

**TITLE**

Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-genome wide association study.

**AUTHORS**

Mike A. Nalls, PhD<sup>1,2,CA\*</sup>, Cornelis Blauwendraat, PhD<sup>1\*</sup>, Costanza L. Vallergera, MS<sup>3,4\*</sup>, Karl Heilbron, PhD<sup>5\*</sup>, Sara Bandres-Ciga, PhD<sup>1\*</sup>, Diana Chang, PhD<sup>6\*</sup>, Manuela Tan, MS<sup>7</sup>, Demis A. Kia, BS<sup>7</sup>, Alastair J. Noyce, MD, PhD<sup>7,8</sup>, Angli Xue, BS<sup>3,4</sup>, Jose Bras, PhD<sup>9,10</sup>, Emily Young, PhD<sup>11</sup>, Rainer von Coelln, MD<sup>12</sup>, Javier Simón-Sánchez, PhD<sup>13,14</sup>, Claudia Schulte, PhD<sup>13,14</sup>, Manu Sharma, PhD<sup>15</sup>, Lynne Krohn, MS<sup>16,17</sup>, Lasse Pihlstrom, MD, PhD<sup>18</sup>, Ari Siitonen, PhD<sup>19,20</sup>, Hirotaka Iwaki, MD PhD<sup>1,2,21</sup>, Hampton Leonard, MS<sup>1,2</sup>, Faraz Faghri, PhD<sup>1,22</sup>, J. Raphael Gibbs, PhD<sup>1</sup>, Dena G. Hernandez, PhD<sup>1</sup>, Sonja W. Scholz, MD, PhD<sup>23,24</sup>, Juan A. Botia, PhD<sup>7,25</sup>, Maria Martinez, PhD<sup>26</sup>, Jean-Christophe Corvol, PhD<sup>27</sup>, Suzanne Lesage, PhD<sup>27</sup>, Joseph Jankovic, MD<sup>11</sup>, Lisa M. Shulman, MD<sup>11</sup>, Margaret Sutherland, PhD<sup>28</sup>, Pentti Tienari, MD PhD<sup>29,30</sup>, Kari Majamaa, MD<sup>19,20</sup>, Mathias Toft, MD PhD<sup>18,31</sup>, Ole A. Andreassen, MD<sup>32,33</sup>, Tushar Bangale, PhD<sup>6</sup>, Alexis Brice, MD, PhD<sup>27</sup>, Jian Yang, PhD<sup>3,4</sup>, Ziv Gan-Or, MD, PhD<sup>16,17,34</sup>, Thomas Gasser, MD<sup>13,14</sup>, Peter Heutink, PhD<sup>13,14</sup>, Joshua M Shulman, MD, PhD<sup>11,35,36</sup>, Nicolas Wood, MD<sup>7</sup>, David A. Hinds, PhD<sup>5</sup>, John A. Hardy, PhD<sup>7</sup>, Huw R Morris, MD, PhD<sup>37,38</sup>, Jacob Gratten, PhD<sup>3,4</sup>, Peter M. Visscher, PhD<sup>3,4</sup>, Robert R. Graham, PhD<sup>6</sup>, Andrew B. Singleton, PhD<sup>1</sup> on behalf of the 23andMe Research Team, System Genomics of Parkinson's Disease (SGPD) Consortium and the International Parkinson's Disease Genomics Consortium.

**AFFILIATIONS**

<sup>1</sup> Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, 20892 USA

<sup>2</sup> Data Tecnica International, Glen Echo, MD, 20812 USA

<sup>3</sup> Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072 Australia

<sup>4</sup> Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072 Australia

<sup>5</sup> 23andMe, Inc., Mountain View, California 94041 USA

<sup>6</sup> Department of Human Genetics, Genentech, South San Francisco, 94080, CA, USA

<sup>7</sup> Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK

<sup>8</sup> Preventive Neurology Unit, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK

<sup>9</sup> Center for Neurodegenerative Science, Van Andel Research Institute, Grand Rapids, Michigan, USA

<sup>10</sup> Department of Neurodegenerative Diseases, UCL Institute of Neurology, University College London, London, UK

<sup>11</sup> Department of Neurology, Baylor College of Medicine, Houston, USA

<sup>12</sup> Department of Neurology, University of Maryland School of Medicine, Baltimore, MD, USA

- <sup>13</sup> Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany
- <sup>14</sup> German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany.
- <sup>15</sup> Centre for Genetic Epidemiology, Institute for Clinical Epidemiology and Applied Biometry, University of Tübingen, Germany
- <sup>16</sup> Department of Human Genetics, McGill University, Montreal, Quebec, Canada
- <sup>17</sup> Montreal Neurological Institute, McGill University, Montreal, Quebec, Canada
- <sup>18</sup> Department of Neurology, Oslo University Hospital, Oslo, Norway
- <sup>19</sup> Institute of Clinical Medicine, Department of Neurology, University of Oulu, Oulu, Finland
- <sup>20</sup> Department of Neurology and Medical Research Center, Oulu University Hospital, Oulu, Finland
- <sup>21</sup> The Michael J. Fox Foundation, New York, New York, 10036 USA
- <sup>22</sup> Department of Computer Science, University of Illinois Urbana-Champaign, Champaign, IL, 61820, USA
- <sup>23</sup> Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA
- <sup>24</sup> Department of Neurology, Johns Hopkins University Medical Center, Baltimore, MD 21287, USA
- <sup>25</sup> Departamento de Ingeniería de la Información y las Comunicaciones, Universidad de Murcia, Spain
- <sup>26</sup> INSERM UMR 1220; and Paul Sabatier University, Toulouse, France
- <sup>27</sup> INSERM U1127, CNRS UMR 7225, Sorbonne Université UMR S1127, APHP, Institut du Cerveau et de la Moelle épinière, ICM, Paris F-75013, France
- <sup>28</sup> National Institute on Neurological Diseases and Stroke, National Institutes of Health, Bethesda, MD 20892 USA
- <sup>29</sup> Clinical Neurosciences, Neurology, University of Helsinki, Helsinki, Finland
- <sup>30</sup> Helsinki University Hospital, Helsinki, Finland
- <sup>31</sup> Institute of Clinical Medicine, University of Oslo, Oslo, Norway
- <sup>32</sup> Jepsen Centre for Psychosis Research, University of Oslo, Oslo, Norway
- <sup>33</sup> Division of Mental Health and Addiction, Oslo University Hospital, Oslo, Norway
- <sup>34</sup> Department of Neurology & Neurosurgery, McGill University, Montreal, Quebec, Canada
- <sup>35</sup> Departments of Molecular & Human Genetics and Neuroscience, Baylor College of Medicine, Houston, USA
- <sup>36</sup> Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston
- <sup>37</sup> Department of Clinical Neuroscience, UCL Institute of Neurology, London UK
- <sup>38</sup> UCL Movement Disorders Centre, UCL Institute of Neurology, London, UK
- \*denotes shared first authorship.
- <sup>CA</sup>denotes corresponding author, mike[at]datatecnica[dot]com

Full consortia membership (PubMed indexed) is available in the supplemental materials (Text S1).

## ACKNOWLEDGEMENTS

See supplemental materials (Text S2).

## **SUMMARY**

### **Background**

Genome-wide association studies (GWASs) in Parkinson's disease (PD) have increased the scope of biological knowledge about the disease over the past decade. We sought to use the largest aggregate of GWAS data to identify novel risk loci and gain further insight into disease etiology.

### **Methods**

We performed the largest meta-GWAS of PD to date, involving the analysis of 7.8M SNPs in 37.7K cases, 18.6K UK Biobank proxy-cases (having a first degree relative with PD), and 1.4M controls. We carried out a meta-analysis of this GWAS data to nominate novel loci. We then evaluated heritable risk estimates and predictive models using this data. We also utilized large gene expression and methylation resources to examine possible functional consequences as well as tissue, cell type and biological pathway enrichments for the identified risk factors. Additionally we examined shared genetic risk between PD and other phenotypes of interest via genetic correlations followed by Mendelian randomization.

### **Findings**

We identified 90 independent genome-wide significant risk signals across 78 genomic regions, including 38 novel independent risk signals in 37 loci. These 90 variants explained 16-36% of the heritable risk of PD depending on prevalence. Integrating methylation and expression data within a Mendelian randomization framework identified putatively associated genes at 70 risk signals underlying GWAS loci for follow-up functional studies. Tissue-specific expression enrichment analyses suggested PD loci were heavily brain-enriched, with specific neuronal cell types being implicated from single cell data. We found significant genetic correlations with brain volumes, smoking status, and educational attainment. Mendelian randomization between cognitive performance and PD risk showed a robust association.

### **Interpretation**

These data provide the most comprehensive understanding of the genetic architecture of PD to date by revealing many additional PD risk loci, providing a biological context for these risk factors, and demonstrating that a considerable genetic component of this disease remains unidentified.

### **Funding**

See supplemental materials (Text S2).

## **RESEARCH IN CONTEXT**

### **Evidence before this study**

Previous studies such as Chang et al. 2017 and its predecessors have utilized GWAS methods to discover 42 independent risk loci associated with PD (1). Some of these loci harboring common risk variants also include rare variants implicated in familial PD risk such as *SNCA*, *LRRK2* or *GBA*. Earlier studies like Keller et al. 2012 have attempted to quantify how much heritable risk is captured by common variation that can be easily imputed using commercial genotyping arrays and estimate the amount of risk explained by GWAS (2). As far back as Nalls et al. 2011, GWAS studies of PD have integrated expression and methylation datasets to evaluate possible candidate genes for follow-up at PD loci (3). Many epidemiological and observational studies have attempted to assess risk of PD and various exposures like smoking, caffeine or occupational hazards, with a mixed track record of success at validating putative associations.

### **Added value of this study**

The primary deliverable of this study was increasing the count of independent common genetic risk factors for PD to 90. We added 38 novel risk variants not identified as genome-wide significant in previous reports. We refined heritability estimates and genetic risk predictions suggesting that common genetic variants account for approximately 22% of PD risk on the liability scale, with a range of 16-36% of that risk being explained by GWAS loci in this study. These updated risk predictions also suggested that polygenic risk scoring can be used to achieve an area under the curve of near 70%, although this prediction uses many more variants than just the 90 independent risk factors identified in this report. Of the 90 risk variants we have characterized here, we have nominated at least one possible candidate gene for follow-up functional studies in 70 of these genomic regions by mining recently available expression and methylation reference datasets on a scale not possible just a few years ago. We have additionally mined single cell RNA sequencing data from mice to identify tissue-specific signatures of enrichment relating to PD genetic risk, showing a major focus on neuronal cell types. We also utilized the massive amount of publicly available GWAS results to survey genetic correlations between PD and other phenotypes showing significant correlations with smoking, education and brain morphology. Subsequent analyses using Mendelian randomization (MR) methods showed that there are likely causal links between increased cognitive performance and PD risk on a genetic level.

### **Implications of all available evidence**

First and foremost, this study increased the scope of our knowledge of PD genetics by adding 38 novel risk factors, directly broadening our knowledge base of disease etiology. Using updated heritability estimates and risk predictions, we took preliminary steps down the long path to early detection. In future studies, combining genetic and clinico-demographic risk factors may

lead to earlier detection and refined diagnostics, which may help improve clinical trials (4). The generation of copious amounts of public summary statistics created by this effort relating to both the GWAS and subsequent analyses of gene expression and methylation patterns may be of use to investigators planning follow-up functional studies in stem cells or other cellular screens, allowing them to prioritize targets more efficiently using our data as additional evidence. We hope our findings may have some downstream clinical impact in the future such as improved patient stratification for clinical trials and genetically informed drug targets.

## **INTRODUCTION**

Parkinson's disease is a neurodegenerative disorder, affecting approximately 1 million individuals in the United States alone (5). PD patients suffer from a combination of progressive motor and non-motor symptoms affecting daily function and quality of life. The prevalence of PD is projected to double in some age groups by 2030, creating a substantial burden on healthcare systems (5).

Early investigations into the role of genetic factors in PD focused on the identification of rare mutations underlying familial disease, (6,7) over the past decade there has been a growing appreciation for the contribution of genetics in sporadic disease(1,8). Genetic studies of sporadic PD have altered the foundational view of disease etiology.

We executed a series of experiments to explore the genetics of PD (summarized in Figure 1). We performed the largest-to-date GWAS for PD, including 7.8M SNPs, 37.7K cases, 18.6K UK Biobank (UKB) "proxy-cases" (individuals without PD that have a family history of PD) and 1.4M controls. We identified mechanistic candidate genes for PD, providing valuable therapeutic targets. We assessed the function of these potential risk genes via Mendelian randomization, expression enrichment, and protein-protein interaction network analysis. We estimated PD heritability, developed a polygenic risk score that predicted a substantial proportion of this heritability, and leveraged these results to inform future studies. Finally, we identified candidate PD biomarkers and risk factors using genetic correlation and Mendelian randomization.

## **SUMMARY OF METHODS**

### **GWAS Study design and risk locus discovery**

Three sources of data were used for discovery analyses, these include three previously published studies, 13 new datasets, and proxy-case data from the UK BioBank (UKB). Previous studies include summary statistics published in Nalls *et al.* 2014, GWAS summary statistics from the 23andMe Web-Based Study of Parkinson's Disease (PDWBS) in Chang *et al.* 2017, and the publicly available NeuroX dataset from the International Parkinson's Disease Genomics Consortium (IPDGC) previously used previously as a replication sample. These cohorts have been reported in detail (1,9). We included 13 new case-control sample series for meta-analyses through either publicly available data or collaborations (please see Supplementary Table S1 for details regarding these studies). All samples from the 13 new datasets underwent similar

standardized quality control for inclusion, mirroring that of previous studies. We attempted to generate summary statistics for GWAS meta-analyses as uniformly as possible. This analysis utilized fixed-effects meta-analyses as implemented in METAL to combine summary statistics across all sources (10).

### **Conditional-joint analysis to nominate variants of interest**

To nominate variants of interest, we employed a conditional and joint analysis strategy (COJO, <http://cns.genomics.com/software/gcta/>) to algorithmically identify variants that best account for the heritable variation within and across loci (11). Additional analyses described below were utilized to further scrutinize putative associated variants and account for possible differential linkage disequilibrium (LD) signatures, including using the massive single site reference data from 23andMe in more conditional analyses. If a variant nominated during the COJO phase of analysis was greater than 1Mb from any of the genome-wide significant loci nominated in Chang *et al.* 2017, we considered this to be novel. We defined nominated risk variants as from a single locus if they were within +/- 250kb of each other. We instituted two filters after fixed-effects and COJO analyses, excluding variants that 1) had a random-effects P value across all datasets > 4.67E-04 and 2) a conditional analysis P > 4.67E-04 using participant level 23andMe genotype data. Please see Supplementary Table S2 summarizing all variants nominated plus the Methods and Results Supplement.

### **Refining heritability estimates and determining extant genetic risk**

We used the R package PRSice2 for risk profiling (12), this carries out polygenic risk score (PRS) profiling in the standard weighted allele dose manner (1,9,13–15). In addition, PRSice incorporates permutation testing where case and control labels are swapped in the withheld samples to generate an empirical P. This workflow identifies the best P thresholds for variant inclusion while carrying out LD pruning. In many cases this best P threshold for PRS construction does not meet what is commonly regarded as genome-wide significance.

A two stage design was also employed, training on the largest single array study (NeuroX-dbGaP) and then tested on the second largest study (HBS) using the same array. These two targeted array studies were chosen for three reasons: precedent in the previous publications where the NeuroX-dbGaP dataset was used in PRS; direct genotyping of larger effect rare variants in *GBA* and *LRRK2*; participant level genotypes for these datasets are publicly available.

To calculate heritability in clinically defined PD datasets, we used LD score regression (LDSC) employing the LD references for Europeans provided with the software (16). This workflow was also repeated on a per cohort level (see Supplementary Appendix).

### **Functional causal inferences via Quantitative Trait Loci (QTL)**

We used MR to test whether changes in DNA methylation and/or RNA expression of genes physically proximal to significant PD risk loci were causally related to PD risk. To nominate genes of interest for MR analyses, we took our putative 90 loci in the large LD reference used for the COJO phase of analysis and identified SNPs in LD with our SNPs at an  $r^2 > 0.5$  within +/- 1MB (Supplementary Table S5). MR was used by integrating discovery phase summary statistics with quantitative trait locus (QTL) association summary statistics across well-curated methylation and expression datasets. We used the curated versions of Qi et al., 2018 brain methylation and expression summary statistics (multi-study and multi-tissue meta-analysis), as well as a specific focus on substantia nigra data (GTEx), we made use of the blood expression data from Vösa et al. 2018 (eQTLGen, all available here <http://cnsgenomics.com/software/smr/#Overview> and here <http://www.eqtlgen.org>) (17–21). For all QTL analyses, we utilized the multi-SNP summary-based Mendelian randomization (SMR) method as a framework to carry out MR. All MR effect estimates are reported on the scale of a standard deviation increase in the exposure variable relating to a similar change in PD risk. Simply, these MR analyses compare the local polygenic risk of an exposure (methylation or expression) to similar polygenic risk in an outcome (PD), inferring causal associations under the assumption that there is no intermediate confounder associated with both parameters and that the association is not simply due to LD.

To further investigate expression enrichment across cell types in PD, we integrated GWAS summary statistics with expression and network data from the FUMA webserver (<https://fuma.ctglab.nl/>, version 1.3.1) (22).

### **Rare coding variant burden tests**

A uniformly quality controlled and imputed dataset from the IPDGC was used to carry out burden tests for all rarer coding variants successfully imputed in an average of 85% of the sample series (17,188 cases and 22,875 controls). These analyses include all variants at a hard call threshold of imputation quality  $> 0.8$ . After annotation with annovar, we had a total of 37,503 exonic coding variants (nonsynonymous, stop or splicing) at MAF  $< 5\%$  and a subset of 29,016 at MAF  $< 1\%$  (23). For inclusion in this phase, a gene must have contained at least 2 coding variants. After assembling this subset of 113 testable genes, we used the optimized sequence kernel association test to generate summary statistics at maximum MAFs of 1% and 5% (24).

### **LD score regression and causal inference**

To investigate correlations of PD genetics with that of multiple traits and diseases, we employed bivariate LDSC (16). These analyses were carried out using data from the 757 GWAS available via LD Hub and biomarker GWAS summary statistics on c-reactive protein and cytokine measures; LD Hub was accessed on June 20th, 2018 (version 1.2.0) (25–27). P values from the bivariate LDSC were adjusted for FDR to account for multiple testing. Traits showing significant genetic correlations with PD were analyzed using MR methods. We excluded the UKB data

when a nominated trait was from summary statistics derived from the UKB or if the UKB was included as part of a meta-analysis.

When complete GWAS summary statistics were available for traits of interest (relating to smoking and education), we used the more powerful bi-directional generalized summary-data-based Mendelian Randomization (GSMR). We analyzed GWAS summary statistics for smoking initialization (453,693 records from a self-report survey with 208,988 regular smokers and 244,705 never regular smokers) and current smoking (CS) within the UKB, CS contrasted 47,419 current smokers versus 244,705 never regular smokers. The same analysis was carried out incorporating recent GWAS data regarding educational attainment (N = 766,345) from self report in the UK and cognitive performance (N = 257,828) as measured by the *g* composite score (28). These were analyzed using methods to mirror that of the UKB PD GWAS dataset. Combined left and right putamen volume from a T2 magnetic resonance imaging GWAS available from Oxford Brain Imaging Genetics (BIG) Server (accessed December 28th, 2018) (29). All MR analyses included GWAS on the scale of tens of thousands of samples and overcame the considerable power demands of the methodology.

**For additional quality control, methods details and ancillary results, see the Methods Supplement.**

**The funder of the study had no role in study design, data collection, data analysis, data interpretation, or writing of the report. The corresponding author had full access to all of the data and the final responsibility to submit for publication.**

## RESULTS

To maximize our power for locus discovery we used a single stage design, meta-analyzing all available GWAS summary statistics. Supporting this design, we found strong genetic correlations using PD cases ascertained by clinicians compared to 23andMe self-reported cases (genetic correlation from LDSC ( $r_G$ ) = 0.85, SE = 0.06) and UKB proxy cases ( $r_G$  = 0.84, SE = 0.134).

We identified a total of 90 independent genome-wide significant association signals through our analyses of 37,688 cases, 18,618 UKB proxy-cases and 1,417,791 controls at 7,784,415 SNPs (Figure 2, Table 1, Supplementary Appendices, Table S1, Table S2). Of these, 38 signals are new and more than 1MB from loci described previously (1) (Table S3).

We detected 10 loci containing more than one independent risk signal (22 risk SNPs in total across these loci), of which nine had been identified by previous GWAS, including multi-signal loci in the vicinity of *GBA*, *NUCKS1/RAB29*, *GAK/TMEM175*, *SNCA* and *LRRK2*. The novel multi-signal locus comprised independent risk variants rs2269906 (*UBTF/GRN*) and rs850738 (*FAM171A2*). Detailed summary statistics on all nominated loci can be found in Table S2, including variants filtered out during additional quality control.



To quantify how much of the genetic liability we have explained and what direction to take with future PD GWAS we generated updated heritability estimates and PRS. Using LDSC on a meta-analysis of all 11 clinically-ascertained datasets from our GWAS and estimated the liability-scale heritability of PD as 0.22 (95% CI 0.18 - 0.26), only slightly lower than a previous estimate derived using GCTA (0.27, 95% CI 0.17 - 0.38) (2,16,30). LDSC is known to be more conservative than GCTA, however, our LDSC heritability estimate does fall within the 95% confidence interval of the GCTA estimate.

To determine the proportion of SNP-based heritability explained by our PD GWAS results using PRS, we used a two-stage design, with variant selection and training in the NeuroX-dbGaP dataset (5,851 cases and 5,866 controls) and then validation in the Harvard Biomarker Study (HBS, 527 cases and 472 controls). Using equations from Wray *et al.* 2010 and our current heritability estimates, the 88 variant PRS explained a minimum 16% of the genetic liability of PD assuming a global prevalence of 0.5% (2,31). The 1805 variant PRS explained roughly 26% of PD heritability. In a high-risk population with a prevalence of 2%, the 1805 variant PRS explained a maximum 36% of PD heritable risk (2,31) (Table S4).

We then attempted to quantify strata of risk in our more inclusive PRS. Compared to individuals with PRS values in the lowest quartile, the PD odds ratio for individuals with PRS values in the highest quartile was 3.74 (95% CI = 3.35 - 4.18) in the NeuroX-dbGaP cohort and 6.25 (95% CI = 4.26 - 9.28) in the HBS cohort (Table 2, Figure 3, Figure S1).

Variants in the range of  $5E-08 < P < 1.35E-03$  (used in the 1805 variant PRS) were rarer and had smaller effect estimates than variants reaching genome-wide significance. These sub-significant variants had a median minor allele frequency of 21.3% and a median effect estimate (absolute value of the log odds ratio of the SNP parameter from regression) of 0.047. Genome-wide significant risk variants were more common with a median minor allele frequency of 25.1%, and had a median effect estimate of 0.081. Here we assume that the lower minor allele frequencies and smaller effect size estimates are typical and representative of variants contributing to our more inclusive PRS and represent future GWAS hits. We performed power calculations to forecast the number of additional PD cases needed to achieve genome-wide significance at 80% power for a variant with a minor allele frequency of 21.3% and an effect estimate of 0.047 (32). Assuming that future data is well-harmonized with current data and that disease prevalence is 0.5%, we estimated that we would need a total of ~99K cases, ~2.3 times more than this work for these to reach genome-wide significance. These variants already contribute towards the current increases in AUC when considering the 1805 variant PRS outperforms the 88 variant PRS. Expanding future studies to this size will invariably identify new loci and improve the AUC for a genetic predictor in PD (maximum potential AUC estimated at 85% using the equations from Wray *et al.* 2010) (31).

There were 305 genes within the 78 GWAS loci. We sought to identify the likely causal gene(s) in each locus using large QTL datasets and summary-data-based Mendelian randomization (Table 3, Table S5, Table S6) (33). This method allows for functional inferences between two

datasets to be made in an analogous framework to a randomized controlled trial, treating the genotype as the randomizing factor.

Of the 305 genes under linkage disequilibrium (LD) peaks around our risk variants of interest, 237 were possibly associated with at least one QTL in public reference datasets and were therefore testable via SMR (Methods Supplement, Table S6). The expression or methylation of 151 of these 237 genes (63.7%) was significantly associated with a possible causal change in PD risk.

Of the 90 PD GWAS risk variants, 70 were in loci containing at least one of these putatively causal genes after multiple test correction (Table 3). For 53 out of these 70 PD GWAS hits (75.7%), the gene nearest to the most significant SNP was a putatively causal gene (Table S2). Most loci tested contained multiple putatively causal genes. Interestingly, the nearest putatively causal gene to the rs850738/*FAM171A2* GWAS risk signal is *GRN*, a gene known to be associated with frontotemporal dementia (FTD) (34). Mutations in *GRN* have also been shown to be connected with another lysosomal storage disorder, neuronal ceroid lipofuscinosis (35).

As an orthogonal approach for nominating genes under GWAS peaks we carried out rare coding variant burden analyses. We performed kernel-based burden tests on the 113 genes out of the 305 under our GWAS peaks that contained two or more rare coding variants (MAF < 5% or MAF < 1%). After Bonferroni correction for 113 genes, we identified 7 significant genes: *LRRK2*, *GBA*, *CATSPER3* (rs11950533/*C5orf24* locus), *LAMB2* (rs12497850/*IP6K2* locus), *LOC442028* (rs2042477/*KCNIP3* locus), *NFKB2* (rs10748818/*GBF1* locus), and *SCARB2* (rs6825004 locus). These results suggest that some of the risk associated with these loci may be due to rare coding variants or that these are pleomorphic risk loci. The *LRRK2* and *NFKB2* associations at MAF < 1% remained significant after correcting for all ~20,000 genes in the human genome ( $P = 2.15E-10$  and  $P = 4.02E-07$ , Table S7, Table S5).

We tested whether genes of interest were enriched in 10,651 biological pathways (from gene ontology annotations) using Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) (22,36). We found 10 significantly enriched pathways (FDR-adjusted  $P < 0.05$ , Table S8), including four related to vacuolar function and three related to known drug targets (calcium transporters: *ikedamir1\_targets\_dn* and *ikedamir30\_targets\_up*, kinase signaling: *kim\_pten\_targets\_dn*). At least three candidate genes within novel loci are involved in lysosomal storage disorders (*GUSB*, *GRN*, and *NEU1*), a pathway of keen interest in PD (37). Our GWAS results also include candidate genes *VAMP4* and *NOD2* from the endocytic pathway (38).

To determine the tissues and cell types most relevant to PD etiology using FUMA (22,36) we tested whether the genes highlighted by our PD GWAS were enriched for expression in 53 tissues from across the body. We found 13 significant tissues, all of which were brain-derived (Figure S2A), in contrast to what has been seen in Alzheimer's disease which shows a strong bias towards blood, spleen, lungs and microglial enrichments (39). To further disentangle the enrichment in brain tissues, we tested whether our PD GWAS genes were enriched for expression in 88 brain cell types using single cell RNA sequencing reference data from mouse

brains (<http://dropviz.org>)(40). After FDR correction we found seven significant brain cell types, all of which were neuronal (Figure S2B). The strongest enrichment was for neurons in the substantia nigra (SN) at  $P = 1.0E-06$ , with additional significant results at  $P < 5.0E-4$  for the globus pallidus (GP), thalamus (TH), posterior cortex (PC), frontal cortex (FC), hippocampus (HC) and entopeduncular nucleus (ENT).

Next, we used cross-trait genetic correlation and MR to identify possible PD biomarkers and risk factors by comparing with 757 other GWAS datasets curated by LD hub (41). We found four significant genetic correlations (FDR-adjusted  $P < 0.05$ , Table S10) including positive correlations with intracranial volume and putamen volume (42), and negative correlations with current tobacco use and “academic qualifications: National Vocational Qualifications (NVQ) or Higher National Diploma (HND) or Higher National Certificate (HNC) or equivalent” (43). The negative association with one’s academic qualifications suggests that individuals without a college education may be at less risk of PD. The correlation between PD and smoking status may not be independent from the correlation between PD and education as smoking status and years of education were significantly correlated (44).

We used MR to assess whether there was evidence of a causal relationship between PD and five phenotypes related to academic qualification, smoking, and brain volumes described above (Figure S4). Cognitive performance had a large, significant causal effect on PD risk (MR effect = 0.213, SE = 0.041, Bonferroni-adjusted  $P = 8.00E-07$ ), while PD risk did not have a significant causal effect on cognitive performance (Bonferroni-adjusted  $P = 0.125$ ). Educational attainment also had a significant causal effect on PD risk (MR effect = 0.162, SE = 0.040, Bonferroni-adjusted  $P = 2.06E-04$ ), but PD risk also had a weak but significant causal effect on educational attainment (MR effect = 0.007, SE = 0.002, Bonferroni-adjusted  $P = 7.45E-3$ ). There was no significant causal relationship between PD and current smoking status (forward analysis: MR effect = -0.069, SE = 0.031, Bonferroni-adjusted  $P = 0.125$ ; reverse analysis: MR effect = 0.004, SE = 0.010, Bonferroni-adjusted  $P = 1$ ). Smoking initiation (the act of ever starting smoking) did not have a causal effect on PD risk (MR effect = -0.063, SE = 0.034, Bonferroni-adjusted  $P = 0.315$ ), whereas PD had a small, but significantly positive causal effect on smoking initiation (MR effect = 0.027, SE = 0.006, Bonferroni-adjusted  $P = 1.62E-05$ ). Intracranial volume could not be tested because its GWAS did not contain any genome-wide significant risk variants. There was no significant causal relationship between PD and putamen volume ( $P > 0.05$  in both the forward and reverse directions).

## DISCUSSION

Our work marks a significant step forward in our understanding of the genetic architecture of PD and provides a genetic reference set for the broader research community. We identified 90 independent common genetic risk factors for PD, nearly doubling the number of known PD risk variants. We re-evaluated the cumulative contribution of genetic risk variants, both genome-wide significant and not-yet discovered, in order to refine our estimates of heritable PD risk. We also nominated likely genes at each locus for further follow-up using QTL analyses and rare variant burden analyses. Our work has highlighted the pathways, tissues, and cell types

involved in PD etiology. Finally, we identified intracranial and putaminal volume as potential future PD biomarkers, and cognitive performance as a PD risk factor. Altogether, the data presented here has significantly expanded the resources available for future investigations into potential PD interventions.

We were able to explain 16%-36% of PD heritability, the range being directly related to prevalence estimates varying (0.5% to 2%). Power estimates suggest that expansions of case numbers to 99K cases will continue to reveal additional insights into PD genetics. While these risk variants will have relatively small effects and/or be quite rare, they will help to further expand our knowledge of the genes and pathways that drive PD risk.

Population-wide screening for individuals who are likely to develop PD is currently not feasible using our 1805 variant PRS alone. There would be roughly 14 false positives per true positive assuming a prevalence of 0.5%. While large-scale genome sequencing and non-linear machine learning methods will likely improve these predictive models, we have previously shown that we will need to incorporate other data sources (*e.g.* smell tests, family history, age, sex) in order to generate algorithms that have more possible value in population-wide screening (4).

Evaluating these results in the larger context of pathway, tissue, and cellular functionality revealed that genes near PD risk variants showed enrichment for expression in the brain, contrasting with previous work in Alzheimer's disease. Strikingly, we showed that the expression enrichment of genes at PD loci occurred exclusively in neuronal cell types. We also found that PD genes were enriched in chemical signaling pathways and pathways involving the response to a stressor. We believe that this contrast, in which the pathway enrichment analyses suggest at least some immune component to PD and the expression enrichment analysis does not suggest any significant immune related tissue component should be viewed with a critical eye. In particular, the marginal P values of most immune related pathways in our recent analyses after multiple test correction reinforce this moderate view. These observations may be informative for disease modeling efforts, highlighting the importance of disease modeling in neurons and possibly incorporating a cellular stress component. This information can help inform and focus stem cell derived therapeutic development efforts that are currently underway.

We found four phenotypes that were genetically correlated with PD. Putamen and intracranial volumes may prove to be valuable in future PD biomarker studies. Our bi-directional GSMR results suggest a complex etiological connection between smoking initiation and PD that will require further follow-up. One of the implications of this work is that PD trials of nicotine or other smoking-related compound(s) may be less likely to succeed. The strong causal effect of cognitive performance on PD is supported by observational studies (45).

While this study marks major progress in assessing genetic risk factors for PD, there remains a great deal to be done. No defined external validation dataset was used, which may be seen as a limitation. Also, external replication of the novel associations we present will be difficult simply due to the sample sizes needed. Simulations have suggested that without replication variants with P values between  $5E-08$  and  $5E-9$  should be interpreted with greater caution (46,47). We

found 16 risk variants in this range, including two known variants near *WNT3* (proximal to the *MAPT* locus) and *BIN3*. To a degree, the fact that we filtered our variants with a secondary random-effects meta-analysis may make our 90 PD GWAS hits somewhat more robust due to the conservative nature of random-effects.

This study focused on PD risk in individuals of European ancestry. Adding datasets from non-European populations would be helpful to further improve our granularity in association testing and ability to fine-map loci through integration of more variable LD signatures while also evaluating population specific associations. Also, risk predictions may not generalize across populations in some cases and ancestry specific PRS should be investigated. Additionally, large ancestry-specific PD LD reference panels, such as those for Ashkenazi Jewish patients, will help us further unravel the genetic architecture of loci such as *GBA* and *LRRK2*. This may be particularly crucial at these loci where LD patterns may be variable within European populations, accentuating the possible influence of LD reference series on conditional analyses in some cases (48). Finally, our work utilized state-of-the-art QTL datasets to nominate candidate genes, but many QTL associations are hampered by both small sample size and low *cis*-SNP density. Larger QTL studies and PD-specific network data from large scale cellular screens would allow us to build a more robust functional inference framework.

As the field moves forward there are some critical next steps that should be prioritized. First, allowing researchers to share participant-level data in a secure environment would facilitate inclusiveness and uniformity in analyses while maintaining the confidentiality of study participants. Our work suggests that GWASes of increasing size will continue to provide useful biological insights into PD. In addition to studies of the genetics of PD risk, studies of disease onset, progression, and subtype will be important and will require large series of well-characterized patients(49). We also believe that work across diverse populations is important, not only to be able to best serve these populations but also to aid in fine mapping of loci. Notably, the use of genome sequencing technologies could further improve discovery by capturing rare variants and structural variants, but with the caveat that very large sample sizes will be required. While there is still much left to do, we believe that our current work represents a significant step forward and that the results and data will serve as a foundational resource for the community to pursue this next phase of PD research.

## **DATA ACCESS**

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](https://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](#) 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). Underlying participant level IPDGC data is available to potential collaborators, please contact [ipdgc.contact@gmail.com](mailto:ipdgc.contact@gmail.com).

## WORKS CITED

1. Chang D, Nalls MA, Hallgrímsdóttir IB, Hunkapiller J, van der Brug M, Cai F, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet.* 2017 Oct;49(10):1511–6.
2. 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 Nov 15;21(22):4996–5009.
3. International Parkinson Disease Genomics Consortium, Nalls MA, Plagnol V, Hernandez DG, Sharma M, Sheerin U-M, et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet.* 2011 Feb 19;377(9766):641–9.
4. Nalls MA, McLean CY, Rick J, Eberly S, Hutten SJ, Gwinn K, et al. Diagnosis of Parkinson's disease on the basis of clinical and genetic classification: a population-based modelling study. *Lancet Neurol.* 2015 Oct;14(10):1002–9.
5. Dorsey ER, Constantinescu R, Thompson JP, Biglan KM, Holloway RG, Kieburtz K, et al. Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology.* 2007 Jan 30;68(5):384–6.
6. Polymeropoulos MH, Higgins JJ, Golbe LI, Johnson WG, Ide SE, Di Iorio G, et al. Mapping of a gene for Parkinson's disease to chromosome 4q21-q23. *Science.* 1996 Nov 15;274(5290):1197–9.
7. Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, et al. alpha-Synuclein locus triplication causes Parkinson's disease. *Science.* 2003 Oct 31;302(5646):841.
8. Fung H-C, Scholz S, Matarin M, Simón-Sánchez J, Hernandez D, Britton A, et al. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol.* 2006 Nov;5(11):911–6.
9. Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet.* 2014 Sep;46(9):989–93.
10. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010 Sep 1;26(17):2190–1.
11. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet.* 2012 Mar 18;44(4):369–75, S1–3.
12. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics.*

2014;31(9):1466–8.

13. International Parkinson Disease Genomics Consortium, Nalls MA, Plagnol V, Hernandez DG, Sharma M, Sheerin U-M, et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet*. 2011 Feb 19;377(9766):641–9.
14. International Parkinson's Disease Genomics Consortium (IPDGC), Wellcome Trust Case Control Consortium 2 (WTCCC2). A two-stage meta-analysis identifies several new loci for Parkinson's disease. *PLoS Genet*. 2011 Jun;7(6):e1002142.
15. Nalls MA, Escott-Price V, Williams NM, Lubbe S, Keller MF, Morris HR, et al. Genetic risk and age in Parkinson's disease: Continuum not stratum. *Mov Disord*. 2015 May;30(6):850–4.
16. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015 Mar;47(3):291–5.
17. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet*. 2013 Jun;45(6):580–5.
18. 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 Jun 11;9(1):2282.
19. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet*. 2016 May;48(5):481–7.
20. Wu Y, Zeng J, Zhang F, Zhu Z, Qi T, Zheng Z, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat Commun*. 2018 Mar 2;9(1):918.
21. 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 [Internet]. *bioRxiv*. 2018 [cited 2018 Oct 25]. p. 447367. Available from: <https://www.biorxiv.org/content/early/2018/10/19/447367.abstract>
22. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun*. 2017 Nov 28;8(1):1826.
23. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc*. 2015;10(10):1556–66.
24. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet*. 2012 Aug 10;91(2):224–37.
25. LD Hub [Internet]. [cited 2018 Jun 20]. Available from: <http://ldsc.broadinstitute.org/ldhub/>
26. Prins BP, Abbasi A, Wong A, Vaez A, Nolte I, Franceschini N, et al. Investigating the

Causal Relationship of C-Reactive Protein with 32 Complex Somatic and Psychiatric Outcomes: A Large-Scale Cross-Consortium Mendelian Randomization Study. *PLoS Med.* 2016 Jun;13(6):e1001976.

27. Ahola-Olli AV, Würtz P, Havulinna AS, Aalto K, Pitkänen N, Lehtimäki T, et al. Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors. *Am J Hum Genet.* 2017 Jan 5;100(1):40–50.
28. Lee JJ, Wedow R, Okbay A, Kong E, Maghziyan O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* [Internet]. 2018 Jul 23; Available from: <http://dx.doi.org/10.1038/s41588-018-0147-3>
29. Elliott LT, Sharp K, Alfaro-Almagro F, Shi S, Miller KL, Douaud G, et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature.* 2018 Oct;562(7726):210–6.
30. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. *Am J Hum Genet.* 2011 Jan 7;88(1):76–82.
31. Wray NR, Yang J, Goddard ME, Visscher PM. The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet.* 2010 Feb 26;6(2):e1000864.
32. 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 Feb;38(2):209–13.
33. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016 May;48(5):481–7.
34. Cruts M, Gijselink I, van der Zee J, Engelborghs S, Wils H, Pirici D, et al. Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. *Nature.* 2006 Aug 24;442(7105):920–4.
35. Smith KR E al. Strikingly different clinicopathological phenotypes determined by progranulin-mutation dosage. - PubMed - NCBI [Internet]. [cited 2019 Jan 10]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/22608501>
36. Wang J, Vasaiakar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* 2017 Jul 3;45(W1):W130–7.
37. Robak LA, Jansen IE, van Rooij J, Uitterlinden AG, Kraaij R, Jankovic J, et al. Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease. *Brain.* 2017 Dec 1;140(12):3191–203.
38. Bandres-Ciga S, Saez-Atienzar S, Bonet-Ponce L, Billingsley K, Vitale D, Blauwendraat C, et al. The endocytic membrane trafficking pathway plays a major role in the risk of Parkinson's disease. *Mov Disord.* 2019 Apr;34(4):460–8.
39. Jansen IE, Savage JE, Watanabe K, Bryois J, Williams DM, Steinberg S, et al. Genome-



wide meta-analysis identifies new loci and functional pathways influencing Alzheimer's disease risk. *Nat Genet* [Internet]. 2019 Jan 7; Available from: <http://dx.doi.org/10.1038/s41588-018-0311-9>

40. Saunders A E al. Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. - PubMed - NCBI [Internet]. [cited 2018 Oct 31]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/30096299>
41. LD Hub [Internet]. [cited 2018 Jun 26]. Available from: <http://ldsc.broadinstitute.org>
42. Hibar DP, Stein JL, Renteria ME, Arias-Vasquez A, Desrivieres S, Jahanshad N, et al. Common genetic variants influence human subcortical brain structures. *Nature*. 2015 Apr 9;520(7546):224–9.
43. Rapid GWAS of thousands of phenotypes for 337,000 samples in the UK Biobank [Internet]. Neale lab. [cited 2018 Jun 24]. Available from: <http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank>
44. Bulik-Sullivan B, ReproGen Consortium, Finucane HK, Anttila V, Gusev A, Day FR, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet*. 2015;47(11):1236–41.
45. Valdés EG, Andel R, Sieurin J, Feldman AL, Edwards JD, Långström N, et al. Occupational complexity and risk of Parkinson's disease. *PLoS One*. 2014 Sep 8;9(9):e106676.
46. Wu Y, Zheng Z, Visscher PM, Yang J. Quantifying the mapping precision of genome-wide association studies using whole-genome sequencing data. *Genome Biol*. 2017 May 16;18(1):86.
47. 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 Feb;41(2):145–51.
48. Rivas MA, Avila BE, Koskela J, Huang H, Stevens C, Pirinen M, et al. Insights into the genetic epidemiology of Crohn's and rare diseases in the Ashkenazi Jewish population. *PLoS Genet*. 2018 May;14(5):e1007329.
49. Iwaki H, Blauwendraat C, Leonard HL, Kim JJ, Liu G, Maple-Grødem J, et al. Genome-wide association study of Parkinson's disease progression biomarkers in 12 longitudinal patients' cohorts [Internet]. bioRxiv. 2019 [cited 2019 Apr 25]. p. 585836. Available from: <https://www.biorxiv.org/content/10.1101/585836v1.abstract>

**Table 1: Novel loci associated with Parkinson's disease.**

Summary statistics for 38 novel genome-wide significant PD variants. Columns include single nucleotide polymorphism ID (SNP), chromosome (CHR), base pair position (BP) based on hg19 build, nearest gene annotation for the variant, effect allele designation and frequency, as well as metrics for the odds ratio (OR), regression coefficient (beta), and standard error of the beta for the SNP from fixed-effects meta-analysis as well as the index of heterogeneity (I<sup>2</sup>). We also include four P values from: fixed-effects meta-analyses, random-effects meta-analyses, standard conditional analyses in 23andMe, and a conditional joint analysis approach (COJO).

**Table 2: Summary of genetic predictive model performance.**

These are estimates of performance for predictive models including single study estimates, estimates from meta-analyses across studies, as well as a two stage design. Here the best P value threshold column denotes the filtering value for SNP inclusion to achieve the maximal pseudo (Nagelkerke's) R<sup>2</sup>. The odds ratio (OR) column is the exponent of the regression coefficient (beta) from logistic regression of the polygenic risk score (PRS) on case status, with the standard error (SE) representing the precision of these estimates. These same metrics are derived across array types and datasets using random-effects meta-analyses. The area under the curve (AUC) is included as the most common metric for predictive model performance. In the table, \* denotes R<sup>2</sup> approximation adjusted for an estimated prevalence of 0.5%, equivalent to roughly half of the unadjusted R<sup>2</sup> estimates for the PRS. All calculations and reported statistics include only the PRS and no other parameters after adjusting for principal components 1-5, age and sex at variant selection in the NeuroX-dbGaP dataset.

**Table 3: Summary of significant functional inferences from QTL associations via Mendelian randomization for nominated genes of interest.**

Multi-SNP eQTL Mendelian randomization results focusing only on the most significant association per nearest gene to the PD risk variant of interest after Bonferroni correction. If a locus was significantly associated with both brain and blood QTLs after multiple test correction, we opted to show the most significant brain tissue derived association here after filtering for possible polygenicity (HEIDI P > 0.01). Effects with significant HEIDI P values may indicate a possible effect complicated by LD and are less likely to be a true causal association. All tested QTL summary statistics can be found in Supplementary Table S6. Effect estimates represent the change in PD odds ratio per one standard deviation increase in gene expression or methylation.

**FIGURES****Figure 1: Workflow and rationale summary.**

This figure describes study design and rationale behind the analyses included in this report.

**Figure 2: Manhattan plot.**

The nearest gene to each of the 90 significant variants are labeled in green for previously-identified loci and in blue for novel loci.  $-\log_{10}$  P values were capped at 40. Variant points are color coded red and orange, with orange representing significant variants at P 5E-08 and 5E-9 and red representing significant variants at P < 5E-9. The X axis represents the base pair position of variants from smallest to largest per chromosome (1-22).

**Figure 3: Predictive model details.**

A. The odds ratio of developing PD for each quartile of polygenic risk score (PRS) compared to the lowest quartile of genetic risk. B. PRS receiver-operator curves for the more inclusive 1805 variant PRS in the validation dataset as well as in the corresponding training dataset that was used for PRS thresholding and SNP selection.

## **AUTHOR CONTRIBUTIONS**

### **Study level analysis**

MAN, CB, CLV, KH, SB-C, DC, MT, DK, LR, JS-S, LK, LP, ABS

### **Additional analysis and data management**

MAN, CB, SB-C, AJN, AX, JY, JG, PMV, ABS

### **Design and funding**

MAN, CB, CLV, KH, SB-C, LP, MS, KM, MT, AB, JY, ZG-O, TG, PH, JMS, NW, DAH, JH, HRM, JG, PMV, RRG, ABS

### **Critical review and writing the manuscript**

MAN, CB, CLV, KH, SB-C, DC, MT, DAK, AJN, AX, JB, EY, RvC, JS-S, CS, MS, LK, LP, AS, HI, HL, FF, JRG, DGH, SWS, JAB, MM, OAA, J-CC, SL, JJ, LMS, MS, PT, KM, MT, AB, JY, ZG-O, TG, PH, JMS, NW, DAH, JH, HRM, JG, PMV, RRG, ABS

### **Disclosures and conflicts of interest**

Dr. Nalls reports that this work was carried out under a consulting contract with NIH, he also consults for Lysosomal Therapeutics Inc, Neuron23 Inc and Illumina Inc. Dr. Heilbron reports other support from 23andMe, during the conduct of the study outside the submitted work. Dr. Chang reports other support from Genentech outside the submitted work. Ms. Tan reports grants from Parkinson's UK during the conduct of the study. Dr. Noyce reports grants from Parkinson's UK, grants from Virginia Kieley benefaction, grants and non-financial support from GE Healthcare, personal fees from Profile, Bial and Britannia, outside the submitted work. Dr. von Coelln reports grants from American Brain Foundation, grants from Michael and Eugenia Brin Foundation, during the conduct of the study. Dr. Pihlstrøm reports grants from Southeastern Regional Health Authority, Norway, during the conduct of the study. Dr. Siitonen reports grants from Sigrid Juselius Foundation, during the conduct of the study. Dr. Scholz is a scientific advisory council member for the Lewy Body Dementia Association. Dr. Corvol reports grants from French Ministry of Health, during the conduct of the study; grants from Sanofi, personal fees from Ever Pharma, personal fees from Denali, personal fees from Biogen, personal fees from Air Liquide, personal fees from BrainEver, personal fees from Theranexus, outside the submitted work. Dr. L Shulman reports grants from NIH, outside the submitted work. Dr. Tienari reports no conflicts of interest. In addition, Dr. Tienari has a patent c9orf72 in the diagnosis and treatment of neurodegenerative disease pending. Dr. Toft reports grants from Research Council of Norway, grants from Regional Health Authority South-Eastern Norway, grants from Michael J. Fox Foundation, personal fees from Roche, outside the submitted work. Dr. Andreassen reports grants from Research Council of Norway, grants from KG Jebsen Stiftelsen, during the conduct of the study; personal fees from Lundbeck, outside the submitted work; In addition, Dr. Andreassen has a patent PCT/US2014/011014 pending. Dr. Bhangale

reports other from Genentech, outside the submitted work. Dr. BRICE reports grants from France Parkinson + FRC, grants from ANR - EPIG - Agence nationale de recherche, grants from ANR - JPND - Agence nationale de recherche, grants from RDS (Roger de Spoelberch Foundation), grants from France Alzheimer, grants from ENP -Ecole des neurosciences Paris, grants from Institut de France, grants from CHU de Nimes, grants from ERA NET, grants from ANR - EPIG, grants from APHP, outside the submitted work. Dr. Gan-Or reports personal fees from Lysosomal Therapeutics Inc, personal fees from Idorsia, personal fees from Inception Sciences, personal fees from Denali, personal fees from Prevail Therapeutics, outside the submitted work. Dr. Gasser reports grants from The Michael J Fox Foundation for Parkinson's Research, personal fees from UCB Pharma, other from "Joint Programming for Neurodegenerative Diseases" program, funded by the European Commission, personal fees from Novartis, personal fees from Teva, personal fees from MedUpdate, outside the submitted work; In addition, Dr. Gasser has a patent Patent Number: EP1802749 (A2) KASPP (LRRK2) gene, its production and use for the detection and treatment of neurodegenerative disorders issued. Dr. Heutink is a consultant for NEURON23, inc. Dr. J Shulman reports grants from Burroughs Wellcome Fund during the conduct of the study; grants from National Institutes of Health and personal fees from Helis Medical Foundation, outside the submitted work. Dr. Hinds reports other from 23andMe, Inc., outside the submitted work. Dr. Morris reports grants from Parkinson's UK, grants from Medical Research Council, grants from Cure Parkinson's Trust, during the conduct of the study; grants from PSP Association, grants from CBD Solutions, personal fees from Teva, personal fees from Boehringer Ingelheim, personal fees from GSK, grants from Drake Foundation, personal fees from Biogen, personal fees from UCB, personal fees from Biohaven, outside the submitted work. Dr. Graham reports other from Genentech, during the conduct of the study; other from Genentech, outside the submitted work. All others have no disclosures or potential conflicts of interest.

Table 1.

SNP	CHR	BP	Nearest Gene	Effect allele	Other allele	Effect allele frequency	OR	Low 95% CI	High 95% CI	Beta	SE	P, fixed-effects	P, COJO	P, conditional	P, random-effects	I2, %
rs6658353	1	161469054	FCGR2A	c	g	0.501	1.07	1.05	1.09	0.065	0.009	6.10E-12	4.69E-12	1.38E-05	3.71E-05	40.2
rs11578699	1	171719769	VAMP4	t	c	0.195	0.93	0.91	0.95	-0.070	0.012	4.47E-09	4.45E-09	2.63E-03	1.09E-07	5.1
rs76116224	2	18147848	KCNS3	a	t	0.904	1.12	1.08	1.16	0.110	0.019	1.27E-08	1.27E-08	3.75E-07	1.27E-08	0
rs2042477	2	96000943	KCNIP3	a	t	0.242	0.94	0.92	0.96	-0.066	0.012	1.38E-08	1.48E-08	3.49E-05	1.38E-08	0
rs6808178	3	28705690	LINC00693	t	c	0.379	1.07	1.05	1.09	0.066	0.010	8.09E-12	7.18E-12	8.84E-05	8.09E-12	0
rs55961674	3	122196892	KPNA1	t	c	0.172	1.09	1.06	1.12	0.086	0.013	9.98E-12	8.30E-12	2.80E-06	9.98E-12	0
rs11707416	3	151108965	MED12L	a	t	0.367	0.94	0.92	0.96	-0.063	0.010	1.13E-10	1.02E-10	2.66E-04	1.77E-07	10.9
rs1450522	3	161077630	SPTSSB	a	g	0.674	0.94	0.92	0.96	-0.062	0.010	5.01E-10	4.90E-10	3.51E-04	2.27E-05	24.6
rs34025766	4	17968811	LCORL	a	t	0.159	0.92	0.90	0.94	-0.084	0.013	2.87E-10	2.82E-10	7.43E-06	2.87E-10	0
rs62333164	4	170583157	CLCN3	a	g	0.326	0.94	0.92	0.96	-0.064	0.010	2.00E-10	1.77E-10	5.10E-05	2.17E-05	21.3
rs26431	5	102365794	PAM	c	g	0.703	1.06	1.04	1.09	0.062	0.010	1.57E-09	1.65E-09	6.00E-03	2.36E-07	7.9
rs11950533	5	134199105	C5orf24	a	c	0.102	0.91	0.88	0.94	-0.092	0.016	7.16E-09	6.73E-09	5.08E-04	2.68E-08	1.9
rs9261484	6	30108683	TRIM40	t	c	0.245	0.94	0.92	0.96	-0.064	0.011	1.62E-08	1.43E-08	1.26E-06	1.62E-08	0
rs12528068	6	72487762	RIMS1	t	c	0.284	1.07	1.05	1.09	0.066	0.010	1.63E-10	1.79E-10	9.80E-06	1.63E-10	0
rs997368	6	112243291	FYN	a	g	0.805	1.07	1.05	1.10	0.071	0.012	1.84E-09	1.97E-09	2.61E-05	1.84E-09	0
rs75859381	6	133210361	RPS12	t	c	0.967	0.80	0.75	0.86	0.221	0.034	1.04E-10	9.67E-11	1.09E-06	1.04E-10	0
rs76949143	7	66009851	GS1-124K5.11	a	t	0.051	0.87	0.82	0.91	-0.143	0.025	1.43E-08	1.51E-08	5.47E-09	2.04E-06	12.3
rs2086641	8	130901909	FAM49B	t	c	0.723	0.94	0.92	0.96	-0.061	0.011	1.81E-08	1.57E-08	6.07E-06	1.81E-08	0
rs6476434	9	34046391	UBAP2	t	c	0.734	0.94	0.92	0.96	-0.062	0.011	6.58E-09	6.56E-09	2.74E-04	6.58E-09	0
rs10748818	10	104015279	GBF1	a	g	0.851	0.92	0.90	0.95	0.079	0.013	1.05E-09	1.23E-09	7.47E-06	1.05E-09	0
rs7938782	11	10558777	RNF141	a	g	0.878	1.09	1.06	1.12	0.087	0.015	2.12E-09	1.97E-09	2.17E-07	2.12E-09	0
rs7134559	12	46419086	SCAF11	t	c	0.404	0.95	0.93	0.97	-0.054	0.010	3.96E-08	3.80E-08	1.69E-02	1.84E-05	25.2

rs11610045	12	133063768	FBRSL1	a	g	0.490	1.06	1.04	1.08	0.060	0.009	1.77E-10	1.62E-10	3.57E-05	8.79E-07	19.5
rs9568188	13	49927732	CAB39L	t	c	0.740	1.06	1.04	1.09	0.062	0.011	1.15E-08	1.11E-08	4.29E-06	2.46E-04	21.4
rs4771268	13	97865021	MBNL2	t	c	0.230	1.07	1.05	1.09	0.068	0.011	1.45E-09	1.67E-09	1.41E-04	1.45E-09	0
rs12147950	14	37989270	MIPOL1	t	c	0.438	0.95	0.93	0.97	-0.053	0.010	3.54E-08	3.58E-08	1.06E-03	3.54E-08	0
rs3742785	14	75373034	RPS6KL1	a	c	0.787	1.07	1.05	1.10	0.071	0.012	1.92E-09	2.08E-09	2.22E-06	8.18E-06	24.8
rs2904880	16	28944396	CD19	c	g	0.309	0.94	0.92	0.96	-0.065	0.011	7.87E-10	8.68E-10	1.39E-05	7.87E-10	0
rs6500328	16	50736656	NOD2	a	g	0.599	1.06	1.04	1.08	0.059	0.010	1.82E-09	1.53E-09	1.43E-03	1.82E-09	0
rs12600861	17	7355621	CHRNA1	a	c	0.648	0.95	0.93	0.96	-0.057	0.010	1.01E-08	1.15E-08	5.10E-03	1.01E-08	0
rs2269906	17	42294337	UBTF	a	c	0.653	1.07	1.04	1.09	0.063	0.010	6.24E-10	8.63E-09	1.17E-05	6.24E-10	0
rs850738	17	42434630	FAM171A2	a	g	0.606	0.93	0.91	0.95	-0.071	0.011	1.29E-11	3.55E-10	4.18E-04	2.17E-07	17
rs61169879	17	59917366	BRIP1	t	c	0.164	1.09	1.06	1.11	0.082	0.013	9.28E-10	9.40E-10	9.07E-07	6.21E-06	16.4
rs666463	17	76425480	DNAH17	a	t	0.833	1.08	1.05	1.11	0.076	0.013	3.20E-09	2.90E-09	1.62E-05	4.17E-04	41
rs1941685	18	31304318	ASXL3	t	g	0.498	1.05	1.04	1.07	0.053	0.009	1.69E-08	1.61E-08	1.64E-08	1.69E-08	0
rs8087969	18	48683589	MEX3C	t	g	0.550	0.94	0.93	0.96	-0.058	0.010	1.41E-08	1.46E-08	1.09E-04	1.41E-08	0
rs77351827	20	6006041	CRLS1	t	c	0.128	1.08	1.05	1.11	0.080	0.014	8.87E-09	7.94E-09	1.84E-05	4.38E-07	11.2
rs2248244	21	38852361	DYRK1A	a	g	0.283	1.07	1.05	1.10	0.071	0.011	2.74E-11	2.51E-11	6.31E-05	8.78E-06	34.3

Table 2.

Study	Max P threshold	pseudo R2 from PRS*	Beta	SE	P	OR, highest quartile PRS	95% CI, highest quartile PRS	N SNPs	N samples	AUC	95% CI (DeLong)	Sensitivity	Specificity	Positive predictive value (PPV)	Negative predictive value (NPV)	Balanced accuracy
Training dataset: IPDGC - Neurox	1.35E-03	0.029	0.553	0.022	8.99E-135	3.74	3.35 - 4.18	1809	11,243	0.640	0.630 - 0.650	0.569	0.632	0.591	0.611	0.601
Test dataset: HBS	4.00E-02	0.054	0.709	0.072	8.28E-23	6.25	4.26 - 9.28	1805	999	0.692	0.660 - 0.725	0.628	0.686	0.691	0.623	0.657

Table 3.

Gene	Probe	CHR	Probe, BP	Top SNP, BP	Top SNP	N SNPs	QTL reference	Effect	SE	P	Bonferroni adjusted P
VAMP4	ENSG00000117533	1	171,690,343	171,717,417	rs10913587	98	Vösa et al. 2018 - blood expression	-0.272	0.005	5.67E-07	1.19E-04
KCNIP3	ENSG00000115041	2	96,007,438	95,989,766	rs3772034	14	Qi et al. 2018 - brain expression	-0.161	0.004	1.12E-05	1.15E-03

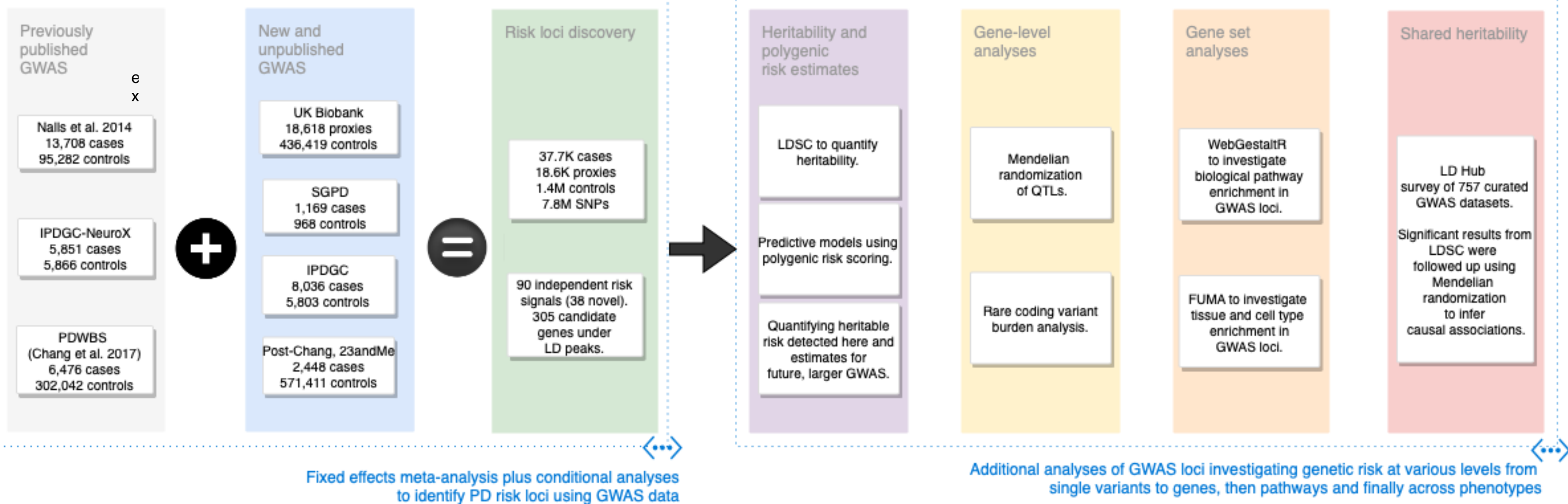
MAP4K4	ENSG00000071054	2	102,410,880	102,338,377	rs6733355	3	Vösa et al. 2018 - blood expression	1.119	0.24	2.32E-06	4.87E-04
TMEM163	ENSG00000152128	2	135,344,950	135,248,544	rs598668	28	Qi et al. 2018 - brain expression	0.074	0.02	3.55E-07	3.65E-05
KPNA1	ENSG00000114030	3	122,187,294	122,201,610	rs73190142	110	Vösa et al. 2018 - blood expression	0.310	0.05	1.56E-06	3.28E-04
GAK	ENSG00000178950	4	884,612	906,131	rs11248057	1	Qi et al. 2018 - brain expression	0.508	0.10	7.47E-07	7.69E-05
CAMK2D	ENSG00000145349	4	114,527,635	114,730,260	rs115671064	146	Vösa et al. 2018 - blood expression	-0.006	0.05	5.74E-06	1.21E-03
PAM	ENSG00000145730	5	102,228,247	102,118,633	rs2432162	679	Vösa et al. 2018 - blood expression	-0.031	0.01	2.08E-06	4.36E-04
LOC100131289	cg21339923	6	27,636,378	27,636,378	rs78149975	2	Qi et al. 2018 - brain methylation	-0.094	0.02	1.53E-06	3.06E-04
TRIM40	cg01641092	6	30,094,300	30,094,315	rs9261443	8	Qi et al. 2018 - brain methylation	0.072	0.01	6.15E-06	1.23E-03
HLA-DRB5	cg26036029	6	32,552,443	32,570,311	rs34039593	8	Qi et al. 2018 - brain methylation	-0.153	0.02	7.53E-10	1.51E-07
GNPMB	ENSG00000136235	7	23,295,156	23,294,668	rs858274	74	Qi et al. 2018 - brain expression	0.090	0.01	2.73E-21	2.81E-19
CTSB	ENSG00000164733	8	11,713,495	11,699,279	rs4631423	33	Qi et al. 2018 - brain expression	0.150	0.04	4.37E-09	4.50E-07
BIN3	ENSG00000147439	8	22,502,296	22,456,517	rs71513892	32	Qi et al. 2018 - brain expression	0.046	0.01	1.43E-06	1.48E-04
SH3GL2	ENSG00000107295	9	17,688,103	17,684,784	rs10756899	15	Qi et al. 2018 - brain expression	0.252	0.05	5.83E-08	6.00E-06
ITGA8	ENSG00000077943	10	15,659,036	15,548,925	rs7910668	6	Qi et al. 2018 - brain expression	-0.201	0.05	6.13E-05	6.32E-03
RNF141	ENSG00000110315	11	10,548,001	10,553,355	rs4910153	120	Vösa et al. 2018 - blood expression	-0.054	0.05	6.25E-07	1.31E-04
IGSF9B	cg25790212	11	133,800,774	133,800,477	rs11223626	1	Qi et al. 2018 - brain methylation	-0.172	0.04	3.24E-06	6.48E-04
FBRSL1	cg03621470	12	133,137,479	133,138,334	rs10781619	16	Qi et al. 2018 - brain methylation	-0.057	0.01	6.35E-05	1.27E-02
CAB39L	ENSG00000102547	13	49,950,524	49,918,175	rs35214871	30	Qi et al. 2018 - brain expression	0.097	0.02	3.51E-08	3.62E-06
GCH1	ENSG00000131979	14	55,339,148	55,348,837	rs3825611	6	Qi et al. 2018 - brain expression	0.113	0.03	2.76E-04	2.85E-02
SYT17	ENSG00000103528	16	19,229,472	19,273,554	rs727747	4	Qi et al. 2018 - brain expression	0.177	0.05	1.54E-04	1.58E-02

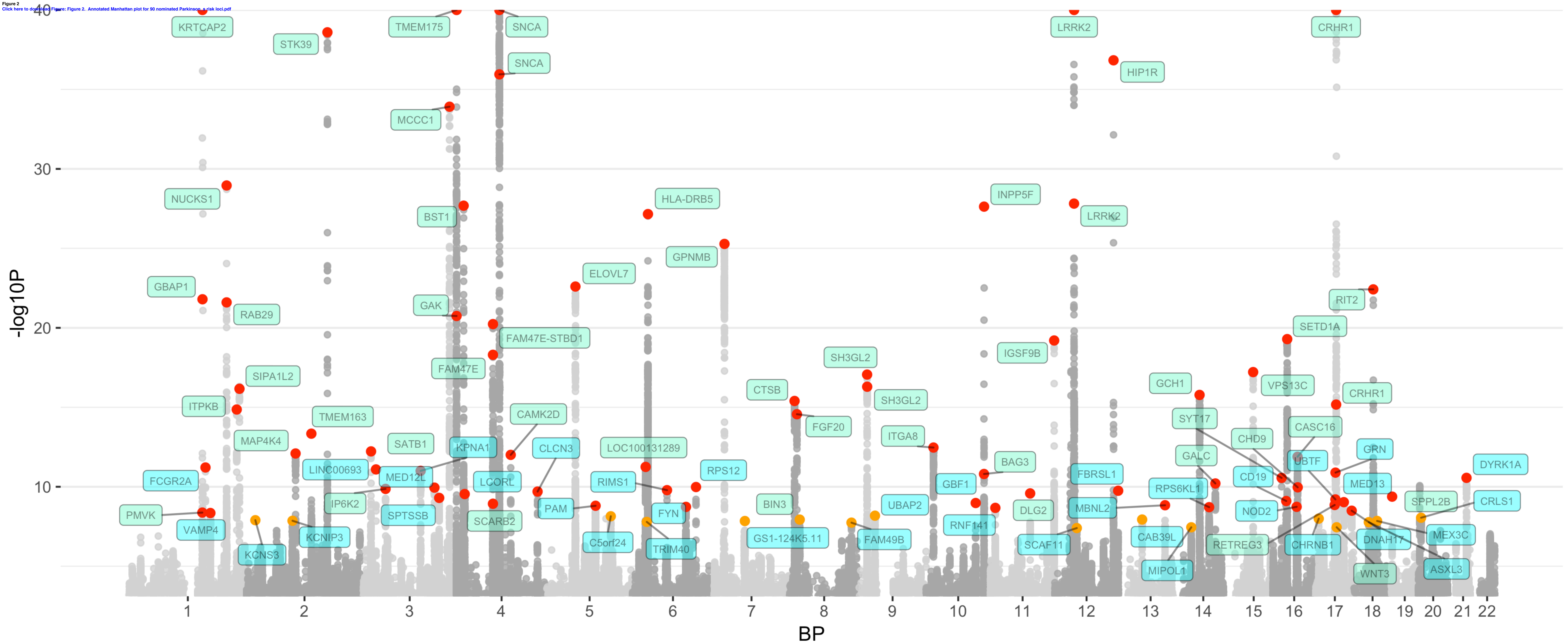
SETD1A	ENSG00000099381	16	30,982,526	30,950,352	rs7206511	34	Vösa et al. 2018 - blood expression	- 0.71 0	0.0 9	2.75E- 13	5.77E-11
CHRN1	ENSG00000170175	17	7,354,703	7,373,595	rs6048885 5	18	Qi et al. 2018 - brain expression	0.11 5	0.0 3	1.67E- 05	1.72E-03
UBTF	ENSG00000108312	17	42,290,697	42,297,631	rs1138447 52	34	Vösa et al. 2018 - blood expression	- 0.46 6	0.0 9	5.68E- 06	1.19E-03
MAPT	ENSG00000186868	17	44,038,724	44,862,347	rs199502	6	Qi et al. 2018 - brain expression	0.26 5	0.0 3	7.13E- 24	7.35E-22
WNT3	ENSG00000108379 .5	17	44,875,148	44,908,263	rs9904865	2	GTEx v7 - substantia nigra brain expression	- 0.08 2	0.0 2	4.01E- 06	4.81E-05
DNAH17	cg09006072	17	76,425,972	76,427,732	rs589582	3	Qi et al. 2018 - brain methylation	0.10 0	0.0 2	2.44E- 05	4.88E-03
MEX3C	ENSG00000176624	18	48,722,797	48,731,131	rs1245891 6	40	Vösa et al. 2018 - blood expression	- 0.29 1	0.0 5	5.28E- 05	1.11E-02

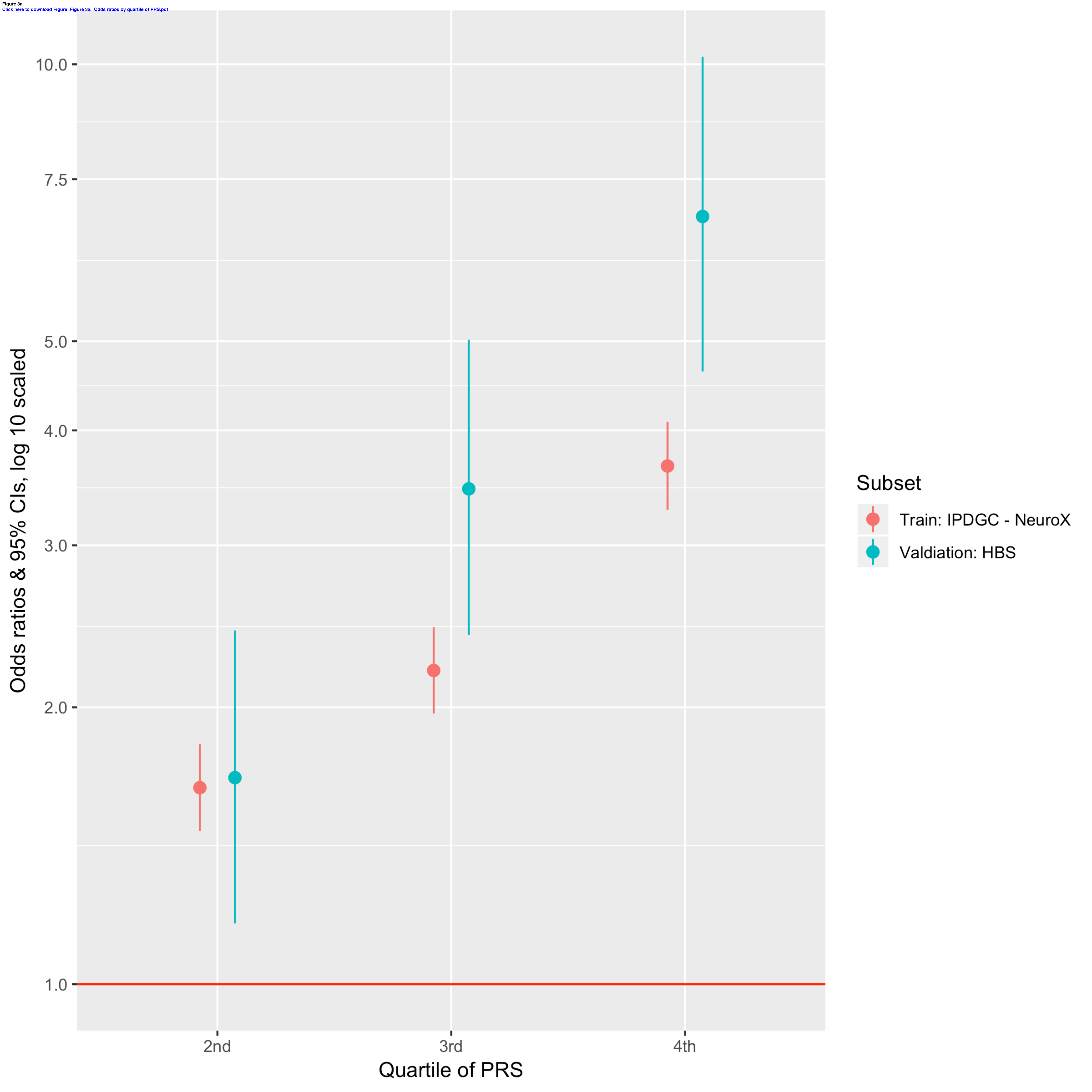


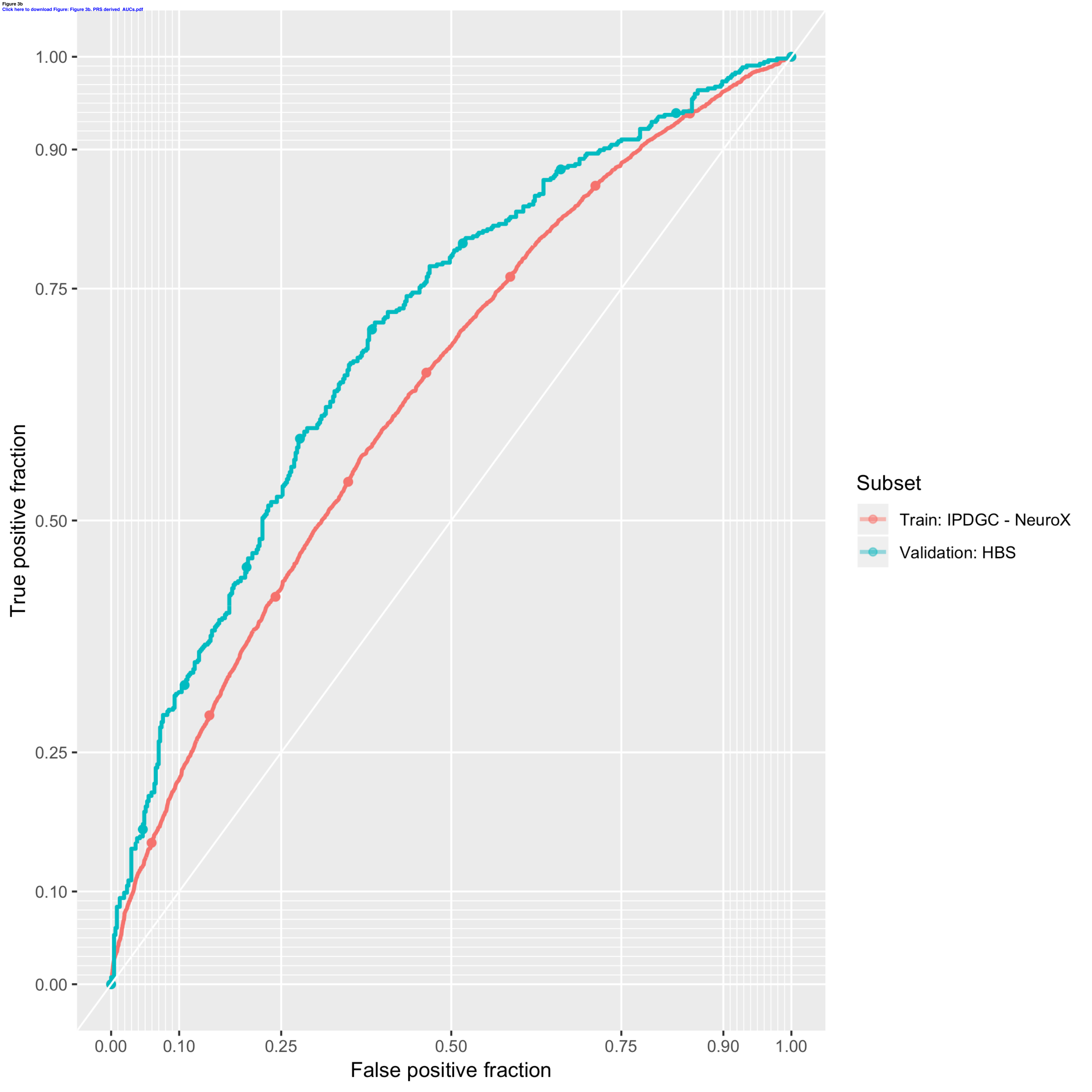
Figure 1

[Click here to download Figure: Figure 1. workflow and rationale.pdf](#)









## SUPPLEMENTARY MATERIALS

### METHODS SUPPLEMENT

#### GWAS Study design and risk locus discovery

Three primary sources of data were used for discovery analyses, these include three previously published studies, 13 new datasets, and proxy-case data from the UK BioBank (UKB). Previous studies include summary statistics from the meta-analyses of GWAS published in Nalls *et al.* 2014, GWAS summary statistics from the 23andMe Web-Based Study of Parkinson's Disease (PDWBS) cohort described in Chang *et al.* 2017, and the publicly available NeuroX dataset from the International Parkinson's Disease Genomics Consortium (IPDGC) that had been used in the two prior publications as a replication sample. These cohorts have been reported in detail elsewhere (1,2). We also included 13 new case-control sample series for meta-analyses through either publicly available data or collaborations (please see Supplementary Table S1 for details regarding these studies). All new samples from the 13 new datasets underwent similar standardized quality control for inclusion, mirroring that of the previous studies (3).

For each dataset we attempted to generate summary statistics for GWAS meta-analyses as uniformly as possible. We used additive allele dosages from imputation in a logistic regression framework adjusting for age at onset in cases (age at most recent examination for controls), biological sex and up to the first five principal components from stepwise modeling per cohort to account for population substructure. The System Genomics of Parkinson's Disease (SGPD) study did not use logistic regression, instead mixed modeling was used due to sample relatedness. UKB data was analyzed slightly differently from standard case-control GWAS because of the use of proxy cases. In the UKB, genome-wide association study by proxy (GWAX) was carried out as per Liu (4), adjusting for age, sex, the first ten principal components, genotyping batch and Townsend index(5). Similar to previous publications, study-level summary statistics were filtered for inclusion criteria of imputation quality score  $> 0.3$  and MAF  $> 1\%$  (1,2,6). We retained 2 variants for study below the MAF minimum of 1%, these include known coding risk factors rs34637584 (*LRRK2*, p.G2019S) and rs76763715 (*GBA*, p.N370S). Our GWAS meta-analyses spanned all three data sources (previously published case-control datasets, new case-control datasets and proxy-case data from the UKB) including 37,688 cases, 18,618 proxy-cases and 1,474,097 controls with data for 7,784,415 SNPs that passed inclusion filtering. This analysis utilized fixed-effects meta-analyses as implemented in METAL to combine summary statistics across all sources (7).

#### Samples and quality control

Initial sample inclusion criteria include: age at disease onset or last examination at 18 years of age or older, minimum sample call rate of  $>95\%$ , majority European ancestry confirmed through principal-components, no genetically ascertained relation to other samples in the meta-analysis (proportional sharing at a maximum of 12.5%) at the cousin level or closer (except in the case of the System Genomics of Parkinson's Disease (SGPD) which utilized mixed modeling to account

for related samples) and no heterozygosity outliers past +/- 15% (quantified by F estimates in PLINK)(8) . Studies from IPDGC collaborators (including previously published series) and those in publicly available databases were checked for relatedness and duplicates using identity-by-descent filtering to remove related samples both within and across datasets contributing to this effort (also using PLINK to generate PI\_HAT estimates of relatedness and excluding samples with > 12.5% PI\_HAT). All publicly available datasets and newly genotyped IPDGC datasets were clustered to identify cryptically related samples between each other as well as the UKB using similar methods in PLINK. These studies screened for relatedness at the National Institute on Aging site include: Baylor College of Medicine / University of Maryland, Finnish Parkinson's, Harvard Biomarker Study (HBS), McGill Parkinson's, Oslo Parkinson's Disease Study, Parkinson's Disease Biomarker's Program (PDBP), Parkinson's Progression Markers Initiative (PPMI), Spanish Parkinson's (from IPDGC), Tubingen Parkinson's Disease cohort (CouragePD), Vance (dbGap phs000394), UK PDMED (CouragePD), UK BioBank (UKB), IPDGC (Nalls et al. 2014 discovery phase) and NeuroX - dbGaP (phs000918.v1.p1). If a sample overlapped in two studies, it was removed from the larger study. Samples derived from 23andMe were checked internally for relatedness both among themselves and publicly available PD datasets using similar methods as previously described in earlier publications before being incorporated into this analysis (studies included in Nalls et al 2014, as well as the NeuroX-dbGaP, HBS, PPMI, PDBP, UKB) (1,9). There is a low likelihood for overlap between the SGPD and IPDGC samples due to geographic differences in sample collection (Australia versus Europe and North America). This is summarized in Supplementary Table S1.

For case diagnosis, all of the new cases, except for the post-Chang *et al.* dataset from 23andMe, conformed to the generally used criteria of a clinic visit and standard UK Brain Bank criteria with a modification to allow the inclusion of cases that had a family history of PD. In the most recent dataset from 23andMe (post-Chang *et al.*) cases were ascertained via the criteria of self-report of diagnosis, similar to previous publications (1,9,10). Status as "post-Chang" is denoted by enrollment as a case at a time after the analysis for the PDWBS closed. Post-Chang 23andMe samples were imputed using a combination of Finch for phasing (an in-house developed fork of Beagle) and miniMac2 for imputation with all-ethnicity samples from the September 2013 release of 1000 Genomes Phase1 as reference haplotypes(11–13). Both the Finnish Parkinson's study and post-Chang sample series used population controls to some degree, with inclusion criteria of no self-report of neurological disease, memory loss, tremor or family history of PD when available. UKB proxy-cases were defined as the report of a first degree relative with PD and no International Classification of Diseases (ICD-10) or self-report of actual PD. UKB cases with self-reported PD were excluded, this includes 727 cases that self-reported PD without family history to keep the GWAX analysis homogenous and parsimonious. UKB controls were free of first degree family history of PD or PD by self-report or ICD-10 designation. No censoring was made on parent age in the UKB. Additional summary statistics were generated for the UKB censoring on those who answered "Do not know" or "Prefer not to answer" to any of the possible illnesses. This reduced control counts down to 369,711. The reduced UKB results were highly correlated with the larger set included in all analyses ( $rG = 0.9865$ ,  $SE = 0.0011$ ), so we opted to utilize the results including more samples. All participants donated DNA samples and provided informed consent for participation in genetics studies.

Prior to imputation, SNPs were filtered using similarly uniform criteria for inclusion, such as: minimum genotype call rate of >95%, a minor allele frequency (MAF) of >0.1%, a Hardy-Weinberg equilibrium P values >1 E-04 in controls (1% MAF and HWE  $p > 1E-06$  for SGPD), and non-random missingness by phenotype or haplotype at P values of >1E-04. Palindromic SNPs were also removed. The Nalls *et al.* 2014 and Chang *et al.* 2017 samples were imputed with Minimac2 using 1000 Genomes phase 1 haplotypes (13). All additional sample series except for the post-Chang *et al.* 2017 samples from 23andMe were imputed using the Haplotype Reference Consortium (HRC) on the University of Michigan imputation server under default settings with Eagle v2.3 phasing and minimac4 imputation based on reference panel HRC r1.1 2016(14,15). UKB genotype data for proxy-cases and controls was downloaded in April 2018 as provided by the analysis group at the Wellcome Trust Centre for Human Genetics at the University of Oxford and is fully detailed at <http://biobank.ctsu.ox.ac.uk>(16).

### **Study-level analyses and meta-analyses**

Summary statistics for each study were generated uniformly using additive allele dosages from imputation in a logistic regression framework adjusting for age at onset in cases (age at most recent examination for controls), biological sex and up to the first five principal components from stepwise modeling per cohort. The SGPD study did not use logistic regression, as an inclusion of related samples required mixed modeling. The Finnish Parkinson's study was unable to include age as a covariate due to collinearity issues as the controls were significantly older than the cases. Age at onset or last exam data was not available for the UK PDMED dataset. Age at last exam was used for the Oslo Parkinson's Disease Study dataset.

UKB data was analyzed slightly differently from standard case-control GWAS because phenotypes were available on relatives of the individuals with genotypes. Genome-wide association study by proxy (GWAX) was carried out as per Liu (4), adjusting for age, sex, the first ten principal components, genotyping batch and Townsend index(5). GWAX is a analysis method shown to carry out reliable and generalizable association analyses on biobank / population scale data utilizing at-risk samples instead of true cases. Summary statistics were also filtered at a MAF > 1% for inclusion. Per SNP effect estimates were transformed using the method from Lloyd-Jones et al. assuming the lifetime probability of disease (K) at 0.02 and then converted to standard case-control scale (with the total proxy cases equivalent to ~4.5K cases) (17).

Similar to previous publications, study-level summary statistics were filtered for inclusion criteria of reasonable beta estimates imputation quality score > 0.3 and MAF > 1% (1,6,9). We retained 2 variants for study below the MAF minimum of 1%, these include known coding risk factors include rs34637584 (LRRK2, p.G2019S) and rs76763715 (GBA, p.N370S). Our GWAS meta-analyses spanned all three data sources (previously published case-control datasets, new case-control datasets and proxy-case data from the UKB) including 37,688 cases, 18,618 proxy-cases and 1,474,097 controls with data for 7,784,415 SNPs that passed inclusion filtering. Inclusion filtering included variants seen in at least five of the 17 sets of summary statistics and

differences in the per variant minimum and maximum MAFs across all studies at less than the 99th percentile (< 15% frequency differences). Four of the studies involved in our meta-analysis were genotyped using the NeuroX array, which is enriched for PD risk loci (see Supplementary Table S1) (18). The motivation for choosing five out of 17 studies was to reduce potential bias due to targeted genotyping in the four NeuroX array studies.

Overall genomic inflation was minimal with a raw lambda estimate of 1.170 and a lambda scaled to 1,000 cases and 1,000 controls at 1.002(18–20). Well behaved lambda estimates in conjunction with an LD score intercept of 0.991 and a number of well replicated risk loci spanning larger tracts of the genome (i.e. *MAPT* and *HLA/MHC*), led us to not implement genomic control on the meta-analysis level. All quality control metrics on a per study basis including lambda estimates across strata of minor allele frequencies are summarized in Supplementary Table S1.

As previously stated, we did not correct for genomic control. Both the previous datasets and new datasets exhibited LD score regression intercepts close to 1, with 0.988 for the previous datasets and 0.975 for the new datasets, suggesting that our results are unlikely to be due to population stratification (21). As detailed in Supplementary Table S1, scaled lambdas were acceptable across the allele frequency spectrum with minor inflations for studies using genotyping arrays with targeted PD-centric content.

We also queried the 17 novel variants of interest from Chang et al., 2017 for their inclusion under GWAS peaks from this report(9). We identified proxy variants tagging genome-wide significant peaks in this report (Supplementary Tables S2). Pairwise linkage disequilibrium was calculated from the European subset of the 1000 genomes data. Proxies are summarized in Supplementary Table S3. Four loci are directly validated using the same SNPs, 6 are tagged by a strong proxy ( $r^2 > 0.8$ ), 4 tagged by a weaker proxy ( $0.2 < r^2 < 0.5$ ) and an additional 3 with no genome-wide significant proxy within 500kb observed (rs353116, rs143918452 and rs78738012). One missing variant, rs78738012, is tagged under one of our peaks of interest by a SNP not passing genome-wide significance near the gene *CAMK2D*. The two untagged variants are likely due to expanding our analysis past genotyped SNPs from the NeuroX array and utilizing updated imputation references.

### **Conditional-joint analysis to nominate variants of interest**

To nominate variants of interest, we employed a conditional and joint analysis strategy (GCTA-COJO, <http://cnsgenomics.com/software/gcta/>) as a means to algorithmically identify variants that best account for the heritable variation within and across nearby loci(22). This is particularly useful in scenarios where only basic summary statistics are available for a majority of samples in a meta-analysis and additional participant level analyses are logistically prohibitive. For this analysis, we used the full meta-analysis summary statistics in conjunction with the largest single site collection of HRC-level imputed PD and control data as a reference for linkage disequilibrium patterns in the conditional-joint workflow (described below and in the Methods Supplement).



Using the consensus IPDGC data cleaning and imputation workflow, we assembled a reference set of 17,188 cases and 22,875 controls at variants overlapping with the locus discovery analysis results that passed quality control on average in 74.3% of samples incorporating soft call genotypes at a minimum imputation quality of 0.30. This set includes data from all samples described in Supplementary Table S1, except the 23andMe post-Chang *et al.*, SGPD, UK PDMED and UKB sample series. This aggregate dataset included previously described samples series such as the Dutch GWAS, German, UK and US IPDGC series plus the NeuroX-dbGaP and Myers-Faroud datasets from dbGaP(18,23,24). We assembled this large PD-specific LD reference to help better ascertain LD patterns at PD loci, particularly the LRRK2 and GBA regions where rarer risk variants are located. The COJO analysis was run using default analysis parameters including a significance threshold of  $P < 5E-8$  and a window specification of 1 megabase. Additional analyses described below were utilized to further scrutinize putative associated variants and account for possible differential linkage disequilibrium (LD) signatures in multiple ways, including utilizing the massive single site reference data from 23andMe in further conditional analyses. If a variant nominated during the COJO phase of analysis was greater than 1 megabase from any of the genome-wide significant loci nominated in Chang *et al.* 2017, we considered this to be a novel locus.

### **Additional filtering of nominated variants**

We instituted two additional filters after fixed-effects and COJO analyses. These additional filters exclude variants that 1) had a random-effects P value across all datasets  $> 4.67E-04$  and 2) a conditional analysis  $P > 4.67E-04$  using participant level 23andMe genotype data. This Bonferroni multiple testing threshold is based on up to 107 nominated variants at this stage of filtering, of which 90 passed these criteria. Random-effects meta-analysis P values were generated under the residual maximum likelihood method using the R package metafor(25). Forest plots for all loci of interest are available in the Supplemental Appendix. Conditional analyses were carried out using 23andMe pooled data analyses including all available 23andMe data (from Nalls *et al.* 2014, PDWBS and the post-Chang *et al.* 2017 datasets combined). For the participant level conditional analyses in 23andMe, all nominated variants per chromosome were included in a single logistic regression model with appropriate covariates, then parameter estimates per variant were extracted. Conditional analyses on a per chromosome interval instead of a locus or megabase interval should adjust for possible longer range LD associations. For more information on variant filtering, please see Supplementary Table S2 summarizing all variants nominated. We defined nominated risk variants as sharing a single locus if they are within +/- 250kb of each other.

### **Additional sensitivity analyses**

Mirroring our previous workflows used in the initial meta-analysis, we conducted 17 “leave-one-out” meta-analyses (LOOMA), excluding one dataset’s summary statistics each time. We also carried out a similar set of meta-analyses separately for previously published data versus the combined set of 14 unpublished case-control and proxy-case datasets.

We employed a number of sensitivity analyses to further investigate our data and results as opposed to using a two-stage study design in order to maximize discovery power. LD score regression was utilized to calculate the genetic correlation between the meta-analysis of previously published datasets and the meta-analysis of the 14 unpublished datasets(21). Additionally, all datasets described in Supplementary Table S1 were meta-analyzed using fixed-effects, stratified by diagnostic criteria of either self-reported PD (post-Chang and PDWBS) or clinically ascertained PD (all other datasets excluding data from the UKB and Nalls et al. 2014). Summary statistics stratified by diagnosis were then used to compare genetic correlations across clinically defined, self-reported and proxy-case derived datasets using LD score regression.

Due to the relatively small sample sizes of many of the datasets involved in this study, we could not run LD score regression on all combinations of LOOMAs. Instead, we opted to calculate linear regression models comparing the beta coefficients for the left-out dataset with the beta coefficients from a meta-analysis of the remaining (non-left out) datasets, stratified by all, novel and known PD risk loci.

Similar to above, we calculated random-effects meta-analysis p-values using the residual maximum likelihood method in the R package metafor. We also generated forest plots from this analysis that reflect the distributions of effect estimates across studies per SNP, summarized by the index of heterogeneity (I<sup>2</sup>) statistics in Table 1 and Supplementary Table S2. This statistic estimates possible variance accounted for by study heterogeneity.

Final sensitivity analyses included “leave-one-out” meta-analyses (LOOMA) comparisons of each dataset to a meta-analysis of the remaining datasets. This analysis focused on comparing the log odds ratios (termed beta here) per SNP identified in the GWAS analyses across all cohorts. After adjusting for multiple test correction for 17 tests ( $P < 0.003$  for significance) in regressions of up to 90 betas per iteration, we noted only 5 departures from significant correlations between the withheld and included datasets. These non-significant results included only novel loci in the Baylor / University of Maryland dataset, the Finnish Parkinson’s dataset, the Harvard Biomarker Study (HBS), the Parkinson’s Disease Biomarkers Program (PDBP) and the Parkinson’s Progression Markers Initiative (PPMI). For these five studies, correlations were significant in the known and all loci strata of variants. This is likely be related to statistical power for detecting recently identified risk variants in this subset of smaller studies. While there may be some caution in utilizing UKB proxy-cases, our data shows that the UKB data was significantly representative of other datasets, with high  $r^2$  estimates across novel ( $r^2 = 0.714$ , 38 variants), known ( $r^2 = 0.897$ , 47 variants) and all variants strata ( $r^2 = 0.866$ , 85 variants) in the LOOMAs. We view these LOOMAs as a means of detecting an outlier study and estimating generalizability in the context of the 90 nominated variants. Forest plots included in the Supplemental Appendix compare each study on a per variant basis.

### **Refining heritability estimates and determining extant genetic risk**

The primary tool for risk profiling used here was the R package PRSice2(26). This package carries out polygenic risk score (PRS) profiling in the standard weighted allele dose manner as we have previously described (1,2,6,27–30). In addition, PRSice incorporates permutation testing where case and control labels are swapped in the withheld samples to generate an empirical P. This workflow identifies the best P thresholds for variant inclusion while simultaneously carrying out LD pruning. In many cases this best P threshold for PRS construction is below what is commonly regarded as genome-wide significant. This workflow also uses P value aware LD pruning to facilitate identifying the best P thresholds for variant inclusion into the PRS below what is commonly regarded as genome-wide significance levels. PRS analyses were conducted in sample series with readily accessible participant level IPDGC GWAS data.

A two stage design was also employed, training on the largest single array study (NeuroX-dbGaP) and then tested on the second largest study (HBS) using the same array. These two targeted array studies were chosen for three reasons: precedent in the previous publications where the NeuroX-dbGaP dataset was used in PRS comparisons; direct genotyping of larger effect rare variants in GBA and LRRK2; participant level genotypes for these datasets are publicly available.

To facilitate risk profiling analyses with less bias, we regenerated the meta-analysis summary statistics excluding the NeuroX-dbGaP and HBS datasets. To select SNPs we ran the PRSice workflow, utilizing beta weights from the meta-analyses excluding our studies of interest. LD clumping was carried out under recommended default settings (window size = 250kb,  $r^2 > 0.1$ ). Next 10,000 permutations were used to generate empirical P estimates for each GWAS derived P threshold ranging from 5E-08 to 1E-04, by increments of 5E-08, then again with GWAS derived P thresholds from 1E-04 to 0.5 by increments of 1E-04. For each iteration of the permutation tests in the training dataset, Nagelkerke's pseudo  $r^2$  estimates between the PRS and PD were estimated, after adjustment for an estimated prevalence of 0.5% and study-specific eigenvectors 1-5, age and sex as covariates; this prevalence was chosen as a conservative estimate based on global estimates of disease (31). All variant clumping and P thresholding was done using the NeuroX-dbGaP dataset before testing the PRS in the HBS dataset. Then the summary statistics for these SNPs of interest (1805 overlapping with the HBS dataset after QC) were extracted from the meta-analysis excluding HBS to generate variant weights for the validation phase of analysis. Next the PRS was tested in the HBS dataset. After this, we also reduced the PRS SNPs to just 90 variants reported as independent GWAS risk variants. We then repeated this workflow using the 88 SNPs passing QC in HBS as an additional test.

Areas under the curve and related metrics for predictive models based on the PRS were generated by utilizing the best threshold of the receiver operator curve per study, denoted by the top-left most point of the curve, thus maximizing classification accuracy. To calculate heritability in clinically defined PD datasets, we also used LD score regression under default settings, also employing the LD references for Europeans provided with the software(21). This

workflow was also repeated on a per cohort level and is detailed in the Supplementary Appendix.

To prevent bias, we estimated the effect size of each SNP contributing to the PRS using a meta-analysis of all PD GWAS datasets except NeuroX-dbGAP and HBS. Using permutation testing in the NeuroX-dbGAP training cohort, we found that the optimal  $P$  threshold for variant inclusion was  $1.35E-03$ , which included 1809 variants after LD pruning. Two PRSs were tested in HBS, one limited to 88 of the 90 genome-wide significant variants (two variants failed to pass quality control in the HBS study), and the other incorporating 1805 variants from the training phase (four variants failed to pass quality control in HBS due to low imputation quality). The 88 variant PRS had an area under the curve (AUC) of 0.651 (95% CI 0.617 - 0.684), while the 1805 variant PRS had an AUC of 0.692 (95% CI 0.660 - 0.725) in the test data from HBS. The AUCs from our 88 variant PRS in both the NeuroX-dbGAP cohort and the HBS cohort were significantly larger than the AUCs in those same cohorts using a published PRS (Chang *et al.* 2017, AUC = 0.624,  $P < 0.002$  from DeLong's test). Although the HBS cohort was used to discover the 90 PD GWAS risk variants, therefore potentially biasing our 88 variant PRS, all 90 variants remained genome-wide significant in a meta-analysis of all GWAS datasets excluding the HBS study. A possible contributor to the higher AUC in the test set compared to training set is the higher frequency of the large effect *LRK2* p.G2019S variant carriers in the HBS dataset (0.50%) versus the NeuroX-dbGaP dataset (0.26%). Extended results for all included studies can be found in the Supplementary Appendix.

### **Functional causal inferences via Quantitative Trait Loci (QTL) Mendelian randomization to infer functional consequences**

We used MR to test whether changes in methylation and RNA expression of genes physically proximal to genome-wide significant PD risk loci were causally related to PD risk. To nominate genes of interest for MR analyses, we took our putative 90 loci of interest in the large LD reference used for the COJO phase of analysis and identified SNPs in LD with our SNPs of interest at an  $r^2 > 0.5$  within  $\pm 1$  MB (Supplementary Table S5). Once these SNPs were identified, nearest genes were queried from the European Bioinformatics Institute (EMBL-EBI, <https://www.ebi.ac.uk>) and compiled into a list of 305 possible genes linked to PD risk loci. Note, because of slight annotation differences, the *MAPT* and *GBA* genes were forced into the list (their nearby pseudogenes were automatically added). This process nominates genes for QTL analyses that contain variants that are in LD with SNPs of interest and therefore are not only spatially proximal but likely associated with disease risk to some degree.

MR was used to make functional inferences by integrating discovery phase summary statistics with quantitative trait locus (QTL) association summary statistics across well-curated methylation and expression datasets. We utilized the curated versions of Qi *et al.*, 2018 brain methylation and expression summary statistics (GTEx derived), as well as a specific focus on GTEx substantia nigra data (GTEx), we also made use of the blood expression data from Vösa *et al.* 2018 (eQTLGen), all available from the website for summary-data-based Mendelian randomization (SMR, <http://cnsgenomics.com/software/smr/#Overview>) or the eQTLgen

consortium (<http://www.eqtigen.org>) (32–36). For all QTL analyses, we utilized the multi-SNP SMR method under default analysis settings as a framework to carry out MR analyses. In the analyses using SMR, the large LD reference set from our COJO phase of analysis was used. For each of the four QTL datasets, Bonferroni correction false discovery rate (FDR) was used to adjust P values and account for multiple testing within each dataset. All MR effect estimates are reported on the scale of a standard deviation increase in the exposure variable relating to a similar change in PD risk. In its simplest description, these MR analyses compare the local polygenic risk of an exposure (significant changes in methylation or expression) to similar polygenic risk in an outcome (Parkinson's disease) to infer causal associations under the assumption that there is no intermediate confounder associated with both parameters and that the association is not simply due to LD.

### **Rare coding variant burden tests**

A uniformly quality controlled and imputed dataset from the IPDGC (described above) was used to carry out burden tests for all rarer coding variants successfully imputed in an average of 85% of the sample series (17,188 cases and 22,875 controls). These analyses include all variants at a hard call threshold of imputation quality > 0.8. After annotation with annovar, we had a total of 37,503 exonic coding variants (nonsynonymous, stop or splicing) at MAF < 5% and a subset of 29,016 at MAF < 1%(37). We then extracted proximal genes for SNPs tagging any of our 90 loci using same the LD reference as in the COJO analysis ( $r^2 > 0.5$  within +/- 1MB, see Supplementary Table S5). For inclusion in this phase of analyses, a gene must have contained at least 2 coding variants. After assembling this subset of 113 testable genes, we used the optimized sequence kernel association test to generate summary statistics at maximum MAFs of 1% and 5%(38). All burden analyses were adjusted for the first 15 principal components (selected by backwards stepwise modeling) based on common unlinked variants, age, sex and study site. Resulting P values were then adjusted via Bonferroni for the numbers of genes tested, we treated each MAF strata as a separate set of tests.

### **Network analyses**

Two network-based approaches were utilized to assess connectivity across loci. The first focuses on integrating GWAS summary statistics with expression data (Functional Mapping and Annotation of Genome-Wide Association Studies, FUMA), the second focuses on protein interactions (webgestaltR).

Functional mapping and annotation based on publicly available gene expression and ontology resources were made using FUMA version 1.3.1 (39). In brief, summary statistics were analyzed using MAGMA gene property tests to compare enrichment of the average gene expression per tissue in GTEx v7 (40,41). Bonferroni correction was applied to tissue enrichment analyses. In total 10,651 gene sets (Curated gene sets: 4734, GO terms: 5917) were tested. Curated gene sets were generated from nine data resources including KEGG, Reactome and BioCarta (see MSigDB for details, <http://software.broadinstitute.org/gsea/msigdb/collections.jsp>)(42). GO terms were comprised of

the three standard categories, biological processes (bp), cellular components (cc) and molecular functions (mf). All parameters were set as default for the competitive test. We employed a more conservative version of the FDR correction for multiple testing by applying it to all pathways from various sources at once. Single cell RNA sequencing data from DropViz (<http://dropviz.org>) was also queried for enrichment using FUMA in an identical manner, spanning 88 possible tissue and cell type combinations (43).

To investigate protein components related to genetic risk loci, we utilized the R package *webgestaltR* to build networks via its network topology analysis and random walk algorithm (44). The input data for this analysis was all genes found under the association peaks in our GWAS based on the LD structure in reference samples as described in the previous subsection. Ontologies were extracted from the Biogrid Protein-Protein Interaction Networks(45). FDR adjustment was used to adjust for multiple testing.

Using *webgestaltR* (39,44) we found that the genes highlighted by our PD GWAS were enriched in six functional ontological networks (FDR-adjusted  $P < 0.1$ ). The majority of these networks were related to chemical signaling pathways or response to some type of stressor. The most significant protein-protein interaction was related to response to interferon-gamma (Table S9, Figure S3A, Figure S3B). The strength of the results for protein-protein interactions should be interpreted with a degree of caution and will benefit from ongoing follow-up studies of high throughput proteomics in PD specific datasets to nominate potential mechanisms of interest.

### **LD score regression and causal inference**

To investigate shared genetic correlations of PD with multiple traits and diseases, we employed bivariate LD score regression (LDSC) (21). These analyses were carried out under the default settings as previously discussed, using data from the 757 GWAS available via LD Hub as well as biomarker GWAS summary statistics from two additional publications of interest focusing on c-reactive protein and cytokine measures; LD Hub was accessed on June 20th, 2018 (version 1.2.0)(46–48). P values from the bivariate LDSC were adjusted for FDR to account for multiple testing. We acknowledge that for some traits of interest there is a minor overlap with samples derived from the CHARGE studies utilized in a small portion of discovery data in Nalls et al. 2014 which may influence results slightly in downstream MR analyses (1). For evaluation of genetic correlations between PD and UKB derived GWAS, we utilized PD summary statistics with the UKB data excluded to reduce bias. The curated data on LD Hub includes GWAS meta-analyses of over 5,000 European ancestry samples each, and are well powered to ascertain genetic correlations.

Traits showing significant genetic correlations with PD were analyzed using MR methods. We excluded the UKB data when a nominated trait was from summary statistics derived from the UKB or if the UKB was included as part of a meta-analysis.

When complete GWAS summary statistics were available for traits of interest (relating to smoking and education), we utilized the more powerful bi-directional generalized summary-data-

based Mendelian Randomization (GSMR). This approach utilized bi-directional GSMR under default settings with the exception of more stringent HEIDI outlier removal options (global HEIDI threshold at 0.05)(49). For all MR results (both here and below), effect estimates where PD is the outcome are interpreted as the change in the log odds ratio (beta) of PD for a single standard deviation increase in the polygenic risk score for the exposure.

Summary statistics for PD excluded UKB data at this stage. The previously described PD reference datasets for estimating LD was used, excluding samples with missing data for SNPs of interest. We analyzed GWAS summary statistics for smoking initialization (453,693 records from a self-report survey with 208,988 regular smokers and 244,705 never regular smokers) and current smoking within the UKB Current smoking (CS) contrasted 47,419 current smokers versus 244,705 never regular smokers. The same analysis was carried out incorporating recent GWAS data regarding educational attainment (N = 766,345) from self report in the UK and cognitive performance (N = 257,828) as measured by the *g* composite score(50). These outcomes were analyzed using methods to mirror that of the UKB PD GWAS dataset. Combined left and right putamen volume from a T2 magnetic resonance imaging GWAS available from Oxford Brain Imaging Genetics (BIG) Server (accessed December 28th, 2018) (51). All MR analyses included GWAS on the scale of tens of thousands of samples and overcame the considerable power demands of the methodology.

## WORKS CITED

1. Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. *Nat Genet.* 2014 Sep;46(9):989–93.
2. Chang D, Nalls MA, Hallgrímsdóttir IB, Hunkapiller J, van der Brug M, Cai F, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet.* 2017 Oct;49(10):1511–6.
3. Blauwendraat C E al. Parkinson's disease age at onset genome-wide association study: Defining heritability, genetic loci, and  $\alpha$ -synuclein mechanisms. - PubMed - NCBI [Internet]. [cited 2019 Apr 19]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/30957308>
4. Liu JZ, Erlich Y, Pickrell JK. Case–control association mapping by proxy using family history of disease. *Nat Genet.* 2017;49(3):325–31.
5. Mackenbach JP. Health and deprivation. *Inequality and the North. Health Policy.* 1988;10(2):207.
6. International Parkinson Disease Genomics Consortium, Nalls MA, Plagnol V, Hernandez DG, Sharma M, Sheerin U-M, et al. Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet.* 2011 Feb 19;377(9766):641–9.
7. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010 Sep 1;26(17):2190–1.
8. 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 Feb 25;4:7.

9. Chang D, Nalls MA, Hallgrímsdóttir IB, Hunkapiller J, van der Brug M, Cai F, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat Genet*. 2017 Oct;49(10):1511–6.
10. Dorsey ER, Ray Dorsey E, Darwin KC, Mohammed S, Donohue S, Tethal A, et al. Virtual research visits and direct-to-consumer genetic testing in Parkinson's disease. *DIGITAL HEALTH*. 2015;1:205520761559299.
11. Browning SR, Browning BL. Rapid and accurate haplotype phasing and missing-data inference for whole-genome association studies by use of localized haplotype clustering. *Am J Hum Genet*. 2007 Nov;81(5):1084–97.
12. 1000 Genomes Project Consortium, Abecasis GR, Altshuler D, Auton A, Brooks LD, Durbin RM, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010 Oct 28;467(7319):1061–73.
13. Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. *Bioinformatics*. 2015 Mar 1;31(5):782–4.
14. Das S, Forer L, Schönherr S, Sidore C, Locke AE, Kwong A, et al. Next-generation genotype imputation service and methods. *Nat Genet*. 2016 Oct;48(10):1284–7.
15. McCarthy S E al. A reference panel of 64,976 haplotypes for genotype imputation. - PubMed - NCBI [Internet]. [cited 2018 Apr 2]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/27548312>
16. Allen NE, Sudlow C, Peakman T, Collins R, UK Biobank. UK biobank data: come and get it. *Sci Transl Med*. 2014 Feb 19;6(224):224ed4.
17. Lloyd-Jones LR, Robinson MR, Yang J, Visscher PM. Transformation of Summary Statistics from Linear Mixed Model Association on All-or-None Traits to Odds Ratio. *Genetics*. 2018 Apr;208(4):1397–408.
18. 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 Mar;36(3):1605.e7–12.
19. Blauwendraat C, Faghri F, Pihlstrom L, Geiger JT, Elbaz A, Lesage S, et al. NeuroChip, an updated version of the NeuroX genotyping platform to rapidly screen for variants associated with neurological diseases. *Neurobiol Aging*. 2017 Sep;57:247.e9–247.e13.
20. de Bakker PIW, Ferreira MAR, Jia X, Neale BM, Raychaudhuri S, Voight BF. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. *Hum Mol Genet*. 2008 Oct 15;17(R2):R122–8.
21. Bulik-Sullivan BK, Loh P-R, Finucane HK, Ripke S, Yang J, Schizophrenia Working Group of the Psychiatric Genomics Consortium, et al. LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat Genet*. 2015 Mar;47(3):291–5.



22. Yang J, Ferreira T, Morris AP, Medland SE, Genetic Investigation of ANthropometric Traits (GIANT) Consortium, DIAbetes Genetics Replication And Meta-analysis (DIAGRAM) Consortium, et al. Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat Genet.* 2012 Mar 18;44(4):369–75, S1–3.
23. Pankratz N, Beecham GW, DeStefano AL, Dawson TM, Doheny KF, Factor SA, et al. Meta-analysis of Parkinson's disease: identification of a novel locus, RIT2. *Ann Neurol.* 2012 Mar;71(3):370–84.
24. Simón-Sánchez J E al. Genome-wide association study confirms extant PD risk loci among the Dutch. - PubMed - NCBI [Internet]. [cited 2018 Apr 5]. Available from: <https://www.ncbi.nlm.nih.gov/pubmed/21248740>
25. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. *J Stat Softw.* 2010 Aug 5;36(1):1–48.
26. Euesden J, Lewis CM, O'Reilly PF. PRSice: Polygenic Risk Score software. *Bioinformatics.* 2014;31(9):1466–8.
27. International Parkinson's Disease Genomics Consortium (IPDGC), Wellcome Trust Case Control Consortium 2 (WTCCC2). A two-stage meta-analysis identifies several new loci for Parkinson's disease. *PLoS Genet.* 2011 Jun;7(6):e1002142.
28. Hernandez DG, Nalls MA, Ylikotila P, Keller M, Hardy JA, Majamaa K, et al. Genome wide assessment of young onset Parkinson's disease from Finland. *PLoS One.* 2012 Jul 24;7(7):e41859.
29. Ripatti S, Tikkanen E, Orho-Melander M, Havulinna AS, Silander K, Sharma A, et al. A multilocus genetic risk score for coronary heart disease: case-control and prospective cohort analyses. *Lancet.* 2010 Oct 23;376(9750):1393–400.
30. Nalls MA, Escott-Price V, Williams NM, Lubbe S, Keller MF, Morris HR, et al. Genetic risk and age in Parkinson's disease: Continuum not stratum. *Mov Disord.* 2015 May;30(6):850–4.
31. Gasser T. Genetics of parkinson's disease. *Ann Neurol.* 1998;44(S1):S53–7.
32. GTEx Consortium. The Genotype-Tissue Expression (GTEx) project. *Nat Genet.* 2013 Jun;45(6):580–5.
33. 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 Jun 11;9(1):2282.
34. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* 2016 May;48(5):481–7.
35. Wu Y, Zeng J, Zhang F, Zhu Z, Qi T, Zheng Z, et al. Integrative analysis of omics summary data reveals putative mechanisms underlying complex traits. *Nat Commun.* 2018 Mar 2;9(1):918.

36. 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 [Internet]. bioRxiv. 2018 [cited 2018 Oct 25]. p. 447367. Available from: <https://www.biorxiv.org/content/early/2018/10/19/447367.abstract>
37. Yang H, Wang K. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat Protoc.* 2015;10(10):1556–66.
38. Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. *Am J Hum Genet.* 2012 Aug 10;91(2):224–37.
39. Watanabe K, Taskesen E, van Bochoven A, Posthuma D. Functional mapping and annotation of genetic associations with FUMA. *Nat Commun.* 2017 Nov 28;8(1):1826.
40. de Leeuw CA, Mooij JM, Heskes T, Posthuma D. MAGMA: generalized gene-set analysis of GWAS data. *PLoS Comput Biol.* 2015 Apr;11(4):e1004219.
41. GTEx Consortium, Laboratory, Data Analysis & Coordinating Center (LDACC)—Analysis Working Group, Statistical Methods groups—Analysis Working Group, Enhancing GTEx (eGTEx) groups, NIH Common Fund, NIH/NCI, et al. Genetic effects on gene expression across human tissues. *Nature.* 2017 Oct 11;550(7675):204–13.
42. GSEA | MSigDB | MSigDB Collections: Details and Acknowledgments [Internet]. [cited 2018 Jun 20]. Available from: [http://software.broadinstitute.org/gsea/msigdb/collection\\_details.jsp#C2](http://software.broadinstitute.org/gsea/msigdb/collection_details.jsp#C2)
43. Saunders A, Macosko EZ, Wysoker A, Goldman M, Krienen FM, de Rivera H, et al. Molecular Diversity and Specializations among the Cells of the Adult Mouse Brain. *Cell.* 2018 Aug 9;174(4):1015–30.e16.
44. Wang J, Vasaikar S, Shi Z, Greer M, Zhang B. WebGestalt 2017: a more comprehensive, powerful, flexible and interactive gene set enrichment analysis toolkit. *Nucleic Acids Res.* 2017 Jul 3;45(W1):W130–7.
45. Chatr-Aryamontri A, Oughtred R, Boucher L, Rust J, Chang C, Kolas NK, et al. The BioGRID interaction database: 2017 update. *Nucleic Acids Res.* 2017 Jan 4;45(D1):D369–79.
46. LD Hub [Internet]. [cited 2018 Jun 20]. Available from: <http://ldsc.broadinstitute.org/ldhub/>
47. Prins BP, Abbasi A, Wong A, Vaez A, Nolte I, Franceschini N, et al. Investigating the Causal Relationship of C-Reactive Protein with 32 Complex Somatic and Psychiatric Outcomes: A Large-Scale Cross-Consortium Mendelian Randomization Study. *PLoS Med.* 2016 Jun;13(6):e1001976.
48. Ahola-Olli AV, Würtz P, Havulinna AS, Aalto K, Pitkänen N, Lehtimäki T, et al. Genome-wide Association Study Identifies 27 Loci Influencing Concentrations of Circulating Cytokines and Growth Factors. *Am J Hum Genet.* 2017 Jan 5;100(1):40–50.

49. Zhu Z, Zheng Z, Zhang F, Wu Y, Trzaskowski M, Maier R, et al. Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat Commun*. 2018 Jan 15;9(1):224.

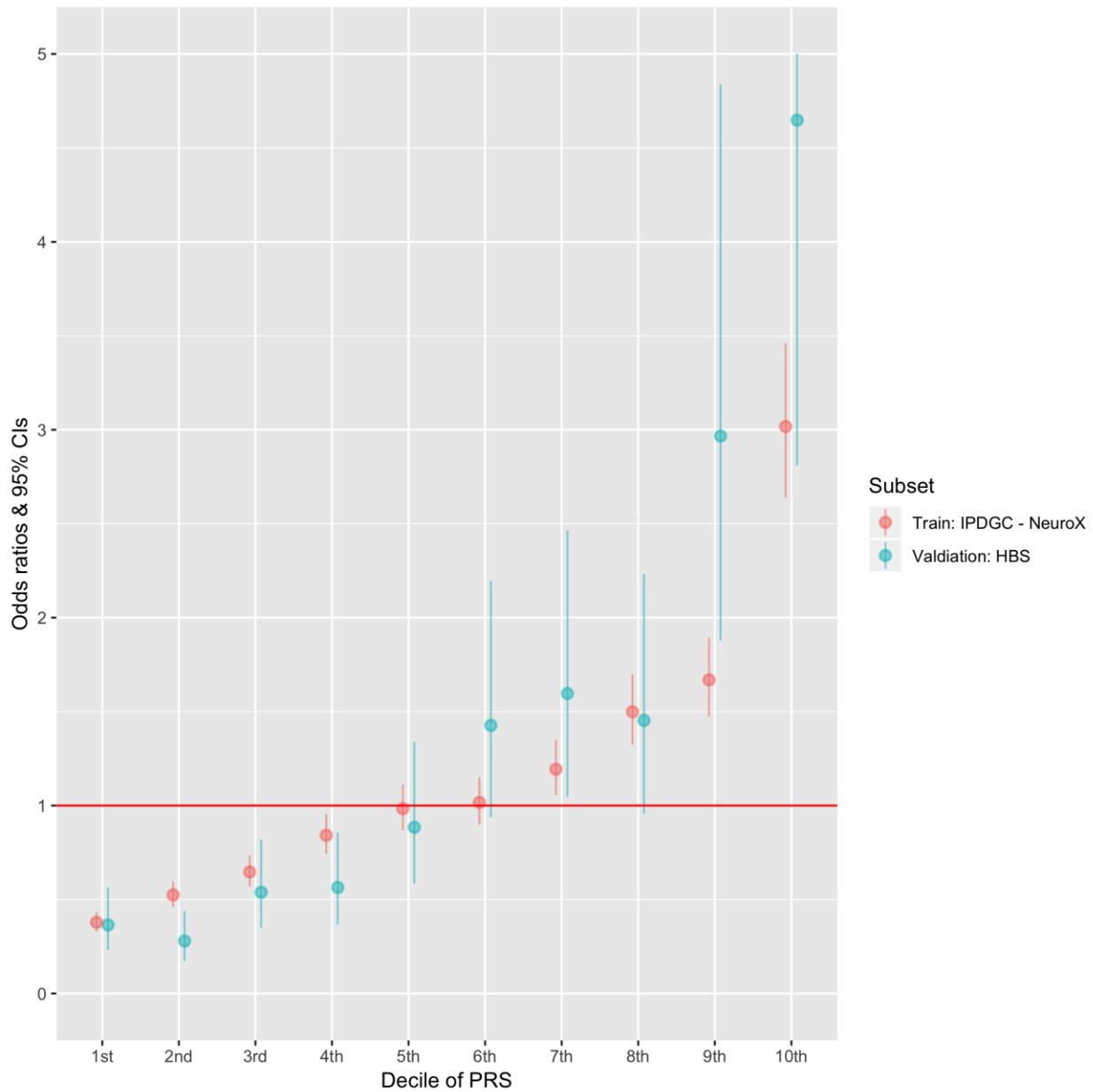
50. Lee JJ, Wedow R, Okbay A, Kong E, Maghzian O, Zacher M, et al. Gene discovery and polygenic prediction from a genome-wide association study of educational attainment in 1.1 million individuals. *Nat Genet* [Internet]. 2018 Jul 23; Available from: <http://dx.doi.org/10.1038/s41588-018-0147-3>

51. Elliott LT, Sharp K, Alfaro-Almagro F, Shi S, Miller KL, Douaud G, et al. Genome-wide association studies of brain imaging phenotypes in UK Biobank. *Nature*. 2018 Oct;562(7726):210–6.

Due to formatting and display issues for large tables as well as facilitating a more detailed examination of the supplements, all additional tables, figures and text are available at the following web address ...

[https://drive.google.com/file/d/1VUU\\_tYI-ew08vupEVRuNEPDpTpWWom/view?usp=sharing](https://drive.google.com/file/d/1VUU_tYI-ew08vupEVRuNEPDpTpWWom/view?usp=sharing).

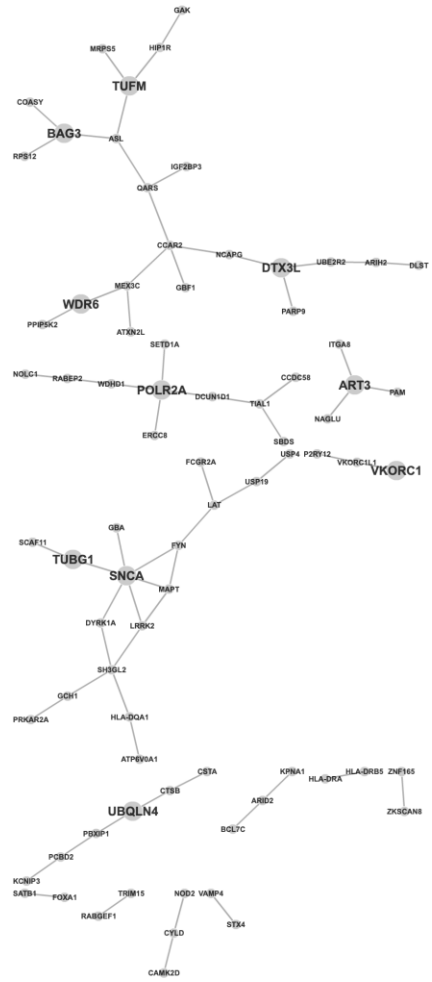
**Figure S1:** The odds ratio of developing PD for each decile of PRS, comparing each decile to all others for all samples in this analysis.



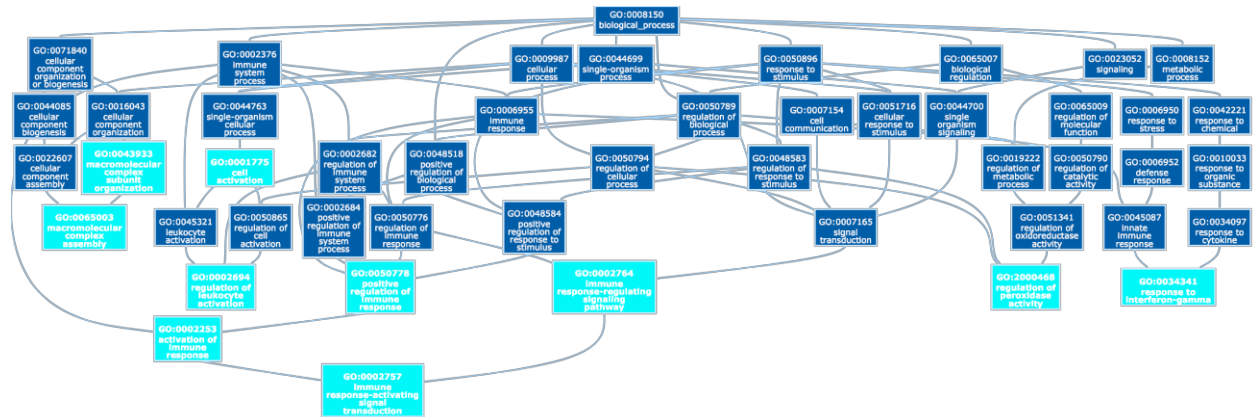


**Figure S3:** Panel A: Gene ontology term connectivity within protein-protein networks. This panel shows network of gene ontology (GO) terms from pathway analyses. Most significant GO terms are shown in green. Panel B: Gene level connectivity within protein-protein networks. This panel shows connectivity between genes across enriched pathways.

A.

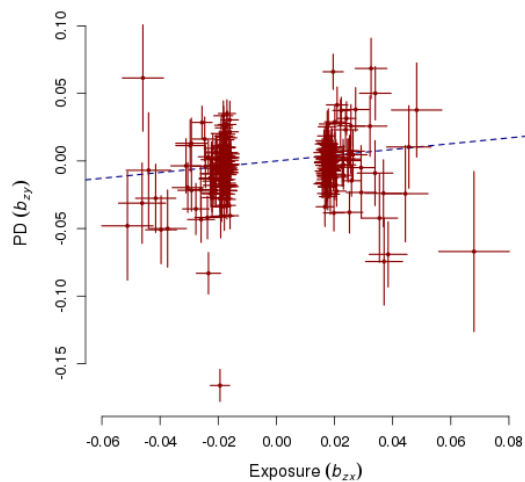


B.

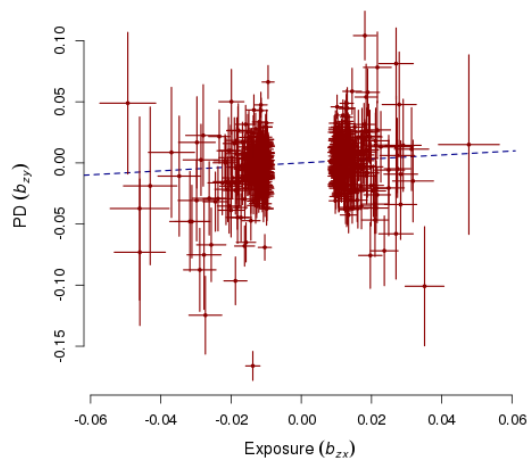


**Figure S4:** Comparison of regression coefficients in Mendelian randomization analyses across traits. Each cross represents a SNP, with the dashed lines representing the trend across all variants. Axes position are regression coefficients from GWAS for significant SNPs from either GWAS. Panel A includes results for cognitive performance, panel B includes results for educational attainment, panel C includes results for putamen volume, panel D includes results for smoking initiation and panel E includes results for current smoking status.

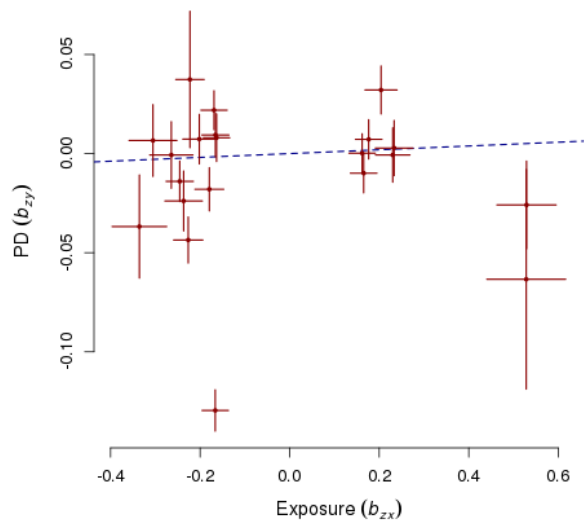
**A.**



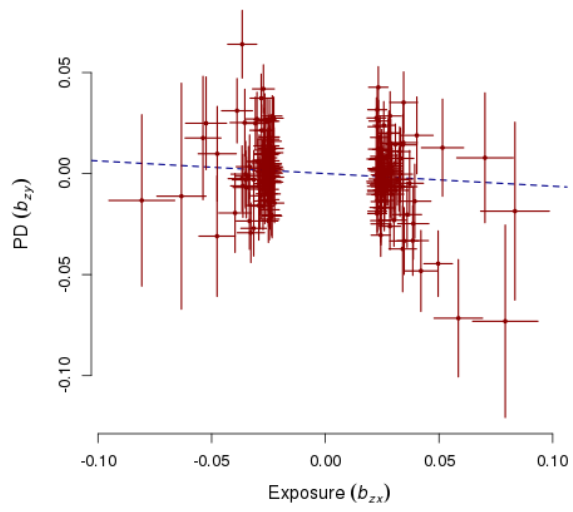
**B.**



**C.**

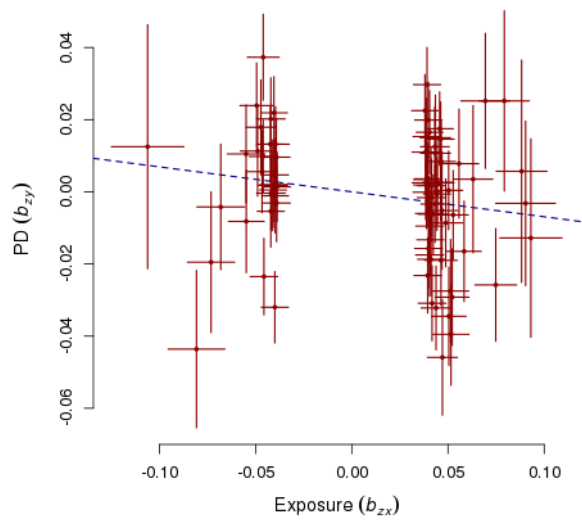


**D.**



**E.**





**Table S1:** Descriptive statistics and quality control summaries for meta-analyzed genome-wide association studies. ! denotes age at exam for both cases and controls. \$ denotes age at death, onset not available. \* based on 599 PD cases and 715 controls. ^ denotes samples checked for overlap across datasets as per Nalls et al. 2014 and Chang et al. 2017, ^^ denotes checked for overlap within IPDGC sample series, ^^denotes a combination of both workflows for identifying sample overlap.

**Table S2:** Summary statistics for all nominated risk variants, known and novel. For binary variables, 0 = negative and 1 = positive. Some specific notes include: delineations of all studies, new studies and previous studies as discussed in the methods section. Betas and standard errors (StdErr) refer to effect estimates per SNP from logistic regression or fixed-effects meta-analyses. I<sub>2</sub> is the index of heterogeneity. QTL Nominated Gene = genes which represent the nearest cis-QTL for that locus significant in MR.

**Table S3:** Comparison with novel results from Chang et al., 2017. This table summarizes linkage disequilibrium estimates between Chang et al., 2017 novel loci and variants passing quality control in this report.

**Table S4:** Estimates of genetic liability explained in different scenarios. Here we compare how different AUC estimates and prevalence rates change the amount of genetic liability ( $h^2$ ) explained by GWAS.

**Table S5:** SNPs of interest tagging genes for functional inferences and networks analysis. Nominated genes and SNPs for follow-up analyses based on minimum  $r^2 > 0.5$  within +/- 1MB of one of our 90 risk loci.

**Table S6:** Complete summary statistics for QTL Mendelian randomization. Output from the SMR package for all QTLs of interest. Additional columns include QTL reference dataset, dataset-level Bonferroni corrected P values and a binary indicator if a candidate association passed multiple test correction. All columns prefixed by SMR indicate multi-SNP SMR results.

**Table S7:** Rare coding variant burden analyses for genes under GWAS peaks. Detailed results of burden tests for genes proximal to risk loci. This includes variant counts, test statistics ( $\rho$ ,  $q$ , P, adjusted P) for each gene of interest.

**Table S8:** FUMA expression pathway enrichment analysis results. Pathway enrichment from collapsed GWAS summary statistics.

**Table S9:** Protein network analysis for linked genes under association peaks. Gene ontology terms passing false discovery rate adjustment.

**Table S10:** Bivariate LDscores. Default output from LD Hub. Abbreviations defined in main text and methods section.

**Text S1:** Authors and affiliations.

**INTERNATIONAL PARKINSON'S DISEASE GENOMICS CONSORTIUM COLLABORATORS:**

**United Kingdom:** Alastair J Noyce (Preventive Neurology Unit, Wolfson Institute of Preventive Medicine, QMUL, London, UK and Department of Molecular Neuroscience, UCL, London, UK), Arianna Tucci (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Ben Middlehurst (Institute of Translational Medicine, University of Liverpool, Liverpool, UK), Demis A Kia (UCL Genetics Institute; and Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Manuela Tan (Department of Clinical Neuroscience, University College London, London, UK), Henry Houlden (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Huw R Morris (Department of Clinical Neuroscience, University College London, London, UK), Helene Plun-Favreau (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Peter Holmans (Biostatistics & Bioinformatics Unit, Institute of Psychological Medicine and Clinical Neuroscience, MRC Centre for Neuropsychiatric Genetics & Genomics, Cardiff, UK), John Hardy (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Daniah Trabzuni (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK; Department of Genetics, King Faisal Specialist Hospital and Research Centre, Riyadh, 11211 Saudi Arabia), Jose Bras (UK Dementia Research Institute at UCL and Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), John Quinn (Institute of Translational Medicine, University of Liverpool, Liverpool, UK), Kin Y Mok (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Kerri J. Kinghorn (Institute of Healthy Ageing, University College London, London, UK), Kimberley Billingsley (Institute of Translational Medicine, University of Liverpool, Liverpool, UK), Nicholas W Wood (UCL Genetics Institute; and Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Patrick Lewis (University of Reading, Reading, UK), Rita Guerreiro (UK Dementia Research Institute at UCL and Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Ruth Lovering (University College London, London, UK), Lea R'Bibo (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Mie Rizig (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Mina Ryten (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Valentina Escott-Price (MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff University School of Medicine, Cardiff, UK), Viorica Chelban (Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK), Thomas Foltynie (UCL Institute of Neurology, London, UK), Nigel Williams (MRC Centre for Neuropsychiatric Genetics and Genomics, Cardiff, UK),

**France:** Alexis Brice (Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS, UMR 7225, Sorbonne Universités, UPMC University Paris 06, UMR S 1127, AP-HP, Pitié-Salpêtrière Hospital, Paris, France), Fabrice Danjou (Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS, UMR 7225, Sorbonne Universités, UPMC University Paris 06, UMR S 1127, AP-HP, Pitié-Salpêtrière Hospital, Paris, France), Suzanne Lesage (Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS, UMR 7225, Sorbonne Universités, UPMC University Paris 06, UMR S 1127, AP-HP, Pitié-Salpêtrière Hospital, Paris, France), Jean-Christophe Corvol (Institut du Cerveau et de la Moelle épinière, ICM, Inserm U 1127, CNRS, UMR 7225, Sorbonne Universités, UPMC University Paris 06, UMR S 1127, Centre d'Investigation Clinique Pitié Neurosciences CIC-1422, AP-HP, Pitié-Salpêtrière Hospital, Paris, France), Maria Martinez (INSERM UMR 1220; and Paul Sabatier University, Toulouse, France),

**Germany:** Anamika Giri (Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, and DZNE, German Center for Neurodegenerative Diseases, Tübingen, Germany), Claudia Schulte (Department for Neurodegenerative Diseases, Hertie Institute for

Clinical Brain Research, University of Tübingen, and DZNE, German Center for Neurodegenerative Diseases, Tübingen, Germany), Kathrin Brockmann (Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, and DZNE, German Center for Neurodegenerative Diseases, Tübingen, Germany), Javier Simón-Sánchez (Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, and DZNE, German Center for Neurodegenerative Diseases, Tübingen, Germany), Peter Heutink (DZNE, German Center for Neurodegenerative Diseases and Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany), Patrizia Rizzu (DZNE, German Center for Neurodegenerative Diseases), Manu Sharma (Centre for Genetic Epidemiology, Institute for Clinical Epidemiology and Applied Biometry, University of Tübingen, Germany), Thomas Gasser (Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, and DZNE, German Center for Neurodegenerative Diseases, Tübingen, Germany),

**United States of America:** Aude Nicolas (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA), Mark R Cookson (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, USA), Sara Bandres-Ciga (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA), Cornelis Blauwendraat (National Institute on Aging and National Institute of Neurological Disorders and Stroke, USA), Faraz Faghri (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, USA; Department of Computer Science, University of Illinois at Urbana-Champaign, Urbana, IL, USA), J Raphael Gibbs (Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA), Dena G Hernandez (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA), Joshua M. Shulman (Departments of Neurology, Neuroscience, and Molecular & Human Genetics, Baylor College of Medicine, Houston, Texas, USA; Jan and Dan Duncan Neurological Research Institute, Texas Children's Hospital, Houston, Texas, USA), Mike A. Nalls (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, USA; CEO/Consultant Data Tecnica International, Glen Echo, MD, USA), Laurie Robak (Baylor College of Medicine, Houston, Texas, USA), Steven Lubbe (Ken and Ruth Davee Department of Neurology, Northwestern University Feinberg School of Medicine, Chicago, IL, USA), Steven Finkbeiner (Departments of Neurology and Physiology, University of California, San Francisco; Gladstone Institute of Neurological Disease; Taube/Koret Center for Neurodegenerative Disease Research, San Francisco, CA, USA), Niccolo E. Mencacci (Northwestern University Feinberg School of Medicine, Chicago, IL, USA), Codrin Lungu (National Institutes of Health Division of Clinical Research, NINDS, National Institutes of Health, Bethesda, MD, USA), Andrew B Singleton (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA), Sonja W. Scholz (Neurodegenerative Diseases Research Unit, National Institute of Neurological Disorders and Stroke, Bethesda, MD, USA), Xylena Reed (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA), Hampton Leonard (Laboratory of Neurogenetics, National Institute on Aging, Bethesda, MD, USA).

**Canada:** Ziv Gan-Or (Montreal Neurological Institute and Hospital, Department of Neurology & Neurosurgery, Department of Human Genetics, McGill University, Montréal, QC, H3A 0G4, Canada), Guy A. Rouleau (Montreal Neurological Institute and Hospital, Department of Neurology & Neurosurgery, Department of Human Genetics, McGill University, Montréal, QC, H3A 0G4, Canada), Lynne Krohan (Department of Human Genetics, McGill University, Montréal, QC, H3A 0G4, Canada),

**The Netherlands:** Jacobus J van Hilten (Department of Neurology, Leiden University Medical Center, Leiden, Netherlands), Johan Marinus (Department of Neurology, Leiden University Medical Center, Leiden, Netherlands)

**Spain:** Astrid D. Adarmes-Gómez (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Miquel Aguilar (Fundació Docència i Recerca Mútua de Terrassa and Movement Disorders Unit, Department of Neurology, University Hospital Mutua de Terrassa, Terrassa, Barcelona.), Ignacio Alvarez (Fundació Docència i Recerca Mútua de Terrassa and Movement Disorders Unit, Department of Neurology, University Hospital Mutua de Terrassa, Terrassa, Barcelona.), Victoria Alvarez (Hospital Universitario Central de Asturias, Oviedo), Francisco Javier Barrero (Hospital Universitario Parque Tecnológico de la Salud, Granada), Jesús Alberto Bergareche Yarza (Instituto de Investigación Sanitaria Biodonostia, San Sebastián), Inmaculada Bernal-Bernal (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Marta Blazquez (Hospital Universitario Central de Asturias, Oviedo), Marta Bonilla-Toribio (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Juan A. Botía (Universidad de Murcia, Murcia), María Teresa Bongiorno (Fundació Docència i Recerca Mútua de Terrassa and Movement Disorders Unit, Department of Neurology, University Hospital Mutua de Terrassa, Terrassa, Barcelona.) Dolores Buiza-Rueda (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Ana Cámara (Hospital Clinic de Barcelona), Maria Carcel (Fundació Docència i Recerca Mútua de Terrassa, Barcelona and Department of Neurology, Movement Disorders Unit, Hospital Universitari Mutua de Terrassa, Barcelona), Fátima Carrillo (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Mario Carrión-Claro (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Debora Cerdan (Hospital General de Segovia, Segovia), Jordi Clarimón (Memory Unit, Department of Neurology, IIB Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona and Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED), Madrid), Yaroslau Compta (Hospital Clinic de Barcelona), Monica Diez-Fairen (Fundació Docència i Recerca Mútua de Terrassa and Movement Disorders Unit, Department of Neurology, University Hospital Mutua de Terrassa, Terrassa, Barcelona.), Oriol Dols-Icardo (Memory Unit, Department of Neurology, IIB Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, and Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED), Madrid), Jacinto Duarte (Hospital General de Segovia, Segovia), Raquel Duran (Centro de Investigación Biomédica, Universidad de Granada, Granada), Francisco Escamilla-Sevilla (Hospital Universitario Virgen de las Nieves, Instituto de Investigación Biosanitaria de Granada, Granada), Mario Ezquerro (Hospital Clinic de Barcelona), Manel Fernández (Hospital Clinic de Barcelona), Rubén Fernández-Santiago (Hospital Clinic de Barcelona), Ciara Garcia (Hospital Universitario Central de Asturias, Oviedo), Pedro García-Ruiz (Instituto de Investigación Sanitaria Fundación Jiménez Díaz, Madrid), Pilar Gómez-Garre (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Maria Jose Gomez Heredia (Hospital Universitario Virgen de la Victoria, Malaga), Isabel Gonzalez-Aramburu (Hospital Universitario Marqués de Valdecilla-IDIVAL, Santander), Ana Gorostidi Pagola (Instituto de Investigación Sanitaria Biodonostia, San Sebastián), Janet Hoenicka (Institut de Recerca Sant Joan de Déu, Barcelona), Jon Infante (Hospital Universitario Marqués de Valdecilla-IDIVAL and University of Cantabria, Santander, and Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED)), Silvia Jesús (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Adriano Jimenez-Escrig (Hospital Universitario Ramón y Cajal, Madrid), Jaime Kulisevsky (Movement Disorders Unit, Department of Neurology, IIB Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, and Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED)), Miguel A. Labrador-Espinosa (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Jose Luis Lopez-Sendon (Hospital Universitario Ramón y Cajal, Madrid), Adolfo López de Munain Arregui (Instituto de Investigación Sanitaria Biodonostia, San

Sebastián), Daniel Macias (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Irene Martínez Torres (Department of Neurology, Instituto de Investigación Sanitaria La Fe, Hospital Universitario y Politécnico La Fe, Valencia), Juan Marín (Movement Disorders Unit, Department of Neurology, IIB Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, and Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED)), María Jose Martí (Hospital Clinic Barcelona), Juan Carlos Martínez-Castrillo (Instituto Ramón y Cajal de Investigación Sanitaria, Hospital Universitario Ramón y Cajal, Madrid), Carlota Méndez-del-Barrio (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Manuel Menéndez González (Hospital Universitario Central de Asturias, Oviedo), Adolfo Mínguez (Hospital Universitario Virgen de las Nieves, Granada, Instituto de Investigación Biosanitaria de Granada), Pablo Mir (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Elisabet Mondragon Rezola (Instituto de Investigación Sanitaria Biodonostia, San Sebastián), Esteban Muñoz (Hospital Clinic Barcelona), Javier Pagonabarraga (Movement Disorders Unit, Department of Neurology, IIB Sant Pau, Hospital de la Santa Creu i Sant Pau, Universitat Autònoma de Barcelona, Barcelona, and Centro de Investigación Biomédica en Red en Enfermedades Neurodegenerativas (CIBERNED)), Pau Pastor (Fundació Docència i Recerca Mútua de Terrassa and Movement Disorders Unit, Department of Neurology, University Hospital Mutua de Terrassa, Terrassa, Barcelona.), Francisco Perez Errazquin (Hospital Universitario Virgen de la Victoria, Malaga), Teresa Perriñán-Tocino (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Javier Ruiz-Martínez (Hospital Universitario Donostia, Instituto de Investigación Sanitaria Biodonostia, San Sebastián), Clara Ruz (Centro de Investigación Biomedica, Universidad de Granada, Granada), Antonio Sanchez Rodriguez (Hospital Universitario Marqués de Valdecilla-IDIVAL, Santander), María Sierra (Hospital Universitario Marqués de Valdecilla-IDIVAL, Santander), Esther Suarez-Sanmartin (Hospital Universitario Central de Asturias, Oviedo), Cesar Taberero (Hospital General de Segovia, Segovia), Juan Pablo Tartari (Fundació Docència i Recerca Mútua de Terrassa and Movement Disorders Unit, Department of Neurology, University Hospital Mutua de Terrassa, Terrassa, Barcelona), Cristina Tejera-Parrado (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Eduard Tolosa (Hospital Clinic Barcelona), Francesc Valldeoriola (Hospital Clinic Barcelona), Laura Vargas-González (Instituto de Biomedicina de Sevilla (IBiS), Hospital Universitario Virgen del Rocío/CSIC/Universidad de Sevilla, Seville), Lydia Vela (Department of Neurology, Hospital Universitario Fundación Alcorcón, Madrid), Francisco Vives (Centro de Investigación Biomedica, Universidad de Granada, Granada).

**Austria:** Alexander Zimprich (Department of Neurology, Medical University of Vienna, Austria)

**Norway:** Lasse Pihlstrom (Department of Neurology, Oslo University Hospital, Oslo, Norway)

**Estonia:** Sulev Koks (Department of Pathophysiology, University of Tartu, Tartu, Estonia; Department of Reproductive Biology, Estonian University of Life Sciences, Tartu, Estonia; Perron Institute for Neurological and Translational Science, Perth, Western Australia, Australia), Pille Taba (Department of Neurology and Neurosurgery, University of Tartu, Tartu, Estonia)

**Finland:** Kari Majamaa (Institute of Clinical Medicine, Department of Neurology, University of Oulu, Oulu, Finland; Department of Neurology and Medical Research Center, Oulu University Hospital, Oulu, Finland), Ari Siitonen (Institute of Clinical Medicine, Department of Neurology, University of Oulu, Oulu, Finland; Department of Neurology and Medical Research Center, Oulu University Hospital, Oulu, Finland)

**Nigeria:** Njideka U. Okubadejo (University of Lagos, Lagos State, Nigeria), Oluwadamilola O. Ojo (University of Lagos, Lagos State, Nigeria),

### **23AndMe RESEARCH TEAM MEMBERS:**

Michelle Agee, Babak Alipanahi, Adam Auton, Robert K. Bell, Katarzyna Bryc, Sarah L. Elson, Pierre Fontanillas, Nicholas A. Furlotte, Karen E. Huber, Barry Hicks, Ethan M. Jewett, Yunxuan Jiang, Aaron Kleinman, Keng-Han Lin, Nadia K. Litterman, Jennifer C. McCreight, Matthew H. McIntyre, Kimberly F. McManus, Joanna L. Mountain, Elizabeth S. Noblin, Carrie A.M. Northover, Steven J. Pitts, G. David Poznik, J. Fah Sathirapongsasuti, Janie F. Shelton, Suyash Shringarpure, Chao Tian, Joyce Y. Tung, Vladimir Vacic, Xin Wang, and Catherine H. Wilson.

### **SYSTEM GENOMICS OF PARKINSON'S DISEASE COLLABORATORS:**

Tim Anderson(New Zealand Brain Research Institute, Christchurch, New Zealand; Department of Medicine, University of Otago, Christchurch, New Zealand),Steven Bentley(Griffith Institute for Drug Discovery, Griffith University, Brisbane, Australia), John Dalrymple-Alford(Dept. Psychology, University of Canterbury, New Zealand Brain Research Institute), Javed Fowdar(Griffith Institute for Drug Discovery, Griffith University, Brisbane, Australia),Jacob Gratten (Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia; Queensland Brain Institute, University of Queensland, Brisbane, Australia), Glenda Halliday(Brain and Mind Centre, Sydney Medical School, University of Sydney, Sydney, Australia),Anjali K. Henders(Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia),Ian Hickie(Brain and Mind Centre, Sydney Medical School, University of Sydney, Sydney, Australia),Irfahan Kassam(Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia),Martin Kennedy(Department of Pathology, University of Otago, Christchurch, New Zealand),John Kwok(Brain and Mind Centre, Sydney Medical School, University of Sydney, Sydney, Australia),Simon Lewis(Brain and Mind Centre, Sydney Medical School, University of Sydney, Sydney, Australia),George Mellick(Griffith Institute for Drug Discovery, Griffith University, Brisbane, Australia),Grant Montgomery(Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia),John Pearson(Department of Pathology, University of Otago, Christchurch, New Zealand),Toni Pitcher(New Zealand Brain Research Institute, Christchurch, New Zealand; Department of Medicine, University of Otago, Christchurch, New Zealand),Julia Sidorenko(Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia),Peter A. Silburn(Queensland Brain Institute, University of Queensland, Brisbane, Australia),Costanza L. Vallerqa(Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia), Peter M. Visscher(Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia; Queensland Brain Institute, University of Queensland, Brisbane, Australia), Leanne Wallace, (Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia), Naomi R. Wray(Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia; Queensland Brain Institute, University of Queensland, Brisbane, Australia),Angli Xue(Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia), Jian Yang (Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia; Queensland Brain Institute, University of Queensland, Brisbane, Australia), Futao Zhang(Institute for Molecular Bioscience, University of Queensland, Brisbane, Australia)

### **OTHER COLLABORATORS:**

**Department of Neurology, University of Maryland School of Medicine:** Lisa M. Shulman, Rainer von Coelln, Stephen Reich, Joseph Savitt

**Institute of Clinical Medicine, Department of Neurology, University of Turku, Turku, Finland;**  
**Division of Clinical Neurosciences, Turku University Hospital, Turku, Finland:** Pauli Ylikotila

**Text S2:** Acknowledgements and Funding.

#### **ACKNOWLEDGMENTS:**

We would like to thank all of the subjects who donated their time and biological samples to be a part of this study. This work was supported in part by the Intramural Research Programs of the National Institute of Neurological Disorders and Stroke (NINDS), the National Institute on Aging (NIA), and the National Institute of Environmental Health Sciences both part of the National Institutes of Health, Department of Health and Human Services; project numbers 1ZIA-NS003154, Z01-AG000949-02 and Z01-ES101986. In addition this work was supported by the Department of Defense (award W81XWH-09-2-0128), and The Michael J Fox Foundation for Parkinson's Research. John Hardy's contribution was in part supported by MR/N026004/1. This work was supported by National Institutes of Health grants R01NS037167, R01CA141668, P50NS071674, American Parkinson Disease Association (APDA); Barnes Jewish Hospital Foundation; Greater St Louis Chapter of the APDA. The KORA (Cooperative Research in the Region of Augsburg) research platform was started and financed by the Forschungszentrum für Umwelt und Gesundheit, which is funded by the German Federal Ministry of Education, Science, Research, and Technology and by the State of Bavaria. This study was also funded by the German Federal Ministry of Education and Research (BMBF) under the funding code 031A430A, the EU Joint Programme - Neurodegenerative Diseases Research (JPND) project under the aegis of JPND -[www.jpnd.eu](http://www.jpnd.eu)- through Germany, BMBF, funding code 01ED1406 and iMed - the Helmholtz Initiative on Personalized Medicine. This study is funded by the German National Foundation grant (DFG SH599/6-1) (grant to M.S), Michael J Fox Foundation, and MSA Coalition, USA (to M.S). The French GWAS work was supported by the French National Agency of Research (ANR-08-MNP-012). This study was also funded by France-Parkinson Association, Fondation de France, the French program "Investissements d'avenir" funding (ANR-10-IAIHU-06) and a grant from Assistance Publique-Hôpitaux de Paris (PHRC, AOR-08010) for the French clinical data. This study was also sponsored by the Landspítali University Hospital Research Fund (grant to SSv); Icelandic Research Council (grant to SSv); and European Community Framework Programme 7, People Programme, and IAPP on novel genetic and phenotypic markers of Parkinson's disease and Essential Tremor (MarkMD), contract number PIAP-GA-2008-230596 MarkMD (to HP and JHu). Institutional research funding IUT20-46 was received of the Estonian Ministry of Education and Research (SK). Finnish exome sequencing study was partly funded by Sigrid Juselius Foundation. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, Md. (<http://biowulf.nih.gov>), and DNA panels, samples, and clinical data from the National Institute of Neurological Disorders and Stroke Human Genetics Resource Center DNA and Cell Line Repository. People who contributed samples are acknowledged in descriptions of every panel on the repository website. We thank the French



Parkinson's Disease Genetics Study Group and the Drug Interaction with genes (DIGPD) study group: Y Agid, M Anheim, F Artaud, A-M Bonnet, C Bonnet, F Bourdain, J-P Brandel, C Brefel-Courbon, M Borg, A Brice, E Broussolle, F Cormier-Dequaire, J-C Corvol, P Damier, B Debilly, B Degos, P Derkinderen, A Destée, A Dürr, F Durif, A Elbaz, D Grabli, A Hartmann, S Klebe, P. Krack, J Kraemmer, S Leder, S Lesage, R Levy, E Lohmann, L Lacomblez, G Mangone, L-L Mariani, A-R Marques, M Martinez, V Mesnage, J Muellner, F Ory-Magne, F Pico, V Planté-Bordeneuve, P Pollak, O Rascol, K Tahiri, F Tison, C Tranchant, E Roze, M Tir, M Vérin, F Viallet, M Vidailhet, A You. We also thank the members of the French 3C Consortium: A Alperovitch, C Berr, C Tzourio, and P Amouyel for allowing us to use part of the 3C cohort, and D Zelenika for support in generating the genome-wide molecular data. We thank P Tienari (Molecular Neurology Programme, Biomedicum, University of Helsinki), T Peuralinna (Department of Neurology, Helsinki University Central Hospital), L Myllykangas (Folkhalsan Institute of Genetics and Department of Pathology, University of Helsinki), and R Sulkava (Department of Public Health and General Practice Division of Geriatrics, University of Eastern Finland) for the Finnish controls (Vantaa85+ GWAS data). We used genome-wide association data generated by the Wellcome Trust Case-Control Consortium 2 (WTCCC2) from UK patients with Parkinson's disease and UK control individuals from the 1958 Birth Cohort and National Blood Service. Genotyping of UK replication cases on ImmunoChip was part of the WTCCC2 project, which was funded by the Wellcome Trust (083948/Z/07/Z). UK population control data was made available through WTCCC1. This study was supported by the Medical Research Council and Wellcome Trust disease centre (grant WT089698/Z/09/Z to NW, JHa, and ASc). As with previous IPDGC efforts, this study makes use of data generated by the Wellcome Trust Case-Control Consortium. A full list of the investigators who contributed to the generation of the data is available from [www.wtccc.org.uk](http://www.wtccc.org.uk). Funding for the project was provided by the Wellcome Trust under award 076113, 085475 and 090355. This study was also supported by Parkinson's UK (grants 8047 and J-0804) and the Medical Research Council (G0700943 and G1100643). Sequencing and genotyping done in McGill University was supported by grants from the Michael J Fox Foundation, the Canadian Consortium on Neurodegeneration in Aging (CCNA), the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives (HBHL) program. The access to part of the participants in the McGill cohort has been made possible thanks to the Quebec Parkinson's Network (<http://rpq-qpn.ca/en/>). Ziv Gan-Or is supported by the FRQS Chercheurs-boursiers Award, granted by the Fonds de recherche du Québec – Santé (FRQS) and Parkinson's Quebec. We thank Jeffrey Barrett and Jason Downing for assistance with the design of the ImmunoChip and NeuroX arrays. DNA extraction work that was done in the UK was undertaken at University College London Hospitals, University College London, who received a proportion of funding from the Department of Health's National Institute for Health Research Biomedical Research Centres funding. This study was supported in part by the Wellcome Trust/Medical Research Council Joint Call in Neurodegeneration award (WT089698) to the Parkinson's Disease Consortium (UKPDC), whose members are from the UCL Institute of Neurology, University of Sheffield, and the Medical Research Council Protein Phosphorylation Unit at the University of Dundee. We thank the Quebec Parkinson's Network (<http://rpq-qpn.org>) and its members. This work was supported by the Medical Research Council grant MR/N026004/1. The Braineac project was supported by the MRC through the MRC Sudden Death Brain Bank Grant (MR/G0901254) to J.H. P.A.L. was

supported by the MRC (grants MR/N026004/1 and MR/L010933/1) and Michael J. Fox Foundation for Parkinson's Research. D.T. was supported by the King Faisal Specialist Hospital and Research Centre, Saudi Arabia, and the Michael J. Fox Foundation for Parkinson's Research and MRC grant (MR/N026004/1). Lasse Pihlstrøm is supported by the Norwegian Health Association and Michael J. Fox Foundation. Mathias Toft is supported by the Research Council of Norway and the South-Eastern Norway Regional Health Authority. We thank Ole Andreassen and the DemGene consortium, Norway for genotyping of the Oslo cohort. Mike A. Nalls' participation is supported by a consulting contract between Data Tecnica International and the National Institute on Aging, NIH, Bethesda, MD, USA, as a possible conflict of interest. Dr. Nalls also consults for SK Therapeutics Inc, Lysosomal Therapeutics Inc, the Michael J. Fox Foundation and Vivid Genomics among others. Joshua M. Shulman was supported by Huffington Foundation, the Jan and Dan Duncan Neurological Research Institute at Texas Children's Hospital, and a Career Award for Medical Scientists from the Burroughs Wellcome Fund. The Baylor College of Medicine and University of Maryland replication cohorts were made possible in part due to assistance from Amanda Stillwell, Christina Griffin, Katrina Schrader, and Weidong Le with sample collection, preparation, and handling. Data used in the preparation of this article were obtained from the Parkinson's Progression Markers Initiative (PPMI) database ([www.ppmi-info.org/data](http://www.ppmi-info.org/data)). For up-to-date information on the study, visit [www.ppmi-info.org](http://www.ppmi-info.org). PPMI – a public-private partnership – is funded by The Michael J. Fox Foundation for Parkinson's Research and funding partners, including Abbvie, Allergan, Avid Radiopharmaceuticals, Biogen, BioLegend, Bristol-Myers Squibb, Denali, GE Healthcare, Genentech, GlaxoSmithKline, Lilly, Lundbeck, Merck, Meso Scale Discovery, Pfizer, Piramal, Roche, Sanofi, Servier, Takeda, Teva, and UCB. . The SGPD's contribution was supported by the Australian Research Council (ARC) (DP160102400) and the Australian National Health and Medical Research Council (NHMRC) (1078037,1078901, 1103418, 1107258, 1127440, 1113400). Support also came from ForeFront, a large collaborative research group dedicated to the study of neurodegenerative diseases and funded by the NHMRC (Program Grant 1132524, Dementia Research Team Grant 1095127, NeuroSleep Centre of Research Excellence 1060992) and ARC (Centre of Excellence in Cognition and its Disorders Memory Program CE10001021). Simon Lewis was supported by an NHMRC-ARC Dementia Fellowship (1110414) and Glenda Halliday was supported by an NHMRC Fellowship (1079679). The Queensland Parkinson's Project (QPP) was supported by a grant from the Australian National Health and Medical Research Council (1084560) to George Mellick. The New Zealand Brain Research Institute (NZBRI) cohort was funded by a University of Otago Research Grant, together with financial support from the Jim and Mary Carney Charitable Trust (Whangarei, New Zealand). We thank Allison Miller for processing and handling of NZBRI samples.

### **Additional acknowledgement for use of LD Hub**

We gratefully acknowledge all the studies and databases that made GWAS summary data available: **ADIPOGen** (Adiponectin genetics consortium), **C4D** (Coronary Artery Disease Genetics Consortium), **CARDIoGRAM** (Coronary ARtery Disease Genome wide Replication and Meta-analysis), **CKDGen** (Chronic Kidney Disease Genetics consortium), **dbGAP** (database of

Genotypes and Phenotypes), **DIAGRAM** (DIAbetes Genetics Replication And Meta-analysis), **ENIGMA**(Enhancing Neuro Imaging Genetics through Meta Analysis), **EAGLE**(EARly Genetics & Lifecourse Epidemiology Eczema Consortium, excluding 23andMe), **EGG** (Early Growth Genetics Consortium), **GABRIEL**(A Multidisciplinary Study to Identify the Genetic and Environmental Causes of Asthma in the European Community), **GCAN** (Genetic Consortium for Anorexia Nervosa), **GEFOS** (GEnetic Factors for OSteoporosis Consortium), **GIANT** (Genetic Investigation of ANthropometric Traits), **GIS**(Genetics of Iron Status consortium), **GLGC** (Global Lipids Genetics Consortium), **GPC** (Genetics of Personality Consortium), **GUGC**(Global Urate and Gout consortium), **HaemGen**(haematological and platelet traits genetics consortium), **HRgene**(Heart Rate consortium), **IIBDGC** (International Inflammatory Bowel Disease Genetics Consortium), **ILCCO** (International Lung Cancer Consortium), **IMSGC** (International Multiple Sclerosis Genetic Consortium), **MAGIC** (Meta-Analyses of Glucose and Insulin-related traits Consortium), **MESA** (Multi-Ethnic Study of Atherosclerosis), **PGC** (Psychiatric Genomics Consortium), **Project MinE** consortium, **ReproGen** (Reproductive Genetics Consortium), **SSGAC** (Social Science Genetics Association Consortium) and **TAG** (Tobacco and Genetics Consortium), **TRICL**(Transdisciplinary Research in Cancer of the Lung consortium), **UK Biobank**. We gratefully acknowledge the contributions of Alkes Price (the systemic lupus erythematosus GWAS and primary biliary cirrhosis GWAS) and Johannes Kettunen (lipids metabolites GWAS).

#### **Additional acknowledgement for the use of the UKBB data**

This research has been conducted using the UK Biobank Resource under Application Number 33601.

#### **Additional 23andMe acknowledgement**

We thank the research participants and employees of 23andMe.

**Supplemental Appendix:** This appendix is split into four sections detailing: first comparisons of effect estimates across GWAS cohorts (beta~beta plots), second forest plots for each significant variant, thirdly locus plots showing regional GWAS results, and QTL and burden associations for each variant, finally the fourth section including extended PRS results. Beta~beta plots compare the regression coefficients for up to 90 of the significant variants in one study to a meta-analysis of all others via linear regression. Forest plots communicate similar sensitivity analyses, for each of the 90 variants of interest. In the forest plots, box size indicates relative sample size for that study, and the width of the diamond representing the meta-analysis effect estimates indicate the 95% confidence interval. The locus plots are a zoomed-in version of Figure 2 for each of the 90 significant variants. These plots are truncated at a  $-\log_{10} P$  value of 50 for display purposes and include the most significant burden test and QTL analysis results per gene denoted by label color-coding in each figure. In each locus plot,  $R^2$  is measured in our in-house LD reference dataset and shows the correlation between the most significant local SNP and all other proximal SNPs. Additional detailed PRS results for a subset of cohorts are available in the appendix summarizing PRS estimates at varied P thresholds. Each cohort

specific PRS in the appendix is based on meta-analyses excluding that cohort when calculating SNP weights. A smaller table summarizing PRS associations at the P threshold with the highest  $r^2$  is also included. Column headers in the PRS section of the appendix mirror that of Table 2.

**Table of authors for pubmed indexing.**

<b>FIRST</b>	<b>LAST</b>	<b>IPDGC</b>	<b>23andMe</b>	<b>SGPD</b>
Astrid D	Adarmes-Gómez	X		
Miquel	Aguilar	X		
Akbota	Aitkulova	X		
Vadim	Akhmetzhanov	X		
Roy N	Alcalay	X		
Ignacio	Alvarez	X		
Victoria	Alvarez	X		
Sara	Bandres-Ciga	X		
Francisco Javier	Barrero	X		
Jesús Alberto	Bergareche Varza	X		
Inmaculada	Bernal-Bernal	X		
Kimberley	Billingsley	X		
Cornelis	Blauwendraat	X		
Marta	Blazquez	X		
Marta	Bonilla-Toribio	X		
Juan A	Botía	X		
María Teresa	Boungiorno	X		
Jose	Bras	X		
Alexis	Brice	X		
Kathrin	Brockmann	X		
Vivien	Bubb	X		
Dolores	Buiza-Rueda	X		
Ana	Cámara	X		
Fátima	Carrillo	X		
Mario	Carrión-Claro	X		
Debora	Cerdan	X		
Viorica	Chelban	X		
Jordi	Clarimón	X		
Carl	Clarke	X		
Yaroslau	Compta	X		
Mark R	Cookson	X		
Jean-Christophe	Corvol	X		
David W	Craig	X		
Fabrice	Danjou	X		
Monica	Diez-Fairen	X		
Oriol	Dols-Icardo	X		
Jacinto	Duarte	X		
Raquel	Duran	X		
Francisco	Escamilla-Sevilla	X		
Valentina	Escott-Price	X		
Mario	Ezquerria	X		
Faraz	Faghri	X		