Parkinson's disease age at onset genome-wide association study: 
Defining heritability, genetic loci, and alpha-synuclein mechanisms 
Blauwendraat, C.; Heilbron, K.; Vallerga, C.L.; Bandres-Ciga, S.; Coelln, R. von; Pihlstrom, L.; ... ; IPDGC

Citation 

Version: Not Applicable (or Unknown) 
License: Leiden University Non-exclusive license 
Downloaded from: https://hdl.handle.net/1887/120799

Note: To cite this publication please use the final published version (if applicable).
Parkinson’s Disease Age at Onset Genome-Wide Association Study: Defining Heritability, Genetic Loci, and α-Synuclein Mechanisms

Cornelis Blauwendraat, PhD,1,2* Karl Heilbron, PhD,3 Costanza L. Vallerga, MSc,4 Sara Bandres-Ciga, PhD,1 Rainer von Coeltn, MD,3,4 Lasse Pihlstrøm, MD, PhD,5 Javier Simón-Sánchez, PhD,7,8 Claudia Schulte, PhD,7,8 Manu Sharma, PhD,3 Lynne Krohn, MSc,10,11 Ari Sitonen, PhD,12,13 Hirotaka Iwaki, MD, PhD,1,14 Hampton Leonard, MSc,1 Alastair J. Noyce, MD, PhD,15,16 Manuela Tan, BA,16 J. Raphael Gibbs, PhD,1 Dena G. Hernandez, PhD,1, Sonja W. Scholz, MD, PhD,2 Joseph Jankovic, MD,7,17 Lisa M. Shulman, MD,5 Mike A. Nalls, PhD,1,30 and Andrew B. Singleton, PhD,1*

1Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, Maryland, USA
2Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, Maryland, USA
323andMe, Inc., Mountain View, California, USA
4Institute for Molecular Bioscience, The University of Queensland, Brisbane, Australia
5Department of Neurology, University of Maryland School of Medicine, Baltimore, Maryland, USA
6Department of Neurology, Oslo University Hospital, Oslo, Norway
7Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany
8German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany
9Centre for Genetic Epidemiology, Institute for Clinical Epidemiology and Applied Biometry, University of Tübingen, Germany
10Department of Human Genetics, McGill University, Montreal, Quebec, Canada
11Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada
12Institute of Clinical Medicine, Department of Neurology, University of Oulu, Oulu, Finland
13Department of Neurology and Medical Research Center, Oulu University Hospital, Oulu, Finland
14The Michael J Fox Foundation for Parkinson’s Research, New York, New York, USA
15Preventive Neurology Unit, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, United Kingdom
16Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, London, United Kingdom
17Parkinson’s Disease Center and Movement Disorders Clinic, Department of Neurology, Baylor College of Medicine, Houston, Texas, USA
18Inserm U1127, Sorbonne Universités, UPMC Univ Paris 06 UMR S1127, Institut du Cerveau et de la Moelle épinière, ICM, Paris, France
19Department of Neurology, Leiden University Medical Center, Leiden, The Netherlands
20Department of Neurology, Helsinki University Hospital, Medical Research Programs Unit, Biomedicum, University of Helsinki, Helsinki, Finland
21Institute of Clinical Medicine, University of Oslo, Oslo, Norway
22Department of Neurology, Institute of Neurological Sciences, Queen Elizabeth University Hospital, Glasgow, United Kingdom
23Institute of Neuroscience & Psychology, University of Glasgow, Glasgow, United Kingdom
24Departments of Molecular & Human Genetics and Neuroscience, Baylor College of Medicine, Houston, Texas, USA
25Jan and Dan Duncan Neurological Research Institute, Texas Children’s Hospital, Houston, Texas, USA
26Department of Neurodegenerative Diseases, UCL Queen Square Institute of Neurology, London, United Kingdom
27Mater Research, Translational Research Institute, Brisbane, Queensland, Australia
28Queensland Brain Institute, The University of Queensland, Brisbane, Australia
29Department of Neurology & Neurosurgery, McGill University, Montreal, Quebec, Canada
30Data Tecnica International, Glen Echo, Maryland, USA

*Correspondence to: Dr. Cornelis Blauwendraat, Laboratory of Neurogenetics, NIA, NIH, Building 35, 35 Convent Drive, Bethesda, MD 20892, USA; E-mail: cornelis.blauwendraat@nih.gov; or Dr. Andrew B. Singleton, Laboratory of Neurogenetics, NIA, NIH, Building 35, 35 Convent Drive, Bethesda, MD 20892, USA; E-mail: singletaa@mail.nih.gov

Received: 30 November 2018; Revised: 2 January 2019; Accepted: 21 January 2019

Authors C.B. and K.H. contributed equally to this work.

Relevant conflicts of interest/financial disclosures: Nothing to report.

Full financial disclosures and author roles may be found in the online version of this article.

Published online 00 Month 2019 in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/mds.27659
ABSTRACT: Background: Increasing evidence supports an extensive and complex genetic contribution to PD. Previous genome-wide association studies (GWAS) have shed light on the genetic basis of risk for this disease. However, the genetic determinants of PD age at onset are largely unknown.

Objectives: To identify the genetic determinants of PD age at onset.

Methods: Using genetic data of 28,568 PD cases, we performed a genome-wide association study based on PD age at onset.

Results: We estimated that the heritability of PD age at onset attributed to common genetic variation was ~0.11, lower than the overall heritability of risk for PD (~0.27), likely, in part, because of the subjective nature of this measure. We found two genome-wide significant association signals, one at SNCA and the other a protein-coding variant in TMEM175, both of which are known PD risk loci and a Bonferroni-corrected significant effect at other known PD risk loci, GBA, INPP5F/BAG3, FAM47E/SCARB2, and MCCC1. Notably, SNCA, TMEM175, SCARB2, BAG3, and GBA have all been shown to be implicated in α-synuclein aggregation pathways. Remarkably, other well-established PD risk loci, such as GCH1 and MAPT, did not show a significant effect on age at onset of PD.

Conclusions: Overall, we have performed the largest age at onset of PD genome-wide association studies to date, and our results show that not all PD risk loci influence age at onset with significant differences between risk alleles for age at onset. This provides a compelling picture, both within the context of functional characterization of disease-linked genetic variability and in defining differences between risk alleles for age at onset, or frank risk for disease. © 2019 International Parkinson and Movement Disorder Society

Key Words: age at onset; GBA; Parkinson’s disease; SNCA; TMEM175

Parkinson’s disease (PD) is the most common neurodegenerative movement disorder. PD is pathologically characterized by the loss of dopaminergic neurons in the SN and α-synuclein (encoded by SNCA) protein aggregates. Current estimates are that in 2015 there were 6.9 million PD patients worldwide, and this number is predicted to be 14.2 million in 2040. PD has a strong age-dependent prevalence and male-female differences, where males are ~1.5 times more likely to develop PD.2,3

The exact cause of PD is unknown; however, there is clear evidence that genetic variability plays a role in both disease development and progression. In the past two decades, mutations in several genes have been identified that cause monogenic forms of PD, accounting for around 1% to 5% of all PD cases.4 The majority of PD cases are therefore referred to as sporadic PD. Genome-wide association studies (GWAS) have successfully identified over 40 loci robustly associated with PD. Interestingly, variability in several genes, including SNCA, LRRK2, VPS13C, GCH1, and GBA, appear to play a role in both monogenic and sporadic disease.5

Substantial evidence suggests that genetic variation plays a role in age at onset (AAO), where, for example, monogenic forms of PD often present as early-onset cases, but the identity of the exact genetic component in sporadic PD remains unclear.6,7 Additionally, there is some evidence that AAO may be different between males and females.8-9 Multiple studies have nominated variants and genes of interest, but replication and large cohort studies are lacking.7,10,11 Additionally, rare coding variants in PD risk genes like GBA and LRRK2 have been associated with an earlier AAO.12-17 Currently, the best predictor of PD AAO is a genetic risk score (GRS) based on cumulative genetic PD risk,7,18,19 implying broad genetic overlap between PD susceptibility and PD AAO.

In this study, we performed the largest PD AAO GWAS to date including 28,568 PD cases. Using this large cohort, we have estimated the heritability of PD AAO and have identified several associated variants. Additionally, we have shown that the latest PD GRS remains highly correlated with of AAO.

Materials and Methods
Processing of International Parkinson’s Disease Genomics Consortium Data Sets

Genotyping data (all Illumina platform based) were obtained from International Parkinson’s Disease Genomics Consortium (IPDGC) members, collaborators, and public resources (Supporting Information Table S1). All data sets underwent quality control separately, both on individual-level data and variant-level data, before imputation. See Supporting Information Methods for a detailed description of data processing. Where possible, AAO was defined based on patient report of initial manifestation of parkinsonian motor signs (tremor, bradykinesia, rigidity, or gait impairment). Where this information was not available, age of diagnosis was used as a proxy for onset age. Note that the correlation was high between age of diagnosis and AAO.
The resulting quality-controlled and imputed data sets of PD cases (total n = 17,415) were analyzed with the formula AGE_AT_ONSET ~ SNP + SEX + PC1-PC5. Analyses were performed per data set with rvttests linear regression using imputed dosages. System Genomics of Parkinson’s Disease (SGPD) data (n = 581) was processed as above with the following minor changes: A linear regression model in PLINK was used to run a GWAS for AAO, and relatedness cutoff of 0.05 was used. The association analysis was adjusted for sex and the first 10 eigenvectors from principal component analysis.

Processing of the 23andMe Data Set

As an independent second cohort, we used the 23andMe PD data set (n = 10,572), which consisted of customers of the personal genetics company, 23andMe, Inc., who had consented to participate in research. PD patients were recruited through a targeted e-mail campaign in conjunction with the Michael J. Fox Foundation or other partners, patient groups, and clinics. PD cases were individuals who self-reported having been diagnosed with PD and who self-reported their age of PD diagnosis. We used age of diagnosis rather than age of symptom onset because symptom onset is often gradual and may be more difficult to self-report accurately. Note that the correlation was high between age of diagnosis and AAO (\(\rho = 0.917, p < 1E-300\)), and the average difference between age of diagnosis and AAO was 2.58 years. We have previously shown that self-reported PD case status is accurate with 50 of 50 cases confirmed by telemedicine interview.\(^{20}\) See Supporting Information Methods for a detailed description of data processing. After quality control, which included \(R^2 > 0.5\) variant removal, we analyzed 11,956,580 single-nucleotide polymorphisms (SNPs). Remaining PD cases (n = 10,572) were analyzed using the formula AGE_AT_DIAGNOSIS ~ SNP + SEX + PC1-PC5 + genotyping platform. The genomic inflation factor was \(\lambda = 1.015\).

Meta-Analyzing Data Set

The 17 IPDGC data sets were meta-analyzed using METAL (v.2011-03-25) fixed effects using default settings.\(^{21}\) We excluded SNPs with an \(I^2\) statistic >50% and variants that were present in fewer than 66.7% of the data sets. One genome-wide significant variant was excluded because of this filter criteria (rs2737029, \(I^2 = 58.1\%\)). This resulted in a total of 6,850,647 variants and a genomic inflation factor of \(\lambda = 1.001\). For the combined GWAS, we meta-analyzed all 17 IPDGC data sets with the 23andMe data set using the same quality-control steps.

Additional Analyses and Figures

PD GWAS loci were obtained from using Table 1 and Table 2 of Chang et al 2017\(^5\) and using PMETA from the provided summary stats (see Supporting Information Table S3). Additionally, to assess the influence of a GRS for PD case-control status on PD AAO, we obtained the GRS from the most recent PD GWAS Supplementary.\(^5\) The GRS was calculated and processed using PLINK for each individual as described previously.\(^7\) Associations were performed using the formula AGE_AT_ONSET ~ GRS + SEX + PC1-PC5. Quantile-quantile (QQ)-plots and Manhattan plots were generated using FUMA.\(^{22}\) Forest plots were generated in R using the package mmeta (https://cran.r-project.org/web/packages/mmeta/index.html). Locus plots were generated for the genome-wide significant loci and were compared to the latest published PD GWAS.\(^3\) GCTA-cojo analyses were performed to identify whether there were multiple independent signals in loci of interest using all IPDGC data sets (excluding SGPD) described in Supporting Information Table S1 as a reference panel.\(^{24,25}\) This reference panel was created using variants passing the following criteria: \(R^2 < 0.8\), minor allele frequency >0.001, Hardy-Weinberg equilibrium \(P < 1E-6\), and a maximum variant missingness of 5%. The genetic correlation between the PD AAO GWAS and the PD GWAS\(^5\) was calculated using linkage disequilibrium score regression (LDSC)\(^{26}\) using default settings for European populations. Heritability was estimated using LDSC and GCTA based on GWAS summary statistics and individual-level genotypes, respectively.

Power Calculations

We performed power calculations using the method of Brion and colleagues:\(^{27}\)

\[
\text{Power} = 1 - P(\chi^2_{NCP} > \chi^2_{0.95})
\]

where \(\chi^2_{NCP}\) is a random variable from a noncentral \(\chi^2\) with one degree of freedom, and \(\chi^2_{0.95}\) is the threshold of a central \(\chi^2\) distribution with one degree of freedom and a type 1 error rate of 0.05. The noncentrality parameter was calculated using the method of Sham and Purcell:\(^{28}\)

\[
\lambda = N \times \frac{2p(1-p) \times \beta^2}{\text{variance}(Y)}
\]

where \(\lambda\) is the noncentrality parameter, \(p\) is the minor allele frequency, \(\beta\) is the effect size in years, and \(\text{variance}(Y)\) is the variance in PD AAO. We estimated variance(\(Y\)) by taking the mean variance across each cohort, weighted by their sample size. We assessed effect sizes between 0.01 and 1 years and reported the minimum effect size that yielded >80% power.

RESULTS

Initial Data Overview

In total, we included 18 data sets: the IPDGC data contain 17 independent cohorts (\(n = 17,996\)) and the 23andMe PD data set using the same quality-control steps.
cohort (n = 10,572; see Supporting Information Table S1 for more details). The average AAO in the IPDGC data set was 62.14 (range, 20–96; standard deviation [SD] = 12.08), whereas in the 23andMe dataset the average age of diagnosis was 60.71 (range, 40–97; SD = 9.98). We found minor differences in AAO in females and males in both the IPDGC data set (females = 62.15, SD = 11.71; males = 62.03, SD = 11.95) and the 23andMe data set (females = 59.95, SD = 9.79; males = 61.20, SD = 10.07).

Heritability of PD Age at Onset

Using LDSC, the heritability of PD AAO was \( h^2 = 0.076 \) (standard error [SE] = 0.0277) in the IPDGC cohort, \( h^2 = 0.0805 \) (SE = 0.0403) in the 23andMe cohort, and \( 0.109 \) (SE = 0.0255) in the complete meta-analysis. These heritability estimates for PD AAO were similar to estimates derived using GCTA in the largest IPDGC data set (IPDGC_NeuroX, \( h^2 = 0.0798; SE = 0.0391; N = 5,428 \)) and in the 23andMe dataset (h2 = 0.1235; SE = 0.0341; N = 10,697). Heritability estimates in the other IPDGC data sets were considered less reliable because of the low number of included cases.

SNCA as Top Associated Loci With PD Age at Onset

Two loci reached genome-wide significance, SNCA and TMEM175, both of which are established PD risk loci based on PD case/control GWASes (Fig. 1). Common variation at the SNCA locus has clearly been established as a risk factor PD, and both rare mutations and whole-gene multiplications have been identified to cause monogenic PD.\(^{29,30}\) Several independent signals have been reported in this locus, where initially a variant in intron 4 was identified in PD;\(^{31}\) subsequently a second independent signal at the 5’ end was identified,\(^{32}\) which is also associated with dementia with Lewy bodies;\(^{33}\) and currently at least three independent signals are present in this locus.\(^{34}\) Here, we identified both signals to be associated with AAO (Fig. 2A). The 3’ end signal is the strongest with rs356203 as the most associated SNP, resulting in a reduction of \( \sim 0.6 \) year in AAO (P_meta = 1.90E-12; beta = -0.626; SE = 0.0890). Using conditional analysis, we also identified the 5’ end signal (Fig. 2B; rs9833631, P_conditional = 6.82E-6; beta = -0.484; SE = 0.108).

A Coding Variant in TMEM175 Is Associated With PD Age at Onset

The TMEM175/GAK locus is another known PD locus where two independent signals have been identified.\(^{32}\) Whereas it is more parsimonious to assume that a single GWAS peak is the product of a single causal gene in the locus, there is evidence that both TMEM175 and GAK may modulate PD risk.\(^{35,36}\) TMEM175 and GAK share a promoter, suggesting that they may be coregulated and may perform related cellular functions. Interestingly, the TMEM175/GAK locus is one of the few PD GWAS loci where a coding variant (TMEM175 p.M393T, exon 11, rs34311866) is among the highest associated variants (Fig. 2A). This coding variant was also the most associated variant in this locus in our PD AAO analysis (P_meta = 9.62E-9; beta = -0.613; SE = 0.107), resulting in an average reduction of \( \sim 0.6 \) years (Figs. 3B and 5). Using conditional analysis based on rs34311866, we did not identify a second independent signal in this locus (P = 0.879; beta = -0.0240; SE = 0.158; Supporting Information Fig. S9).

PD High-Risk Variants in GBA and LRRK2 and Age at Onset

LRRK2 p.G2019S and GBA variants are the most commonly identified genetic risk factors for PD. LRRK2 p.G2019S is identified in \( \sim 1\% \) of the general European PD population, but higher frequencies have been reported in other populations.\(^{37}\) GBA variants are more commonly identified in Ashkenazi Jewish populations, but are also present in the European population.\(^{38}\)

LRRK2 p.G2019S (rs34637584) has already been described to have a large range of AAO and reduced penetrance.\(^{12,14}\) Because of the low frequency and low imputation quality of this variant, it did not pass the variant-level quality control and was only tested in 9 of the 18 data sets. Similar to previous studies,\(^{37}\) we identified 140 carriers in the IPDGC data sets (\( \sim 0.8\% \) of all PD cases). The average AAO for p.G2019S carriers was 66.58 (range = 36–95; SD = 11.66), which is \( \sim 4 \) years later compared to the average AAO of noncarriers in the IPDGC data sets (62.14; range, 20–96; SD = 12.08). However, this is based on a relatively small amount of cases, and some cohorts excluded LRRK2 p.G2019S carriers pregenotyping or excluded familial PD cases. Besides, the noncoding variant at the 5’ end of LRRK2 (rs76904798), which is also identified as a risk factor PD did not show an association with AAO (P = 0.1267; beta = -0.184; SE = 0.121).

For GBA, there have been several reports that variants in GBA affect AAO in PD.\(^{39,41}\) When solely looking at the GBA locus, we identified three significant hits for the GBA region including p.N370S variant (rs76763715, P = 2.628E-6; beta = -2.600; SE = 0.553), p.E326K (rs2230288, P = 0.00123; beta = -0.929; SE = 0.287 and p.T369M (rs75548401: P = 0.01153; beta = -1.281; SE = 0.507). All resulted in relatively large reduction AAO ranging from of 2.6 to 0.9 year. Note that the p.N370S variant was only tested in 13 of the 18 data sets because of the low frequency or low imputation quality. Using conditional analysis, we did not identify other clear significant signals in the GBA gene (Supporting Information Fig. S11).
PD Genetic Risk Loci Are Associated With Age at Onset

Previously, a GRS based on PD GWAS loci has been identified as the main genetic predictor for AAO. Here, we tested the association between the updated 47-SNP GRS used in the latest PD GWAS study. For each IPDGC data set, the GRS was calculated and we identified a consistent association between the Chang and colleagues GRS and PD AAO. After meta-analyzing results from the individual cohorts, we found that each 1-SD increase in GRS led to an earlier AAO by 0.8 years (summary effect = –0.801; 95% confidence interval CI: –0.959, –0.643; I² = 11.6%; Supporting Information Fig. S12). Although this effect was highly significant, the genetic proportion explained by the GRS was ~0.56%, and the SNCA and TMEM175 loci combined explained ~0.3%, most likely attributed to the lower heritability. Furthermore, there was a significantly negative genetic correlation between the AAO GWAS summary statistics and the most recent PD GWAS using LDSC (rg = –0.5511; se = 0.103; P = 9.027E-8).

When solely looking at the SNPs that were genome-wide significant in the most recent PD GWAS, we identified a significant effect in six loci after Bonferroni correction: SNCA, TMEM175, BST1, INPP5F/BAG3, FAM47E/SCARB2, and MCCC1 (Fig. 4; Supporting Information Table S3). Interestingly, for some well-established loci, no significant P values were identified in any of the data sets, including GCH1 (P_meta = 0.9769; 80% power to detect

FIG. 1. Manhattan plot of Parkinson’s disease age at onset GWAS. Based on meta-analyses of data sets (n = 28,568) using 7,426,111 SNPs. Two genome-wide significant loci were identified: SNCA and the TMEM175/GAK. Additionally, one borderline significant locus was found, APOE. [Color figure can be viewed at wileyonlinelibrary.com]

FIG. 2. SNCA association with Parkinson’s disease age at onset. (A) Locus zoom plot of the association signal of the SNCA locus. The age at onset association signal is primarily based at the 3’ end of the SNCA gene, highly similar as the PD GWAS signal of Chang and colleagues (Supporting Information Fig. S8). (B) Conditional analysis based on the most associated SNP (rs356203) shows that there is a secondary signal at the 5’ end of the SNCA gene rs6532192, as well similar as previously reported for the PD GWAS. [Color figure can be viewed at wileyonlinelibrary.com]
a change in AOO >0.28 years), RAB7L1/NUCKS1 (P_meta = 0.124; 80% power to detect a change in AOO >0.26 years), and MAPT (P_meta = 0.745; 80% power to detect a change in AOO >0.32 years; Fig. 5; see Supporting Information Table S3 for all power calculations).

**APOE E4 Allele as Positive Control Longevity Marker**

Interestingly, the variant representing the APOE E4 allele (rs429358) was among the borderline significant loci with a P_meta value of 5.696E-8 resulting in a 0.7 year earlier of AAO (beta = −0.707; SE = 0.130). The APOE E4 allele is the most common genetic risk factor for Alzheimer’s disease, but has no effect on PD (P value in latest GWAS = 0.625). Interestingly, it has already been previously identified as an aging marker. To investigate this in our data, we performed a GWAS on the reported age of controls from the IPDGC data sets and this demonstrated a result consistent with that observed in the PD cases: P value of 1.49E-5 (effect = −0.644; SE = 0.149), suggesting that associations observed at this locus are general age-related effects (Fig. 5). Conditional tests based on rs429358 showed that the whole signal is coming from the APOE E4 allele (Supporting Information Fig. S10).

**FIG. 3.** TMEM175/GAK locus association with Parkinson’s disease age at onset and Parkinson’s disease. (A) Locus zoom plot of the association signal of the TMEM175/GAK locus with PD age at onset. The association signal is primarily based on the coding variant TMEM175 p.M393T, rs34311866. (B) Locus zoom plot of the association signal of the TMEM175/GAK locus with PD versus controls. Similarly, as in (A), the main signal is based on the coding variant TMEM175 p.M393 T, rs34311866. [Color figure can be viewed at wileyonlinelibrary.com]

**FIG. 4.** P-value and beta-value correlation plot between Parkinson’s disease GWAS and age at onset GWAS. (A) Log-transformed P values were plotted of the Chang and colleagues’ PD GWAS (x-axis) and the current age at onset GWAS (y-axis). SNPs are annotated by their closest gene. Green dots are loci that pass Bonferroni correction and red are loci that did not pass Bonferroni correction. (B) Beta values were plotted of the Chang and colleagues’ PD GWAS (x-axis) and the current age at onset GWAS (y-axis). SNPs are annotated by their closest gene. Green dots are loci that pass Bonferroni correction, and red are loci that did not pass Bonferroni correction. [Color figure can be viewed at wileyonlinelibrary.com]
Additionally, none of the PD loci in the control dataset had a significant P value after correction for multiple testing (Supporting Information Table S5).

**Replication of Previous Associated Loci and Potential Sex Effects**

In the past decade, several studies have performed PD AAO analyses and nominated many genes and variants of interest (Supporting Information Table S2). Many of these variants and genes were identified in smaller data sets and were reported with nominal P values or could not be replicated in other data sets. In our current GWAS, we replicate the findings in SNCA, GBA, and TMEM175 as described above (Supporting Information Table S2).

Additionally, some evidence exists that there might be sex differences in AAO. To investigate this, we divided the IPDGC data sets (excluding the SGPD data set) into a males-only data set (N = 11,411) and a females-only data set (N = 6,621) and performed sex-specific AAO GWASes and meta-analyses. No genome-wide-significant hits were identified in either the male or the female meta-analysis, and the effect sizes for the most recent PD GWAS variants Chang and colleagues were similar (Pearson correlation = 0.532; P = 0.00034; Supporting Information Table S4). Interestingly, we did identify a similar sex-specific effect of the COMT coding variant rs4680 (p.V158M) as reported previously: P = 0.0133 in males (effect = 0.370; SE = 0.150) and P = 0.922 in females (effect = -0.0188; SE = 0.193).46

**Discussion**

Here, we performed a genome-wide AAO analysis using 28,568 PD cases and discovered several PD loci associated with AAO. We identified two genome-wide significantly associated loci—SNCA (P_meta = 1.90E-12) and TMEM175 (P_meta = 9.62E-9)—as well as several subsignificant loci including GBA, SCARB2, and BAG3. Notably, this is the first study reporting genome-wide significance for PD GWAS loci for PD AAO. The heritability estimate of PD AAO was ~0.11, which is much less than the heritability of PD case-control status ~0.27.47 Our heritability estimate was lower than an
Several previous studies have performed PD AAO analyses in smaller data sets. We were unable to replicate the majority of these previous reported variants, likely because of the lack of power to detect a true signal in the smaller studies (<2,000 samples; see Supporting Information Table S2). We may have been unable to replicate previous work by Hill-Burns and colleagues, given that much of their analysis was restricted to familial PD cases. Their analysis of sporadic PD included almost 4,000 cases, but unlike our study and several others, they did not find significant associations with SNCA (rs356203, P = 0.355) or TMEM175 (rs34311866, P = 0.525).

Although we included a very large amount of data, there were still several limitations to our study. First, we had limited power to detect rare variants that modulate PD AAO because of the use of genotyping arrays and large number of cohorts with a relatively small size. Several reports have shown that rare variants may influence AAO of PD. For example, a recent report showed that rare variants in LRRK2 lowered AAO. This likely also explains the relatively moderate P values for rare coding variants in GBA and LRRK2 on age onset.

Second, because of the lack of reliable genotype data, we excluded chromosomes X and Y from our analysis. Both of these limitations could be addressed by using genome sequencing data, and large genome sequencing projects are currently underway. Third, the heritability of PD AAO (~0.11) was much lower than the heritability of PD case-control status, which, in part, explains the lack of genome-wide significantly associated loci in our study. Fourth, there was a reasonable amount of heterogeneity in the mean AAO of the different cohorts. Age of PD diagnosis was self-reported in some of our cohorts and assessed by physicians in other cohorts. Some cohorts were specifically designed to include younger-onset cases. Other cohorts provided age of diagnosis whereas others used AAO, although these are likely to be highly correlated (r > 0.9 in our data). By analyzing each cohort separately, we were able to mitigate many of these interstudy sources of variation. Indeed, there was relatively little heterogeneity of SNP effect-size estimates between studies (Supporting Information Table S3). However, future data collection and studies would probably benefit from more specific and more predefined structured and symptom-specific AAO diagnostic criteria. Implementation of such criteria in large studies and cohorts or even in health care systems could significantly improve the understanding of the genetics of PD AAO.

Overall, we have performed the largest AAO of PD GWAS to date, and our results reveal an interesting and significant genetic component. Our results show that not all PD risk loci influence AAO with significant differences between risk alleles for AAO, which implies different mechanisms for risk for developing PD and PD AAO. This provides a compelling picture, both within the context of functional characterization of disease-linked genetic variability and in defining differences between risk alleles for AAO, or frank risk for another pathway by a different mechanism, but more experiments are needed to confirm this.

It has been shown that PD case-control GWAS loci are associated with PD AAO. We found that a GRS based on the latest PD GWAS was also highly associated with an earlier age of PD onset (Supplementary Figure 12). Furthermore, there was a significant negative genetic correlation between the PD case-control GWAS and the PD AAO GWAS. Six individual loci remained significant after Bonferroni correction, including two well-established PD loci—SNCA and TMEM175—and four less-studied PD loci—BST1, INPP5F/BAG3, FAM47E/SCARB2, and MCCC1. Interestingly, no effect was identified for several well-established loci, including GCH1, RAB7L1 (PARK16), and MAPT (Supporting Information Table S3). For these three loci, our study had 80% power (at P = 0.05) to detect changes greater than 0.28, 0.26, and 0.32 years, respectively. For comparison, the two genome-wide significant hits in SNCA and TMEM175 modified AAO by ~0.6 year. We believe this provides compelling evidence that only a subset of PD case-control risk SNPs also modulates PD AAO, and notably that variability at MAPT, a major GWAS identified risk factor for PD, does not influence AAO. We believe these data may suggest that different PD loci modulate risk through different pathways and that this likely has therapeutic consequences.

Interestingly, three of our top hits (SNCA, TMEM175, and GBA) are strongly implicated in α-synuclein mechanisms of PD pathogenesis pathology. SNCA encodes for the α-synuclein protein, which is the major constituent of Lewy bodies, the defining pathology of PD. Both common and rare variants at SNCA are associated with increased PD risk, including duplications or triplications of the genomic locus. One of the distinguishing features of autosomal-dominant PD caused by SNCA gene multiplications is an early AAO. The common variants at SNCA associated with PD risk and AAO have also been demonstrated to enhance SNCA gene expression. This, based on the genetics of both Mendelian and sporadic PD, SNCA dosage appears to be a major driver of disease onset age. TMEM175 has been shown to impair lysosomal and mitochondrial function and increases α-synuclein aggregation. GBA has been shown to be involved in several α-synuclein aggregation mechanisms, including autophagy and enhancing α-synuclein cell-to-cell transmission.

Other loci identified by our PD AAO analyses have also been linked to α-synuclein clearance, for example for SCARB2, encoding the endoplasmic reticulum–lysosome transporter of GBA, and BAG3. Taken together, it is tempting to speculate about the direct link between PD AAO loci influencing α-synuclein accumulation and clearance pathways and the other nonsignificant PD loci acting in another pathway by a different mechanism, but more experiments are needed to confirm this.

GWAS. Six individual loci remained signiﬁcant after Bonferroni correction, including two well-established PD loci—SNCA and TMEM175—and four less-studied PD loci—BST1, INPP5F/BAG3, FAM47E/SCARB2, and MCCC1. Interestingly, no effect was identified for several well-established loci, including GCH1, RAB7L1 (PARK16), and MAPT (Supporting Information Table S3). For these three loci, our study had 80% power (at P = 0.05) to detect changes greater than 0.28, 0.26, and 0.32 years, respectively. For comparison, the two genome-wide significant hits in SNCA and TMEM175 modified AAO by ~0.6 year. We believe this provides compelling evidence that only a subset of PD case-control risk SNPs also modulates PD AAO, and notably that variability at MAPT, a major GWAS identified risk factor for PD, does not influence AAO. We believe these data may suggest that different PD loci modulate risk through different pathways and that this likely has therapeutic consequences.

Interestingly, three of our top hits (SNCA, TMEM175, and GBA) are strongly implicated in α-synuclein mechanisms of PD pathogenesis pathology. SNCA encodes for the α-synuclein protein, which is the major constituent of Lewy bodies, the defining pathology of PD. Both common and rare variants at SNCA are associated with increased PD risk, including duplications or triplications of the genomic locus. One of the distinguishing features of autosomal-dominant PD caused by SNCA gene multiplications is an early AAO. The common variants at SNCA associated with PD risk and AAO have also been demonstrated to enhance SNCA gene expression. This, based on the genetics of both Mendelian and sporadic PD, SNCA dosage appears to be a major driver of disease onset age. TMEM175 has been shown to impair lysosomal and mitochondrial function and increases α-synuclein aggregation.

GBA has been shown to be involved in several α-synuclein aggregation mechanisms, including autophagy and enhancing α-synuclein cell-to-cell transmission.

Other loci identified by our PD AAO analyses have also been linked to α-synuclein clearance, for example for SCARB2, encoding the endoplasmic reticulum–lysosome transporter of GBA, and BAG3. Taken together, it is tempting to speculate about the direct link between PD AAO loci influencing α-synuclein accumulation and clearance pathways and the other nonsignificant PD loci acting in another pathway by a different mechanism, but more experiments are needed to confirm this.
IPDGC GWAS summary statistics are available on the IPDGC website (http://pdgenetics.org/resources), and 23andMe GWAS summary statistics will be made available to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please visit http://research.23andme.com/collaborate/#publication for more information and to apply to access the data.

References


Supporting Data

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