TITLE

Parkinson's disease genetics: identifying novel risk loci, providing causal insights and improving estimates of heritable risk.

AUTHORS

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

AFFILIATIONS

[1] Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, 20892 USA
[2] Data Tecnica International, Glen Echo, MD, 20812 USA
[3] Queensland Brain Institute, The University of Queensland, Brisbane, QLD 4072 Australia
[4] Institute for Molecular Bioscience, The University of Queensland, Brisbane, QLD 4072 Australia
[5] 23andMe, Inc., Mountain View, California 94041 USA
[6] Department of Human Genetics, Genentech, South San Francisco, 94080, CA, USA
[7] Department of Molecular Neuroscience, UCL Institute of Neurology, London, UK
[8] Preventive Neurology Unit, Wolfson Institute of Preventive Medicine, Queen Mary University of London, London, UK
[9] UK Dementia Research Institute at UCL, London, UK
[10] Department of Neurodegenerative Diseases, UCL Institute of Neurology, University College London, London, UK
[11] Department of Neurology, Baylor College of Medicine, Houston, USA
[12] Department of Neurology, University of Maryland School of Medicine, Baltimore, MD, USA
[13] Department for Neurodegenerative Diseases, Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany
[14] German Center for Neurodegenerative Diseases (DZNE), Tübingen, Germany.
[15] Centre for Genetic Epidemiology, Institute for Clinical Epidemiology and Applied Biometry, University of Tubingen, Germany

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

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

ACKNOWLEDGEMENTS AND FUNDING

See supplemental materials (Text S2).

ABSTRACT

We performed the largest genetic study of Parkinson's disease to date, involving analysis of 11.4M SNPs in 37.7K cases, 18.6K 'proxy-cases' and 1.4M controls, discovering 39 novel risk loci. In total, we identified 92 putative independent genome-wide significant signals including 53 at previously published loci. Next, we dissected risk within these loci, identifying 22 candidate independent risk variants in close proximity to one another representing multiple risk signals in one locus (20 variants proximal to known risk loci). We then employed tests of causality within a Mendelian randomization framework to infer functional genomic consequences for genes within loci of interest in concert with protein-centric network analyses to nominate likely candidates for follow-up investigation. This report also shows expression network signatures of PD loci to be heavily brain enriched and different in comparison to Alzheimer's disease. We also used risk scoring methods to improve genetic predictions of disease risk, and show that GWAS signals explain 11-15% of the heritable risk of PD at thresholds below genome-wide significance. Additionally, these data also suggest genetic correlations relating to risk overlapping with brain morphology, smoking status and educational attainment. Further analyses of smoking initiation and cognitive performance relating to PD risk in more comprehensive datasets show complex etiological links between PD risk and these traits. These data in sum provide the most comprehensive understanding of the genetic architecture of PD to date, revealing a large number of additional loci, and demonstrating that there remains a considerable genetic component of this disease that has not yet been discovered.

INTRODUCTION

Parkinson's disease (PD) is a common neurodegenerative movement disorder, affecting 1-2% of the population older than 60 years. PD patients suffer from different combinations of motor and non-motor symptoms, which ultimately have a drastic effect on daily function and quality of life[1]. With the aging population, the social and economic burden of PD will increase dramatically over the next 30 years, creating a substantial burden on healthcare systems with its prevalence in some age groups likely to double by 2030[1,23]. Thus far, despite our increasing understanding of PD, there is no neuroprotective treatment for PD, only treatments that provide some degree of symptomatic relief.

Since its description in 1817, PD was long thought to have no heritable component; the point of change in this notion occurred with the discovery of deleterious rare genetic variants in the mid 1990s and early 2000s [4–6]. The identification of rare genetic forms of disease served as the mainstay of PD genetics for many years; however, over the past decade, collaborative groups have worked together to investigate the genetic basis of apparently sporadic disease. These studies have grown from the first study of slightly more than 500 samples yielding no genome-wide significant risk loci to studies including tens of thousands of samples and defining dozens of definitive genetic risk factors[7,8]. In the most recent genome wide association study (GWAS), collaborators found genetic factors that are common in the population make a substantial contribution to PD, with heritability estimates of 20.9% explained by common variants [8]. That GWAS identified or confirmed 41 risk loci with 48 independent genetic risk factors, of which 29 were in linkage disequilibrium with cis-expression quantitative trait loci

(eQTL), suggesting that genetic risk for PD is mediated by the regulation of gene expression to some degree. Notably, as more risk loci have been resolved, it has become clear that a number of these risk loci harbor genes that also contain mutations that are likely causal or confer high-risk for PD. These include *SNCA, GBA, LRRK2, VPS13C* and *GCH1*; this phenomenon further demonstrates the potential strength of GWAS in identifying specific genes and loci with pathophysiological relevance to PD [5,9–12].

In addition to identifying specific genes or variants involved in PD, GWAS can enhance our understanding of the mechanism underlying PD pathogenesis. For example, many of the identified genes within the GWAS loci are involved in the autophagy lysosomal pathway, highlighting its importance in PD, and providing novel targets for therapeutics development [13]. Another major importance of GWAS is its ability to map the overall genetic susceptibility for PD to better quantify risk predictions, which may assist in early detection of PD. An increasing recognition that early detection is most likely required for successful treatment highlights the importance of such efforts.

In the current study, we performed the largest-to-date GWAS in PD, including over 11.4M SNPs, 37.7K cases, 18.6K “proxy-cases” and 1.4M controls, aiming to address all the issues above. Furthermore, we aimed to gain additional insight into nominated risk loci as potential contributors to actual disease processes via Mendelian randomization (MR) and protein-protein interaction network analyses[14,15]. We examined if more of the heritable risk can be explained by using a lower significance threshold for variant inclusion in PD risk profiling. Lastly, we surveyed genetic correlations between PD and other phenotypes of interest. Implicitly this work highlights the need for further investigation of the genetic basis of typical apparently sporadic PD, through future larger GWAS and genome sequencing studies.

METHODS

See Supplementary Methods

RESULTS

Novel loci identified and multiple signals in known loci

We identified a total of 92 genome-wide significant independent association signals through our meta-analysis and conditional analyses of 37,688 cases, 18,618 proxy-cases and 1,417,791 controls at 11,477,547 SNPs (Figure 1, Table 1, Supplementary Appendices and Supplementary Tables S1/S2 for details). Of these, 39 signals are new and more than 1MB from loci described in a previous report by Chang *et al*. 2017.

To maximize our power for locus discovery we used a single stage design, meta-analyzing all available summary statistics. We performed a variety of additional analyses to assess the

compatibility of cohorts comprising the meta-analysis. There was a strong genetic correlation between previously published datasets and new/proxy-case datasets using LD score regression (genetic correlation = 0.923, SE = 0.054), consistent with homogeneity of the phenotype. Both the previous datasets and new datasets when separately meta-analyzed 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 [16]. Additionally no dataset suffered from substantial lambda inflations, both with regard to raw lambdas (range for new datasets: 0.898 to 1.061) and lambas scaled to 1000 cases and 1000 controls ($lambda_{1000}$, range for new datasets: 0.741 to 1.044). The overall discovery meta-analysis was also quite well behaved with a raw lambda of 1.072, a scaled $lambda_{1000}$ of 1.000 and an LD score intercept of 0.991.

The implementation of conditional and joint analysis (GCTA-COJO, http://cnsgenomics.com/software/gcta/) analysis with a large study-specific reference genotype series plus participant level conditional analyses in 23andMe has facilitated the estimation of signal independence within risk loci [17]. If we define multi-signal loci as loci with variants within 250kb of the proximal nominated variants, then we have detected 22 separate genetic risk factors sharing loci (these can be annotated to 18 nearest gene regions, with 20 (90.9%) of these proximal risk factors within regions identified by previous GWAS). These include but are not limited to: two variants in the *GAK/TMEM175* region which remain independent of each other using this method, similarly two variants in the *NUCKS1/RAB29* region, two signals within the *SNCA* gene, two proximal *LRRK2* signals and another three in proximity to *GBA*. Detailed summary statistics on all nominated loci can be found in Supplementary Table S2.

Final sensitivity analyses included “leave-one-out” meta-analyses (LOOMA) comparisons of each dataset to a meta-analysis of the remaining datasets. After adjusting for multiple testing correction for 17 tests ($P < 0.003$ for significance) in regressions of up to 92 beta coefficients per iteration, we noted only 5 departures from significant correlations of betas 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 may be related to statistical power for detecting recently identified risk variants in this subset of smaller studies. While there may be some trepidation in utilizing proxy-cases in some instances, our data shows that the UKBB 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.

To further examine our study design and data we used a similar paradigm as the LOOMAs to compare risk beta coefficients to age at onset GWAS coefficients from an unpublished study of over 28K cases (Blauwendraat et al. 2018 under review, Supplementary Appendix). Associations between PD risk variants and age at onset are well documented[18,19]. We found strong correlations between betas across novel, known and all loci identified in this report and

those in the upcoming PD age at onset report, with r^2s ranging from 0.27 in the novel variants to 0.684 in the known variants (all $P < 0.017$, adjusting for 3 tests), suggesting a roughly 1.11 - 1.99 year earlier onset per unit increase in the cumulative log odds ratio across each variant strata.

Functional causal inferences via QTLs

Using two-sample MR methods, we gained insight into the potential biological underpinnings of genes underlying nominated risk loci (summarized in Table 2 and Supplementary Tables S3-S4). Our goal here was to interrogate gene level data for candidates in linkage disequilibrium under nominated risk peaks to help inform high throughput functional studies. In turn, these upcoming functional studies will aid in the fine mapping of these loci to putative functional variants.

We surveyed 349 gene regions in LD with our risk variants of interest, of which 282 genes had testable QTLs in either expression or methylation datasets we queried. Across four large QTL datasets in varied tissues of interest from blood to brain regions, stomach, and nerves assayed for mRNA expression or methylation, we identified 184 of these genes (65.2%) that may have some functional genomic aspects associated with PD risk. For all 88 genes annotated as nearest to our 92 SNPs of interest, we could test QTL associations and show that 53 of the 88 genes (60.2%) show some functional consequence via Mendelian randomization. It should be further noted that 40 loci only exhibited significant QTL associations with a single gene (Table S2), and in all but seven instances, the nearest gene was representative of the only QTL. Comparing rates of QTLs between genes nearest our risk variants of interest and the other genes under the LD-defined association peaks, there is no significant difference in rates of QTL associations overall (chi-squared $P = 0.290$). It is interesting to note that the only gene with significant functional consequences under the rs850738 / *FAM171A2* is *GRN*, a known gene associated with frontotemporal dementia[20]. This analysis also nominates *TOX3*, a candidate gene for restless leg syndrome as the likely functionally relevant gene under the rs3104783 / *CASC16* association peak[21].

Protein-protein networks and enriched expression pathways

We analyzed protein-protein interaction data using webgestaltR and also gene expression enrichment data using Functional Mapping and Annotation of Genome-Wide Association Studies (FUMA) to infer risk networks in PD [15,22]. Our goal was to connect genes underlying risk loci with similar biological functionality.

In our analyses of protein-protein interaction networks, we identified 10 functional networks sharing ontological and gene content overlaps that are significantly enriched for PD GWAS loci. Thematically, a majority of the networks identified are associated with response to some type of stressor or chemical signaling pathways. These significant networks are defined by the GO terms including: response to stress, response to interferon-gamma, response to organic

substance, cellular response to heat, regulation of protein stability, immune response-activating signal transduction, cell activation, regulation of proteasomal protein catabolic process and regulation of proteasomal ubiquitin-dependent protein catabolic process (see Table 3, Supplementary Figures S1). Connectivity of the genes comprising these 10 pathways is found in Supplementary Figure S2.

Additional expression derived analyses from FUMA suggest that among PD risk regions, expression in various brain regions is highly over represented (Supplementary Figure S3). Based on our PD GWAS data, all 12 tissues significantly enriched for expression at risk loci are brain derived, in contrast to what has been seen in Alzheimer's disease which shows a strong bias towards blood, spleen, lungs and microglial enrichments [23]. We do acknowledge that this may be somewhat distorted by long-range LD in regions such as *APOE*. 27 out of 10,651 tested pathways were enriched for PD associations after multiple test correction (Supplementary Table S5). Among these 27 enriched pathways, six annotations were related to vacuolar functionality and autophagy, three pathways for endosomal trafficking, two pathways for catabolism related functions, and two lysosomal pathways. Novel loci include nominations for at least 4 lysosomal storage disorder related genes *NAGLU*, *GUSB*, *GRN* and *NEU1*, a pathway of interest in recent PD studies[24].

For brevity, we focus only on a subset of genes with significant QTL associations via MR that are also nominated multiple times by our network analyses in the discussion section.

Risk profiling

Utilizing permutation testing to identify optimal *P* thresholds for variant inclusion in risk profiling shows that there are more meaningful GWAS variants to discover, with each new locus further expanding the breadth of biological knowledge and potentially increasing prediction accuracy. After surveying 11 cohorts from the LOOMAs, best thresholds for inclusion in the PRS were inclusive of over 4,000 unique variants below the standard P < 5E-08 used in many prior publications (1.25E-03 for targeted genotyping arrays like NeuroX and 9.85E-05 for standard GWAS arrays). Results of these analyses are detailed in Table 4 and Figure 2a/2b/2c.

After adjusting for appropriate covariates and accounting for an estimated PD prevalence of 0.5%, we see an overall pseudo r^2 between PRS and PD to be approximately 3% of the disease variance accounted for, corresponding to an overall AUC of 0.634 (95% CI 0.626 - 0.641) across 11 cohorts for the PRS alone[1]. This AUC for the PRS itself is significantly higher than had been published in Chang et al. 2017 based on DeLong's test for correlated receiver operator curves (ROC) at a p-value of 0.002 (previous PRS-only AUC = 0.624) in the IPDGC-NeuroX dataset. Using equations from Wray *et al*. 2010 and heritability estimates from Keller et al. 2012 our PRS at this AUC explains roughly 11% of the genomic liability in PD risk at a prevalence of 0.5%; in a high-risk population with a prevalence of 1-2% we could expect to explain 13-15% of the genomic liability with the PRS alone [25,26] .

For each standard deviation from the population mean of the PRS, we detected risk estimated to be an odds ratio of 1.763 (from random-effects across all 11 cohorts, beta = 0.567, SE = 0.035 and P = 2.98E-60). When comparing the lowest versus highest quartile of PRS estimated risk, membership in the highest risk quartile was associated with an odds ratio of 3.51 (95% CI = 3.26 - 3.79) as displayed in Figure 2a. There was some heterogeneity in effect estimates across studies relating to genotyping platforms with I2 estimates of heterogeneity for all arrays, targeted arrays (NeuroX) and GWAS genotyping are respectively 70.06%, 0% and 75.65% (Supplementary Appendix 2). Additionally we carried out a joint analysis of odds ratios across each decile of PRS per study compared to the remainder of samples. We show an OR of 2.86 in the 10th risk decile (95% CI 2.60 - 3.16), non-significant ORs around 1 at the 5th and 6th deciles, and a significant protective OR at the 1st decile of PRS estimated risk, with an OR of 0.41 (95% CI 0.37 - 0.45). When comparing the top 10% and bottom 10% of PRS across all samples, we see membership in the top 10% being associated with an OR of 5.83 (95% CI 5.14 - 6.62). We also acknowledge that there might be some effects of selection bias by evaluating P thresholds per study and then meta-analyzing, although effect estimates all show at least similar trends and remain significant in random-effects meta-analyses.

Genetic correlations across phenotypes

We analyzed cross-trait genetic correlations between PD and 757 other GWAS datasets of interest curated by LDhub [27]. After adjusting for multiple testing via false discovery rate, 4 genetic correlations remained significant (Supplementary Table S6). These include positive correlations with intracranial volume and mean putamen volume from Hribar et al. 2015 (for the former RG = 0.351, SE = 0.077, P = 4.64E-06 and for the latter RG = 0.248, SE = 0.064, P = 9.55E-05 respectively) [28]. We also note significant negative correlations between current tobacco use and PD as well as “academic qualifications: NVQ or HND or HNC or equivalent” and PD in the UKBB dataset, suggesting that there is some mitigating effect for smoking status and educational attainment as they relate to PD risk (for the former RG = -0.134, SE = 0.034, P = 7.92E-05 and for the latter RG = -0.169, SE 0.045, P = 2.00E-04 respectively) [29].

After noting these significant genetic correlations, we utilized MR methods to infer causality, using these phenotypes as exposures. The inverse variance weighted method was used to combine Wald ratios (or in the case of educational attainment by “Qualifications: NVQ or HND or HNC or equivalent”, a single Wald ratio was used since only 1 variant passed our pre-analysis clumping threshold of $r^2 = 0.001$, 10,000 kb) to ascertain significance of putative causal associations. Intracranial volume could not be tested due to the stringent nature of our pre-analysis filtering for this phase of MR. When these thresholds were relaxed to less stringent P thresholds for inclusion of SNPs from the intracranial volume GWAS, no significant associations were detected. In an analysis incorporating 16 variants of interest, current tobacco smoking status was not causally associated with PD risk (P > 0.05), suggesting shared genetic factors but no likely causative mechanism in a reduced SNP set from MR base. Educational attainment, quantified by attaining qualifications of NVQ or HND or HNC was associated with a decreased risk of PD (beta = -5.971, SE = 1.847, P = 0.001), although this should be viewed

with some guarded optimism as this analysis only includes a single variant (rs968050) and therefore sensitivity analyses such as MR Egger and radial analyses to further dissect the association are not possible. To note, the variant of interest in this case, rs968050, is implicated as a proxy for a genome-wide significant locus associated with bipolar disorder further complicating any possible association [30]. Putamen volume exhibited a small but significant risk association in inverse variance weighted MR analysis (beta = 7.81E-04, SE = 3.07E-04, P = 0.011). In addition, radial analyses using both inverse variance weighted and MR Egger models suggest significant heterogeneity in the risk estimates (Cochran's test for heterogeneity, P < 0.001 for tests of heterogeneity across variants), decreasing the likelihood of a truly causative association between putamen volume and PD, with likely violations of the assumptions of the MR paradigm. In radial analyses, one outlier variant in the putamen volume data was identified, rs62097986 (see Supplementary Figure S4). After excluding this variant, we saw a persistent association via inverse variance weighting (beta = 0.001, SE = 1.28E-04, P = 8.35E-16) although the MR Egger estimate was not significant (P > 0.05). This evidence suggests that putamen volume may have a causal relationship with PD, although more detailed future research is needed for a definitive answer. We also evaluated the possibility of reverse causality for these four GWAS, using our own PD derived summary statistics as an exposure. In this reverse causality analysis, no iterations of the inverse variance weighted or MR Egger analyses showed any significant associations (all P > 0.05).

Additional smoking analyses using bi-directional GSMR and expanded exposure GWAS data provided interesting results. These include no significant association signal in either direction for current smoking (beta = -0.032, SE = 0.031, P = 0.302 at 72 SNPs for current smoking status as an exposure for PD and beta = 0.016, SE = 0.010, P = 0.113 at 136 SNPs for the reverse). For smoking initiation (status as ever regularly smoked), we see a protective association when using smoking as an exposure for PD (beta = -0.081, SE = 0.034, P = 0.016 at 180 SNPs) although this does not pass multiple test correction after other Mendelian randomization analyses described earlier. When looking in the reverse direction, the association for PD as an exposure for smoking initiation is significant and is contrary to what is commonly expected based on the results described above. We see that PD is a risk factor with a very small but significant effect estimate associated with smoking initiation (beta = 0.027, SE = 0.006, P = 5.03E-06 at 136 SNPs), at roughly one third of the estimated effect size of its paired association in the opposite direction.

In our expanded Mendelian randomization analysis of recently published GWAS focusing on educational attainment and cognitive performance, we note interesting associations related to putative causal relationships between traits. Educational attainment is associated with risk of PD (beta = 0.125, SE = 0.038, P = 1.16E-03 at 549 SNPs) while a much smaller effect size but more significant result exists in the reverse direction (beta = 0.010, SE = 0.002, P = 2.38E-05 at 125 SNPs). Cognitive performance is a relatively large effect and significant risk factor for PD (beta = 0.242, SE = 0.042, P = 5.88E-09 at 197 SNPs) in our analyses, and the association in the reverse direction is not significant (P = 0.287 at 129 SNPs). The additional MR analyses of smoking and education exposures are summarized in Supplementary Figure S5.

DISCUSSION

Risk genes highlight new biology and candidate therapeutic targets

A key yield of GWAS is widening the scope of our biological knowledge base for diseases of interest, as well as making connections across diseases and other biological processes. We have identified 92 independent common genetic risk factors for a disease that was previously thought to be almost entirely environmental in its etiology a few decades ago. Through two-sample MR and network analysis of genes in linkage disequilibrium under association peaks, we have also nominated most likely regional targets for follow-up functional screening studies.

Bcl2-associated athanogene 3 (*BAG3*) is a candidate gene containing two risk signals on chromosome 10. BAG3 is a component of the proteolytic stress response and protein stability networks identified. It is also a consistently significant QTL across multiple datasets including increased risk of PD associated with elevated expression of *BAG3* in blood as well as increased risk associated with decreased expression in muscle and elevated methylation in blood.

GCH1, HLA-DRB5 and *SNCA* have been implicated in a network that may have some utility in drug development and neuroinflammatory studies, relating to response to interferon gamma. These three well-studied genes also comprise networks associated to stress responses and responses to organic substances (*i.e.* chemical stimuli and likely cytokines as per Supplementary Figure S1). *GCH1* expression in the caudate basal ganglia and tibial nerve tissues exhibits a strong inverse association with PD risk, while increased expression in blood and the cerebellum are associated with increased risk of PD. *GCH1* is GTP cyclohydrolase, a rate limiting enzyme co-factor for dopamine synthesis and it is biologically plausible that lower *GCH1* could increase risk PD. Coding mutations in GCH1 cause dopa-responsive dystonia, and have also been implicated as PD risk variants[12]. Increased methylation at two probes at *HLA-DRB5* are associated with increased PD risk. Decreased activity in terms of both methylation and expression at *SNCA* are associated with PD risk in blood as well as expression in the basal caudate ganglia.

PAM and *BRIP1* are novel loci identified in this report and both involved in networks related to stress and stimuli response on a cellular level. Peptidylglycine α-amidating monooxygenase (PAM) functions in neuropeptide secretion and interacts with copper deficiencies causing temperature dysregulation, seizures and anxiety in mouse models[31]. *BRIP1,* also known as *BACH1*, is associated with autosomal dominant breast cancers, ovarian cancers and Fanconi anemia[32–34]. This gene is a potential target for research into neuroprotective compounds as it is associated with oxidative stress and activation of nuclear factor erythroid 2-related factor 2 (Nrf2)[35]. Decreased expression of *PAM* is associated with PD risk across all cerebellar brain

regions and in all blood expression datasets. Increased expression of *BRIP1* in the GTEx basal ganglia tissues is associated with increased risk of PD.

FYN is a novel risk locus; *CTSB*, *DDRGK1* and *MAP3K14* are previously described risk loci, all are involved in response pathways described above. The *FYN* region identified as a novel PD risk locus in this report has been implicated in microglial inflammatory response [36]. Additionally, *FYN* in an activated state is known to phosphorylate α-synuclein and play a key role in dopamine trafficking [37–39]. *FYN* has also been implicated as a *MAPT* kinase involved in Alzheimer's disease [40]. *FYN*, *CTSB* and *MAP3K14* all share immune network connections. *MAP3K14* is also involved in interferon gamma response and appears in multiple overlapping enriched networks along with *DDRGK1* and *FYN*. Methylation levels at *FYN* are inversely associated with PD risk while *MAP3K14* shows the opposite directionality. *CTSB* and *DDRGK1* show increased risk of PD associated with blood expression levels, while *CTSB* also shows the inverse of that association in brain, nerve and stomach, and a subset of basal ganglia tissues [41].

Of additional interest is *GAK* and *TMEM175* both exhibiting significant associations with gene expression in the tibial nerve, and in opposite directions. This region has been of high interest in drug development but is complicated due to the independence of risk loci and the two genes sharing a promoter (Jinn et al. 2018, under review). Although, the effect size for the association at *TMEM175* (β = -0.3707, SE = 0.0582) has an absolute magnitude of effect almost twice that of *GAK* (β = 0.1465, SE = 0.0637). Also, between the two genes, *TMEM175* is the only one that is also associated with causal changes to expression in the cortical regions of the brain.

In its current state, the data presented in Tables 2 and 3 plus Supplementary Tables S3-S5 are suggestive or confirmatory of both known and novel biological pathways within PD etiology. It is of particular interest that this method can shed light on two nearby, but independent loci, such as *GAK*/*TMEM175*. While both variants remained significant and independent across analyses in this and in previous work, the MR analysis suggested more widespread impact in disease etiology associated with functional consequences related to the risk signals at *TMEM175* coding variation in tissues of interest as opposed to intronic variation in *GAK* . Additionally, this type of integrative data analysis has supported multiple novel loci that might have future therapeutic potential such as *BRIP1, PAM and FYN*. Significant functional insights described above are also graphically summarized with regard to significance (FDR adjusted P) and directionality across different tissues in Supplementary Figure S6.

Using LD score regression for cross trait analyses, we have highlighted new possible pathways of interest and confirmed previous hypotheses. The use of putamen and intracranial volumes as biomarkers may prove efficacious in future multi-modal modeling efforts. We also suggest at least some degree of possible functional connectivity in a causal pathway relating educational attainment and/or putamen volume to PD risk. The bi-directional GSMR results suggest a complex etiological connection between smoking initiation and PD as well as a strong association between cognitive performance and PD, each possibly relevant to multiple different mechanisms during disease manifestation.

Increased prediction strength and future prospects in GWAS

From a genetic perspective, we have succeeded in improving our predictive efforts by simply moving P thresholds and increasing sample sizes as has been shown previously for other complex traits and common diseases with a polygenic architecture. Current AUC estimates from our PRS suggest we have explained between 11-15% of Parkinson's genetic liability at an AUC approaching 64% from only common SNP variation. This is accomplished through current GWAS applications by including potential risk factors at up to P < 1E-03 in some datasets and using over 4,000 semi-independent risk variants. The same calculations provide a ceiling for predictions derived from heritable risk in PD at a maximum possible AUC of roughly 85%. These calculations coupled with the biological insights provided above suggest that GWAS and related methods have a great deal more insight to provide and should further help move us down a path towards quality etiologic-based interventions and diagnostics. Put simply, more samples, more data, more understanding, and more potential.

Current PRS derived predictive models, while explaining a respectable amount of disease liability, do not have positive predictive values that would be feasible to screen large populations. At current levels, on a population scale there would be roughly 14 false positives per real case using genetic data alone at an estimated prevalence of 0.5%. To improve on this, adding multimodal data such as smell tests, family history and similar low cost / non-invasive factors could be useful and one day provide a strong tool for prospective study and trial recruitment [42]. As an additional note, current collaborative efforts underway to utilize large-scale genome sequencing and nonlinear machine learning methods to build genetic classifiers may be able to push beyond the predictive boundaries detailed in this manuscript through the incorporation of rare genetic variation.

Limitations of the study

While we have made progress in assessing genetic risk factors for PD in this study, there are a number of limitations to our study.

One of the limitations of this study is the use of multiple imputation panels, due to logistic constraints. 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. Additionally, ancestry specific PD linkage disequilibrium reference panels of substantial size such as those for Ashkenazi Jewish participants at 23andMe will help us unravel further levels of detail in interesting loci such as *GBA* and *LRRK2*. This may be particularly evident at strongly associated loci such as *LRRK2* and *GBA* whose LD patterns may be quite variable within European populations, accentuating the possible influence of LD reference series on conditional analyses in some cases[43]. Moreover, larger QTL studies and PD-specific network data from large scale cellular screenings would allow us to build more robust functional inference particularly when many of our QTL associations are hampered by

both sample size and *cis*-SNP density. One major limitation in particular is the implementation of QTL analyses that only include one variant per cis-QTL tested. This can affect MR results by not only decreasing power due to lower variance explained, but also by preventing the use of tests such as MR egger or weighted median, which can be used to ascertain the likelihood of violating the inherent assumptions of MR.

To overcome these weaknesses and push the field forward, there are a few solutions that we should work toward. First of all, data transparency and openness, allowing all researchers to share participant-level data in a secure environment as this would facilitate inclusiveness and uniformity in analyses while maintaining the confidentiality of study participants. Also, the use of genome sequencing technologies could improve this effort with greater accuracy for rarer variants that are more difficult to impute, and better capture structural variations, although due to the need for very large sample sizes for analysis of rarer variants, well-powered analyses are likely quite far in the future. Outreach to underrepresented and diverse ancestry populations to build additional collaborations and include samples from non-European backgrounds could prove extremely valuable. Finally, current solutions such as federated (collaborative) learning methods could be applied to genomics to make strides for both the GWAS and machine learning approaches on a global scale while still maintaining privacy of individual sites when needed[44,45]. Tools like federated learning approaches could be key in building global analytics communities and more productive collaborations in genomics despite growing privacy restrictions. It is our goal that the next global scale efforts in PD genetics will be even larger, more open, better harmonized, and much more inclusive.

WORKS CITED

1. Gasser, T. Genetics of parkinson's disease. *Ann. Neurol.* **44,** S53–S57 (1998).
2. Dorsey, E. R. *et al.* Projected number of people with Parkinson disease in the most populous nations, 2005 through 2030. *Neurology* **68,** 384–386 (2007).
3. Dorsey, E. R., Ray Dorsey, E. & Bloem, B. R. The Parkinson Pandemic—A Call to Action. *JAMA Neurol.* **75,** 9 (2018).
4. Polymeropoulos, M. H. *et al.* Mapping of a gene for Parkinson's disease to chromosome 4q21-q23. *Science* **274,** 1197–1199 (1996).
5. Singleton, A. B. *et al.* alpha-Synuclein locus triplication causes Parkinson's disease. *Science* **302,** 841 (2003).
6. Parkinson, J. An Essay on the Shaking Palsy. *J. Neuropsychiatr.* **14,** 223–236 (2002).
7. Fung, H.-C. *et al.* Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. *Lancet Neurol.* **5,** 911–916 (2006).
8. Chang, D. *et al.* A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. *Nat. Genet.* **49,** 1511–1516 (2017).
9. Sidransky, E. *et al.* Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N. Engl. J. Med.* **361,** 1651–1661 (2009).
10. Paisán-Ruíz, C. *et al.* Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. *Neuron* **44,** 595–600 (2004).
11. Lesage, S. *et al.* Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy. *Am. J. Hum. Genet.* **98,** 500–513 (2016).
12. Mencacci, N. E. *et al.* Parkinson's disease in GTP cyclohydrolase 1 mutation carriers. *Brain*

137, 2480–2492 (2014).

13. Gan-Or, Z., Dion, P. A. & Rouleau, G. A. Genetic perspective on the role of the autophagy-lysosome pathway in Parkinson disease. *Autophagy* **11,** 1443–1457 (2015).
14. Noyce, A. J. & Nalls, M. A. Mendelian Randomization - the Key to Understanding Aspects of Parkinson's Disease Causation? *Mov. Disord.* **31,** 478–483 (2016).
15. 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.* **45,** W130–W137 (2017).
16. Bulik-Sullivan, B. K. *et al.* LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47,** 291–295 (2015).
17. Yang, J. *et al.* Conditional and joint multiple-SNP analysis of GWAS summary statistics identifies additional variants influencing complex traits. *Nat. Genet.* **44,** 369–75, S1–3 (2012).
18. Nalls, M. A. *et al.* Genetic risk and age in Parkinson's disease: Continuum not stratum. *Mov. Disord.* **30,** 850–854 (2015).
19. Escott-Price, V. *et al.* Polygenic risk of Parkinson disease is correlated with disease age at onset. *Ann. Neurol.* **77,** 582–591 (2015).
20. Cruts, M. *et al.* Null mutations in progranulin cause ubiquitin-positive frontotemporal dementia linked to chromosome 17q21. *Nature* **442,** 920–924 (2006).
21. Winkelmann, J. *et al.* Genome-wide association study identifies novel restless legs syndrome susceptibility loci on 2p14 and 16q12.1. *PLoS Genet.* **7,** e1002171 (2011).
22. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8,** 1826 (2017).
23. Jansen, I. *et al.* Genetic meta-analysis identifies 10 novel loci and functional pathways for

Alzheimer's disease risk. *bioRxiv* 258533 (2018). doi:10.1101/258533

24. Robak, L. A. *et al.* Excessive burden of lysosomal storage disorder gene variants in Parkinson's disease. *Brain* **140,** 3191–3203 (2017).
25. Keller, M. F. *et al.* Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. *Hum. Mol. Genet.* **21,** 4996–5009 (2012).
26. Wray, N. R., Yang, J., Goddard, M. E. & Visscher, P. M. The genetic interpretation of area under the ROC curve in genomic profiling. *PLoS Genet.* **6,** e1000864 (2010).
27. LD Hub. Available at: http://ldsc.broadinstitute.org. (Accessed: 26th June 2018)
28. Hibar, D. P. *et al.* Common genetic variants influence human subcortical brain structures. *Nature* **520,** 224–229 (2015).
29. Rapid GWAS of thousands of phenotypes for 337,000 samples in the UK Biobank. *Neale lab* Available at: http://www.nealelab.is/blog/2017/7/19/rapid-gwas-of-thousands-of-phenotypes-for-337000-samples-in-the-uk-biobank. (Accessed: 24th June 2018)
30. Mühleisen, T. W. *et al.* Genome-wide association study reveals two new risk loci for bipolar disorder. *Nat. Commun.* **5,** 3339 (2014).
31. Bousquet-Moore, D., Mains, R. E. & Eipper, B. A. Peptidylglycine α-amidating monooxygenase and copper: A gene-nutrient interaction critical to nervous system function. *J. Neurosci. Res.* **88,** 2535–2545 (2010).
32. Cantor, S. B. *et al.* BACH1, a novel helicase-like protein, interacts directly with BRCA1 and contributes to its DNA repair function. *Cell* **105,** 149–160 (2001).
33. Seal, S. *et al.* Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nat. Genet.* **38,** 1239–1241 (2006).
34. Litman, R. *et al.* BACH1 is critical for homologous recombination and appears to be the

Fanconi anemia gene product FANCJ. *Cancer Cell* **8,** 255–265 (2005).

35. Gazaryan, I. G. & Thomas, B. The status of Nrf2-based therapeutics: current perspectives and future prospects. *Neural Regeneration Res.* **11,** 1708–1711 (2016).
36. Panicker, N. *et al.* Fyn Kinase Regulates Microglial Neuroinflammatory Responses in Cell Culture and Animal Models of Parkinson's Disease. *J. Neurosci.* **35,** 10058–10077 (2015).
37. Nakamura, T., Yamashita, H., Takahashi, T. & Nakamura, S. Activated Fyn phosphorylates alpha-synuclein at tyrosine residue 125. *Biochem. Biophys. Res. Commun.* **280,** 1085–1092 (2001).
38. Dunah, A. W. Dopamine D1-Dependent Trafficking of Striatal N-Methyl-D-aspartate Glutamate Receptors Requires Fyn Protein Tyrosine Kinase but Not DARPP-32. *Mol. Pharmacol.* **65,** 121–129 (2004).
39. Phamluong, K., Darcq, E., Wu, S., Sakhai, S. A. & Ron, D. Fyn Signaling Is Compartmentalized to Dopamine D1 Receptor Expressing Neurons in the Dorsal Medial Striatum. *Front. Mol. Neurosci.* **10,** 273 (2017).
40. Lee, G. *et al.* Phosphorylation of tau by fyn: implications for Alzheimer's disease. *J. Neurosci.* **24,** 2304–2312 (2004).
41. McGlinchey, R. P. & Lee, J. C. Cysteine cathepsins are essential in lysosomal degradation of α-synuclein. *Proc. Natl. Acad. Sci. U. S. A.* **112,** 9322–9327 (2015).
42. Nalls, M. A. *et al.* Diagnosis of Parkinson's disease on the basis of clinical and genetic classification: a population-based modelling study. *Lancet Neurol.* **14,** 1002–1009 (2015).
43. Rivas, M. A. *et al.* Insights into the genetic epidemiology of Crohn's and rare diseases in the Ashkenazi Jewish population. *PLoS Genet.* **14,** e1007329 (2018).
44. Hofmann, T. & Basilico, J. Collaborative Machine Learning. in *From Integrated Publication and Information Systems to Information and Knowledge Environments* 173–182 (Springer,

Berlin, Heidelberg, 2005).

45. Google. Federated Learning: Collaborative Machine Learning without Centralized Training Data. *Research Blog* Available at: https://research.googleblog.com/2017/04/federated-learning-collaborative.html. (Accessed: 31st March 2018)

TABLES

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

SNP	CHR	BP	Nearest Gene	Effect allele	Other allele	Effect allele frequency	OR	Beta	SE	P, fixed-effects	P, COJO	P, conditional	P, random-effects	I2, %
rs6658353	1	161469054	FCGR2A	c	g	0.501	1.07	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	-0.070	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	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	-0.066	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	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	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	-0.063	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	-0.062	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	-0.084	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	-0.064	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	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	-0.092	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	-0.064	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	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	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	-0.221	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	-0.143	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	-0.061	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	-0.062	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	-0.079	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	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	-0.054	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	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	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	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	-0.053	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	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	-0.065	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	0.059	0.010	1.82E-09	1.53E-09	1.43E-03	1.82E-09	0
rs200564078	16	58587672	CNOT1	t	c	0.003	2.36	0.859	0.127	1.39E-11	1.40E-11	1.12E-04	1.39E-11	0
rs12600861	17	7355621	CHRNB1	a	c	0.648	0.95	-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	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	-0.071	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	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	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	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	-0.058	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	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	0.071	0.011	2.74E-11	2.51E-11	6.31E-05	8.78E-06	34.3

Table 2: Summary of significant functional inferences from QTL associations via Mendelian randomization for nominated genes of interest.

Gene	Beta	SE	P, FDR adjusted	Best analyte	Data source	Top QTL SNP	CHR, top QTL SNP	BP, top QTL SNP	Summary of significant MR associations (directionality, analyte and data source summarized)
ASXL3	-0.066	0.024	1.22E-02	Brain Cerebellum	GTEx expression	rs17747003	18	31332460	- Brain Cerebellum in GTEx
BAG3	0.324	0.102	3.12E-03	ILMN_1659766	Westra et al. expression	rs196345	10	121420922	+ ILMN_1659766 in Westra et al., - Muscle Skeletal in GTEx, + cg17076667 in ARIES
BIN3	0.123	0.023	3.03E-07	Brain Nucleus accumbens basal ganglia	GTEx expression	rs13264187	8	22506098	+ Brain Nucleus accumbens basal ganglia in GTEx, + Stomach in GTEx, + Brain Cerebellar Hemisphere in GTEx, + Whole Blood in GTEx, + ILMN_1708485 in Lloyd-Jones et al.
BRIP1	0.067	0.026	1.97E-02	Brain Nucleus accumbens basal ganglia	GTEx expression	rs12935923	17	59867775	+ Brain Nucleus accumbens basal ganglia in GTEx, + Brain Caudate basal ganglia in GTEx
BST1	-0.179	0.017	4.30E-24	ILMN_1770161	Westra et al. expression	rs4698412	4	15737348	- ILMN_1770161 in Westra et al., - ILMN_1770161 in Lloyd-Jones et al.
CAB39L	0.071	0.015	5.54E-06	Brain Cortex	GTEx expression	rs35214871	13	49918175	+ Brain Cortex in GTEx, + Brain Caudate basal ganglia in GTEx, + Brain Anterior cingulate cortex BA24 in GTEx, - cg19238349 in ARIES, + Muscle Skeletal in GTEx
CHRNB1	0.118	0.021	1.53E-07	Whole Blood	GTEx expression	rs67328311	17	7352976	+ Whole Blood in GTEx, + Muscle Skeletal in GTEx, + ILMN_1741131 in Westra et al., + ILMN_1741131 in Lloyd-Jones et al., + Brain Frontal Cortex BA9 in GTEx
CLCN3	0.104	0.025	1.91E-04	cg26072437	ARIES methylation	rs13435226	4	170550825	+ cg26072437 in ARIES
CRLS1	-0.099	0.017	9.80E-08	cg08376824	ARIES methylation	rs77351827	20	6006041	- cg08376824 in ARIES, - ILMN_1737685 in Lloyd-Jones et al., - Whole Blood in GTEx
CTSB	-0.179	0.022	6.71E-15	Brain Putamen basal ganglia	GTEx expression	rs1736103	8	11712639	- Brain Putamen basal ganglia in GTEx, - Brain Caudate basal ganglia in GTEx, - Nerve Tibial in GTEx, - Stomach in GTEx, + ILMN_1696360 in Lloyd-Jones et al., + ILMN_2359742 in Lloyd-Jones et al., - Brain Cerebellum in GTEx, + ILMN_2359742 in Westra et al.
DDRGK1	0.198	0.040	2.45E-06	Whole Blood	GTEx expression	rs2326083	20	3183348	+ Whole Blood in GTEx, + ILMN_1797828 in Lloyd-Jones et al.
DNAH17	0.120	0.020	2.63E-08	Nerve Tibial	GTEx expression	rs12942703	17	76420373	+ Nerve Tibial in GTEx, + Whole Blood in GTEx, + Muscle Skeletal in GTEx, - cg23053525 in ARIES
DYRK1A	0.350	0.092	3.72E-04	ILMN_2374293	Westra et al. expression	rs7281389	21	38712719	+ ILMN_2374293 in Westra et al., + ILMN_2374293 in Lloyd-Jones et al.
ELOVL7	-0.258	0.029	9.14E-18	Nerve Tibial	GTEx expression	rs921898	5	60089436	- Nerve Tibial in GTEx
FAM47E	-0.132	0.016	3.46E-16	Nerve Tibial	GTEx expression	rs11934863	4	77198791	- Nerve Tibial in GTEx, - Brain Putamen basal ganglia in GTEx, - Brain Nucleus accumbens basal ganglia in GTEx, + Muscle Skeletal in GTEx, - Brain Cerebellar Hemisphere in GTEx, - Brain Cortex in GTEx, - Brain Hippocampus in GTEx, - Brain Cerebellum in GTEx
FAM47E-STBD1	0.386	0.054	9.76E-12	Muscle Skeletal	GTEx expression	rs4859665	4	77240184	+ Muscle Skeletal in GTEx, + Stomach in GTEx, - Brain Cerebellum in GTEx, - Brain Cerebellar Hemisphere in GTEx, - Brain Caudate basal ganglia in GTEx
FBRSL1	0.149	0.030	5.93E-06	ILMN_1655694	Lloyd-Jones et al. CAGE	rs77773608	12	133066392	+ ILMN_1655694 in Lloyd-Jones et al., + cg05053403 in ARIES, + cg14078579 in ARIES, + cg04594507 in ARIES, + cg20454518 in ARIES, + cg06791426 in ARIES, - cg07116732 in ARIES, + cg24780570 in ARIES, - ILMN_1760305 in Lloyd-Jones et al., + Whole Blood in GTEx
FCGR2A	0.156	0.061	2.07E-02	ILMN_1706523	Lloyd-Jones et al. CAGE	rs1256286	1	161652717	+ ILMN_1706523 in Lloyd-Jones et al.
FYN	-0.082	0.016	1.97E-06	cg05517541	ARIES methylation	rs9481189	6	112113938	- cg05517541 in ARIES, - cg20706496 in ARIES, + Brain Nucleus accumbens basal ganglia in GTEx, - cg26846943 in ARIES
GAK	0.208	0.055	3.69E-04	Nerve Tibial	GTEx expression	rs17781378	4	880094	+ Nerve Tibial in GTEx, + Whole Blood in GTEx, - ILMN_1813775 in Westra et al., - ILMN_1813775 in Lloyd-Jones et al.
GALC	0.159	0.027	2.16E-08	ILMN_1799744	Westra et al. expression	rs12881373	14	88391333	+ ILMN_1799744 in Westra et al., + ILMN_2415572 in Lloyd-Jones et al., + ILMN_2415572 in Westra et al., + ILMN_1799744 in Lloyd-Jones et al.
GCH1	-0.129	0.020	1.32E-09	Brain Caudate basal ganglia	GTEx expression	rs72713460	14	55297043	- Brain Caudate basal ganglia in GTEx, - Nerve Tibial in GTEx, + ILMN_2335813 in Lloyd-Jones et al., + Whole Blood in GTEx, + Brain Cerebellar Hemisphere in GTEx, + ILMN_1812759 in Westra et al., + ILMN_1812759 in Lloyd-Jones et al.
GPNMB	0.158	0.015	2.20E-24	Brain Hippocampus	GTEx expression	rs858240	7	23284237	+ Brain Hippocampus in GTEx, + Brain Caudate basal ganglia in GTEx, + Brain Frontal Cortex BA9 in GTEx, + Brain Cortex in GTEx, + Brain Cerebellum in GTEx, + Brain Anterior cingulate cortex BA24 in GTEx, + Brain Putamen basal ganglia in GTEx, + Brain Nucleus accumbens basal ganglia in GTEx, + Brain Cerebellar Hemisphere in GTEx, + Stomach in GTEx, + Brain Hypothalamus in GTEx, + ILMN_1801205 in Westra et al., + Muscle Skeletal in GTEx
GS1-124K5.11	0.046	0.017	1.41E-02	Brain Putamen basal ganglia	GTEx expression	rs1167395	7	65573518	+ Brain Putamen basal ganglia in GTEx
HIP1R	-0.326	0.025	8.89E-36	Brain Cerebellum	GTEx expression	rs10847864	12	123326598	- Brain Cerebellum in GTEx, - Brain Cerebellar Hemisphere in GTEx
HLA-DRB5	0.207	0.027	1.58E-13	cg18816397	ARIES methylation	rs114879339	6	32587350	+ cg18816397 in ARIES
IGSF9B	-0.228	0.034	1.20E-10	Nerve Tibial	GTEx expression	rs329669	11	133795017	- Nerve Tibial in GTEx, + cg15086884 in ARIES, - cg19961173 in ARIES, - cg06067394 in ARIES, - cg27509693 in ARIES
INPP5F	-0.185	0.030	9.46E-09	cg10149021	ARIES methylation	rs1265535	10	121538407	- cg10149021 in ARIES
IP6K2	-0.111	0.033	1.64E-03	Brain Cerebellar Hemisphere	GTEx expression	rs73077157	3	49749705	- Brain Cerebellar Hemisphere in GTEx, + Muscle Skeletal in GTEx, + Brain Cerebellum in GTEx, + Brain Hypothalamus in GTEx, + ILMN_2328835 in Lloyd-Jones et al.
ITGA8	-0.082	0.014	7.21E-08	Brain Caudate basal ganglia	GTEx expression	rs7910668	10	15548925	- Brain Caudate basal ganglia in GTEx
KANSL1	0.174	0.023	8.16E-13	Brain Cerebellum	GTEx expression	rs62074125	17	44852612	+ Brain Cerebellum in GTEx, + Whole Blood in GTEx, + Muscle Skeletal in GTEx, + Brain Hippocampus in GTEx, + Nerve Tibial in GTEx, + Brain Cortex in GTEx, + Brain Nucleus accumbens basal ganglia in GTEx, - Brain Caudate basal ganglia in GTEx
KCNIP3	-0.126	0.023	4.77E-07	cg03595348	ARIES methylation	rs4283455	2	95999215	- cg03595348 in ARIES, - Brain Frontal Cortex BA9 in GTEx, - Brain Cerebellar Hemisphere in GTEx
KPNA1	0.429	0.065	2.60E-10	Whole Blood	GTEx expression	rs112787849	3	122213955	+ Whole Blood in GTEx, - Brain Cerebellar Hemisphere in GTEx, + ILMN_1676005 in Westra et al.
LINC00693	-0.115	0.045	2.24E-02	Muscle Skeletal	GTEx expression	rs7652455	3	28679411	- Muscle Skeletal in GTEx
LRRK2	0.283	0.026	9.41E-27	Nerve Tibial	GTEx expression	rs76904798	12	40614434	+ Nerve Tibial in GTEx, + Muscle Skeletal in GTEx, + Stomach in GTEx
MAP3K14	0.116	0.033	1.84E-03	cg15639951	ARIES methylation	rs7222094	17	43367653	+ cg15639951 in ARIES
MAP4K4	0.291	0.084	1.21E-03	ILMN_2375002	Westra et al. expression	rs2056400	2	102546801	+ ILMN_2375002 in Westra et al., + ILMN_2375002 in Lloyd-Jones et al.
MBNL2	-0.107	0.045	2.97E-02	Nerve Tibial	GTEx expression	rs9556704	13	97989911	- Nerve Tibial in GTEx
MCCC1	0.375	0.037	6.28E-23	Whole Blood	GTEx expression	rs10937106	3	182738287	+ Whole Blood in GTEx, + Muscle Skeletal in GTEx, + Nerve Tibial in GTEx, + ILMN_1760174 in Westra et al., - ILMN_1760174 in Lloyd-Jones et al.
MED12L	-0.124	0.050	3.51E-02	cg09432154	ARIES methylation	rs1565574	3	151067730	- cg09432154 in ARIES
MEX3C	-0.309	0.058	3.88E-07	Nerve Tibial	GTEx expression	rs62092947	18	48778064	- Nerve Tibial in GTEx, - ILMN_1658182 in Lloyd-Jones et al.
NSF	-0.910	0.108	5.67E-16	ILMN_1680353	Westra et al. expression	rs183211	17	44788310	- ILMN_1680353 in Westra et al., - ILMN_1680353 in Lloyd-Jones et al., - ILMN_2330845 in Lloyd-Jones et al., + Muscle Skeletal in GTEx, + Brain Cerebellar Hemisphere in GTEx
NUCKS1	-0.312	0.065	5.56E-06	Whole Blood	GTEx expression	rs823137	1	205738302	- Whole Blood in GTEx, - Brain Putamen basal ganglia in GTEx, - Muscle Skeletal in GTEx, - Brain Caudate basal ganglia in GTEx, - Nerve Tibial in GTEx, - ILMN_1680692 in Westra et al., - ILMN_1680692 in Lloyd-Jones et al., - ILMN_1869109 in Lloyd-Jones et al., - Stomach in GTEx
PAM	-0.059	0.018	3.00E-03	Brain Cerebellum	GTEx expression	rs2657437	5	102002577	- Brain Cerebellum in GTEx, - Brain Cerebellar Hemisphere in GTEx, - ILMN_2313901 in Westra et al., - Whole Blood in GTEx, - ILMN_2313901 in Lloyd-Jones et al.
RNF141	0.097	0.041	3.41E-02	Brain Hippocampus	GTEx expression	rs7109947	11	10887294	+ Brain Hippocampus in GTEx
SATB1	-0.070	0.027	1.97E-02	Brain Nucleus accumbens basal ganglia	GTEx expression	rs9859229	3	17991910	- Brain Nucleus accumbens basal ganglia in GTEx
SCARB2	0.387	0.061	2.51E-09	ILMN_1814726	Westra et al. expression	rs13122345	4	77179898	+ ILMN_1814726 in Westra et al., + Whole Blood in GTEx, - cg13870520 in ARIES, + ILMN_1814726 in Lloyd-Jones et al.
SEMA4A	-0.105	0.027	2.83E-04	Stomach	GTEx expression	rs7529564	1	156189793	- Stomach in GTEx, - Brain Cerebellar Hemisphere in GTEx, - Muscle Skeletal in GTEx
SH3GL2	0.171	0.039	3.36E-05	Stomach	GTEx expression	rs34816089	9	17537911	+ Stomach in GTEx, + Brain Cerebellar Hemisphere in GTEx, + Brain Cerebellum in GTEx
SNCA	-0.411	0.024	4.03E-66	cg14346243	ARIES methylation	rs1372518	4	90757294	- cg14346243 in ARIES, - cg26578617 in ARIES, - cg01966878 in ARIES, - cg15133208 in ARIES, - Brain Caudate basal ganglia in GTEx, - ILMN_1701933 in Westra et al., - ILMN_1766165 in Westra et al., - ILMN_1766165 in Lloyd-Jones et al., - ILMN_1701933 in Lloyd-Jones et al.
SPPL2B	0.203	0.037	2.44E-07	Brain Nucleus accumbens basal ganglia	GTEx expression	rs730960	19	2360806	+ Brain Nucleus accumbens basal ganglia in GTEx, + Brain Frontal Cortex BA9 in GTEx, + Brain Cerebellar Hemisphere in GTEx, + Brain Cerebellum in GTEx, + cg05660874 in ARIES, + Whole Blood in GTEx
SPTSSB	0.060	0.010	5.11E-08	Nerve Tibial	GTEx expression	rs13073147	3	161045237	+ Nerve Tibial in GTEx, + Brain Cerebellum in GTEx, + Brain Cerebellum in GTEx
STK39	-0.397	0.037	7.15E-25	Brain Putamen basal ganglia	GTEx expression	rs4668059	2	169166282	- Brain Putamen basal ganglia in GTEx, - ILMN_1791328 in Westra et al., + cg16875057 in ARIES, - ILMN_1791328 in Lloyd-Jones et al., - Muscle Skeletal in GTEx
SYT17	-0.350	0.082	5.73E-05	ILMN_1657760	Westra et al. expression	rs6497338	16	19274046	- ILMN_1657760 in Westra et al., - ILMN_1657760 in Lloyd-Jones et al.
TMEM175	-0.214	0.022	1.20E-21	cg01437235	ARIES methylation	rs11724898	4	935933	- cg01437235 in ARIES, - cg09279566 in ARIES, - cg15105011 in ARIES, - Nerve Tibial in GTEx, - Brain Cortex in GTEx, - ILMN_1654629 in Lloyd-Jones et al., - cg20814179 in ARIES, - cg06222618 in ARIES, - cg23939001 in ARIES
UBTF	-0.096	0.020	8.71E-06	cg13607699	ARIES methylation	rs2072081	17	42327493	- cg13607699 in ARIES
VAMP4	-0.247	0.043	6.85E-08	Whole Blood	GTEx expression	rs11590680	1	171716574	- Whole Blood in GTEx, - Brain Cerebellum in GTEx, - ILMN_1761363 in Westra et al., - ILMN_1804676 in Lloyd-Jones et al., - ILMN_1804676 in Westra et al.
WNT3	-0.382	0.023	1.94E-60	Nerve Tibial	GTEx expression	rs199530	17	44836653	- Nerve Tibial in GTEx, - Muscle Skeletal in GTEx, - Brain Cortex in GTEx, - Brain Cerebellar Hemisphere in GTEx, - Brain Cerebellum in GTEx, - Brain Nucleus accumbens basal ganglia in GTEx, - Brain Anterior cingulate cortex BA24 in GTEx, - Brain Frontal Cortex BA9 in GTEx, - Brain Caudate basal ganglia in GTEx, - Brain Hypothalamus in GTEx, - Brain Putamen basal ganglia in GTEx, - Brain Hippocampus in GTEx

Table 3: Protein network analysis for linked genes under association peaks.

GO ID	GO Name	Total number of genes in the network	Number of nominated genes within the network from Parkinson's loci	P	P, FDR adjusted	Genes comprising Parkinson's sub-network
GO:1900034	regulation of cellular response to heat	75	7	7.97E-07	0.012	NUPL2, HSPA1A, HSPA1L, MAPT, CCAR2, CAMK2D, BAG3
GO:0034341	response to interferon-gamma	139	8	5.05E-06	0.026	GCH1, HLA-DQA1, HLA-DRA, HLA-DRB5, SNCA, TRIM26, CAMK2D, PARP9
GO:0061136	regulation of proteasomal protein catabolic process	140	8	5.33E-06	0.026	ARIH2, USP19, LRRK2, HSPA1A, UBQLN4, CCAR2, DDRGK1, BAG6
GO:0032434	regulation of proteasomal ubiquitin-dependent protein catabolic process	105	7	7.73E-06	0.026	ARIH2, LRRK2, HSPA1A, UBQLN4, CCAR2, DDRGK1, BAG6
GO:0034605	cellular response to heat	107	7	8.76E-06	0.026	NUPL2, HSPA1A, HSPA1L, MAPT, CCAR2, CAMK2D, BAG3
GO:0031647	regulation of protein stability	213	9	1.57E-05	0.032	USP19, HSPA1A, CCAR2, SNCA, CCT3, USP4, BAG6, HIP1R, BAG3
GO:0010033	response to organic substance	2547	35	1.74E-05	0.032	USP19, LRRK2, CTSB, CYLD, CNOT1, FYN, GCH1, RABGEF1, HLA-DQA1, HLA-DRA, HLA-DRB5, FOXA1, HSPA1A, HSPA1L, LMNA, NFKB2, PAM, ATP6V0A1, POLR2A, PRKAR2A, NOD2, P2RY12, DDRGK1, SNCA, STX4, TIAL1, TUFM, UBTF, TRIM26, BAG6, CAMK2D, PARP9, BRIP1, ITGA8, MAP3K14
GO:0006950	response to stress	3249	41	1.78E-05	0.032	USP19, NUPL2, ERCC8, CLIC1, LRRK2, CTSB, DTX3L, CYLD, VKORC1L1, DYRK1A, CNOT1, FYN, GCH1, LAT, RABGEF1, HLA-DQA1, HLA-DRA, HLA-DRB5, HSPA1A, HSPA1L, LMNA, MAPT, NFKB2, PAM, POLR2A, PRKAR2A, CCAR2, NOD2, P2RY12, DDRGK1, SNCA, STX4, TIAL1, TRIM26, VKORC1, BAG6, CAMK2D, PARP9, BRIP1, TRIM15, BAG3
GO:0002757	immune response-activating signal transduction	401	12	1.91E-05	0.032	CTSB, CYLD, FCGR2A, FYN, LAT, RABGEF1, HLA-DQA1, HLA-DRA, HLA-DRB5, HSPA1A, NOD2, TRIM15
GO:0001775	cell activation	798	17	2.77E-05	0.042	CLIC1, CYLD, FYN, LAT, RABGEF1, HLA-DQA1, HLA-DRA, HLA-DRB5, MAPT, NFKB2, SATB1, NOD2, P2RY12, SNCA, STX4, GBF1, MAP3K14

Table 4: Summary of genetic predictive model performance.

Study	Best P threshold	pseudo R2 from PRS*	OR	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
Finnish Parkinson's	7.00E-07	0.038	1.87	0.628	0.084	9.94E-14	4.76	3.19 - 7.17	294	879	0.665	0.629 - 0.701	0.624	0.641	0.577	0.686	0.633
Harvard Biomarker Study!	1.50E-04	0.031	1.73	0.551	0.073	6.52E-14	3.31	2.30 - 4.80	951	999	0.639	0.604 - 0.673	0.677	0.555	0.630	0.607	0.616
McGill Parkinson's	3.50E-04	0.040	1.85	0.617	0.061	1.10E-23	4.28	3.13 - 5.90	2399	1487	0.660	0.632 - 0.688	0.677	0.579	0.508	0.736	0.628
IPDGC - Neurox!	1.50E-03	0.029	1.73	0.549	0.022	2.12E-133	3.66	3.28 - 4.09	2137	11243	0.640	0.630 - 0.645	0.638	0.560	0.575	0.623	0.599
Oslo Parkinson's Disease Study	2.00E-04	0.028	1.66	0.508	0.074	5.00E-12	3.84	2.63 - 5.65	1711	938	0.642	0.607 - 0.677	0.565	0.652	0.626	0.593	0.608
Parkinson's Disease Biomarker's Program!	5.50E-06	0.039	1.89	0.635	0.088	6.89E-13	5.70	3.57 - 9.26	398	794	0.672	0.634 - 0.710	0.602	0.681	0.774	0.485	0.641
Parkinson's Progression Markers Initiative!	1.20E-06	0.041	2.01	0.699	0.122	9.65E-09	4.20	2.41 - 7.50	291	528	0.645	0.594 - 0.695	0.642	0.594	0.777	0.430	0.618
Baylor College of Medicine / University of Maryland	2.49E-05	0.044	2.10	0.741	0.104	9.03E-13	7.31	4.18 - 13.63	707	964	0.682	0.642 - 0.721	0.572	0.718	0.889	0.299	0.645
Spanish Parkinson's	4.12E-05	0.019	1.53	0.423	0.039	1.21E-27	2.98	2.43 - 3.66	810	3443	0.611	0.592 - 0.630	0.461	0.701	0.710	0.451	0.581
Tubingen Parkinson's Disease cohort (CouragePD)	2.55E-06	0.017	1.48	0.393	0.062	2.62E-10	2.70	1.94 - 3.76	380	1208	0.603	0.571 - 0.635	0.803	0.343	0.600	0.587	0.573
Vance (dbGap phs000394)$	1.00E-04	0.044	2.08	0.731	0.092	2.08E-15	4.83	3.19 - 7.41	1231	919	0.663	0.625 - 0.701	0.742	0.495	0.753	0.481	0.619
All arrays$	7.69E-04	0.030	1.76	0.567	0.035	2.98E-60	3.51	3.26 - 3.79	4162	23402	0.634	0.626 - 0.641	0.608	0.583	0.623	0.567	0.595
Targeted arrays$	1.25E-03	0.030	1.75	0.558	0.021	1.67E-163	3.66	3.31 - 4.04	2490	13564	0.640	0.630 - 0.649	0.636	0.565	0.597	0.605	0.600
GWAS arrays$	9.85E-05	0.029	1.75	0.561	0.053	1.13E-26	3.37	3.00 - 3.80	2868	9838	0.627	0.616 - 0.638	0.588	0.594	0.658	0.521	0.591

Notes: * denotes R2 approximation adjusted for an estimated prevalence of 0.5%, equivalent to roughly half of the unadjusted R2 estimates for the PRS. ! denotes targeted genotyping series using NeuroX arrays. $ denotes effect estimates and summaries from either random effects meta-analysis or weighted means, although for N SNPs it is the number of unique SNPs across all studies. I2 estimates of heterogeneity for all arrays, targeted arrays and GWAS genotyping are respectively 70.06%, 0% and 75.65%. All calculations and reported statistics include only the PRS and no other parameters after adjusting for principal components 1-5, age and sex when possible.

Table 5: Significant genetic correlations across traits.

Trait of interest	PMID	Genetic correlation, RG	SE, RG	Z, RG	P, RG	P, FDR adjusted	Observed H2	SE, H2	H2 intercept	SE, H2 intercept	Cross trait intercept	SE, Cross trait intercept
Intracranial volume	25607358	0.351	0.077	4.580	4.64E-06	3.51E-03	0.166	0.045	1.003	0.007	-0.013	0.005
Current tobacco smoking	Not available (UKBB)	-0.134	0.034	-3.947	7.92E-05	2.41E-02	0.055	0.003	1.014	0.010	0.004	0.007
Mean Putamen	25607358	0.248	0.064	3.902	9.55E-05	2.41E-02	0.282	0.047	0.952	0.007	-0.007	0.006
Qualifications: NVQ or HND or HNC or equivalent	Not available (UKBB)	-0.169	0.045	-3.726	2.00E-04	3.79E-02	0.015	0.002	1.011	0.007	0.005	0.005

FIGURES

Figure 1: Iris plots of the current meta-GWAS data. Radius position denotes -log10 converted p-values (truncated for display) with the outer ring denoting chromosomal positions. Red points indicate SNPs reaching genome-wide significance (P < 5E-08.)

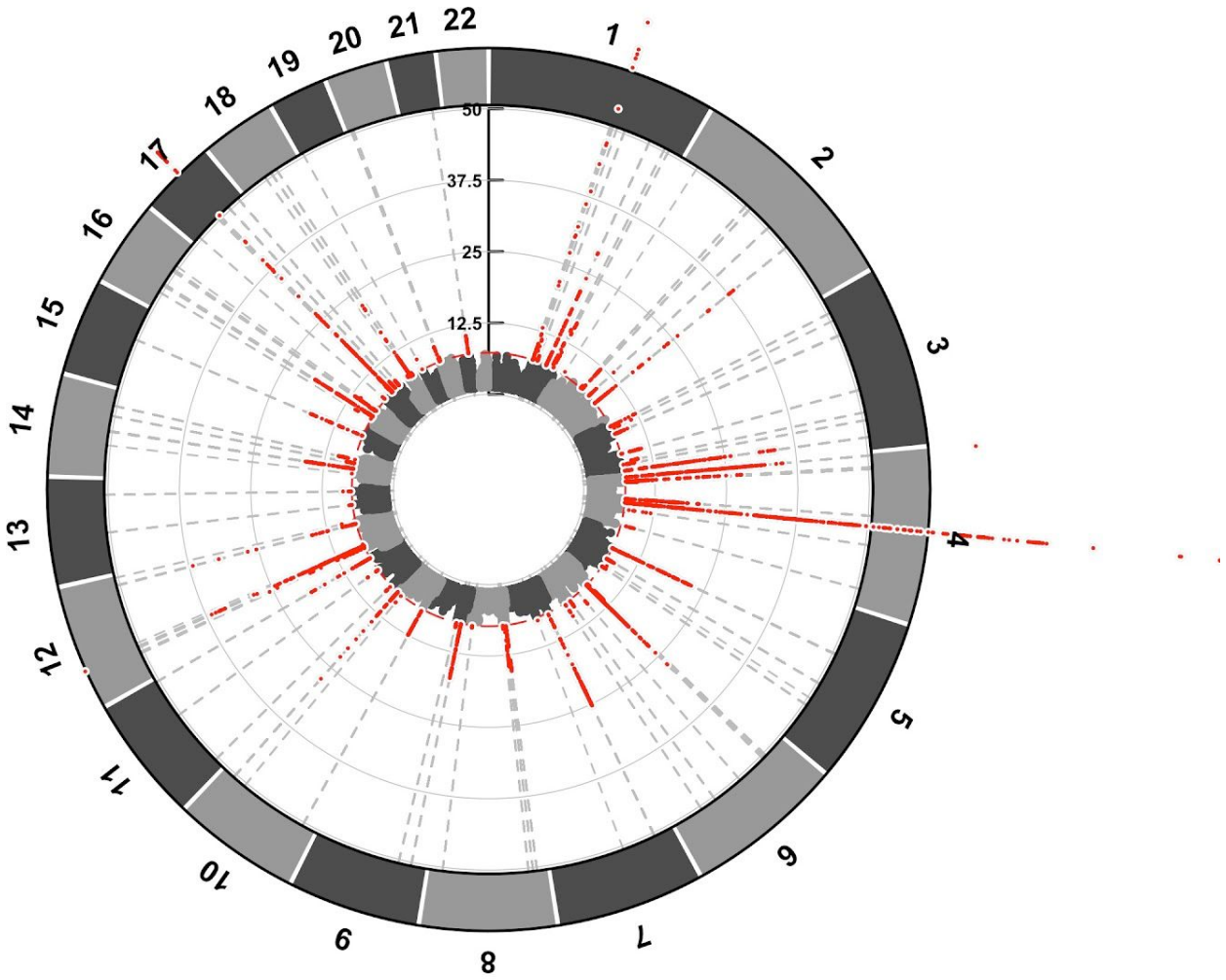
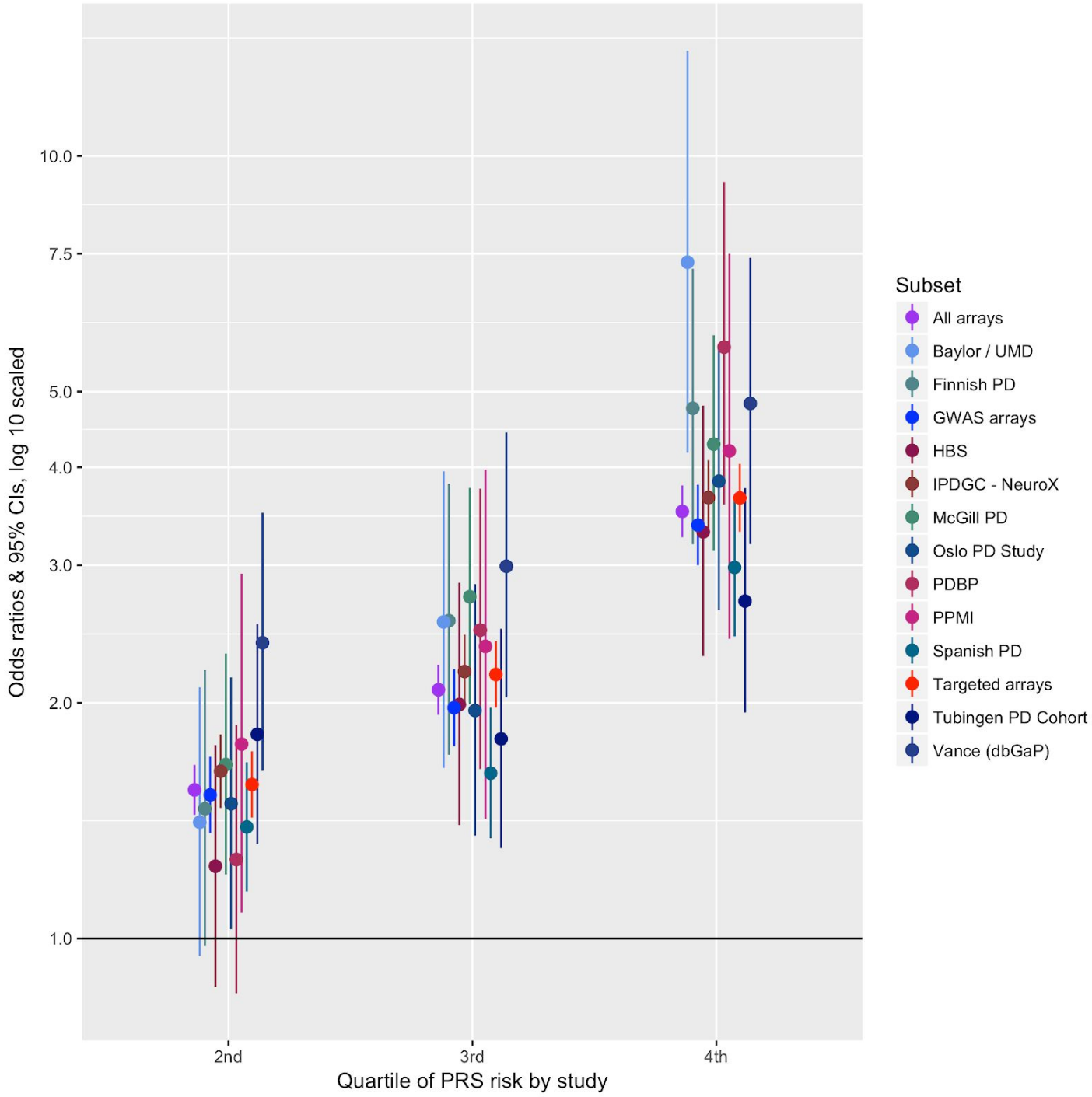
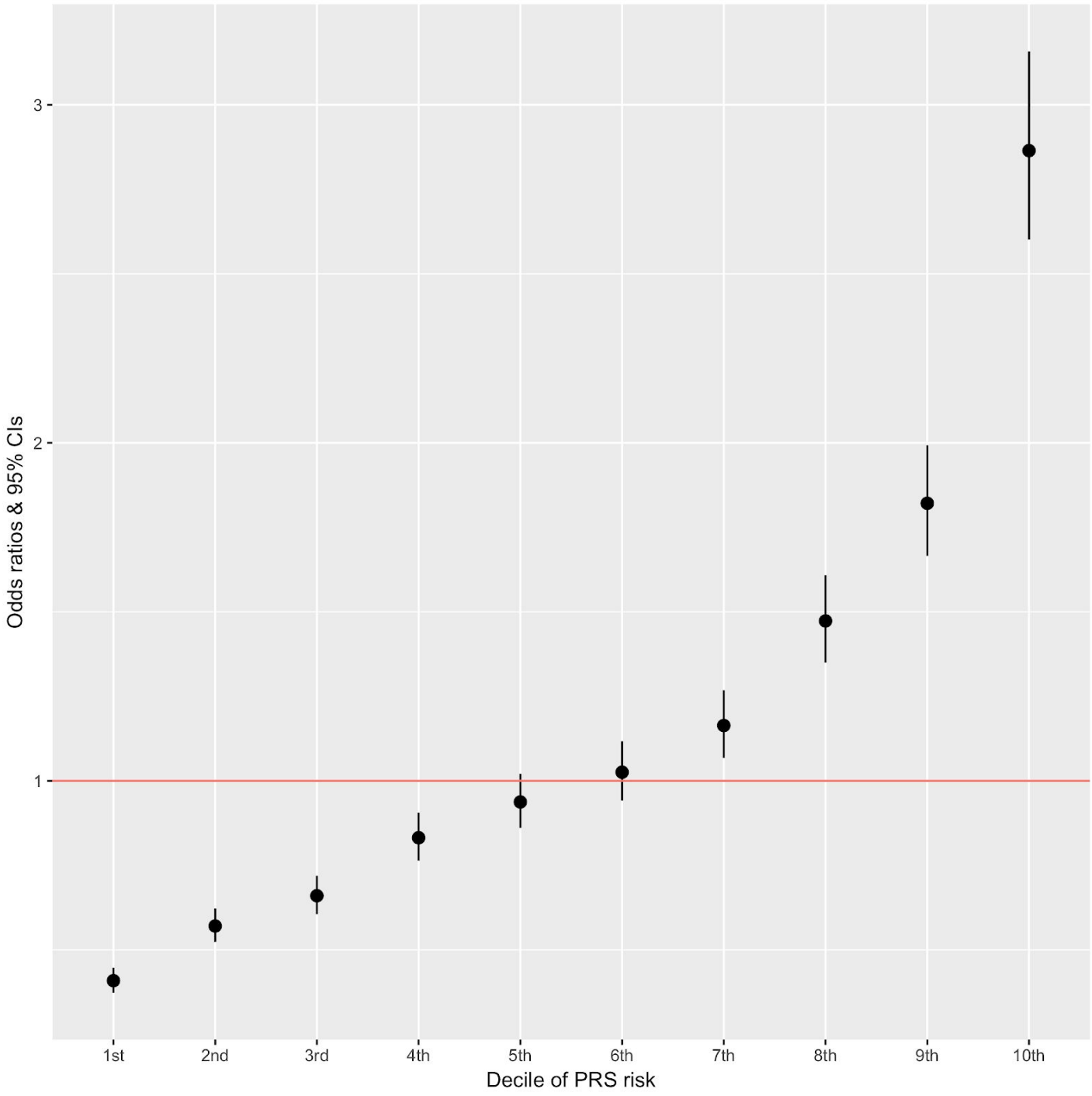


Figure 2: Predictive model details.

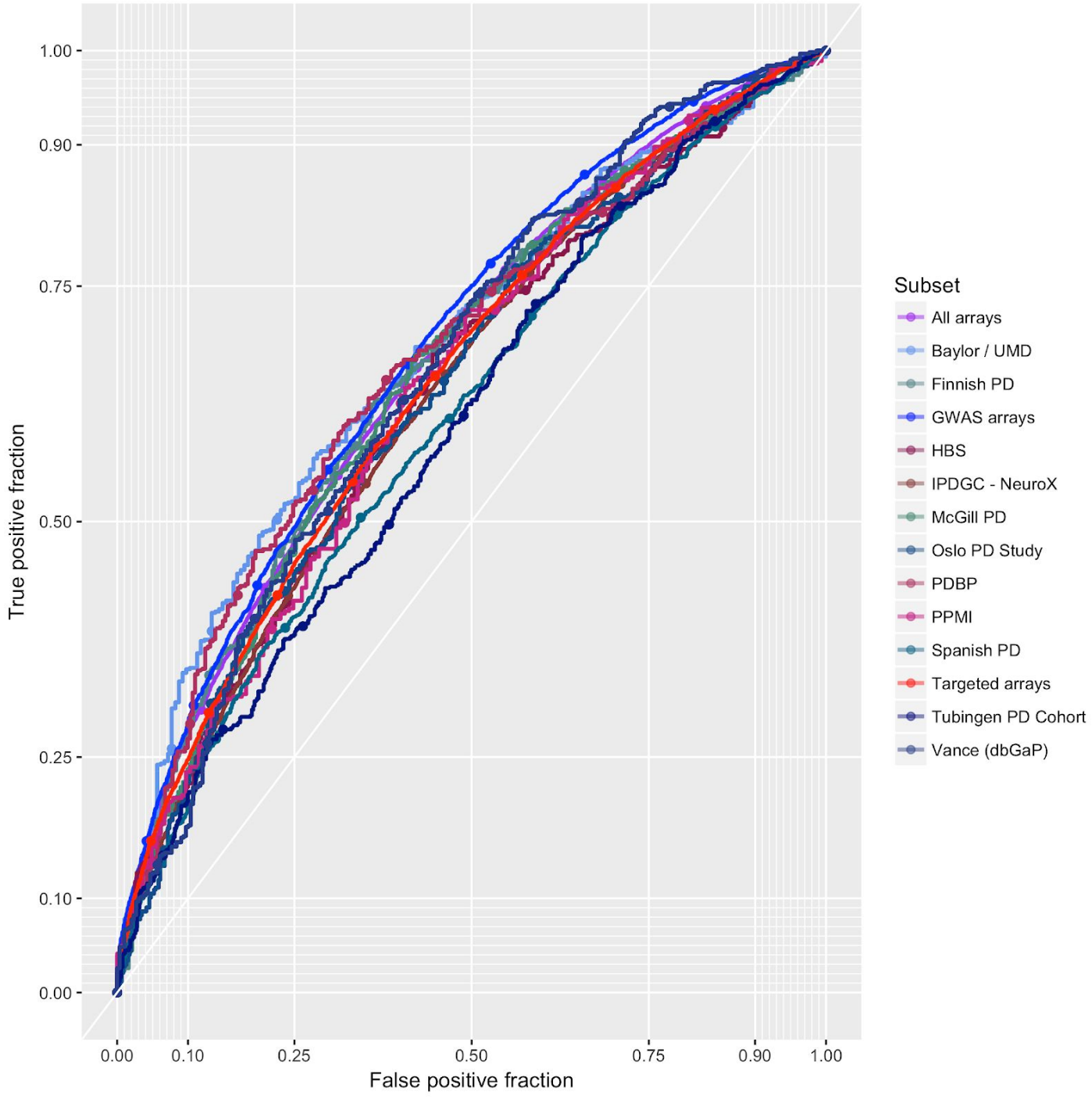
A. Odds ratios by quartile of PRS.



B. Odds ratios by decile of PRS, comparing each decle to all others.



C. PRS derived area under the curve estimates for the predictive models from receiver operator curve analyses.



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, J-CC, SL, JJ, LMS, MS, PT, KM, MT, AB, JY, ZG-O, TG, PH, JMS, NW, DAH, JH, HRM, JG, PMV, RRG, ABS

DATA ACCESS (POST PEER REVIEW)

GWAS summary statistics for 23andMe datasets (post-Chang and data included in Chang et al. 2017 and Nalls et al. 2014) will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please visit research.23andme.com/collaborate/#publication for more information and to apply to access the data.

Summary statistics excluding Nalls et al. 2014, 23andMe post-Chang et al. 2017 and PDWBS are available from https://github.com/neurogenetics/meta5.