

Multi-ancestry genome-wide meta-analysis in Parkinson's disease

Jonggeol Jeffrey Kim^{1, 2#}, Dan Vitale^{1, 3, 4#}, Diego Véliz Otani^{5, 6#}, Michelle Lian^{7, 8#}, Karl Heilbron⁹, the 23andMe Research Team⁹, Hirotaka Iwaki^{1, 3, 4}, Julie Lake¹, Caroline Warly Solsberg^{10, 11, 12}, Hampton Leonard^{1, 3, 4}, Mary B. Makarios^{1, 14, 15}, Eng-King Tan¹⁶, Andrew B. Singleton^{1, 4}, Sara Bandres-Ciga^{1, 4}, Alastair J Noyce², Cornelis Blauwendraat^{1, 4&}, Mike A. Nalls^{1, 3, 4&}, Jia Nee Foo^{7, 8&}, Ignacio Mata^{13&}

Affiliations

1. Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
2. Preventive Neurology Unit, Centre for Prevention Diagnosis and Detection, Wolfson Institute of Population Health, Queen Mary University of London, UK
3. Data Tecnica International, Glen Echo, MD, USA
4. Center for Alzheimer's and Related Dementias, National Institutes of Health, Bethesda, MD, USA
5. Neurogenetics Research Center, Instituto Nacional de Ciencias Neurológicas, 1271 Ancash St., Barrios Altos, Lima, Peru
6. Institute for Genome Sciences, University of Maryland, Baltimore, Baltimore, MD, USA
7. Lee Kong Chian School of Medicine, Nanyang Technological University Singapore, Singapore, Singapore
8. Laboratory of Neurogenetics, Genome Institute of Singapore, A*STAR, Singapore, Singapore
9. 23andMe, Inc., Sunnyvale, CA, USA
10. Pharmaceutical Sciences and Pharmacogenomics, UCSF, San Francisco, CA, USA.
11. Department of Neurology and Weill Institute for Neurosciences, University of California, San Francisco, San Francisco, CA, USA
12. Memory and Aging Center, UCSF, San Francisco, CA, USA
13. Genomic Medicine, Lerner Research Institute, , Cleveland Clinic Foundation, Cleveland, OH, USA
14. Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, London, UK
15. UCL Movement Disorders Centre, University College London, London, UK
16. Department of Neurology, National Neuroscience Institute, Duke NUS Medical School, Singapore

denotes co-first authorship

& denotes co-corresponding authorship

Keywords: Parkinson's disease, meta-analysis, multi-ancestry, genetics, genomics, population genetics

Abstract

Although over 90 independent risk variants have been identified for Parkinson's disease using genome-wide association studies, all studies have been performed in just one population at the time. Here we performed the first large-scale multi-ancestry meta-analysis of Parkinson's disease with 49,049 cases, 18,785 proxy cases, and 2,458,063 controls including individuals of European, East Asian, Latin American, and African ancestry. In a single joint meta-analysis, we identified 78 independent genome-wide significant loci including 12 potentially novel loci (*MTF2*, *RP11-360P21.2*, *ADD1*, *SYBU*, *IRS2*, *USP8:RP11-562A8.5*, *PIGL*, *FASN*, *MYLK2*, *AJ006998.2*, *Y_RNA*, *PPP6R2*) and finemapped 6 putative causal variants at 6 known PD loci. By combining our results with publicly available eQTL data, we identified 23 genes near these novel loci whose expression is associated with PD risk. This work lays the groundwork for future efforts aimed at identifying PD loci in non-European populations.

Introduction

Parkinson's disease (PD) is a neurodegenerative disease pathologically defined by Lewy body inclusions in the brain and the death of dopaminergic neurons in the midbrain. The identification of genetic risk factors is imperative to mitigating the global burden of PD, one of the fastest growing age-related neurodegenerative diseases. The largest PD Genome-Wide Association Study (GWAS) meta-analysis to date uncovered 90 independent genetic risk variants in individuals of European ancestry¹. Similarly, large-scale PD GWAS meta-analyses of East Asian² and a single GWAS of Latin American³ individuals have each identified 2 risk loci that were not previously identified in Europeans. For PD, there are now large-scale efforts to sequence and analyze genomic data in under-represented populations with the goal of both identifying novel associated loci, fine mapping known loci, and addressing the inequality that exists in current precision medicine efforts^{4,5}. Here we performed a large-scale multi-ancestry GWAS meta-analysis of PD by including individuals from four ancestral populations. This effort can serve as a guide for future genetic analyses to increase ancestral representation.

Methods

In brief, we used a single joint meta-analysis study design (Figure 1) due its efficiency over a two-stage GWAS design⁶. All datasets underwent quality control filters and were harmonized to hg19. We used datasets representing four different ancestry groups: European, East Asian, Latin American, and African. In addition to previously described results from European¹, East Asian², and Latin American³ cohorts, we also used FinnGen and additional datasets for East Asian, Latin American and African cohorts from 23andMe, Inc. All multi-ancestry meta-analyses were conducted using a union of a random-effects model and a multi-ancestry meta-regression model using PLINK 1.9⁷ and MR-MEGA⁸. MR-MEGA has the greatest power to detect heterogeneous effects across the different cohorts, while the random-effects model has greater power to detect homogenous allelic effects⁸. In total, we included 49,049 PD cases, 18,618 proxy cases (first-degree relative with PD), and 2,458,063 neurologically-healthy controls.

Bonferroni-adjusted alpha for genome-wide significant variants was set to a more stringent 5×10^{-9} for all multi-ancestry meta-analyses to account for the larger number of haplotypes resulting from the ancestrally diverse datasets⁹. Fine-mapping was also performed using MR-MEGA, which uses ancestry heterogeneity to increase fine-mapping resolution.

We ran our meta-analysis results through the Proportion of Population-specific and Shared Causal Variants (PESCA)¹⁰ software to estimate the population specific and shared burden on the variants found in the meta-analysis. Only East Asian and European populations were compared due to the smaller sample size of the other populations. We used the Functional Mapping and Annotation (FUMA) software^{11,12} to functionally annotate the random effect results and run Multi-marker Analysis of GenoMic Annotation (MAGMA)¹³ for gene-ontology, tissue-level, and single-cell expression data. We also searched our novel loci through GTEx v8 brain tissue eQTLs and multi-ancestry eQTL meta-analysis of the brain¹⁴ to correlate novel loci with gene expression data, then searched the significant-eQTL genes and genes near the loci with previously completed summary-based mendelian randomization (SMR)¹⁵ results to correlate said genes with PD risk. Detailed methods are available in the Online Methods and the Data and Code Availability sections.

Results

Multi-ancestry meta-analyses confirm 62 known loci and nominate 12 novel loci

All together we found 12 novel PD risk loci and 66 hits in known risk loci from single ancestry GWAS (Supplementary Table S2). 9 of the novel loci found in the random-effect method showed homogeneous effects across the different ancestries. The 3 loci identified exclusively in MR-MEGA showed ancestrally heterogeneous effects. All three loci showed evidence for ancestral heterogeneity ($P_{\text{ANC-HET}} < 0.05$) but no significant residual heterogeneity ($P_{\text{RES-HET}} > 0.148$), supporting the idea that the signals are due to population structural differences rather than other confounding factors. The *IRS2* locus ($P_{\text{ANC-HET}} = 5.3 \times 10^{-3}$) shows the biggest departure in allelic effects in the Finnish cohort (Supplementary Figure 4). The *MYLK2* locus has the main European cohort at odds with the Latin American and Finnish cohorts ($P_{\text{ANC-HET}} = 0.035$). While this is a novel single-trait GWAS locus, its lead SNP was previously discovered as a potential pleiotropic locus in a multi-trait conditional/conjunctional false discovery rate (FDR) study between schizophrenia and PD¹⁶. Lastly, the *AJ006998.2* locus had the most significant ancestral heterogeneity ($P_{\text{ANC-HET}} = 4.74 \times 10^{-5}$) and its effects were specific to European and African cohorts.

We found 17 suggestive loci that failed to meet the significance threshold at random-effect but met the more common threshold of $P < 5 \times 10^{-8}$ in fixed-effect and $P < 1 \times 10^{-6}$ in random-effect (Supplementary Table S2). 14 of these regions qualified as novel loci. Two loci near *RP11-182I10.3* and *HS1BP3* were exclusively found in the 23andMe Latin American and African cohorts. The lead SNPs (rs578139575, rs73919910) for these loci are very rare in European

populations but more common in Africans and Latin Americans (gnomAD v3.1.2 EUR: 0.0001616, 0.002307; AFR: 0.01637, 0.08837; AMR: 0.004063, 0.01905). If confirmed, these loci would confer a large amount of PD protective effect (beta: -1.3, -0.54). These loci merit further studies in the African and Latino populations.

Fine mapping finds 6 putative SNPs in 6 loci

6 putative causal variants were found in 6 loci in *TMEM163*, *TMEM175*, *SNCA*, *HIP1R*, *MAPT*, and *LSM7* (Table 4). All putative causal variants had posterior probabilities (PP) of causality > 0.99 and overlapped with known PD loci. Our results confirmed the previous results of the *TMEM175* M393T coding variant as the likely causal variant¹⁷. We reported only single-variant 99% credible sets as MR-MEGA assumes a single causal variant in a GWAS signal. 99% credible sets in 57 additional loci can be found in Supplementary Table S3.

MAGMA Gene Ontology analysis finds enrichment in brain and pituitary tissues

Out of 15,479 gene ontology sets in MsigDB c5, 111 gene sets were significantly enriched among the random effect hits, a significant increase from previous 11 gene sets in Nalls et al. 2019. Out of the 11, only 3 gene ontology terms were replicated in the multi-ancestry: vacuole ($P_{FDR} = 6.9 \times 10^{-4}$), vacuolar part ($P_{FDR} = 0.009$), and vacuolar membrane ($P_{FDR} = 0.01$). At the tissue level the genes of interest were enriched in all brain cells as well as pituitary tissue (Supplementary Figure 9), consistent with the results from Nalls et al. 2019¹.

When analyzing single-cell RNAseq data, there was no expression enrichment across 88 brain cell types in mouse brain data DropViz¹⁸ (Supplementary Figure 13). When narrowed down to substantia nigra tissue, glutamate receptor (*Grin2c*) in neurons were enriched (Supplementary Figure 13). In human midbrain data¹⁹ oculomotor and trochlear nucleus (OMTN), dopaminergic (DA1), neuroblast GABAergic (NbGABA), and GABA were enriched (Supplementary Figure 13).

Out of 15,479 gene ontology sets in MsigDB c5, 110 gene sets were significantly enriched among the random effect hits (Supplementary Table S6), a significant increase from previous 11 gene sets in Nalls et al. 2019. Out of the 11, only 3 gene ontology terms were replicated in the multi-ancestry: vacuole ($P_{FDR} = 6.9 \times 10^{-4}$), vacuolar part ($P_{FDR} = 0.009$), and vacuolar membrane ($P_{FDR} = 0.01$). At the tissue level the genes of interest were enriched in all brain cells as well as pituitary tissue (Supplementary Figure 9), consistent with the results from Nalls et al. 2019¹.

eQTL and SMR nominate 23 putative genes near novel loci

When comparing the SNPs in novel loci with multi-ancestry brain eQTL results¹⁴, 28 genes showed high levels of association (Supplementary Figures 8). SMR found 23 genes near four novel loci associated with PD risk. Interestingly, *PPP6R2* and *CENPV* expression changes in

substantia nigra were associated with PD risk. *PPP6R2* or protein phosphatase 6 regulatory subunit 2 is a regulatory protein for Protein Phosphatase 6 Catalytic Subunit (*PPP6C*), which is involved in the vesicle-mediated transport pathway. *CENPV* or Centromere Protein V is involved in centromere formation and cell division. Detailed results for eQTL and the SMR can be found in Supplementary Table S8, S9, and S10.

Discussion

This is the first large-scale GWAS meta-analysis of PD that incorporates multiple diverse ancestry populations. From the joint cohort analysis, we were able to confirm 65 independent known risk loci and identify 12 potentially novel risk loci, 9 with homogeneous effect and 3 with heterogeneous effect across the different cohorts. We further found 17 suggestive loci at the more traditional method of fixed-effect at 5×10^{-8} and fine-mapped 11 putative SNPs in 11 of the known loci by leveraging the diverse ancestry populations. We highlighted tissues and cell types involved with the loci, which were consistent with previous findings¹. Finally we used SMR to nominate 23 putative genes near our novel loci.

Novel loci contain genes previously implicated with PD. The *MTF2* and *PPP6R2* loci contain genes *TMED5* and *PPP6R2*, which intersect with golgi body²⁰ and vesicular transport pathways^{21,22}, both implicated in PD pathogenesis^{23–28}. Because *substantia nigra* deterioration is a hallmark pathogenic feature of PD, *PPP6R2*, and *CENPV* expression in *substantia nigra* and their association with PD merit additional investigation. Within a known locus, a new independent signal was found in *RILPL1* (rs28659953), which interacts with LRRK2-phosphorylated Rab10 to block primary cilia generation²⁹. All potentially novel PD loci identified in this analysis will require additional replication and functional validation to elucidate their role in PD pathogenesis. Previous findings in European populations found that the known variants explained 16–36% of the heritable risk¹, and while we did not test for genetic liability of the novel loci, they have been added to the list of variants that explain potential heritable PD risk.

We found that 26 of the 68 detected known PD loci have nominal ancestral heterogeneity ($P_{\text{ANC-HET}} < 0.05$) and 5 significant after bonferroni correction ($P_{\text{ANC-HET}} < 0.05/4,512$ significant SNPs) (Supplementary Table S2). This heterogeneity may be caused by significant differences in allele effect sizes and frequencies between the different ancestry populations and thus should be studied as potential loci with ancestrally divergent risk. 18 known loci from single-ancestry GWASes failed to replicate (Supplementary Table S4) which implies that the variants associated in these loci may only be relevant in certain populations. However, it is worth noting that there are large differences in statistical power for each included ancestry. Additional population-specific loci will likely reach significance when larger sample sizes are available for non-European datasets.

Our fine mapping isolated several putative causal variants in previously discovered loci. *TMEM175*-rs34311866 has been previously identified as functionally relevant to PD risk¹⁷, which is consistent with our fine-mapping results. While the fine-mapping results provided by MR-MEGA are sufficient to identify putative causal variants for loci driven by one independent signal, multiple variants in a locus can contribute to complex traits. The additive and epistatic effects of multiple causal variants in a locus can be difficult to interpret when the effects associated with each independent signal are small.

Although this is the largest multi-ancestry PD meta-analysis GWAS to date, the European population is still overrepresented in this analysis. Around 79% of effective PD cases are of European descent. Individuals of African descent were particularly underpowered at just 0.5% of the effective PD cases. The discoveries in our study warrant future efforts to expand the studies in more diverse populations. The Global Parkinson's Genetics Program (GP2) is partnering with institutions that care for underrepresented populations to generate data for these underserved communities all over the world⁵, and we will continue the ongoing analysis as more participants are genotyped. Just as the first PD GWASs failed to identify significant signals^{30,31}, we are confident that future diverse-ancestry GWAS will produce impactful association results as sample sizes increase. Further efforts in multi-ancestry and non-European GWAS will identify loci that are more relevant to the global population and will continue to facilitate fine mapping efforts to identify the genetic variants that drive these associations.

Data and code availability

The analysis pipeline code is available on GP2 github: <https://github.com/GP2code/GP2-WorkingGroups/tree/main/CD-DAWG-Complex-Data-Analysis/PROJECTS/GP2-Multiancestry-PD>

GWAS summary statistics for 23andMe datasets (post-Chang and data included in Chang et al. 2017 and Nalls et al. 2014) will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please visit research.23andme.com/collaborate/#publication for more information and to apply to access the data. An immediately accessible version of the summary statistics is available here <https://drive.google.com/file/d/1TmDZNFgyQvsOZ0xu-aZmBpVCpeUUa0UX/> excluding Nalls et al. 2014, 23andMe post-Chang et al. 2017 and Web-Based Study of Parkinson's Disease (PDWBS) but including all analyzed SNPs. After applying with 23andMe, the full summary statistics including all analyzed SNPs and samples in this GWAS meta-analysis will be accessible to the approved researcher(s).

Acknowledgements

M.A.N.'s research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging (NIA), National Institutes of Health, Department of Health and Human Services; project number ZO1 AG000535, as well as the National Institute of Neurological Disorders and Stroke (NINDS), and 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.

I.F.M.'s research is supported by grants from the National Institute of Neurological Disorders and Stroke under (R01NS112499), the Parkinson's Foundation (Stanley Fahn Junior Faculty Award and an International Research Grants Program award), the Michael J Fox Foundation,

Aligning Science Across Parkinson's Global Parkinson's Genetic Project (ASAP-GP2) and the American Parkinson's Disease Association.

E.K.T. is supported by a grant from the National Medical Research Council Singapore, Open Fund Large Collaborative Grant (MOH-OFLCG-000207). J.N.F. is supported by the Singapore National Research Foundation Fellowship (NRF-NRFF2016-03) and the National Medical Research Council Singapore, Open Fund Individual Research Grant (MOH-000559).

A.J.N. reports grants from Parkinson's UK, Barts Charity, Cure Parkinson's, NIHR, Innovate UK, Virginia Keiley benefaction, Alchemab, Aligning Science Across Parkinson's Global Parkinson's Genetics Program (ASAP-GP2) and Michael J Fox Foundation.

This research has been conducted using the UK Biobank Resource under Application Number 33601. We want to acknowledge the participants and investigators of FinnGen study.

We thank the research participants and employees of 23andMe. The following members of the 23andMe Research Team contributed to this study:

Stella Aslibekyan, Adam Auton, Elizabeth Babalola, Robert K. Bell, Jessica Bielenberg, Katarzyna Bryc, Emily Bullis, Paul Cannon, Daniella Coker, Gabriel Cuellar Partida, Devika Dhamija, Sayantan Das, Sarah L. Elson, Nicholas Eriksson, Teresa Filshtein, Alison Fitch, Kipper Fletez-Brant, Pierre Fontanillas, Will Freyman, Julie M. Granka, Karl Heilbron, Alejandro Hernandez, Barry Hicks, David A. Hinds, Ethan M. Jewett, Yunxuan Jiang, Katelyn Kukar, Alan Kwong, Keng-Han Lin, Bianca A. Llamas, Maya Lowe, Jey C. McCreight, Matthew H. McIntyre, Steven J. Micheletti, Meghan E. Moreno, Priyanka Nandakumar, Dominique T. Nguyen, Elizabeth S. Noblin, Jared O'Connell, Aaron A. Petrakovitz, G. David Poznik, Alexandra Reynoso, Madeleine Schloetter, Morgan Schumacher, Anjali J. Shastri, Janie F. Shelton, Jingchunzi Shi, Suyash Shringarpure, Qiaojuan Jane Su, Susana A. Tat, Christophe Toukam Tchakouté, Vinh Tran, Joyce Y. Tung, Xin Wang, Wei Wang, Catherine H. Weldon, Peter Wilton, Corinna D. Wong

This research was funded by Aligning Science Across Parkinson's through the Michael J. Fox Foundation for Parkinson's Research (MJFF).

Potential Conflict of Interest

K.H. and members of the 23andMe Research Team are employed by and hold stock or stock options in 23andMe, Inc. M.A.N.'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. A.J.N. reports consultancy and personal fees from AstraZeneca, AbbVie, Profile, Roche, Biogen, UCB, Bial, Charco Neurotech, uMedeor, Alchemab, and Britannia, outside the submitted work.

Online Methods

Study Design and Cohort Descriptions

We used a single joint meta-analysis study design to maximize statistical power⁶. We used datasets representing four different ancestry groups: European, East Asian, Latin American, and African. The meta-analysis included 49,049 PD cases, 49,049 PD proxy cases (participant with a parent with PD) and 2,458,063 neurologically normal controls (Table 1, Supplementary Table S1). GWAS results of European¹, East Asian², and Latin American³ populations were previously reported. African dataset as well as the additional Latin American and East Asian PD GWAS summary statistics were provided by 23andMe. The Finnish PD GWAS summary statistics was acquired from FinnGen Release 4 (G6_PARKINSON_EXMORE). For the FinnGen data, we chose the endpoint “Parkinson’s Disease (more controls excluded)” (G6_PARKINSON_EXMORE), which excludes control participants with psychiatric diseases or neurological diseases. While some FinnGen GWAS results also include UK Biobank participants, our FinnGen data did not include any UK Biobank participants.

23andMe diverse ancestry data

All self-reported PD cases and controls from 23andMe provided informed consent and participated in the research online, under a protocol approved by the external AAHRPP-accredited IRB, Ethical & Independent Review Services (E&I Review). Participants were included in the analysis on the basis of consent status as checked at the time data analyses were initiated. The name of the IRB at the time of the approval was Ethical & Independent Review Services. Ethical & Independent Review Services was recently acquired, and its new name as of July 2022 is Salus IRB (<https://www.versiticlinicaltrials.org/salusirb>). Samples were genotyped on one of five genotyping platforms: V1 and V2, which are variants of Illumina HumanHap550+ BeadChip; V3, Illumina OmniExpress+ BeadChip; V4, Illumina custom array that includes SNPs overlapping V2 and V3 chips; or V5, Illumina Infinium Global Screening Array. For inclusion, samples needed a minimal call rate of 98.5%. Genotyped samples were then phased using either Finch or Eagle2³² and imputed using Minimac3 and a reference panel of 1000 Genomes Phase III³³ and UK10K data³⁴. For this replication study, samples were classified as African, East Asian, or Latino using a genotype-based pipeline³⁵ consisting of a support vector machine and a hidden Markov model, followed by a logistic classifier to differentiate Latinos from African Americans. Unrelated individuals were included in the analysis, as determined via identity-by-descent (IBD). Variants were tested for association with PD status using logistic regression, adjusting for age, sex, the first five PCs, and genotyping platform. Reported p-values were from a likelihood ratio test.

Multi-ancestry Meta-Analysis

We performed a multi-ancestry meta-analysis of GWAS results using Meta-Regression of Multi-Ethnic Genetic Association (MR-MEGA)⁸ and PLINK 1.9. MR-MEGA performs a meta-

regression by generating axes of genetic variation for each cohort, which are then used as covariates in the meta-analysis to account for differences in population structure. While MR-MEGA was able to generate 4 principal components as axes of genetic variation, 3 principal components visibly separated the super population ancestries and explained 98% of the population variance (Supplementary Figure 7). Therefore we used 3 principal components to minimize overfitting. MR-MEGA has reduced power to detect associations for variants with homogeneous effects across populations. It is therefore recommended to run MR-MEGA alongside another meta-analysis method. PLINK 1.9 was used to perform random-effect (RE) meta-analysis to detect homogenous allelic effects.

Before the analysis, all datasets were harmonized to genome build hg19 using CrossMap³⁶ and all variants were filtered by imputation score ($r^2 > 0.3$) and minor allele frequency (MAF ≥ 0.001). Only autosomal variants were kept in the final results. In total 20,590,839 variants met the inclusion criteria. Only 5,662,641 SNPs were analyzed in the MR-MEGA analysis due to cohort-number requirements. Bonferroni-adjusted alpha was set to a more stringent 5×10^{-9} for all multi-ancestry meta-analyses to account for the larger number of haplotypes resulting from the ancestrally diverse datasets⁹. Genomic inflations were measured for all cohorts and the meta-analysis. Inflation for cohorts with large discrepancy between the case and control numbers were normalized to 1000 cases and 1000 controls. All inflation was nominal and below 1.02 (Supplementary Figure 3, Supplementary Table S1). No genomic control was applied prior to meta-analysis.

We identified genomic risk loci within our meta-analysis results using Functional Mapping and Annotation (FUMA) v1.3.8^{11,12}. In brief, FUMA first identifies independent significant SNPs in the GWAS results by clumping all significant variants with the r^2 threshold < 0.6 , then a locus is defined by merging linkage disequilibrium (LD) blocks of all independent significant SNPs within 250kb of each other. Start and end of a locus is defined by identifying SNPs in LD with the independent significant SNPs ($r^2 \geq 0.6$) and defining a region that encompasses all SNPs within the locus. The 1000 Genome reference panel with all ancestries were used to calculate the r^2 .

To determine if any associated loci in the meta-analysis were not previously identified, all significant SNPs were compared to the 92 known PD risk variants found in the previous two major meta-analyses^{1,2}. Two variants identified in the Latin American admixture population³ could not be replicated as the variants and their proxies were removed during quality control of the discovery cohort. When a genomic risk loci was scanned for significant variants in previous European and East Asian results. If the locus contained a significant hit in either results, the locus was annotated with the closest previously nominated risk SNP within 250kb. Any loci that did not have previously significant variants or failed the annotation step was considered a novel hit.

Fine-Mapping

Fine-mapping was performed using MR-MEGA⁸, which approximates a single-SNP Bayes' factor in favor of association. This is reported as the natural log of Bayes' factor (lnBF) per SNP in the MR-MEGA meta-analysis summary statistics. SNPs were selected at meta-GWAS

significance level ($P < 5 \times 10^{-9}$). Posterior probabilities (PPs) of driving the association signal at each locus were calculated from the Bayes' factor as follows:

$$\pi_j = \frac{\Lambda_j}{\sum_{j=1}^n \Lambda_j}$$

Where Λ_j is the Bayes' factor of the j th SNP within a locus with n number of SNPs. Credible set of SNPs with PP (π_j) greater than 0.99 were accepted as putative causal variants.

Putative variant burden estimation by population

Proportion of population-specific and shared causal variants (PESCA)¹⁰ was used to estimate the population-specific and shared burden of the variants tested in the meta-analysis. In brief, genome-wide heritability was estimated for the European and East Asian GWAS summary statistics using LD Score regression (LDSC)^{37,38}. Summary statistics of both populations were intersected with common variants with the 1000 Genome reference panels provided by PESCA, which have already been LD pruned ($R^2 > 0.95$) and low frequency SNPs removed (minor allele frequency < 0.05). The intersected variants were further split according to independent LD regions from the European and East Asian populations. The genome-wide prior probabilities of population-specific and shared causal variants were calculated using default parameters or as otherwise recommended by PESCA, then the results were used to calculate the PP for each variant. When the lead SNP was unavailable in the results, proxy variants ($r^2 > 0.8$) were used to approximate the PP for each variant. Other cohorts were not included due to sample size constraints for this method.

Functional Annotation

Functional annotation of the discovery results utilizing publicly available gene expression and ontology data was done using FUMA v1.3.8^{11,12}. The summary statistics were annotated by ANNOVAR³⁹ through the FUMA platform. Our meta-analysis results were analyzed using Multi-marker Analysis of GenoMic Annotation (MAGMA)¹³ to compare gene expression data from gene sets and tissues in GTEx v8⁴⁰. FUMA tested 15,479 gene sets and gene ontology terms from MsigDB v7⁴¹ as well as single-cell RNAseq expression data from mouse brain samples in DropViz¹⁸ and human ventral midbrain samples¹⁹. Test parameters were set to default. Results were adjusted for multiple tests using Benjamini–Hochberg False Discovery Rate (FDR) correction with the alpha of 0.05. Additional pathway analyses of genes mapped by FUMA SNP2GENE were performed through GENE2FUNC with default parameters.

SNPs in the novel loci were searched in multi-ancestry brain eQTL meta-analysis results¹⁴. We used a P-value cutoff of 10^{-6} as previously described¹⁴. eQTL and GWAS comparison plots were generated using LocusCompare⁴². Multi-SNP Summary-Based Mendelian Randomization (SMR) was used to test if DNA methylation and/or RNA expression of genes near the novel loci were associated with PD risk¹⁵. The nearest genes from the lead SNPs, significant genes in the multi-ancestry meta-analysis brain eQTL results, and significant genes in GTEx v8 brain tissue were chosen for SMR. In total 44 genes near the novel loci were searched in a list of previously completed PD SMR results^{15,43–51}. The association signals were

adjusted using FDR correction with the alpha of 0.05 and all signals with $P_{\text{HEIDI}} > 0.05$ were removed due to heterogeneity.

Figures/Tables

Table 1. Cohort Description

Study	Ancestral Population	Cases/Proxy/Controls
Nalls et al. 2019	European (EUR)	37,688 / 18,618 / 1,411,006
Foo et al. 2020	East Asian (EAS)	6,724 / 0 / 24,851
LARGE-PD ³	Latin American (AMR)	807 / 0 / 690
FinnGen Release 4	European-Finnish (EUR)	1,587 / 0 / 94,096
23andMe - African	African (AFR)	288 / 0 / 193,985
23andMe - East Asian	East Asian (EAS)	322 / 0 / 151,905
23andMe - Latino	Latin American (AMR)	1,633 / 0 / 583,163
Multi-ancestry meta-analysis		49,049 / 18,618 / 2,458,063



Figure 1 - Study design of Multi-Ancestry Meta-Analysis

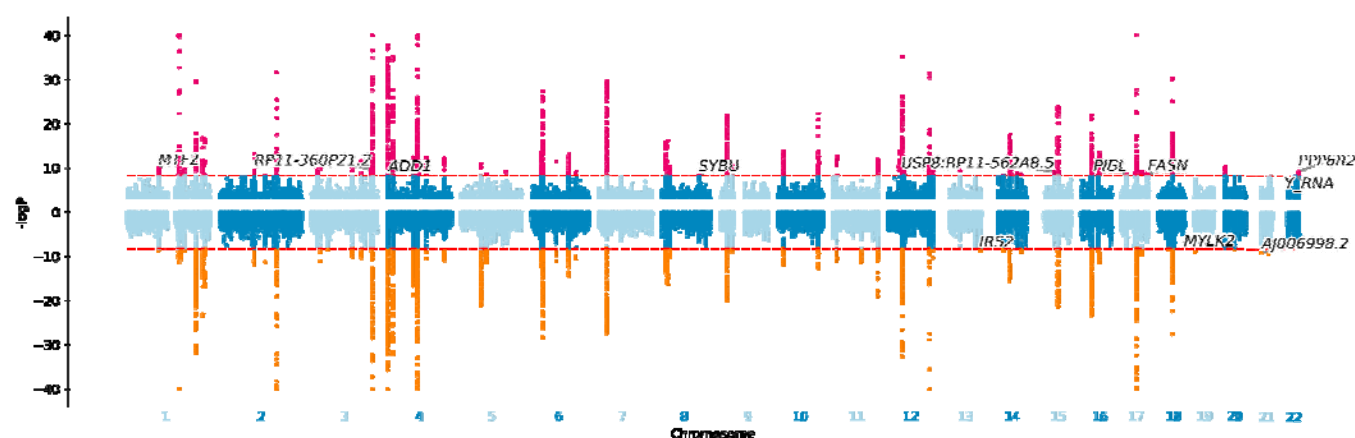


Figure 2: Miami plot of random-effect and MR-MEGA meta-analyses. Top: random effect; Bottom: MR-MEGA. Dotted lines indicate the bonferroni adjusted significant threshold of $P < 5 \times 10^{-9}$. All $-\log_{10}P$ values greater than 40 were truncated to 40 for visual clarity. SNPs significant in random-effect and MR-MEGA models are colored red and orange respectively. Novel loci are indicated by the nearest gene.

Table 2. Meta-analysis results of lead SNPs in the novel loci. RE – random effect, A1 – effect allele, A2 – alternate allele, ANC-HET – ancestry heterogeneity, RES-HET – residual heterogeneity

rsID	Nearest Gene	CHR:BP:A1:A2	BETA(RE)	SE	P(RE)	P(MR-MEGA)	P(ANC-HET)	P(RES-HET)
rs11164870	<i>MTF2</i>	1:93552187:C:G	0.054	0.009	1.15E-10	2.64E-09	0.229	0.928
rs6806917	<i>RP11-360P21.2</i>	3:178861417:T:C	-0.070	0.011	1.65E-10	3.43E-09	0.215	0.762
rs16843452	<i>ADD1</i>	4:2849168:T:C	-0.068	0.012	4.11E-09	3.19E-07	0.747	0.687
rs6469271	<i>SYBU</i>	8:110644774:T:C	-0.056	0.010	3.62E-09	2.04E-07	0.590	0.954
rs1078514	<i>IRS2</i>	13:110463168:T:C	0.068	0.026	0.004817	2.30E-09	5.30E-03	0.261
rs28648524	<i>USP8:RP11-562A8.5</i>	15:50787409:A:T	0.064	0.010	6.45E-10	2.58E-08	0.406	0.661
rs11650438	<i>PI3L</i>	17:16234260:A:G	0.050	0.009	2.93E-09	1.46E-07	0.528	0.288
rs4485435	<i>FASN</i>	17:80045086:C:G	0.082	0.014	2.61E-09			
rs6060983	<i>MYLK2</i>	20:30420924:T:C	0.069	0.037	0.03221	3.86E-09	0.035	0.149
rs1736020	<i>AJ006998.2</i>	21:16812552:A:C	0.006	0.005	0.8853	1.12E-09	4.74E-05	0.638
rs73174657	<i>Y_RNA</i>	22:41434158:A:G	-0.059	0.010	3.81E-09	4.90E-07	0.983	0.655
rs10775809	<i>PPP6R2</i>	22:50808017:A:T	0.092	0.015	4.09E-10	5.61E-08	0.943	0.903

Table 3. MR-MEGA fine mapping results for loci with a single SNP within the 99% credible set. Known PD genes are either known PD risk genes (*SNCA*, *TMEM175*, *MAPT*) or genes with the highest score in the nearest known PD locus by the PD GWAS Locus Browser⁵²

Locus	Number of significant SNPs	Nominated variant	CHR:BP:A1:A2	Nearest Gene	Known PD Genes +/- 1MB	Functional Consequence	CADD	RDB
11	6	rs57891859	2:135464616:A:G	<i>TMEM163</i>	<i>TMEM163</i>	intronic	6.746	4
19	926	rs34311866	4:951947:C:T	<i>TMEM175</i>	<i>TMEM175</i>	exonic	11.09	NA
23	1483	rs356182	4:90626111:A:G	<i>RP11-115D19.1</i>	<i>SNCA</i>	ncRNA intronic	8.962	NA
45	1371	rs10847864	12:123326598:G:T	<i>HIP1R</i>	<i>HIP1R</i>	intronic	2.403	2b
56	1060	rs62053943	17:43744203:C:T	<i>CRHR1:RP11-105N13.4</i>	<i>MAPT</i>	ncRNA intronic	0.183	5
60	1	rs55818311	19:2341047:C:T	<i>SPPL2B</i>	<i>LSM7</i>	ncRNA exonic	1.096	5

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(12):1091-1102.
2. Foo JN, Chew EGY, Chung SJ, et al. Identification of Risk Loci for Parkinson Disease in Asians and Comparison of Risk Between Asians and Europeans: A Genome-Wide Association Study. *JAMA Neurol.* 2020;77(6):746-754.
3. Loesch DP, Horimoto ARVR, Heilbron K, et al. Characterizing the Genetic Architecture of Parkinson's Disease in Latinos. *Ann Neurol.* 2021;90(3):353-365.
4. Rajan R, Divya KP, Kandadai RM, et al. Genetic Architecture of Parkinson's Disease in the Indian Population: Harnessing Genetic Diversity to Address Critical Gaps in Parkinson's Disease Research. *Front Neurol.* 2020;11. doi:10.3389/fneur.2020.00524
5. Program TGPG, The Global Parkinson's Genetics Program. GP2 : The Global Parkinson's Genetics Program. *Movement Disorders.* 2021;36(4):842-851. doi:10.1002/mds.28494
6. Skol AD, Scott LJ, Abecasis GR, Boehnke M. Joint analysis is more efficient than replication-based analysis for two-stage genome-wide association studies. *Nat Genet.* 2006;38(2):209-213.
7. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience.* 2015;4:7.

8. Mägi R, Horikoshi M, Sofer T, et al. Trans-ethnic meta-regression of genome-wide association studies accounting for ancestry increases power for discovery and improves fine-mapping resolution. *Hum Mol Genet.* 2017;26(18):3639-3650.
9. Pulit SL, de With SAJ, de Bakker PIW. Resetting the bar: Statistical significance in whole-genome sequencing-based association studies of global populations. *Genet Epidemiol.* 2017;41(2):145-151.
10. Shi H, Burch KS, Johnson R, et al. Localizing Components of Shared Transethnic Genetic Architecture of Complex Traits from GWAS Summary Data. *Am J Hum Genet.* 2020;106(6):805-817.
11. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun.* 2017;8(1):1826.
12. Watanabe K, Umičević Mirkov M, de Leeuw CA, van den Heuvel MP, Posthuma D. Genetic mapping of cell type specificity for complex traits. *Nat Commun.* 2019;10(1):1-13.
13. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: Generalized Gene-Set Analysis of GWAS Data. *PLoS Comput Biol.* 2015;11(4). doi:10.1371/journal.pcbi.1004219
14. Zeng B, Bendl J, Kosoy R, Fullard JF, Hoffman GE, Roussos P. Multi-ancestry eQTL meta-analysis of human brain identifies candidate causal variants for brain-related traits. *Nat Genet.* Published online January 20, 2022. doi:10.1038/s41588-021-00987-9
15. Wu Y, Zeng J, Zhang F, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat Commun.* 2018;9(1):918.
16. Smeland OB, Shadrin A, Bahrami S, et al. Genome-wide association analysis of Parkinson's disease and schizophrenia reveals shared genetic architecture and identifies novel risk loci. *Biol Psychiatry.* 2021;89(3):227.
17. Jinn S, Blauwendraat C, Toolan D, et al. Functionalization of the TMEM175 p.M393T variant as a risk factor for Parkinson disease. *Hum Mol Genet.* 2019;28(19):3244-3254.
18. Saunders A, Macosko EZ, Wysoker A, et al. Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell.* 2018;174(4):1015-1030.e16.
19. La Manno G, Gyllborg D, Codeluppi S, et al. Molecular Diversity of Midbrain Development in Mouse, Human, and Stem Cells. *Cell.* 2016;167(2). doi:10.1016/j.cell.2016.09.027
20. Koegler E, Bonnon C, Waldmeier L, Mitrovic S, Halbeisen R, Hauri HP. p28, a novel ERGIC/cis Golgi protein, required for Golgi ribbon formation. *Traffic.* 2010;11(1):70-89.
21. Gaudet P, Livstone MS, Lewis SE, Thomas PD. Phylogenetic-based propagation of functional annotations within the Gene Ontology consortium. *Brief Bioinform.* 2011;12(5). doi:10.1093/bib/bbr042
22. "ER to Golgi Anterograde Transport". Reactome, release 73, <https://reactome.org/content/detail/R-HSA-199977>: R-HSA-199977 (Access date: March 29, 2021)
23. Bonet-Ponce L, Beilina A, Williamson CD, et al. LRRK2 mediates tubulation and vesicle

sorting from lysosomes. *Science Advances*. 2020;6(46). doi:10.1126/sciadv.abb2454

24. Bandres-Ciga S, Saez-Atienzar S, Bonet-Ponce L, et al. The endocytic membrane trafficking pathway plays a major role in the risk of Parkinson disease. *Mov Disord*. 2019;34(4):460.
25. Beilina A, Rudenko IN, Kaganovich A, et al. Unbiased screen for interactors of leucine-rich repeat kinase 2 supports a common pathway for sporadic and familial Parkinson disease. *Proc Natl Acad Sci U S A*. 2014;111(7). doi:10.1073/pnas.1318306111
26. MacLeod DA, Rhinn H, Kuwahara T, et al. RAB7L1 interacts with LRRK2 to modify intraneuronal protein sorting and Parkinson's disease risk. *Neuron*. 2013;77(3). doi:10.1016/j.neuron.2012.11.033
27. Fujita Y, Ohama E, Takatama M, Al-Sarraj S, Okamoto K. Fragmentation of Golgi apparatus of nigral neurons with alpha-synuclein-positive inclusions in patients with Parkinson's disease. *Acta Neuropathol*. 2006;112(3):261-265.
28. Martínez-Menárguez JÁ, Tomás M, Martínez-Martínez N, Martínez-Alonso E. Golgi Fragmentation in Neurodegenerative Diseases: Is There a Common Cause? *Cells*. 2019;8(7):748. doi:10.3390/cells8070748
29. Sobu Y, Wawro PS, Dhekne HS, Yeshaw WM, Pfeffer SR. Pathogenic LRRK2 regulates ciliation probability upstream of tau tubulin kinase 2 via Rab10 and RILPL1 proteins. *Proc Natl Acad Sci U S A*. 2021;118(10). doi:10.1073/pnas.2005894118
30. Fung HC, Scholz S, Matarin M, et al. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol*. 2006;5(11):911-916.
31. Maraganore DM, de Andrade M, Lesnick TG, et al. High-resolution whole-genome association study of Parkinson disease. *Am J Hum Genet*. 2005;77(5):685-693.
32. Loh PR, Palamara PF, Price AL. Fast and accurate long-range phasing in a UK Biobank cohort. doi:10.1101/028282
33. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74.
34. UK10K Consortium, Walter K, Min JL, et al. The UK10K project identifies rare variants in health and disease. *Nature*. 2015;526(7571):82-90.
35. Durand EY, Do CB, Mountain JL, Michael Macpherson J. Ancestry Composition: A Novel, Efficient Pipeline for Ancestry Deconvolution. *bioRxiv*. Published online October 18, 2014:010512. doi:10.1101/010512
36. Zhao H, Sun Z, Wang J, Huang H, Kocher JP, Wang L. CrossMap: a versatile tool for coordinate conversion between genome assemblies. *Bioinformatics*. 2014;30(7):1006-1007.
37. Bulik-Sullivan BK, Loh PR, Finucane HK, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015;47(3):291-295.

38. Finucane HK, Bulik-Sullivan B, Gusev A, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. *Nat Genet*. 2015;47(11):1228-1235.
39. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Research*. 2010;38(16):e164-e164. doi:10.1093/nar/gkq603
40. Aguet F, Barbeira AN, Bonazzola R, et al. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *bioRxiv*. Published online October 3, 2019:787903. doi:10.1101/787903
41. Liberzon A, Subramanian A, Pinchback R, Thorvaldsdottir H, Tamayo P, Mesirov JP. Molecular signatures database (MSigDB) 3.0. *Bioinformatics*. 2011;27(12):1739-1740. doi:10.1093/bioinformatics/btr260
42. Liu B, Gloudemans MJ, Rao AS, Ingelsson E, Montgomery SB. Abundant associations with gene expression complicate GWAS follow-up. *Nat Genet*. 2019;51(5):768-769.
43. de Klein N, 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*. Published online March 5, 2021:2021.03.01.433439. doi:10.1101/2021.03.01.433439
44. Bryois J, Garrett ME, Song L, et al. Evaluation of chromatin accessibility in prefrontal cortex of individuals with schizophrenia. *Nat Commun*. 2018;9(1):3121.
45. 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(9):1300-1310.
46. McRae AF, Marioni RE, Shah S, et al. Identification of 55,000 Replicated DNA Methylation QTL. *Sci Rep*. 2018;8(1):17605.
47. Wingo AP, Liu Y, Gerasimov ES, et al. Integrating human brain proteomes with genome-wide association data implicates new proteins in Alzheimer's disease pathogenesis. *Nature Genetics*. 2021;53(2):143-146. doi:10.1038/s41588-020-00773-z
48. Elsworth B, Lyon M, Alexander T, et al. The MRC IEU OpenGWAS data infrastructure. doi:10.1101/2020.08.10.244293
49. Hemani G, Zheng J, Elsworth B, et al. The MR-Base platform supports systematic causal inference across the human phenome. *Elife*. 2018;7. doi:10.7554/eLife.34408
50. Nightingale Health and UK Biobank announces major initiative to analyse half a million blood samples to facilitate global medical research. Accessed March 11, 2022. <https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/news/nightingale-health-and-uk-biobank-announces-major-initiative-to-analyse-half-a-million-blood-samples-to-facilitate-global-medical-research>
51. Consortium TG, The GTEx Consortium. The GTEx Consortium atlas of genetic regulatory effects across human tissues. *Science*. 2020;369(6509):1318-1330.

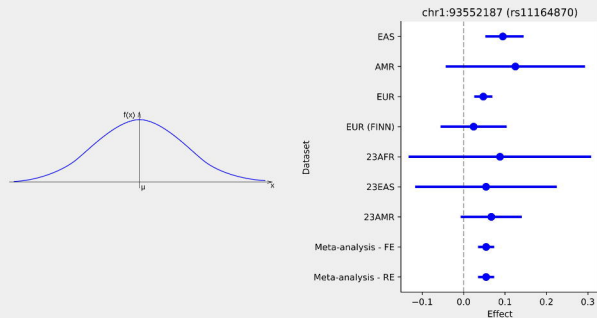
doi:10.1126/science.aaz1776

52. Grenn FP, Kim JJ, Makarios MB, et al. The Parkinson's Disease Genome-Wide Association Study Locus Browser. *Mov Disord.* 2020;35(11):2056-2067.

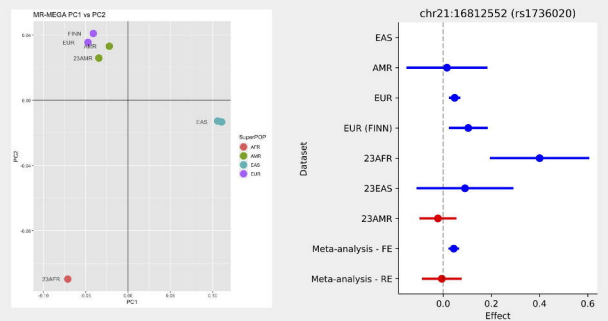
1. Multi-ancestry cohorts



2a. Random effects - Homogeneous effects



2b. MR-MEGA - Heterogeneous effects



3c. Putative burden estimation



3d. Functional annotations

