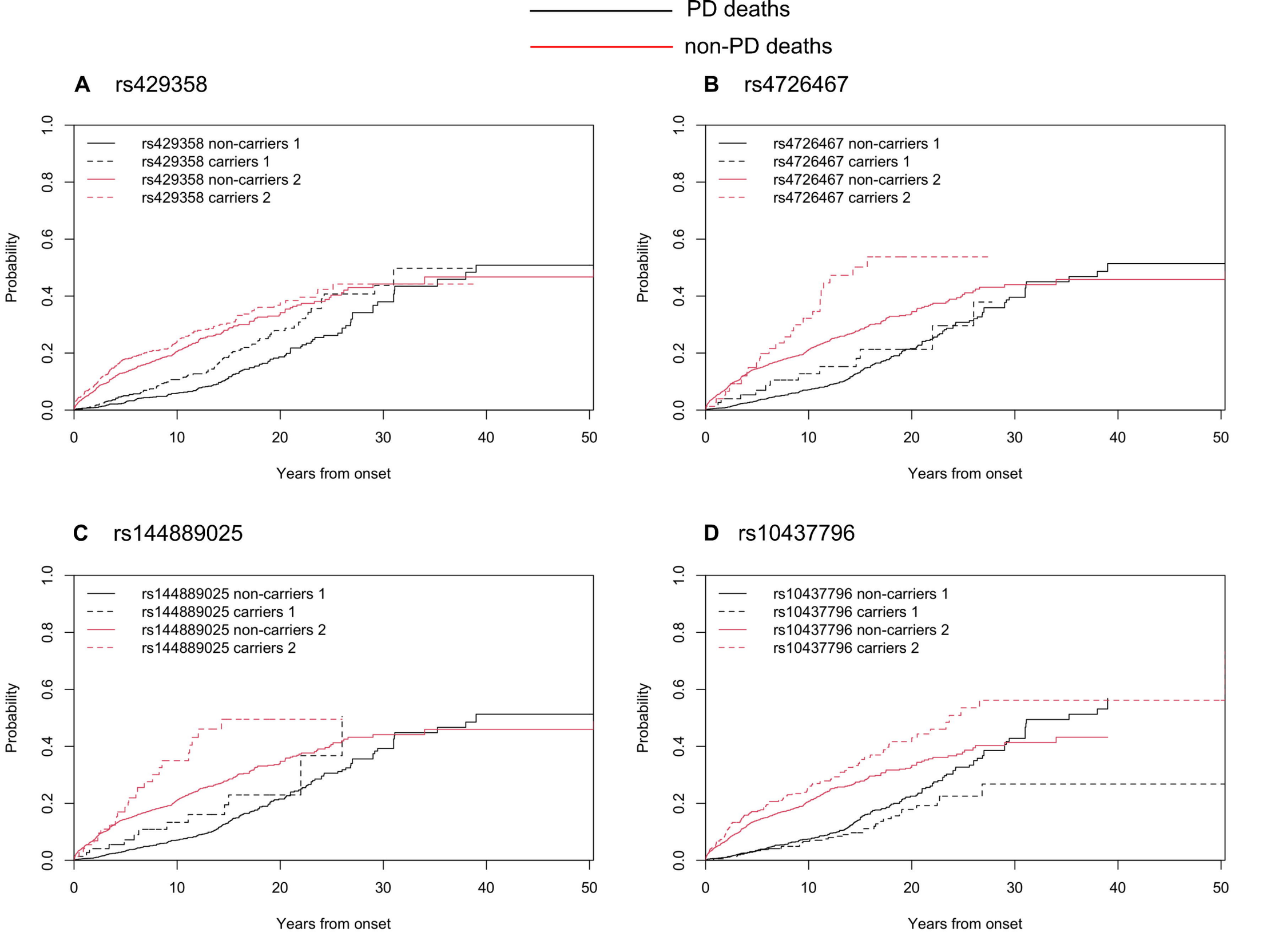
Study	HR	95% CI
Tracking Parkinson's	2.44	[1.26-4.72]
Oxford Discovery	2.94	[1.52-5.72]
QSBB	1.98	[1.21-3.24]
UKB_incident	1.61	[1.02-2.55]
UKB_prevalent	2.08	[1.24-3.48]
Calypso	2.08	[0.79-5.47]
CamPaIGN	6.91	[0.85-56.02]
DIGPD	1.29	[0.39-4.19]
Oslo	1.98	[0.77-5.09]
Meta-analysis	**2.04**	**[1.63-2.56]**

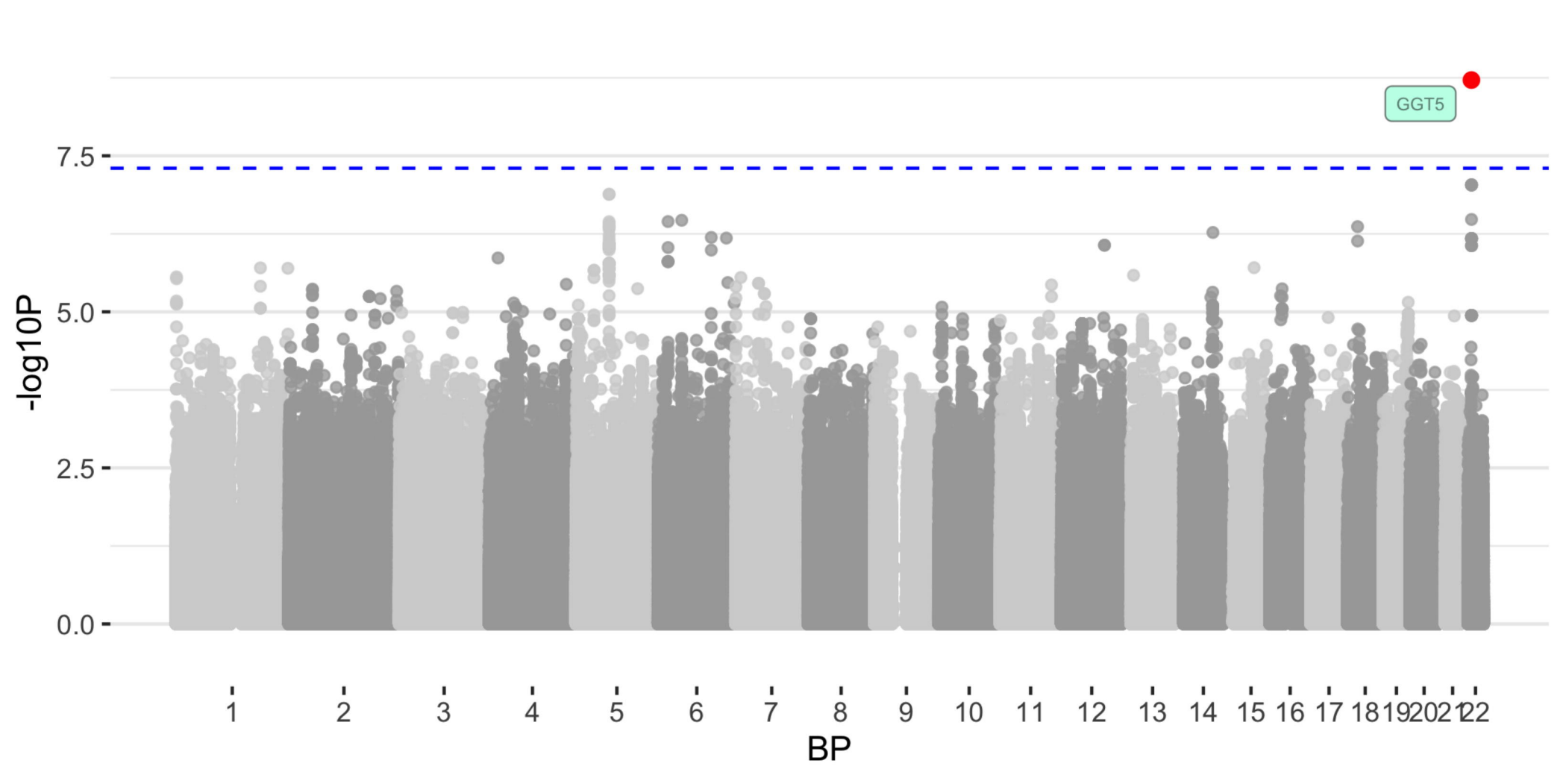


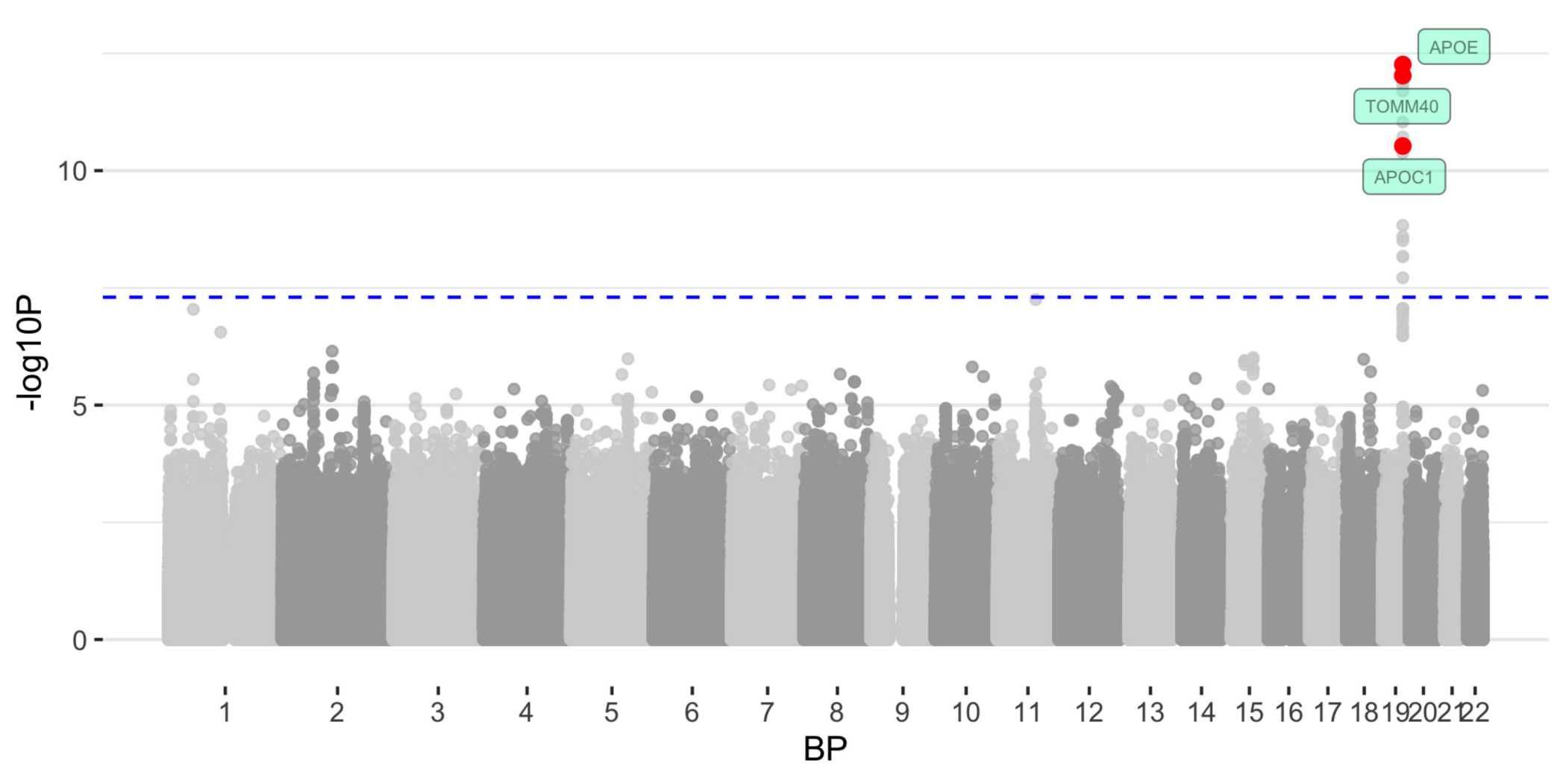
D **rs10437796 (*SYT10*)**

Study	HR	95% CI
Tracking Parkinson's	1.41	[0.98-2.03]
Oxford Discovery	1.29	[0.86-1.93]
QSBB	1.11	[0.82-1.51]
UKB_incident	1.30	[1.03-1.63]
UKB_prevalent	1.42	[1.09-1.84]
Calypso	0.73	[0.47-1.15]
CamPaIGN	1.10	[0.60-2.02]
DIGPD	1.86	[1.10-3.16]
Trondheim	1.76	[1.25-2.47]
Oslo	1.75	[1.26-2.44]
Meta-analysis	**1.36**	**[1.21-1.51]**




PD deaths
non-PD deaths
A rs429358
rs429358 non-carriers 1
rs429358 carriers 1
rs429358 non-carriers 2
rs429358 carriers 2
Probability
Years from onset
B rs4726467
rs4726467 non-carriers 1
rs4726467 carriers 1
rs4726467 non-carriers 2
rs4726467 carriers 2
Probability
Years from onset
C rs144889025
rs144889025 non-carriers 1
rs144889025 carriers 1
rs144889025 non-carriers 2
rs144889025 carriers 2
Probability
Years from onset
D rs10437796
rs10437796 non-carriers 1
rs10437796 carriers 1
rs10437796 non-carriers 2
rs10437796 carriers 2
Probability
Years from onset


GGT5
7.5
5.0
2.5
0.0
-log10P
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
BP


APOE
TOMM40
APOC1
10
5
0
-log10P
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
BP