

Title:

# Genome-wide association study of Parkinson's disease progression biomarkers in 12 longitudinal patients' cohorts

Authors:

Hirota Iwaki M.D. Ph.D.<sup>1,2</sup>, Cornelis Blauwendraat Ph.D.<sup>1</sup>, Hampton L. Leonard M.S.<sup>1,2</sup>, Jonggeol J. Kim BA<sup>1</sup>, Ganqiang Liu Ph.D.<sup>3,4,5</sup>, Jodi Maple-Grødem Ph.D.<sup>6,7</sup>, Jean-Christophe Corvol M.D. Ph.D.<sup>8</sup>, Lasse Pihlstrøm M.D. Ph.D.<sup>9</sup>, Marlies van Nimwegen Ph.D.<sup>10</sup>, Samantha J. Hutten Ph.D.<sup>11</sup>, Khanh-Dung H. Nguyen Ph.D.<sup>12</sup>, Jacqueline Rick Ph.D.<sup>13</sup>, Shirley Eberly M.S.<sup>14</sup>, Faraz Faghri M.S.<sup>1,15</sup>, Peggy Auinger M.S.<sup>16</sup>, Kirsten M. Scott MRCP, M.Phil.<sup>17</sup>, Ruwani Wijeyekoon MRCP, Ph.D.<sup>17</sup>, Vivianna M. Van Deerlin M.D. Ph.D.<sup>18</sup>, Dena G. Hernandez Ph.D.<sup>1</sup>, J. Raphael Gibbs Ph.D.<sup>1</sup>, Kumaraswamy Naidu Chitralla Ph.D.<sup>19</sup>, Aaron G. Day-Williams Ph.D.<sup>20,21</sup>, Alexis Brice M.D.<sup>22,23,24</sup>, Guido Alves M.D., Ph.D.<sup>6,7,25</sup>, Alastair J. Noyce MRCP, Ph.D.<sup>26,27</sup>, Ole-Bjørn Tysnes M.D., Ph.D.<sup>28,29</sup>, Jonathan R. Evans MRCP, Ph.D.<sup>30</sup>, David P. Breen MRCP, Ph.D.<sup>31,32,33</sup>, Karol Estrada Ph.D.<sup>12</sup>, Claire E. Wegel MPH<sup>34</sup>, Fabrice Danjou M.D., Ph.D.<sup>22</sup>, David K. Simon M.D., Ph.D.<sup>35,36</sup>, Ole Andreassen MD PhD<sup>37,38</sup>, Bernard Ravina M.D.<sup>39,40</sup>, Mathias Toft M.D., Ph.D.<sup>9,41</sup>, Peter Heutink Ph.D.<sup>42,43</sup>, Bastiaan R. Bloem M.D., Ph.D.<sup>10</sup>, Daniel Weintraub M.D.<sup>44,45</sup>, Roger A. Barker MRCP, Ph.D.<sup>46</sup>, Caroline H. Williams-Gray MRCP, Ph.D.<sup>46</sup>, Bart P. van de Warrenburg M.D., Ph.D.<sup>10</sup>, Jacobus J. Van Hilten M.D., Ph.D.<sup>47</sup>, Clemens R. Scherzer M.D.<sup>3,4</sup>, Andrew B. Singleton Ph.D.<sup>1</sup>, Mike A. Nalls Ph.D.<sup>1,2</sup>

Affiliations:

<sup>1</sup> Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA

<sup>2</sup> Data Tecnica International, Glen Echo, MD, USA

<sup>3</sup> Advanced Center for Parkinson's Disease Research, Brigham and Women's Hospital and Harvard Medical School, Boston, MA, USA

<sup>4</sup> Precision Neurology Program, Harvard Medical School, Brigham and Women's Hospital, Boston, MA, USA

<sup>5</sup> School of Medicine, Sun Yat-sen University, Guangzhou, Guangdong, China

<sup>6</sup> The Norwegian Centre for Movement Disorders, Stavanger University Hospital, Stavanger, Norway

<sup>7</sup> Department of Chemistry, Bioscience and Environmental Engineering, University in Stavanger,

Stavanger, Norway

<sup>8</sup> Assistance-Publique Hôpitaux de Paris, ICM, INSERM UMRS 1127, CNRS 7225, ICM, Department of Neurology and CIC Neurosciences, Pitié-Salpêtrière Hospital, Paris, France

<sup>9</sup> Department of Neurology, Oslo University Hospital, Oslo, Norway

<sup>10</sup> Radboud University Medical Centre; Donders Institute for Brain, Cognition, and Behaviour; Department of Neurology, Nijmegen, The Netherlands

<sup>11</sup> The Michael J. Fox Foundation for Parkinson's Research, New York, NY, USA

<sup>12</sup> Translational Genome Sciences, Biogen, Cambridge, MA, USA

<sup>13</sup> Department of Neurology University of Pennsylvania, Philadelphia, PA, USA

<sup>14</sup> Department of Biostatistics and Computational Biology, University of Rochester, Rochester, NY, USA

<sup>15</sup> Department of Computer Science, University of Illinois Urbana-Champaign, Champaign, IL, USA

<sup>16</sup> Department of Neurology, Center for Health + Technology, University of Rochester, Rochester, NY, USA

<sup>17</sup> Department of Clinical Neurosciences, University of Cambridge, John van Geest Centre for Brain Repair, Cambridge, UK

<sup>18</sup> Department of Pathology and Laboratory Medicine, Center for Neurodegenerative Disease Research, Parelman School of Medicine at the University of Pennsylvania, Philadelphia, PA, USA

<sup>19</sup> Laboratory of Epidemiology and Population Sciences, National Institute on Aging, National Institutes of Health, Baltimore, MD, USA

<sup>20</sup> Flagship Labs 60 Inc, Cambridge, MA, USA

<sup>21</sup> Statistical Genetics, Biogen, Cambridge, MA, USA

<sup>22</sup> Institut du cerveau et de la moelle épinière ICM, Paris, France

<sup>23</sup> Sorbonne Université SU, Paris, France

<sup>24</sup> INSERM UMR1127, Paris, France

<sup>25</sup> Department of Neurology, Stavanger University Hospital, Stavanger, Norway

<sup>26</sup> Preventive Neurology Unit, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK

<sup>27</sup> Department of Clinical and Movement Neurosciences, UCL Institute of Neurology, London, UK

<sup>28</sup> Department of Neurology, Haukeland University Hospital, Bergen, Norway

<sup>29</sup> University of Bergen, Bergen, Norway

<sup>30</sup> Department of Neurology, Nottingham University NHS Trust, Nottingham, UK

<sup>31</sup> Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, Scotland

<sup>32</sup> Anne Rowling Regenerative Neurology Clinic, University of Edinburgh, Edinburgh, Scotland

<sup>33</sup> Usher Institute of Population Health Sciences and Informatics, University of Edinburgh, Edinburgh,

Scotland

- <sup>34</sup> Department of Medical and Molecular Genetics, Indiana University, Indianapolis, IN, USA
- <sup>35</sup> Department of Neurology, Beth Israel Deaconess Medical Center, Boston, MA, USA
- <sup>36</sup> Harvard Medical School, Boston, MA, USA
- <sup>37</sup> NORMENT; Institute of Clinical Medicine, University of Oslo, Oslo, Norway, Norway
- <sup>38</sup> Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway, Norway
- <sup>39</sup> Voyager Therapeutics, Cambridge, MA, USA
- <sup>40</sup> Department of Neurology, University of Rochester School of Medicine, Rochester, NY, USA
- <sup>41</sup> Institute of Clinical Medicine, University of Oslo, Oslo, Norway
- <sup>42</sup> German Center for Neurodegenerative Diseases-Tuebingen, Tuebingen, Germany
- <sup>43</sup> HIH Tuebingen, Tuebingen, Tuebingen, Germany
- <sup>44</sup> Department of Psychiatry, University of Pennsylvania School of Medicine, Philadelphia, PA, USA
- <sup>45</sup> Department of Veterans Affairs, Philadelphia, PA, USA
- <sup>46</sup> Department of Clinical Neurosciences, University of Cambridge, Cambridge, UK
- <sup>47</sup> Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands

## Abstract

**Background:** Several reports have identified different patterns of Parkinson's disease progression in individuals carrying missense variants in the *GBA* or *LRRK2* genes. The overall contribution of genetic factors to the severity and progression of Parkinson's disease, however, has not been well studied.

**Objectives:** To test the association between genetic variants and the clinical features and progression of Parkinson's disease on a genome-wide scale.

**Methods:** We accumulated individual data from 12 longitudinal cohorts in a total of 4,093 patients with 25,254 observations over a median of 3.81 years. Genome-wide associations were evaluated for 25 cross-sectional and longitudinal phenotypes. Specific variants of interest, including 90 recently-identified disease risk variants, were also investigated for the associations with these phenotypes.

**Results:** Two variants were genome-wide significant. Rs382940(T>A), within the intron of *SLC44A1*, was associated with reaching Hoehn and Yahr stage 3 or higher faster (HR 2.04 [1.58, 2.62], P-value = 3.46E-8). Rs61863020(G>A), an intergenic variant and eQTL for *ADRA2A*, was associated with a lower prevalence of insomnia at baseline (OR 0.63 [0.52, 0.75], P-value = 4.74E-8). In the targeted analysis, we found nine associations between known Parkinson's risk variants and more severe motor/cognitive symptoms. Also, we replicated previous reports of *GBA* coding variants (rs2230288: p.E365K, rs75548401: p.T408M) being associated with greater motor and cognitive decline over time, and *APOE* E4 tagging variant (rs429358) being associated with greater cognitive deficits in patients.

**Conclusions:** We identified novel genetic factors associated with heterogeneity of progression in Parkinson's disease. The results provide new insights into the pathogenesis of Parkinson's disease as well as patient stratification for clinical trials.

## Body

### Introduction

Parkinson's disease (PD) is clinically defined by its motor features of rigidity, bradykinesia, gait disturbance, and tremor. Although these prominent features are important for diagnosis, patients with PD also suffer from many non-motor features such as constipation, urinary incontinence, orthostatic hypotension, REM sleep behavior disorder (RBD), apathy, hyposmia, and cognitive impairment (Postuma *et al.*, 2015). Moreover, patients develop motor complications, including wearing off and dyskinesia, as side effects of medication. The onset, intensity and progression of these different PD clinical features vary among individuals, and the mechanisms underlying this heterogeneity are not well understood.

Recent genome-wide studies have identified 90 common variants associated with the risk of PD, with an overall heritability estimated to be between 22-27% (Keller *et al.*, 2012; Nalls *et al.*, 2019). While previous studies have indicated the importance of genetic contributions to disease risk, the contribution of genetic factors to PD progression and heterogeneity has not been well studied. Investigating genetic factors associated with disease progression and heterogeneity in disease presentation is an important step in elucidating the underlying molecular mechanisms and identifying better patient stratification in clinical trials (Leonard *et al.*, 2018).

Longitudinal patient cohorts are powerful resources that can be used to explore the impact of genetics on the trajectory of PD-related phenotypes; the inherent precision of repeated measurements over time provides more power to detect these associations. However, the available number of participants in each study is usually not enough to conduct a genome-wide association study (GWAS). In this study, we accumulated 25,254 follow-up visits from 4,093 patients across 12 cohorts (Table 1) and performed meta-analyses of longitudinal GWAS on the progression markers of Parkinson's disease. Using the results from this meta-analysis, we evaluated how known risk variants, including the 90 recently identified variants for PD (Nalls *et al.*, 2019), *GBA* protein coding mutations, and *APOE* tagging variants were associated with the progression of phenotypes. To maximize the utility of this work to other researchers, we have made all results from this study publicly searchable and available for download.

[\(https://pdgenetics.shinyapps.io/pdprogmetagwasbrowser/\)](https://pdgenetics.shinyapps.io/pdprogmetagwasbrowser/)

## Methods

### Cohorts

Twelve longitudinal cohorts of PD patients recruited across North America, Europe and Australia were included in our study. The following observational studies were included: the Drug Interaction with Genes in Parkinson's Disease (DIGPD), the Harvard Biomarkers Study (HBS), the Oslo Parkinson's Disease study (partly including retrospective data), the Norwegian ParkWest study (PARKWEST), the Parkinson's Disease Biomarker Program (PDBP), the Parkinsonism Incidence and Cognitive and Non-motor heterogeneity In Cambridgeshire (PICNICS), the Parkinson's Progression Markers Initiative (PPMI), and the Profiling Parkinson's disease study (PROPARK). The four cohorts included were randomized clinical trials: the Deprenyl and Tocopherol Antioxidative Therapy of Parkinsonism (DATATOP), the NIH Exploratory Trials in Parkinson's Disease Large Simple Study 1 (NET-PD\_LS1), the ParkFit study (PARKFIT), and the Parkinson Research Examination of CEP-1347 Trial study with its subsequent prospective study (PreCEPT/PostCEPT). More details of these cohorts are described in Appendix. Participants' information and genetic samples were obtained under appropriate written consent and with local institutional and ethical approvals.

### Phenotyping

For continuous outcomes, we collected the scores of Hoehn and Yahr staging scale (HY) (Goetz *et al.*, 2004), total and sub-scores of the Unified Parkinson's Disease Rating Scale (UPDRS) or the Movement Disorder Society revised UPDRS version (MDS-UPDRS) (Goetz *et al.*, 2007), Mini-Mental State Examination (MMSE), Montreal Cognitive Assessment (MoCA) (Nasreddine *et al.*, 2005) and the modified Schwab and England Activities of Daily Living Scale (SEADL). With the exception of the subscores of UPDRS/MDS-UPDRS part 4, total scores and the subscores of UPDRS and MDS-UPDRS were normalized to the population-baseline mean and standard deviation and converted to Z values. The subscores of UPDRS/MDS-UPDRS part 4, measuring complication of treatment, were normalized to the mean and standard deviation of all observations because the score was 0 at the baseline for the de-novo PD cohorts. We also determined whether subjects were recorded as presenting the following binomial outcomes during participant visits: constipation, cognitive impairment, depression, daytime sleepiness, Hoehn and Yahr stage of 3 or worse (HY3), hyposmia, insomnia, motor fluctuation, REM sleep behavior disorder (RBD), restless legs syndrome (RLS), and an SEADL of 70 or less (SEADL70). Because study-specific criteria for these binomial outcomes were not consistent amongst the studies, we tried to use the

common criteria for these binomial outcomes if we had access to the raw data from the studies. The details of the definitions of binomial outcomes are provided in the Supplemental Table 1.

## Genetics data

The genotyping was conducted with NeuroX, a targeted chip for neurodegenerative disease (Nalls *et al.*, 2015), for NET-PD\_LS1, a part of DIGPD (DIGPD\_neuroX), HBS, PDBP, and PRECEPT. The rest of DIGPD (DIGPD\_chip) were genotyped using Illumina Multi-Ethnic Genotyping Array. Participants in DATATOP, OSLO, PARKFIT, PARKWEST, PICNICS, and PROPARK were genotyped using Illumina Infinium OmniExpress array. Whole genome sequencing data was used for PPMI, with the detailed methods for genome sequencing provided on the PPMI website (<https://www.ppmi-info.org/>).

Variant inclusion criteria consisted of call rate  $> 0.95$ , MAF  $> 0.01$ , and Hardy-Weinberg equilibrium test statistic  $> 1E-4$ . Participants were excluded due to the following criteria: high-missingness ( $> 5\%$  for genotyped variants), sex discordance, extreme heterozygosity (F statistics  $> 0.15$ ), Non-European ancestry confirmed by joint analysis with HapMap 3 data using principal component (Outside of mean  $\pm 6$  SD in PC1 or PC2 for European reference samples) (International HapMap 3 Consortium *et al.*, 2010), and excessive relatedness (pairwise kinships  $> 0.125$ ). We used PLINK version 1.9 for the above filtering (Purcell *et al.*, 2007).

For all samples and variants passing quality control, imputation was conducted for chromosome 1 to 22 using 1000 genome European reference panel at the Michigan Imputation Server (Das *et al.*, 2016) at the default setting, with the exception of the whole genome sequenced PPMI dataset. SNPs with an imputation quality of less than 0.3 and MAF  $< 1\%$  were excluded. After quality control, the number of variants were approximately 2.6 - 2.9 million in NET-PD\_LS1, DIGPD\_neuroX, HBS, PDBP, and PRECEPT; 7.7 - 7.8 million in PICNICS, PROPARK, PARKWEST, DATATOP, PARKFIT, DIGPD, and OSLO; and 8.6 million in PPMI.

## Cohort-level analyses

We conducted a separate GWAS for each cohort per phenotype of interest. In addition, DIGPD cohorts were analyzed separately according to the genotyping array used (DIGPD\_neuroX cohort and DIGPD\_chip cohort). Each outcome was analyzed by an additive model with covariates. For the binomial outcomes at baseline, when the outcomes were positive for more than 5% of participants and  $> 20$  counts, logistic regression analyses were conducted. Those without the binomial outcome at baseline were

followed-up until either censored or the development of the outcome. If more than 20 events were observed during follow-ups, the outcome was analyzed using cox proportional hazard models with time-varying covariates. For the analysis of continuous traits, linear mixed models were used to evaluate the variants' association for the mean difference over time. Age at diagnosis, year from diagnosis to the observation, and sex were adjusted for in all analyses. In addition, the following covariates were associated with the outcome of interest in a backwards stepwise manner: quadratic age, quadratic years from diagnosis, years of education, medication status (levodopa usage, dopamine agonist usage, using either dopamine agonist or levodopa), and a Hoehn and Yahr score of 2 or more at the first observation (except for the models regressing for Hoehn and Yahr score itself or UPDRS motor score). These covariates were selected per study using Akaike's Information Criteria (AIC) for logistic models and Cox survival models, and conditional AIC (cAIC) for linear mixed effect models. The cohort level analyses were conducted with R (version 3.5.0 <https://www.r-project.org/>) and *rvtests* (Zhan *et al.*, 2016). R package 'cAIC4' (Säfken *et al.*, 2018) was used to calculate cAIC.

## Meta-Analyses

The results from cohort-level analyses were combined using an inverse variance weighted fixed effect model. If the study-specific genomic inflation factor was more than 1.2, the study was excluded from the meta-analysis. Five of the 204 GWAS were excluded based on these criteria. For other cohorts, the overall alpha error was corrected using the genomic inflation factor before the meta-analysis. Meta-analyses were carried out with METAL (Willer *et al.*, 2010). From the meta-analysis results, we only evaluated variants with MAF > 0.05 due to statistical power constraints. We also excluded variants with minor allele frequency variability greater than 15% across cohorts. Further exclusions at the meta-analysis level include variants with Cochran's Q-test for heterogeneity < 0.05 and a total participant N < 1000. The null hypothesis was tested with a significance level of 5E-8 on a two-sided test. For genome-wide signals, additional visualization and functional analyses were conducted using LocusZoom (Pruim *et al.*, 2010), FUMA (<http://fuma.ctglab.nl/snp2gene/>, version 1.3.3d) (Watanabe *et al.*, 2017). FUMA is a web-based annotation tool using MAGMA to conduct gene-based tests, a gene-set analysis and a tissue expression analysis. We applied a default setting. Also, we explored in eQTLGen database (<http://www.eqtlgen.org/>) (Võsa *et al.*, 2018) and meta-analyzed expression data in the brain accessible from the study by Qi *et al.* (Qi *et al.*, 2018).

Associations with the variants of interest, including the recently identified 90 risk variants for PD, known *LRRK2* and *GBA* variants, and *APOE*, were extracted from the meta-analysis results. We evaluated the



associations of these variants and clinical features based on the significance level of 0.05, applying the Bonferroni adjustment of a maximum of 25 tests per variant (raw P-value < 0.002).

The summary of analytical processes is shown in figure 1.

## Results

### Novel GWAS associations with PD progression markers

The cohort characteristics are provided in Table 1. Overall, we analyzed 4,093 participants with 25,254 longitudinal data points over a median of 3.81 years. These cohorts varied in the years between enrollment and diagnosis, as well as follow-up durations. DATATOP, ParkWest, PPMI, and PreCEPT/PostCEPT enrolled untreated PD patients while others enrolled both treated and untreated patients. Considering the difference in design and recruitment strategies in the cohorts (Appendix), it is important to adjust for baseline characteristics as well as the follow-up lengths per cohort-level. All cohort-specific models for analysis are listed in Supplemental Table 2.

In total, 204 GWAS were conducted and combined into 33 meta-analyses. Eight meta-analyses were not evaluated because of the small number of total participants in the analyses (N total <1000). Those excluded were baseline analyses for RBD, RLS and SEADL70; and longitudinal analyses for constipation, daytime sleepiness, hyposmia, RBD, and RLS. Therefore, we investigated 9 binomial traits at baseline, 7 binomial traits for survival, and 9 continuous traits over the follow-ups. The genomic inflation factor was the mean value of 0.993, SD of 0.023, and the range was [0.951, 1.031] across meta-analyses. The summary statistics of the meta-analysis results, including the ones which were not evaluated in this manuscript, are publicly available for convenient browsing and downloading (<https://pdgenetics.shinyapps.io/pdprogmetagwasbrowser/>)

One association with the progression of PD was of genome-wide significance (P-value < 5.00E-08). The minor allele of rs382940 (chr9:108058562T>A), an intronic variant of *SLC44A1*, was associated with a higher hazard ratio (HR) of reaching Hoehn and Yahr stage 3.0 or greater (HR 2.04 [1.58, 2.62], P-value = 3.46E-8). When considering the baseline observations, the minor allele of rs61863020 (chr10:112956055G>A), an intergenic variant, was also significantly associated with the lower baseline OR of having insomnia (OR 0.63 [0.52, 0.75], P-value = 4.74E-8). Locus plots and forest plots for these two associations are shown in Figure 2. Cochran's Q statistics, I-square and forest plots all showed no evidence of heterogeneity for these associations. (Figure 2)

To evaluate the potential molecular mechanism for the two genome-wide signals, we explored eQTL datasets in blood and brain (Qi *et al.*, 2018; Vösa *et al.*, 2018), and functional annotation of the GWAS summary statistics using FUMA. Although it is in a regulatory region of *SLC44A*, rs382940 itself was not reported to be an eQTL in blood or brain. Gene-based tests using the GWAS summary statistics for reaching HY3 showed that *SLC44A1* was significant gene-wise (P-value =  $5.8E-07 < \text{Bonferroni correction threshold} = 2.7E-6$ , supplemental figure 1). Rs61863020 was a significant eQTL for *ADRA2A* ( $\alpha$ -2A adrenergic receptor) (P-value =  $7.2E-4$ , the Bonferroni corrected P-value =  $6.5E-3$ , up-regulation for A allele) in the brain.

In the meta-analysis results from the other clinical outcomes, rs382940 was associated with higher scores in the UPDRS part 2 and part 3 (UPDRS2\_scaled: 0.36 [0.15, 0.57], P-value =  $8.21E-04$ ; UPDRS3\_scaled: 0.29 [0.14, 0.45], P-value =  $2.18E-04$ ). These findings are consistent with the primary association of rs382940 and reaching HY3, which is a significant motor milestone (bilateral signs on clinical examination and the emergence of postural instability). Except for the association with having insomnia at baseline, rs61863020 was not significantly associated with other clinical variables in this analysis after adjusting for 25 tests.

## Targeted assessment for the PD risk variants

Of the 90 risk variants from the recently published PD GWAS, rs34637584 (*LRRK2* p.G2019S) and rs76763715 (*GBA* p.N370S) were not available in the meta-analyses because of their minor allele frequency (MAF)  $< 0.01$ . The remaining 88 PD GWAS risk SNPs were assessed in our 25 GWAS summary sets, resulting in evaluations of 2022 candidate associations. 112 associations between known genetic risk variants and clinical markers had raw p-values less than 0.05. After Bonferroni correction for all evaluated candidate associations, nine surpassed the threshold of the analyses-wide significance for the maximum of 25 analyses per variant (raw P-value  $< 0.002$ ). The directions of these associations generally indicated that having the higher risk allele was associated with more severe deficits in both the cognitive and motor domains of PD, but not for sleeping problems. Having the risk allele (A) of rs1293298 (intron variant of *CTSB*) was associated with a lower risk of developing insomnia (HR 0.79 [0.69, 0.91], P-value =  $1.2E-3$ ), and the risk allele (A) of rs6500328 (intron variant of *NOD2*) and (A) of rs76116224 (intergenic variant close to 3' end of *KCNS3*) were associated with a lower prevalence of daytime sleepiness at baseline (OR 0.76 [0.64, 0.90], P-value =  $1.4E-3$ ; OR 0.47 [0.32, 0.68], P-value =  $8.4E-5$ ; respectively). Among the 10 associations with analysis-wide significance, two were significant after adjusting for 88 variants (raw P-value  $< 5.68E-4$ ), and one had test-wide significance (raw P-value  $<$

2.47E-5). Figure 3 shows the strength of the associations for the selected variants with associations of analyses-wide significance in at least one analysis. This figure suggests that some risk variants were associated with specific clinical features. For example, rs35749011 was associated with both the HR of cognitive impairment at test-wide significance (HR 2.45 [1.64, 3.65] for the minor allele, P-value = 1.1E-5) and lower MoCA score over time at analyses-wide significance (-1.16 [-1.89, -0.43], P-value = 0.0018). Although it is an intergenic variant whose closest gene is *KRTCAP2*, the variant is in high LD ( $r^2 = 0.78$ ) with rs2230288 (*GBA* p.E365K) (Berge-Seidl *et al.*, 2017; Blauwendraat *et al.*, 2018), and has a similar spectrum of phenotype associations as rs2230288. Other notable variants with variant-wide significance were rs76904798, the intergenic variant close to the 5' end of *LRRK2*, for reaching HY3 (HR 1.32 [1.14, 1.54] with the minor allele of T, P-value = 3.0E-4), and rs76116224 and the baseline OR of having daytime sleepiness mentioned above. The detailed information for all of the test results is provided as supplemental material. (Supplemental Table and Supplemental Figure)

### *GBA* protein coding variants and APOE tagging variants

In the focused analyses for *GBA* coding variants, rs75548401, *GBA* p.T408M, was associated with the faster development of HY3 (HR 2.35 [1.58, 3.49], P-value = 2.5E-5). rs2230288, *GBA* p.E365K, was associated with the higher odds of having cognitive impairment at baseline (OR 2.05 [1.33, 3.18], P-value = 1.3E-3), faster development of cognitive impairment (HR 2.58 [1.71, 3.89] P-value = 5.5E-6), and lower MoCA score at the analysis-wide significance (Beta -1.23 [-1.97, -0.50], P-value = 1.0E-3). We previously reported these associations (under revision at *Neurology Genetics*) and we were able to confirm them in our updated analysis with more stringent multiple testing correction (FDR vs Bonferroni).

The C allele of rs429358, the tagging variant for the *APOE* E4 allele, was associated with lower MMSE (Beta -0.20 [-0.33, -0.07], P-value = 2.8E-3) and lower MoCA (Beta -0.52 [-0.86, -0.17], P-value = 3.4E-3) as expected. Moreover, it was associated with higher UPDRS part 1 scores (Beta 0.12 [0.04, 0.20] in Z score, P-value = 4.5E-3). We did not have enough evidence to conclude that the *APOE* E4 allele was associated with the prevalence of cognitive impairment at baseline (P-value = 0.4) or its development during follow-ups (P-value = 0.034). The T allele of rs7412 showed no association with these measurements, also predicted as this variant tagging *APOE* E2.

## Discussion

We conducted GWAS using longitudinal data from multiple PD cohorts to investigate markers of PD progression and heterogeneity. Of the 25 meta-analyses that we evaluated, we identified two variant-phenotype associations with genome-wide significance.

We also evaluated the summary statistics to assess clinical value of the variants of interest.

One of our genome-wide hits, rs382940, in the intron of *SLC44A1*, was associated with a faster rate of progression to reach HY3. *SLC44A1*, soluble carrier 44A1, is also referred to as choline transporter-like protein 1 (CTL1). The gene is ubiquitously expressed in the brain, colon, thyroid and other organs and is involved in choline transport. No associations with PD and this variant or the gene itself have been reported so far although it has been studied in several *in vitro* and *in vivo* studies (Machová *et al.*, 2009; Schenkel *et al.*, 2015; Heffernan *et al.*, 2017; Gao *et al.*, 2018). Further investigation is warranted. The search of the Brain eQTL database suggested that another GWAS-signal, rs61863020, was associated with *ADRA2A* expression, a gene reported to be associated with arousal/sleep state (Gelegen *et al.*, 2014). *ADRA2A* is consistently expressed in locus coeruleus as well as nigral dopamine neurons and pyramidal neurons of the human brain (<http://www.humanbraincode.org/>, (Dong *et al.*, 2018). The *ADRA2A*-encoded alpha2 adrenoreceptor modulates norepinephrine levels. Interestingly, norepinephrine (Tong *et al.*, 2006) and its receptors (Srinivasan and Schmidt, 2004; Mittal *et al.*, 2017) have been linked to PD in multiple model systems.

In the targeted assessments, we confirmed the previous results of the associations between *GBA* risk variants and motor and cognitive aspects of PD (Winder-Rhodes *et al.*, 2013; Brockmann *et al.*, 2015; Davis *et al.*, 2016a, b; Liu *et al.*, 2016). In contrast with *GBA* variants, association studies of *APOE* and cognitive function in PD have yielded mixed results (Huang *et al.*, 2006; Kurz *et al.*, 2009; Federoff *et al.*, 2012; Mata *et al.*, 2014; Paul *et al.*, 2016). Our data supported the association of *APOE* and cognitive function on two measurements; MMSE and MoCA.

The strength of the current study is the hypothesis-free approach of GWAS, which can be powerful in identifying new associations and expanding our biological knowledge-base. While the associations here should be replicated and further investigated with *in vivo/vitro* experiments, these findings suggest the prioritization of the two variants and loci for future validations. We have reported all of the summary results on our publicly accessible site to benefit researchers so that they may conduct/replicate the analysis of variants of interest in their own research.

The major limitation of this study, and studies like it, is the heterogeneity of the cohorts, which is apparent in several ways: baseline characteristics, definitions of binomial outcomes, patterns for clinical care over the course of follow-up, the platforms for genotyping/sequencing, and sample

acquisition/enrollment practices. By meta-analyzing at the dataset-level and exercising careful quality control throughout, we tried to extract the most generalizable and reliable results across cohorts. Another limitation is the power of the study. Although we have aggregated the largest collection of longitudinal data in PD genetics so far, more data would be needed to identify relatively small differences expected within PD patients compared to the case-vs-control setting. Finally, the study participants were restricted to individuals with European ancestry. We are now striving to collect more data, including from populations that are under-represented in this study, to improve our understanding of this topic in future studies.

## Conclusion

With 4,093 participants and 25,254 longitudinal data points over a median of 3.81 years, we performed 25 GWAS meta-analyses. We found two genome-wide significant signals: the rate to reach HY3 during the disease course and rs382940; and the prevalence of insomnia at baseline and rs61863020. We also conducted targeted assessments of previously published variants of interest using the GWAS results. These results provide valuable insights into how genetic factors contribute to the heterogeneity of PD and disease progression.

## References

- Berge-Seidl V, Pihlstrøm L, Maple-Grødem J, Forsgren L, Linder J, Larsen JP, et al. The GBA variant E326K is associated with Parkinson's disease and explains a genome-wide association signal. *Neurosci. Lett.* 2017; 658: 48–52.
- Blauwendraat C, Bras JM, Nalls MA, Lewis PA, Hernandez DG, Singleton AB, et al. Coding variation in GBA explains the majority of the SYT11-GBA Parkinson's disease GWAS locus. *Mov. Disord.* 2018; 33: 1821–1823.
- Brockmann K, Sruļijes K, Pflederer S, Hauser A-K, Schulte C, Maetzler W, et al. GBA-associated Parkinson's disease: reduced survival and more rapid progression in a prospective longitudinal study. *Mov. Disord.* 2015; 30: 407–411.
- Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat. Genet.* 2016; 48: 1284–1287.
- Davis AA, Andruska KM, Benitez BA, Racette BA, Perlmutter JS, Cruchaga C. Variants in GBA, SNCA, and MAPT influence Parkinson disease risk, age at onset, and progression. *Neurobiol. Aging* 2016a; 37:

209.e1–209.e7.

Davis MY, Johnson CO, Leverenz JB, Weintraub D, Trojanowski JQ, Chen-Plotkin A, et al. Association of GBA Mutations and the E326K Polymorphism With Motor and Cognitive Progression in Parkinson Disease. *JAMA Neurol.* 2016b; 73: 1217–1224.

Dong X, Liao Z, Gritsch D, Hadzhiev Y, Bai Y, Locascio JJ, et al. Enhancers active in dopamine neurons are a primary link between genetic variation and neuropsychiatric disease. *Nat. Neurosci.* 2018; 21: 1482–1492.

Federoff M, Jimenez-Rolando B, Nalls MA, Singleton AB. A large study reveals no association between APOE and Parkinson's disease. *Neurobiol. Dis.* 2012; 46: 389–392.

Gao P, He M, Zhang C, Geng C. Integrated analysis of gene expression signatures associated with colon cancer from three datasets. *Gene* 2018; 654: 95–102.

Gelegen C, Gent TC, Ferretti V, Zhang Z, Yustos R, Lan F, et al. Staying awake--a genetic region that hinders  $\alpha 2$  adrenergic receptor agonist-induced sleep. *Eur. J. Neurosci.* 2014; 40: 2311–2319.

Goetz CG, Fahn S, Martinez-Martin P, Poewe W, Sampaio C, Stebbins GT, et al. Movement Disorder Society-sponsored revision of the Unified Parkinson's Disease Rating Scale (MDS-UPDRS): Process, format, and clinimetric testing plan. *Mov. Disord.* 2007; 22: 41–47.

Goetz CG, Poewe W, Rascol O, Sampaio C, Stebbins GT, Counsell C, et al. Movement Disorder Society Task Force report on the Hoehn and Yahr staging scale: status and recommendations. *Mov. Disord.* 2004; 19: 1020–1028.

Heffernan C, Jain MR, Liu T, Kim H, Barretto K, Li H, et al. Nectin-like 4 Complexes with Choline Transporter-like Protein-1 and Regulates Schwann Cell Choline Homeostasis and Lipid Biogenesis in Vitro. *J. Biol. Chem.* 2017; 292: 4484–4498.

Huang X, Chen P, Kaufer DI, Tröster AI, Poole C. Apolipoprotein E and dementia in Parkinson disease: a meta-analysis. *Arch. Neurol.* 2006; 63: 189–193.

International HapMap 3 Consortium, Altshuler DM, Gibbs RA, Peltonen L, Altshuler DM, Gibbs RA, et al. Integrating common and rare genetic variation in diverse human populations. *Nature* 2010; 467: 52–58.

Keller MF, Saad M, Bras J, Bettella F, Nicolaou N, Simón-Sánchez J, et al. Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Hum. Mol. Genet.* 2012; 21: 4996–5009.

Kurz MW, Dekomien G, Nilsen OB, Larsen JP, Aarsland D, Alves G. APOE Alleles in Parkinson Disease and Their Relationship to Cognitive Decline: A Population-based, Longitudinal Study. *J. Geriatr. Psychiatry Neurol.* 2009; 22: 166–170.

Leonard H, Blauwendraat C, Krohn L, Faghri F, Iwaki H, Furgeson G, et al. Genetic variability and potential effects on clinical trial outcomes: perspectives in Parkinson's disease. *bioRxiv* 2018: 427385.

Liu G, Boot B, Locascio JJ, Jansen IE, Winder-Rhodes S, Eberly S, et al. Specifically neuropathic Gaucher's mutations accelerate cognitive decline in Parkinson's. *Ann. Neurol.* 2016; 80: 674–685.

Machová E, O'Regan S, Newcombe J, Meunier F-M, Prentice J, Dove R, et al. Detection of choline transporter-like 1 protein CTL1 in neuroblastoma× glioma cells and in the CNS, and its role in choline uptake. *J. Neurochem.* 2009; 110: 1297–1309.

Mata IF, Leverenz JB, Weintraub D, Trojanowski JQ, Hurtig HI, Van Deerlin VM, et al. APOE, MAPT, and SNCA genes and cognitive performance in Parkinson disease. *JAMA Neurol.* 2014; 71: 1405–1412.

Mittal S, Bjørnevik K, Im DS, Flierl A, Dong X, Locascio JJ, et al.  $\beta$ 2-Adrenoreceptor is a regulator of the  $\alpha$ -synuclein gene driving risk of Parkinson's disease. *Science* 2017; 357: 891–898.

Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. Expanding Parkinson's disease genetics: novel risk loci, genomic context, causal insights and heritable risk [Internet]. 2019 Available from: <http://dx.doi.org/10.1101/388165>

Nalls MA, Bras J, Hernandez DG, Keller MF, Majounie E, Renton AE, et al. NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases. *Neurobiol. Aging* 2015; 36: 1605.e7–12.

Nasreddine ZS, Phillips NA, Bédirian V, Charbonneau S, Whitehead V, Collin I, et al. The Montreal Cognitive Assessment, MoCA: a brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* 2005; 53: 695–699.

Paul KC, Rausch R, Creek MM, Sinsheimer JS, Bronstein JM, Bordelon Y, et al. APOE, MAPT, and COMT and Parkinson's Disease Susceptibility and Cognitive Symptom Progression. *J. Parkinsons. Dis.* 2016; 6: 349–359.

Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov. Disord.* 2015; 30: 1591–1601.

Pruim RJ, Welch RP, Sanna S, Teslovich TM, Chines PS, Gliedt TP, et al. LocusZoom: regional visualization of genome-wide association scan results. *Bioinformatics* 2010; 26: 2336–2337.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* 2007; 81: 559–575.

Qi T, Wu Y, Zeng J, Zhang F, Xue A, Jiang L, et al. Identifying gene targets for brain-related traits using transcriptomic and methylomic data from blood. *Nat. Commun.* 2018; 9: 2282.

Säfken B, Rügamer D, Kneib T, Greven S. Conditional Model Selection in Mixed-Effects Models with cAIC4 [Internet]. arXiv [stat.CO] 2018 Available from: <http://arxiv.org/abs/1803.05664>

Schenkel LC, Singh RK, Michel V, Zeisel SH, da Costa K-A, Johnson AR, et al. Mechanism of choline deficiency and membrane alteration in postural orthostatic tachycardia syndrome primary skin fibroblasts. *FASEB J.* 2015; 29: 1663–1675.

Srinivasan J, Schmidt WJ. Treatment with alpha2-adrenoceptor antagonist, 2-methoxy idazoxan, protects 6-hydroxydopamine-induced Parkinsonian symptoms in rats: neurochemical and behavioral evidence. *Behav. Brain Res.* 2004; 154: 353–363.

Tong J, Hornykiewicz O, Kish SJ. Inverse relationship between brain noradrenaline level and dopamine loss in Parkinson disease: a possible neuroprotective role for noradrenaline. *Arch. Neurol.* 2006; 63: 1724–1728.

Võsa U, Claringbould A, Westra H-J, Bonder MJ, Deelen P, Zeng B, et al. Unraveling the polygenic architecture of complex traits using blood eQTL meta-analysis. *bioRxiv* 2018: 447367.

Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* 2017; 8: 1826.

Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; 26: 2190–2191.

Winder-Rhodes SE, Evans JR, Ban M, Mason SL, Williams-Gray CH, Foltynie T, et al. Glucocerebrosidase mutations influence the natural history of Parkinson's disease in a community-based incident cohort. *Brain* 2013; 136: 392–399.

Zhan X, Hu Y, Li B, Abecasis GR, Liu DJ. RVTESTS: an efficient and comprehensive tool for rare variant association analysis using sequence data. *Bioinformatics* 2016; 32: 1423–1426.



Figures:

Figure 1: Graphical overview of the analysis strategy.

\* DIGPD cohort was analyzed separately depending on the genotyping system.

Rsq, R square; MAF, Minor allele frequency; HWE, Hardy–Weinberg equilibrium test; OR, Odds ratio; HR, Hazard ratio; PC, Principal components; AAD, Age at diagnosis; YfD, Years from diagnosis to observation; HY score, the score on the Hoehn and Yahr scale;

Figure 2: Locuszoom plots and forest plots of the two genome-wide significant hits. A: The locus plot for rs382940 which is associated with HY3. B: The locus plot for rs61863020 which is associated with insomnia. C: The forest plot for rs382940. D: The forest plot for rs61863020.

Figure 3: Heatmap of the Parkinson’s disease GWAS loci associated with progression markers  
Cream, P-value > 0.05; light green, P-value < 0.05; green: P-value < 0.002; blue, P-value < 5.68E-4; dark blue, P-value < 2.47E-5).

CONST, constipation; COGi, cognitive impairment; DEPR, depression; HY3, Hoehn and Yahr score; INS, insomnia; MMSE, Mini-Mental State Examination; MOCA, Montreal Cognitive Assessment; SEADL70, the modified Schwab and England Activities of Daily Living Scale; SLEEP, daytime sleepiness; UPDRS, Unified Parkinson’s Disease Rating Scale (UPDRS) or the Movement Disorder Society revised UPDRS, scaled at the baseline (UPDRS1-3) or during the course.

Suffix of ‘base’ indicates the logistic regression model at baseline, ‘surv’ for the survival analysis over the course, and ‘cont’ for the mean difference overtime analyzed by linear mixed model.

Figure 4: Heatmap of the GBA and APOE variants associated with progression markers

Cream, P-value > 0.05; light green, P-value < 0.05; green: Bonferroni corrected P-value < 0.05;

CONST, constipation; COGi, cognitive impairment; DEPR, depression; HY3, Hoehn and Yahr score; INS, insomnia; MMSE, Mini-Mental State Examination; MOCA, Montreal Cognitive Assessment; SEADL, the modified Schwab and England Activities of Daily Living Scale; SLEEP, daytime sleepiness; UPDRS, Unified Parkinson’s Disease Rating Scale (UPDRS) or the Movement Disorder Society revised UPDRS, scaled at the baseline (UPDRS1-3) or during the course.

Suffix of ‘base’ indicates the logistic regression model at baseline, ‘surv’ for the survival analysis over the course, and ‘cont’ for the mean difference overtime analyzed by linear mixed model.

Supplementary Materials:

Appendix: The description of study cohorts

Supplemental table 1: Cohort specific definitions of binomial outcomes

Supplemental table 2: Analytical models per datasets

Supplemental table 3: The meta-analysis results of the association between risk variants and the clinical features and progression of Parkinson's disease

Supplemental figure 1: Gene-based test for reaching Hoehn and Yahr stage 3 or higher

Supplemental figure 2: Heatmap for the meta-analysis results of the association between risk variants and the clinical features and progression of Parkinson's disease

**Table 1. Summary of 13 datasets (12 cohorts)**

|                                      | DATATOP      | DIGPD_chip    | DIGPD_neuroX  | HBS           | NET-PD_LS1   | OSLO          | PARKFIT      | PARKWEST     | PDBP          | PICNICS      | PPMI         | PRECEPT<br>/POSTCEPT | PROPARK       |
|--------------------------------------|--------------|---------------|---------------|---------------|--------------|---------------|--------------|--------------|---------------|--------------|--------------|----------------------|---------------|
| n                                    | 428          | 108           | 235           | 503           | 340          | 318           | 332          | 147          | 515           | 117          | 441          | 318                  | 291           |
| Number of observations               | 7.09 (1.66)  | 2.48 (1.58)   | 3.45 (1.27)   | 2.40 (0.76)   | 6.50 (1.54)  | 6.56 (4.04)   | 2.00 (0.00)  | 2.00 (0.00)  | 4.58 (3.19)   | 3.06 (1.12)  | 12.18 (4.21) | 15.67 (2.18)         | 5.49 (1.13)   |
| Follow up in years                   | 1.22 (0.41)  | 1.52 (1.60)   | 2.54 (1.30)   | 1.53 (0.84)   | 4.47 (1.48)  | 12.08 (6.91)  | 1.97 (0.00)  | 3.04 (0.09)  | 1.97 (1.73)   | 3.03 (1.62)  | 4.88 (2.05)  | 6.78 (0.95)          | 4.64 (1.13)   |
| Age at diagnosis                     | 58.73 (9.12) | 59.04 (10.14) | 60.44 (9.43)  | 62.27 (10.40) | 60.84 (9.07) | 54.43 (10.14) | 60.85 (8.62) | 67.16 (9.29) | 58.59 (10.18) | 68.95 (9.38) | 61.05 (9.82) | 59.45 (9.18)         | 53.11 (10.58) |
| Year from diagnosis                  | 1.15 (1.16)  | 2.78 (1.56)   | 2.52 (1.57)   | 4.02 (4.62)   | 1.48 (0.97)  | 1.76 (4.69)   | 5.14 (4.40)  | 0.14 (0.12)  | 6.31 (5.44)   | 0.23 (0.48)  | 0.55 (0.54)  | 0.80 (0.84)          | 6.61 (4.67)   |
| Female (%)                           | 142 (33.2)   | 44 (40.7)     | 91 (38.7)     | 174 (34.6)    | 122 (35.9)   | 106 (33.3)    | 109 (32.8)   | 55 (37.4)    | 202 (39.2)    | 43 (36.8)    | 159 (36.1)   | 105 (33.0)           | 105 (36.1)    |
| Having HY2 or larger at baseline (%) | 197 (46.0)   | 68 (65.4)     | 151 (64.3)    | 420 (84.5)    | 0 ( NaN)     | 22 (100.0)    | 314 (94.6)   | 85 (57.8)    | 426 (83.0)    | 62 (53.0)    | 242 (54.9)   | 198 (62.5)           | 271 (96.1)    |
| Years of education                   | 14.19 (3.29) | 12.27 (2.92)  | 12.28 (3.17)  | 15.14 (1.72)  | 15.46 (2.23) | NaN (NA)      | NaN (NA)     | 11.31 (3.24) | 15.24 (4.56)  | 12.17 (2.89) | 15.58 (2.98) | 15.93 (3.15)         | 11.95 (4.10)  |
| Use of levodopa (%)                  | 0 ( 0.0)     | 64 (59.8)     | 161 (68.5)    | 353 (70.2)    | 171 (50.6)   | 0 ( NaN)      | 0 ( NaN)     | 0 ( 0.0)     | 409 (83.1)    | 35 (29.9)    | 3 ( 0.9)     | 0 ( 0.0)             | 199 (68.4)    |
| Use of agonist (%)                   | 0 ( 0.0)     | 88 (81.5)     | 163 (69.4)    | 198 (39.4)    | 231 (68.3)   | 0 ( NaN)      | 0 ( NaN)     | 0 ( 0.0)     | 277 (56.3)    | 22 (18.8)    | 0 ( 0.0)     | 1 ( 0.3)             | 218 (74.9)    |
| Hyposmia (%)                         | 0 ( 0.0)     | 33 (30.8)     | 66 (28.4)     | 0 ( NaN)      | 0 ( NaN)     | 0 ( NaN)      | 0 ( NaN)     | 53 (36.1)    | 295 (63.6)    | 0 ( NaN)     | 185 (44.6)   | 0 ( NaN)             | 170 (63.7)    |
| Cognitive impairment (%)             | 24 ( 5.6)    | 1 ( 0.9)      | 2 ( 0.9)      | 59 (11.9)     | 27 ( 7.9)    | 0 ( NaN)      | 55 (16.6)    | 27 (18.4)    | 119 (23.2)    | 11 ( 9.4)    | 35 ( 7.9)    | 3 ( 0.9)             | 76 (27.1)     |
| Motor fluctuations (%)               | 0 ( NaN)     | 14 (13.1)     | 32 (13.6)     | 198 (39.9)    | 87 (25.7)    | 0 ( NaN)      | 0 ( NaN)     | 4 ( 2.7)     | 176 (37.3)    | 0 ( 0.0)     | 0 ( 0.0)     | 0 ( NaN)             | 92 (32.3)     |
| Dyskinesias (%)                      | 3 ( 0.7)     | 4 ( 3.7)      | 13 ( 5.5)     | 168 (33.8)    | 5 ( 1.5)     | 0 ( NaN)      | 0 ( NaN)     | 2 ( 1.4)     | 118 (25.0)    | 0 ( 0.0)     | 0 ( 0.0)     | 0 ( NaN)             | 80 (27.8)     |
| Depression (%)                       | 11 ( 2.6)    | 29 (27.1)     | 77 (33.3)     | 28 (10.3)     | 31 ( 9.2)    | 0 ( NaN)      | 0 ( NaN)     | 20 (13.6)    | 59 (12.5)     | 26 (22.2)    | 141 (33.8)   | 72 (22.6)            | 48 (16.6)     |
| Restless legs syndrome (%)           | 0 ( NaN)     | 16 (15.7)     | 34 (14.5)     | 30 (10.4)     | 0 ( NaN)     | 0 ( NaN)      | 0 ( NaN)     | 0 ( NaN)     | 107 (24.4)    | 0 ( NaN)     | 27 (0.65)    | 0 ( NaN)             | 0 ( NaN)      |
| Constipation (%)                     | 9 ( 2.1)     | 17 (16.0)     | 48 (20.7)     | 0 ( NaN)      | 0 ( NaN)     | 0 ( NaN)      | 0 ( NaN)     | 17 (11.6)    | 255 (54.0)    | 27 (23.1)    | 149 (33.8)   | 0 ( NaN)             | 137 (47.1)    |
| REM sleep behavior disorder (%)      | 0 ( NaN)     | 0 ( NaN)      | 0 ( NaN)      | 0 ( NaN)      | 0 ( NaN)     | 0 ( NaN)      | 0 ( NaN)     | 0 ( NaN)     | 217 (49.4)    | 0 ( NaN)     | 104 (24.9)   | 0 ( NaN)             | 0 ( NaN)      |
| Daytime sleepiness (%)               | 3 ( 0.7)     | 51 (48.6)     | 104 (44.3)    | 0 ( NaN)      | 0 ( NaN)     | 0 ( NaN)      | 0 ( NaN)     | 23 (15.6)    | 178 (37.7)    | 24 (20.5)    | 61 (14.7)    | 0 ( NaN)             | 125 (43.0)    |
| Insomnia (%)                         | 11 ( 2.6)    | 31 (30.4)     | 86 (36.6)     | 170 (33.9)    | 0 ( NaN)     | 0 ( NaN)      | 0 ( NaN)     | 45 (30.6)    | 332 (70.3)    | 59 (50.4)    | 109 (24.7)   | 0 ( NaN)             | 83 (28.5)     |
| Hoehn Yahr scale 3 or greater (%)    | 0 ( 0.0)     | 2 ( 1.9)      | 3 ( 1.3)      | 59 (11.9)     | 10 ( 2.9)    | 0 ( 0.0)      | 17 ( 5.1)    | 11 ( 7.5)    | 81 (15.8)     | 12 (10.3)    | 2 ( 0.5)     | 0 ( 0.0)             | 115 (40.8)    |
| SEADL 70 or less (%)                 | 3 ( 0.7)     | 30 (30.0)     | 12 ( 5.1)     | 0 ( NaN)      | 2 ( 0.6)     | 0 ( NaN)      | 0 ( NaN)     | 5 ( 3.4)     | 75 (15.9)     | 0 ( NaN)     | 2 ( 0.5)     | 1 ( 0.3)             | 0 ( NaN)      |
| Hoehn Yahr scale                     | 1.61 (0.53)  | 1.76 (0.56)   | 1.77 (0.54)   | 2.13 (0.63)   | NaN (NA)     | 3.27 (0.55)   | 2.08 (0.33)  | 1.87 (0.58)  | 2.03 (0.73)   | 1.63 (0.66)  | 1.50 (0.50)  | 1.75 (0.48)          | 2.51 (0.79)   |
| UPDRS_scaled                         | -0.07 (0.98) | 0.08 (1.00)   | -0.03 (0.96)  | -0.00 (1.00)  | 0.00 (1.01)  | NaN (NA)      | -0.06 (0.96) | -0.02 (1.01) | -0.07 (1.01)  | -0.01 (0.98) | NaN (NA)     | 0.02 (1.00)          | NaN (NA)      |
| UPDRS1_scaled                        | NaN (NA)     | -0.10 (0.99)  | -0.02 (0.90)  | 0.01 (1.01)   | 0.01 (1.06)  | NaN (NA)      | NaN (NA)     | -0.03 (1.00) | 0.03 (1.01)   | NaN (NA)     | 0.00 (4.31)  | -0.03 (0.96)         | NaN (NA)      |
| UPDRS2_scaled                        | NaN (NA)     | 0.13 (1.01)   | -0.06 (0.98)  | 0.00 (1.00)   | 0.01 (0.98)  | NaN (NA)      | NaN (NA)     | -0.03 (1.00) | -0.00 (1.01)  | NaN (NA)     | 0.00 (4.19)  | -0.01 (1.00)         | NaN (NA)      |
| UPDRS3_scaled                        | NaN (NA)     | 0.09 (0.99)   | -0.02 (0.96)  | -0.01 (1.00)  | 0.00 (1.00)  | 0.61 (1.21)   | NaN (NA)     | -0.01 (1.00) | -0.09 (1.00)  | NaN (NA)     | -0.00 (8.87) | 0.04 (0.99)          | NaN (NA)      |
| UPDRS4_scaled                        | NaN (NA)     | -0.09 (1.55)  | -0.17 (0.81)  | -0.20 (0.86)  | -0.33 (0.71) | NaN (NA)      | NaN (NA)     | -0.25 (0.86) | -0.12 (0.94)  | NaN (NA)     | NaN (NA)     | NaN (NA)             | NaN (NA)      |
| MMSE                                 | 29.01 (1.32) | 28.59 (1.66)  | 28.21 (1.75)  | 28.39 (2.17)  | NaN (NA)     | NaN (NA)      | 28.09 (1.61) | 27.86 (2.29) | NaN (NA)      | 28.70 (1.44) | NaN (NA)     | 29.28 (1.08)         | 27.05 (2.50)  |
| MOCA                                 | NaN (NA)     | NaN (NA)      | NaN (NA)      | NaN (NA)      | NaN (NA)     | NaN (NA)      | NaN (NA)     | NaN (NA)     | 25.40 (3.57)  | NaN (NA)     | 27.10 (2.31) | NaN (NA)             | NaN (NA)      |
| SEADL                                | 91.56 (6.49) | 64.96 (41.02) | 88.54 (14.94) | NaN (NA)      | 91.88 (5.84) | NaN (NA)      | NaN (NA)     | 89.39 (7.42) | 84.58 (14.84) | NaN (NA)     | 93.79 (6.11) | 92.78 (5.26)         | NaN (NA)      |

Continuous variables were summarized in Mean (SD). MMSE, Mini Mental State Examination; Montreal Cognitive Assessment, SEADL, Schwab and England Activities of Daily Living Scale; UPDRS, Unified Parkinson Disease Rating Scale; MDS-UPDRS, Movement Disorder Society revised version of UPDRS.

Figure 1. Graphical overview of the analysis strategy.

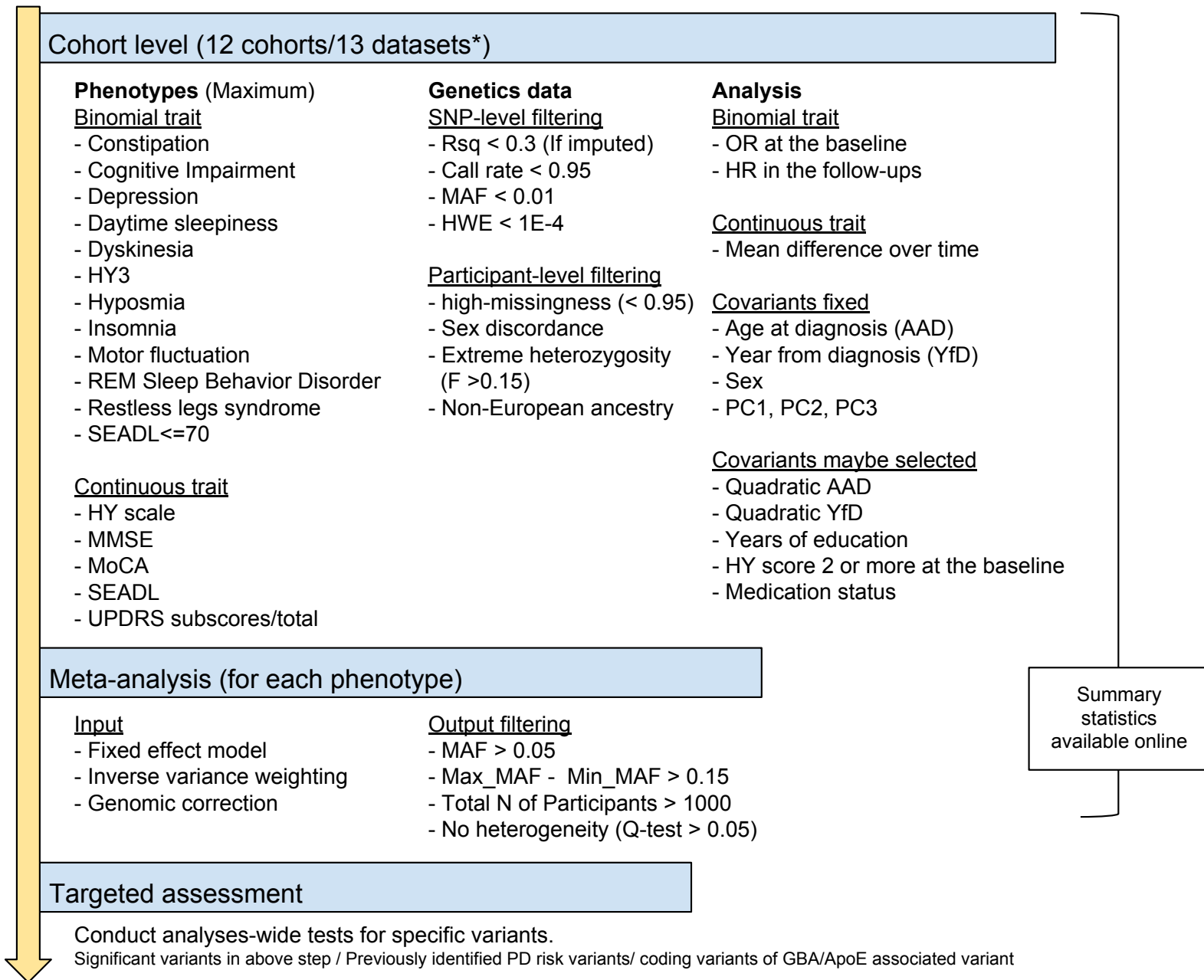
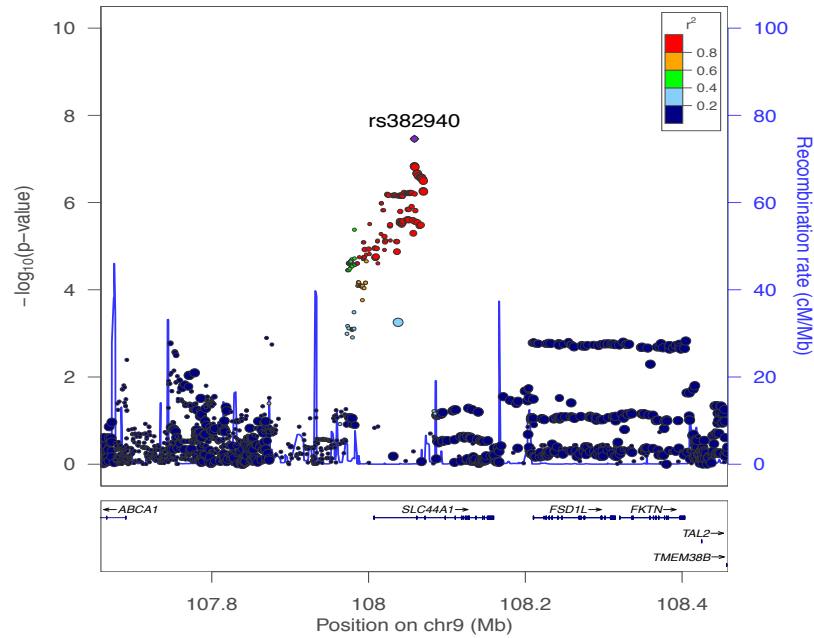
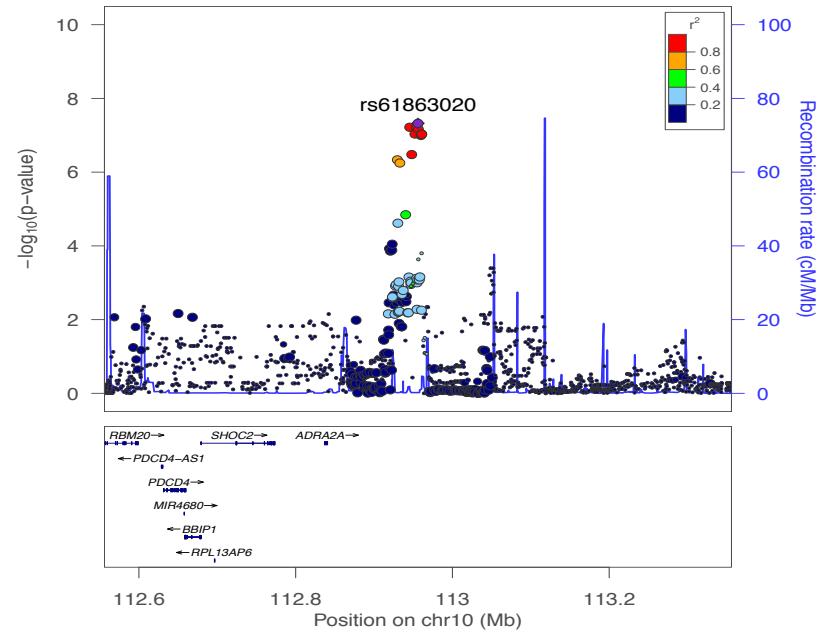


Figure 2. LocusZoom plots and forest plots of the two genome-wide significant hits



Logarithm of HR for rs382940 on HY3

P-value = 3.46E-08  
 I-sq = 44.8%, Q-test = 0.093  
 Total number of events = 527  
 Total number of participants in analysis = 1890



Logarithm of OR for rs61863020 on Insomnia

P-value = 4.74E-08  
 I-sq = 0.0%, Q-test = 0.724  
 Total number of participants in analysis = 2220

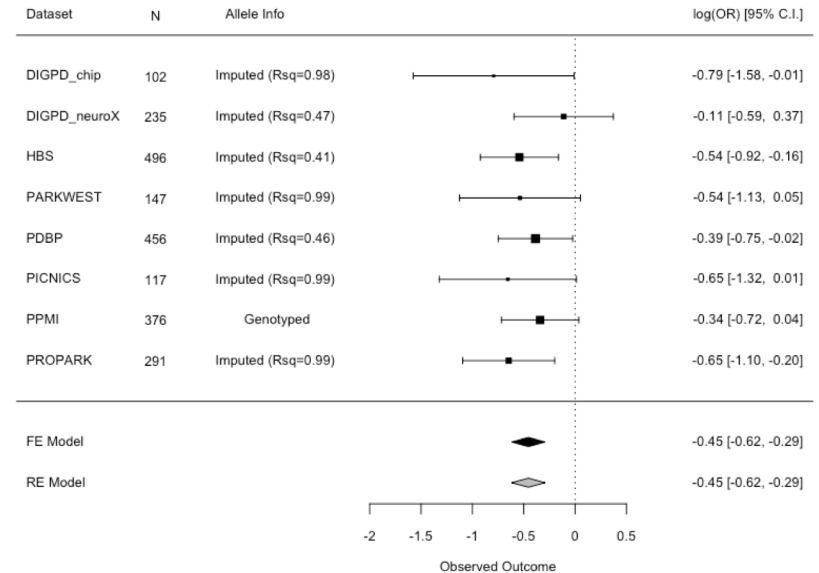
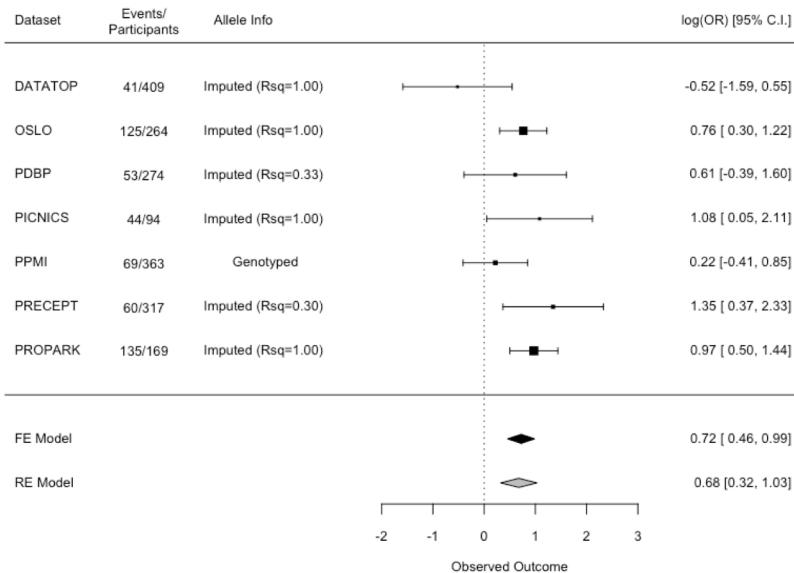


Figure 3. Heatmap of the Parkinson's disease GWAS loci associated with progression markers.

