HHS Public Access

Author manuscript

Nat Genet. Author manuscript; available in PMC 2018 February 14.

Published in final edited form as:

Nat Genet. 2017 October ; 49(10): 1511–1516. doi:10.1038/ng.3955.

A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci

Diana Chang[1], **Mike A Nalls**[2,3], **Ingileif B Hallgrímsdóttir**[4,6], **Julie Hunkapiller**[1], **Marcel van der Brug**[1,6], **Fang Cai**[1], **International Parkinson's Disease Genomics Consortium**[5], **23andMe ResearchTeam**[5], **Geoffrey A Kerchner**[1], **Gai Ayalon**[1], **Baris Bingol**[1], **Morgan Sheng**[1], **David Hinds**[4], **Timothy W Behrens**[1], **Andrew B Singleton**[2], **Tushar R Bhangale**[1,7], and **Robert R Graham**[1,7,iD]

[1]Genentech, Inc., South San Francisco, California, USA

[2]Laboratory of Neurogenetics, National Institute on Aging, US National Institutes of Health, Bethesda, Maryland, USA

[3]Data Tecnica International, Glen Echo, Maryland, USA

[4]23andMe Inc., Mountain View, California, USA

Abstract

Common variant genome-wide association studies (GWASs) have, to date, identified >24 risk loci for Parkinson's disease (PD). To discover additional loci, we carried out a GWAS comparing 6,476 PD cases with 302,042 controls, followed by a meta-analysis with a recent study of over 13,000 PD cases and 95,000 controls at 9,830 overlapping variants. We then tested 35 loci ($P < 1 \times 10^{-6}$) in a replication cohort of 5,851 cases and 5,866 controls. We identified 17 novel risk loci ($P < 5 \times 10^{-8}$) in a joint analysis of 26,035 cases and 403,190 controls. We used a neurocentric strategy to assign candidate risk genes to the loci. We identified protein-altering or *cis*–expression quantitative trait locus (*cis*-eQTL) variants in linkage disequilibrium with the index variant in 29 of the 41 PD loci. These results indicate a key role for autophagy and lysosomal biology in PD risk, and suggest potential new drug targets for PD.

Correspondence should be addressed to R.R.G. (graham.robert@gene.com).

[5]A list of members appears in Supplementary Note 1

[6]Present addresses: Amgen, South San Francisco, California, USA (I.B.H.); E-Scape Bio, South San Francisco, California, USA (M.v.d.B.).

[7]These authors contributed equally to this work

Robert R Graham http://orcid.org/0000-0001-7151-4277

URLs. PDGene, http://pdgene.org/; LDScore, https://github.com/bulik/ldsc; INRICH, https://atgu.mgh.harvard.edu/inrich/; GWAS catalog, https://www.ebi.ac.uk/gwas/; GTEx portal, http://gtexportal.org/; STRING, http://string-db.org/.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper

AUTHOR CONTRIBUTIONS

D.C., M.A.N., I.B.H., the 23andMe Research Team, G.A.K., B.B., M.S., D.H., T.W.B., A.B.S., T.R.B., and R.R.G. contributed to the study design. D.C., M.A.N., I.B.H., T.R.B., and D.H. contributed to analysis and methods. D.C., M.A.N., T.W.B., A.B.S., T.R.B., and R.R.G. wrote the manuscript. D.C., M.A.N., I.B.H., J.H., M.v.d.B., F.C., the International Parkinson's Disease Genomics Consortium (IPDGC), the 23andMe Research Team, G.A.K., G.A., B.B., M.S., D.H., T.W.B., A.B.S., T.R.B., and R.R.G. reviewed the manuscript. M.v.d.B., F.C., IPDGC and the 23andMe Research Team provided samples or data.

COMPETING FINANCIAL INTERESTS

The authors declare competing financial interests: details are available in the online version of the paper.

PD is the second most common neurodegenerative disorder[1,2], with a prevalence of 3–4% in individuals over 80 years of age[3]. PD is characterized by the loss of dopaminergic neurons in the substantia nigra and the presence of Lewy bodies[1,2]. These neuropathologies manifest in affected individuals primarily as motor-related symptoms, but the involvement of other brain regions can lead to nonmotor symptoms[4].

Early-onset, familial PD (onset at <60 years of age) accounts for a small fraction of cases[5], but the identified associated genes, including *LRRK2*, *GBA*, and *SNCA*, provide insight into disease pathogenesis[6,7]. For the later-onset, common form of PD, at least 24 loci have been associated at a genome-wide significant level with disease risk in individuals of European ancestry[8]. The narrow-sense heritability (h^2) explained by the confirmed PD risk loci is low (0.033)[9]; however, the heritability explained by common variants is estimated at 0.227 (s.d.: 0.08)[9], which suggests that additional loci with smaller effect sizes remain to be discovered.

We carried out a GWAS of 6,476 subjects from a 23andMe PD cohort (PDWBS (Web-Based Study of Parkinson's Disease)) and 302,042 controls genotyped on custom Illumina arrays (Fig. 1). The 6,476 PD cases of European ancestry were independent from those previously reported[8] but met the same inclusion criteria, except that carriers of the *LRRK2* G2019S mutation were not removed[8,10]. The 302,042 controls did not report having PD and were of similar ancestry as the cases. The data were imputed with Minimac2 using 1000 Genomes phase 1 haplotypes[11,12]. Single-nucleotide polymorphisms (SNPs) with low imputation quality or that failed general quality control metrics were removed (Online Methods). After correcting for age, sex, and the top principal components (Online Methods), we observed minimal inflation for *P* values genome-wide ($\lambda_{gc} = 1.057$; $\lambda_{1000} = 1.004$; Supplementary Fig. 1).

A total of 12 loci had $P < 5 \times 10^{-8}$ in the PDWBS analysis, including 11 of the loci that were reported in a previous GWAS in individuals of European ancestry[8] (Table 1). For the remaining 13 previously reported loci, we observed $P < 0.05$ for 11 loci, with no significant evidence for association observed in the PDWBS sample for *CHMP2B* (rs115185635) or *TMEM229B* (rs155399). The remaining novel locus in the PDWBS analysis, rs9468199 ($P = 1.77 \times 10^{-9}$), is more than 4 Mb from the nearest PD association in the HLA class II region and is independent of rs9275326 ($P_{conditional} = 2.64 \times 10^{-9}$).

Using genome-wide summary statistics from the PDWBS analysis, we estimated the h^2 value for PD explained by common variants as 0.209 (95% confidence interval (CI): 0.148–0.271, assuming a prevalence of 0.01), which is similar to the h^2 value reported previously[9,10]. Regions contributing to PD heritability were significantly enriched for acetylation of histone H3 at lysine 27 ($P = 0.001$; Supplementary Table 1), a mark of active regulatory regions. PD heritability was also enriched for histone marks in central nervous system, adrenal, and pancreatic cell types (Supplementary Table 2), in agreement with a previous study[13].

We next carried out a meta-analysis between the PDWBS GWAS and results for the top 10,000 variants available from a large-scale meta-analysis for PD with over 13,000 cases and 95,000 controls[8] (PDGene) (Fig. 1). For the 9,830 overlapping SNPs between the PDWBS

and PDGene studies, we used an inverse-variance weighted method to combine association statistics for meta-analysis[14]. The odds ratios and *P* values for the 9,830 overlapping SNPs in the PDWBS and PDGene studies were correlated ($\rho_{-\log10(P\text{value})} = 0.85$, $\rho_{OR} = 0.58$). Furthermore, quantile–quantile (Q-Q) plots indicated an increase in the number of variants with low *P* values (Supplementary Fig. 1), even after the exclusion of variants in regions previously reported as associated with PD risk at a genome-wide significant level (Supplementary Table 3).

The meta-analysis identified 35 loci associated at $P < 1 \times 10^{-6}$, including 15 loci with $P < 5 \times 10^{-8}$ (Fig. 2, Supplementary Figs. 2 and 3, Supplementary Table 4). Only two of the previously reported loci (*BCKDK*:rs14235 and *MAPT*:rs17649553) and 2 of the 20 suggestive loci (*FCGR2A*:rs4657041, *ITGA2B*:rs5910) were in linkage disequilibrium (LD) ($r^2 > 0.8$) with variants associated (at $P < 5 \times 10^{-8}$) with any phenotype in the NHGRI GWAS catalog[15]. Significant pleiotropy of PD risk loci with other complex diseases has not been identified[16], but this pleiotropy landscape may change as more modest effects are uncovered.

We next sought validation of these 35 candidate loci in an independent cohort of 5,851 cases and 5,866 controls of European ancestry genotyped with the semi-customized NeuroX Illumina array[8,17] (Fig. 1). Twenty-nine of the 35 loci either were directly genotyped on the NeuroX array or had suitable proxies ($r^2 > 0.9$ with the original SNP; Supplementary Table 5). Weaker proxies at four additional SNPs ($r^2 > 0.5$) were available but were not used for validation in this study (Supplementary Table 5). In a replication-phase joint analysis of these 29 loci (meta-analysis of PDGene, PDWBS, and NeuroX), 16 had $P < 5 \times 10^{-8}$ (Table 2). Of these 16, all but 3 (rs4073221, rs10906923, and rs9468199) were also nominally associated in the NeuroX study (one-sided $P < 0.05$). A genetic risk score[8,18,19] defined by these 16 loci, in addition to the previously reported loci, had a non-negligible ability to predict PD case status (area under the curve, 0.6518; 95% CI, 0.6419–0.6616). This represents a significant improvement over the predictive power of risk scores defined by previously reported loci alone ($P = 6 \times 10^{-8}$) (Supplementary Note 1). In sum, we identified 16 independent PD risk loci with a joint $P < 5 \times 10^{-8}$ and 1 locus (rs601999) with $P < 5 \times 10^{-8}$ in the discovery cohort with no suitable proxy for replication in the NeuroX cohort (Table 2).

Overall, 11 of 17 novel loci were in high LD ($r^2 > 0.8$) with at least one variant predicted to affect transcription factor binding (Supplementary Table 6). Of the 17 novel loci and 24 previously reported loci, 10 contained residual associations with $P < 1 \times 10^{-3}$ after conditioning on each region's most significant SNP in the PDWBS data (Supplementary Table 7). These regions included three of the four independent secondary signals reported by Nalls *et al.*[8], as well as one variant previously reported at a non-genome-wide significant level ($P = 5.15 \times 10^{-7}$)[20].

We note that the HLA region association with PD is particularly complex. Two candidate genes from the HLA region were nominated on the basis of support from either a protein-coding variant or an eQTL (Fig. 3). This is in line with a previous study that suggested that the PD association in the HLA region may point to multiple HLA factors, including

independent regulatory factors[21]. The association pattern observed at this locus may be reminiscent of the HLA association observed in schizophrenia and linked to C4 copy number[22].

The identification of the causal variants and genes underlying regions associated with common, complex disease is a major challenge[23]. Several statistical methods have been proposed for the fine-mapping of causal variants[23–25]. Alternatively, some studies have narrowed down lists of candidate genes by combining multiple levels of evidence with scoring-based strategies[26,27]. Here we implemented a neurocentric strategy to nominate candidate genes for PD-associated loci.

We incorporated seven sources of data to annotate the index variant and linked variants from PD-associated loci (including eQTLs and expression data from GTEx[28], as well as expression data from brain cell types in mice[29]; a full list is provided in the Online Methods). We used a two-stage approach to assign candidate genes to each locus (see the Online Methods for further details, and Supplementary Fig. 4 for a graphical visualization). In the first stage, we assigned a gene to a locus if (i) the index SNP or linked variants ($r^2 > 0.6$) altered the protein sequence or (ii) the index variant was a *cis*-eQTL for the gene. When no candidate genes were identified by the first stage, we ranked neighboring gene(s) on the basis of neurologically related phenotypes and expression and assigned the gene with the highest score to the locus (Online Methods).

With this strategy we identified a single candidate gene for 28 loci, and multiple candidate genes with similar levels of supporting evidence for 13 loci (Fig. 3, Supplementary Figs. 5 and 6). The candidate-gene nomination strategy confirmed several known PD risk genes, including *GBA*, *LRRK2*, *SNCA*, and *MAPT*. Among the 41 PD risk loci, a total of 29 loci (71%) had either a protein-altering or a *cis*-eQTL variant linked to the index SNP (Supplementary Tables 8 and 9). In addition, we carried out a colocalization analysis to determine whether the GWAS signal and the eQTL signal pointed to the same causal variant[30] (Supplementary Note 1). Seven candidate genes also had evidence for protein–protein interaction (Online Methods, Supplementary Table 10). Further studies are needed to experimentally determine the causal genes in the PD risk loci; however, the identification of candidate genes provides testable hypotheses for functional studies.

To gain insight into the biology, we tested the identified candidate genes in the 41 PD risk loci for association with any pathways or gene sets compared with a background gene list (Online Methods). We investigated whether candidate genes were enriched for pathways previously implicated in PD: autophagy, lysosomal, and mitochondrial biology[1]. PD-associated signals were enriched (at a threshold of $P < 0.05/3 = 0.017$) for lysosomal and autophagy genes ($P = 3.35 \times 10^{-6}$ and $P = 5.71 \times 10^{-3}$, respectively). The addition of candidate genes more than doubled the number of lysosomal genes observed in PD loci and improved the enrichment significance ($P_{\text{all\_loci}} = 3.35 \times 10^{-6}$, $P_{\text{novel\_loci}} = 3.64 \times 10^{-5}$). We also observed that one previously identified gene (*MCCC1*) and two novel candidate genes (*COQ7* and *ALAS1*) mapped to the mitochondrial gene set (Supplementary Table 11).

Lysosomal biology and its role in the degradation of protein aggregates emerged as a highly significant pathway in PD risk. Among the five candidate genes linked to lysosomal biology, two were previously identified candidate genes (*GBA* (glucocerebrosidase) and *TMEM175* (transmembrane protein 175)), and three were newly identified candidate genes (*CTSB* (cathepsin B), *ATP6V0A1* (ATPase H^+ transporting V0 subunit a1), and *GALC* (galactosylceramidase)). Glucocerebrosidase is required for normal lysosomal activity and α-synuclein degradation. In addition, *GBA* loss-of-function alleles are a common PD risk factor[31]. *TMEM175* was recently shown to encode a potassium channel that can regulate lysosomal function[32], and the missense variant *TMEM175* M393T is strongly linked to the index variant in the region (Supplementary Table 8). *CTSB* is a lysosomal cysteine protease. A PD risk allele is linked to a *cis*-eQTL for *CTSB* in multiple tissues (Supplementary Table 9), where the risk allele is associated with reduced levels of *CTSB* mRNA. Double-knockout mice for *Ctsb* and *Ctsl* (cathepsin L) show a tremor phenotype with cerebral and cerebellar atrophy[33]. CTSB is also capable of degrading membrane-bound and soluble α-synuclein in mice[34].

Autophagy is the catabolic process that targets long-lived proteins and dysfunctional organelles for lysosomal degradation. Autophagy and lysosomal degradation have been implicated in PD by rare familial and common GWAS-associated *GBA* variants. We note that a strong *cis*-eQTL for lysine acetyltransferase 8 (*KAT8*) is associated with PD risk, with lower levels of *KAT8* mRNA linked to increased PD risk. Inhibition of KAT8 was recently shown to decrease autophagic flux[35].

Next, we used INRICH[36] to investigate whether PD-associated regions were enriched for gene sets in an unbiased fashion. Once again, we found significant enrichment of the lysosomal pathway ($P_{adjusted} = 0.02$) (Supplementary Table 12). We further examined the expression of the PD candidate genes in a brain-specific cell-type expression data set in mice[29]; however, we observed broad expression across the major brain cell types, and no clear cell-type-specific pattern was evident (Supplementary Fig. 7).

Among the candidate genes newly identified in this study is *SH3GL2* (SH3 domain-containing GRB2-like 2, endophilin A1), a gene recently demonstrated to be phosphorylated by LRRK2 and which may have a role in clathrin-mediated endocytosis of synaptic vesicles[37]. Dysregulation of *Elovl7* (elongation of very long chain fatty acids protein 7) in mice results in several neurological phenotypes, including inflammatory astrocytosis and microgliosis in the brain, and neuronal degeneration[38]. Upregulation of the candidate gene *SCN3A* (sodium voltage-gated channel α-subunit 3) enhances neuronal excitability and is associated with epilepsy in both humans and animal models[39].

The new loci also encode three transcription factors: SATB1, ZNF184, and TOX3. TOX3 has been implicated in neuronal survival[40], and SATB1 has been associated with T cell function, particularly the development of regulatory T cells[41].

Several of the PD candidate genes are within the ‘druggable’ genome[42], including the previously identified serine/threonine kinase 39 (*STK39*) and the novel candidate gene inositol 1,4,5-trisphosphate kinase B (*ITPKB*). An in-frame deletion of *ITPKB*

(rs147889095) is linked to a PD-associated variant, and complete loss of ITPKB was reported in a patient with common-variable immunodeficiency[43]. STK39 is a kinase linked to hypertension[44], regulation of K^+ levels, and the cellular stress response.

In summary, this study presents what to our knowledge is the largest meta-analysis of PD so far, involving a total of 26,035 cases and 403,190 controls. We identified 17 novel PD loci and, using a neurocentric candidate-gene nomination pipeline, found that several of the newly identified PD risk genes have a role in lysosomal biology and autophagy. The identification of these candidate genes allows for the prioritization of functional studies to determine causal genes for PD and possible therapeutic targets.

ONLINE METHODS

PDWBS GWAS

The PDWBS is a genome-wide analysis of 6,476 PD cases and 302,042 control subjects, all of whom were customers of 23andMe Inc. and consented to participate in research. The study protocol was approved by the external AAHRPP-accredited institutional review board, Ethical and Independent Review Services (E&I Review). Cases and controls were designated on the basis of surveys[10]. Controls were selected from 23andMe Inc. research participants who did not self-report as having been diagnosed with PD. Although the use of self-reported controls can result in a reduction of power, the effect of this on the current study was probably minimal (Supplementary Note 1). Any samples present in the PDGene study[8] were removed from the PDWBS analysis. The average age of cases and controls was 67.6 and 50.8 years, respectively. The study also included 147 cases (2.3%) and 554 controls (0.18%) that were *LRRK2* G2019S carriers. Removing *LRRK2* G2019S carriers from the analysis removed genome-wide significant associations at the *LRRK2* locus.

DNA extraction and genotyping were performed on saliva samples by CLIA-certified CAP-accredited clinical laboratories of the Laboratory Corporation of America. Samples were genotyped on one of the following four platforms: V1 and V2, two variants of the Illumina HumanHap550+ BeadChip, with ~25,000 custom SNPs and ~950,000 total SNPs; V3, Illumina OmniExpress+BeadChip with custom SNPs to increase overlap with the V2 chip, with a total of ~950,000 SNPs; and V4, a custom chip that included SNPs overlapping V2 and V3 chips, low-frequency coding variants and ~570,000 SNPs. Samples with a call rate lower than 98.5% were reanalyzed, and research participants with samples that failed repeatedly were re-contacted and asked to provide additional samples.

Research participants were restricted to those of mainly (>97%) European ancestry[10,45]. All research participants in the study were also required to share <700 cM identity by descent (IBD) (estimated by a segmental IBD estimation algorithm[46]), corresponding approximately to the sharing expected between first cousins. We additionally excluded individuals who shared >700 cM IBD with any 23andMe research participant whose data was used in the PDGene GWAS. Data were imputed on 1000 Genomes phase 1 haplotypes (September 2013 release) with Minimac2 on default settings[11,47]. Imputation was run separately on data from each genotyping platform.

For genotyped SNPs, SNPs were removed if they were genotyped on only the V1 and/or the V2 chip, if they failed a parent–offspring transmission test on trio data, if they were not in Hardy–Weinberg equilibrium ($P < 10^{-20}$), or if they had a call rate < 0.90. For imputed SNPs, SNPs were removed if they had an average $r^2 < 0.5$ or minimum $r^2 < 0.3$ in any imputation batch, or failed a test for imputation batch effect (testing imputation dosage with imputation batch; $P < 10^{-50}$).

We applied logistic regression assuming an additive model to test for association between case/control status and either genotypes or imputed dosages (for imputed SNPs). Only SNPs with minor allele frequency (MAF) > 0.1% were analyzed. Covariates were added to adjust for age, sex, the first five principal components, and genotyping platform version. A total of 12,896,220 variants (11,933,700 SNPs) were analyzed. The genomic inflation factor was calculated from the median *P* value of analyzed variants. Scaling of the genomic inflation factor by sample size was carried out as described previously for 1,000 cases and 1,000 controls[48].

Meta-analysis of PD GWASs

Summary odds ratios, 95% CIs, and *P* values of the 10,000 most significant GWAS meta-analysis results were obtained from PDGene (“URLs”). Cohort descriptions, quality control, and meta-analysis for this study have been described previously[8]. SNP s.e. was derived from the reported *P* values and odds ratios. More specifically, the *z*-statistic was calculated as the square root of the inverse χ-square transformation of the *P* value, and the s.e. was calculated as follows: s.e.m. = ln(odds ratio)/absolute(*z*-score).

There were 9,830 SNPs in common between the PDGene and the PDWBS data sets. A fixed-effects model based on inverse-variance weighting, as implemented in METAL, was used to combine summary statistics from the two studies[14]. Heterogeneity values (I^2 and Q) were obtained with PLINK[49]. Novel signals of association were defined as genome-wide significant associations in the meta-analysis that did not overlap loci associated with PD at genome-wide significant thresholds in the PDGene data (35 loci with $P < 1 \times 10^{-6}$).

Joint analysis with NeuroX

The NeuroX cohort was previously described[8,17]. Briefly, 5,851 cases and 5,866 controls of European ancestry were genotyped on a semi-custom NeuroX array. A logistic regression was carried out to test for association, with covariates to adjust for age, sex, and population ancestry (the first five principal components). Twenty-five of the 35 novel loci were directly genotyped on the chip, and four additional SNPs had suitable proxies ($r^2 > 0.9$). At these 29 SNPs, we carried out a fixed-effects inverse-variance weighted meta-analysis14 for all three studies (PDGene, PDWBS, and NeuroX) as described above.

Conditional analysis

Conditional analysis was run on all 17 loci that were significantly associated with PD in the joint meta-analysis ($P < 5 \times 10^{-8}$) and the 24 previously reported PD loci using the PDWBS study. For each locus, SNPs within 500 kb of the index SNP (the SNP with the most

significant *P* value) were tested for association by the same methods as described above for the PDWBS GWAS with the index SNP added as an additional covariate.

Heritability estimates

We used LD score regression (LDSC)[50,51] to compute the narrow-sense heritability (h^2) estimates of PD in the PDWBS GWAS data (described above). Several methods exist for estimating h^2 with GWAS data[50–52]. We used LDSC to estimate h^2 in this study because it requires only summary-level data and is more computationally efficient for larger data sets. Reference LD scores were computed with the European ancestry subset of the 1000 Genomes data for SNPs within 500 kb of the SNP to be scored. Strict filtering was applied to ensure the robustness of heritability estimates as recommended[50,51]. After filtering, *Z*-scores for 7,629,099 SNPs from the 23andMe study were used as input to LDSC. We further used the stratified LD-score regression approach to partition heritability into 24 different cell-type-agnostic annotation categories including conserved regions, histone marks, DNase I hypersensitivity sites, ENCODE chromatin states, and enhancers[51], as well as 10 different cell-type-specific histone annotations. Significant enrichment was assessed at a strict Bonferroni threshold of 0.0021 (0.05/24) for the 24 general categories, and 0.005 for the cell-type-specific enrichment.

Pleiotropy analysis: overlap with EBI-NHGRI GWAS catalog

Data were downloaded from the EBI-NHGRI catalog15 (version available on 17 April 2016). If a variant in the meta-analysis was within 500 kb and in LD ($r^2 > 0.8$) with an association ($P < 5 \times 10^{-8}$) in the catalog, the meta-analysis signal was considered to be overlapping the reported signal.

A neurocentric strategy to identify candidate causal variants and genes

Associated index SNPs were paired to candidate genes on the basis of two broad levels of evidence: variant-level support and gene-level support (see Supplementary Fig. 4 for a graphic representation). In the former category, index SNPs were paired with candidate genes if there was evidence that the index SNP or an SNP in LD ($r^2 > 0.6$) with the index SNP was annotated with a putative high-impact variant (chromosome number variation, exon loss variant, frame-shift variant, rare amino acid variant, splice donor or acceptor variant, start-lost, stop-gained or stop-lost, and transcript ablation) or moderate-impact variant (3′ or 5′ UTR truncation and exon loss, coding sequence variant, disruptive in-frame deletion or insertion, in-frame deletion or insertion, missense variant, regulatory region ablation, splice region variant, and transcription factor binding-site ablation). We obtained variant annotations by running SnpEff[53] on dbSNP build 142. A second source of variant-level support consisted of *cis*-eQTL evidence. *Cis*-eQTLs as pre-computed by GTEx (v6)[28] were downloaded directly from the GTEx portal (“URLs”). Although eQTL results were available for 46 tissues, including ten regions from the brain, our search for eQTLs was limited by the sampled tissues and cell types, and therefore we might have missed any eQTLs that are cell-type or tissue specific, in addition to eQTLs that are present only under certain stimuli (for example, ‘response’ eQTLs). The index SNP was tested for significant association with any gene where the TSS was within 250 kb of the index SNP. As roughly 90% of eQTLs are within 250 kb of a gene[28], it is likely that we captured the majority of

eQTLs while missing rarer, more distal events. For brain eQTLs, a strict Bonferroni correction was applied to the raw eQTL *P* values to adjust for multiple testing of genes within 250 kb of the index SNP. For other tissues, only eQTLs with a false discovery rate of <0.05 as determined by GTEx[28] were considered. We weighted brain and non-brain eQTLs equally.

Gene-level support was used when an index SNP had no candidate genes supported by variant-level data as described above. A list of genes within 250 kb of the index SNP was obtained (gene models used by GTEx were downloaded from the GTEx portal), and each gene was scored for neurological relevant features or annotations. Genes were first weighted for neurological relevant phenotypic annotations. Genes were (i) scored for being differentially expressed between PD patients and healthy controls (see Supplementary Note 1 for further details) (311 genes total genome-wide); (ii) annotated with 'neuro'-associated phenotypes in FlyBase[54] (1,521 genes); (iii) scored for behavioral, neurological, and olfactory phenotypes annotated in MGI55 (3,890 genes); and (iv) annotated with any phenotypes related to neurological disorders or the brain in OMIM[56] (521 genes). Lastly, genes were scored for being expressed (median expression across samples > 2 reads per kilobase per million mapped reads) in any cohort of GTEx brain region samples (15,197 genes) or in at least one brain cell type in the mouse expression data set[29] (astrocyte, microglia, neuron, or oligodendrocyte) (12,092 genes). For the gene-level support, we used a tiered scoring scheme to weight phenotypic annotations more heavily than expression in the brain (scores demarked in Supplementary Fig. 4) to enrich for genes with demonstrated neurological related roles. At each locus, the gene (or tied genes) with the highest score was nominated as the candidate gene for the region.

Protein–protein and coexpression analysis

All protein-coding genes within 250 kb of PD-associated loci (Supplementary Table 13) were used as input to STRING[57]. Gene pairs that were either coexpressed or involved in experimentally validated protein–protein interactions with a medium score or higher (score ≥ 0.4) are reported in Supplementary Table 10.

Pathway enrichment analysis

Previously reported PD loci and novel PD associations were tested for enrichment in particular pathways or gene sets. First, the nominated candidate genes for these PD-associated loci were tested for enrichment in several targeted gene sets by a hypergeometric test. The background list of genes for comparison was matched to the neurological-centric candidate-gene nomination pipeline. The background list thus consisted of genes that had mouse knockout phenotypes, had fly mutant phenotypes, had OMIM-related phenotype annotations, had nominally significant *cis*-eQTLs in GTEx, were differentially expressed in PD patients versus controls, and were expressed in GTEx brain tissue or mouse brain cell types in the Barres data set.

We obtained mitochondrial genes from MitoMiner[58] using the MitoCarta[59,60] reference set after excluding genes that mapped to the mitochondria (genes that map to the mitochondria were not included in this metaanalysis). Lysosomal genes were obtained from the hlGDB[61]

using only the proteomics and literature resources. Finally, we obtained autophagy genes from the Human Autophagy Database[62], as well as ten additional genes reported in a recent siRNA screen of autophagic flux modulators[35]. A list of all genes in each pathway is provided in Supplementary Note 1. The minimum *P* value per gene (for genes that an SNP within the 9,830 variants assayed in this meta-analysis mapped to) is provided in Supplementary Tables 14–16.

Second, we applied a non-targeted gene-set enrichment approach using INRICH[36] to assess whether regions associated with PD were enriched for genes in KEGG[63] and Gene Ontology (GO)[64] gene sets. The 24 previously reported PD index variants and the 17 novel PD-associated variants reported in this study were used as input into PLINK's[65] "show-tags" function. The European 1000 Genomes[12] samples were used for reference LD patterns. An interval for each PD-associated variant was defined as the region from the leftmost tag variant to the rightmost tag variant in the 1000 Genomes data.

We ran INRICH on these 41 intervals with the default settings, with the exception of increasing the number of replicates and bootstraps to 5,000 (-r 5000 --q 5000) and setting the pre-compute feature to false for software stability (-c). Enrichment for KEGG and GO gene sets was assessed separately.

Data availability

A **Life Sciences Reporting Summary** for this paper is available. Summary statistics for the 9,830 variants presented in the discovery phase meta-analysis are available at http://research-pub.gene.com/chang_et_al_2017. The full GWAS summary statistics for PDWBS will be made available through 23andMe and Genentech to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants and an agreement with Genentech for data sharing. Please contact D.H. (dhinds23andme.com) for more information and to apply to access the data.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank all of the subjects who donated their time and biological samples to be a part of this study. Funding details and additional acknowledgments are provided in Supplementary Note 1.

References

1. Corti O, Lesage S, Brice A. What genetics tells us about the causes and mechanisms of Parkinson's disease. Physiol. Rev. 2011; 91:1161–1218. [PubMed: 22013209]
2. Verstraeten A, Theuns J, Van Broeckhoven C. Progress in unraveling the genetic etiology of Parkinson disease in a genomic era. Trends Genet. 2015; 31:140–149. [PubMed: 25703649]
3. Nussbaum RL, Ellis CE. Alzheimer's disease and Parkinson's disease. N. Engl. J. Med. 2003; 348:1356–1364. [PubMed: 12672864]
4. Shulman JM, De Jager PL, Feany MB. Parkinson's disease: genetics and pathogenesis. Annu. Rev. Pathol. 2011; 6:193–222. [PubMed: 21034221]

5. Klein C, Westenberger A. Genetics of Parkinson's disease. Cold Spring Harb. Perspect. Med. 2012; 2:a008888. [PubMed: 22315721]
6. Hardy J. Genetic analysis of pathways to Parkinson disease. Neuron. 2010; 68:201–206. [PubMed: 20955928]
7. Singleton AB, Farrer MJ, Bonifati V. The genetics of Parkinson's disease: progress and therapeutic implications. Mov. Disord. 2013; 28:14–23. [PubMed: 23389780]
8. Nalls MA, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat. Genet. 2014; 46:989–993. [PubMed: 25064009]
9. Keller MF, et al. Using genome-wide complex trait analysis to quantify 'missing heritability' in Parkinson's disease. Hum. Mol. Genet. 2012; 21:4996–5009. [PubMed: 22892372]
10. Do CB, et al. Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. PLoS Genet. 2011; 7:e1002141. [PubMed: 21738487]
11. Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. Bioinformatics. 2015; 31:782–784. [PubMed: 25338720]
12. Abecasis GR, et al. An integrated map of genetic variation from 1,092 human genomes. Nature. 2012; 491:56–65. [PubMed: 23128226]
13. Gagliano SA, et al. Genomics implicates adaptive and innate immunity in Alzheimer's and Parkinson's diseases. Ann. Clin. Transl. Neurol. 2016; 3:924–933. [PubMed: 28097204]
14. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics. 2010; 26:2190–2191. [PubMed: 20616382]
15. Welter D, et al. The NHGRI GWAS Catalog, a curated resource of SNP-trait associations. Nucleic Acids Res. 2014; 42:D1001–D1006. [PubMed: 24316577]
16. Pickrell JK, et al. Detection and interpretation of shared genetic influences on 42 human traits. Nat. Genet. 2016; 48:709–717. [PubMed: 27182965]
17. Nalls MA, et al. NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases. Neurobiol. Aging. 2015; 36:1605.e7–1605.e12.
18. Nalls MA, 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; 377:641–649. [PubMed: 21292315]
19. International Parkinson's Disease Genomics Consortium & Wellcome Trust Case Control Consortium 2. A two-stage meta-analysis identifies several new loci for Parkinson's disease. PLoS Genet. 2011; 7:e1002142. [PubMed: 21738488]
20. Pankratz N, et al. Meta-analysis of Parkinson's disease: identification of a novel locus. RIT2. Ann. Neurol. 2012; 71:370–384. [PubMed: 22451204]
21. Wissemann WT, et al. Association of Parkinson disease with structural and regulatory variants in the HLA region. Am. J. Hum. Genet. 2013; 93:984–993. [PubMed: 24183452]
22. Sekar A, et al. Schizophrenia risk from complex variation of complement component 4. Nature. 2016; 530:177–183. [PubMed: 26814963]
23. Kichaev G, et al. Integrating functional data to prioritize causal variants in statistical fine-mapping studies. PLoS Genet. 2014; 10:e1004722. [PubMed: 25357204]
24. Maller JB, et al. Bayesian refinement of association signals for 14 loci in 3 common diseases. Nat. Genet. 2012; 44:1294–1301. [PubMed: 23104008]
25. Chen W, et al. Fine mapping causal variants with an approximate Bayesian method using marginal test statistics. Genetics. 2015; 200:719–736. [PubMed: 25948564]
26. Okada Y, et al. Genetics of rheumatoid arthritis contributes to biology and drug discovery. Nature. 2014; 506:376–381. [PubMed: 24390342]
27. Bentham J, et al. Genetic association analyses implicate aberrant regulation of innate and adaptive immunity genes in the pathogenesis of systemic lupus erythematosus. Nat. Genet. 2015; 47:1457–1464. [PubMed: 26502338]
28. GTEx Consortium. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. Science. 2015; 348:648–660. [PubMed: 25954001]
29. Zhang Y, et al. An RNA-sequencing transcriptome and splicing database of glia, neurons, and vascular cells of the cerebral cortex. J. Neurosci. 2014; 34:11929–11947. [PubMed: 25186741]

30. Giambartolomei C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet. 2014; 10:e1004383. [PubMed: 24830394]
31. Sidransky E, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N. Engl. J. Med. 2009; 361:1651–1661. [PubMed: 19846850]
32. Cang C, Aranda K, Seo YJ, Gasnier B, Ren D. TMEM175 is an organelle K^+ channel regulating lysosomal function. Cell. 2015; 162:1101–1112. [PubMed: 26317472]
33. Felbor U, et al. Neuronal loss and brain atrophy in mice lacking cathepsins B and L. Proc. Natl. Acad. Sci. USA. 2002; 99:7883–7888. [PubMed: 12048238]
34. McGlinchey RP, Lee JC. Cysteine cathepsins are essential in lysosomal degradation of α-synuclein. Proc. Natl. Acad. Sci. USA. 2015; 112:9322–9327. [PubMed: 26170293]
35. Hale CM, et al. Identification of modulators of autophagic flux in an image-based high content siRNA screen. Autophagy. 2016; 12:713–726. [PubMed: 27050463]
36. Lee PH, O'Dushlaine C, Thomas B, Purcell SM. INRICH: interval-based enrichment analysis for genome-wide association studies. Bioinformatics. 2012; 28:1797–1799. [PubMed: 22513993]
37. Arranz AM, et al. LRRK2 functions in synaptic vesicle endocytosis through a kinase-dependent mechanism. J. Cell Sci. 2015; 128:541–552. [PubMed: 25501810]
38. Shin D, Shin JY, McManus MT, Ptácek LJ, Fu YH. Dicer ablation in oligodendrocytes provokes neuronal impairment in mice. Ann. Neurol. 2009; 66:843–857. [PubMed: 20035504]
39. Tan NN, et al. Epigenetic downregulation of *Scn3a* expression by valproate: a possible role in its anticonvulsant activity. Mol. Neurobiol. 2016; 54:2831–2842. [PubMed: 27013471]
40. Dittmer S, et al. TOX3 is a neuronal survival factor that induces transcription depending on the presence of CITED1 or phosphorylated CREB in the transcriptionally active complex. J. Cell Sci. 2011; 124:252–260. [PubMed: 21172805]
41. Kondo M, et al. SATB1 plays a critical role in establishment of immune tolerance. J. Immunol. 2016; 196:563–572. [PubMed: 26667169]
42. Hopkins AL, Groom CR. The druggable genome. Nat. Rev. Drug Discov. 2002; 1:727–730. [PubMed: 12209152]
43. Louis AG, Yel L, Cao JN, Agrawal S, Gupta S. Common variable immunodeficiency associated with microdeletion of chromosome 1q42.1–q42.3 and inositol 1,4,5-trisphosphate kinase B (ITPKB) deficiency. Clin. Transl. Immunology. 2016; 5:e59. [PubMed: 26900472]
44. Wang Y, et al. Whole-genome association study identifies *STK39* as a hypertension susceptibility gene. Proc. Natl. Acad. Sci. USA. 2009; 106:226–231. [PubMed: 19114657]
45. Durand, EY., Do, CB., Mountain, JL., Macpherson, JM. Ancestry composition: a novel, efficient pipeline for ancestry deconvolution. bioRxiv. 2014. Preprint at http://www.biorxiv.org/content/early/2014/10/18/010512
46. Henn BM, et al. Cryptic distant relatives are common in both isolated and cosmopolitan genetic samples. PLoS One. 2012; 7:e34267. [PubMed: 22509285]
47. Howie B, Fuchsberger C, Stephens M, Marchini J, Abecasis GR. Fast and accurate genotype imputation in genome-wide association studies through prephasing. Nat. Genet. 2012; 44:955–959. [PubMed: 22820512]
48. de Bakker PI, et al. Practical aspects of imputation-driven meta-analysis of genome-wide association studies. Hum. Mol. Genet. 2008; 17:R122–R128. [PubMed: 18852200]
49. Purcell S, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am. J. Hum. Genet. 2007; 81:559–575. [PubMed: 17701901]
50. Bulik-Sullivan BK, et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. Nat. Genet. 2015; 47:291–295. [PubMed: 25642630]
51. Finucane HK, et al. Partitioning heritability by functional annotation using genome-wide association summary statistics. Nat. Genet. 2015; 47:1228–1235. [PubMed: 26414678]
52. Lee SH, Yang J, Goddard ME, Visscher PM, Wray NR. Estimation of pleiotropy between complex diseases using single-nucleotide polymorphism-derived genomic relationships and restricted maximum likelihood. Bioinformatics. 2012; 28:2540–2542. [PubMed: 22843982]

53. Cingolani P, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. Fly (Austin). 2012; 6:80–92. [PubMed: 22728672]

54. dos Santos G, et al. FlyBase: introduction of the *Drosophila melanogaster* Release 6 reference genome assembly and large-scale migration of genome annotations. Nucleic Acids Res. 2015; 43:D690–D697. [PubMed: 25398896]

55. Eppig JT, Blake JA, Bult CJ, Kadin JA, Richardson JE. The Mouse Genome Database (MGD): facilitating mouse as a model for human biology and disease. Nucleic Acids Res. 2015; 43:D726–D736. [PubMed: 25348401]

56. McKusick VA. MENDELIAN Inheritance in Man and its online version, OMIM. Am. J. Hum. Genet. 2007; 80:588–604. [PubMed: 17357067]

57. Szklarczyk D, et al. The STRING database in 2017: quality-controlled protein-protein association networks, made broadly accessible. Nucleic Acids Res. 2017; 45:D362–D368. [PubMed: 27924014]

58. Smith AC, Robinson AJ. MitoMiner v3.1, an update on the mitochondrial proteomics database. Nucleic Acids Res. 2016; 44:D1258–D1261. [PubMed: 26432830]

59. Calvo SE, Clauser KR, Mootha VK. MitoCarta2.0: an updated inventory of mammalian mitochondrial proteins. Nucleic Acids Res. 2016; 44:D1251–D1257. [PubMed: 26450961]

60. Pagliarini DJ, et al. A mitochondrial protein compendium elucidates complex I disease biology. Cell. 2008; 134:112–123. [PubMed: 18614015]

61. Brozzi A, Urbanelli L, Germain PL, Magini A, Emiliani C. hLGDB: a database of human lysosomal genes and their regulation. Database (Oxford). 2013; 2013:bat024. [PubMed: 23584836]

62. Moussay E, et al. The acquisition of resistance to TNFα in breast cancer cells is associated with constitutive activation of autophagy as revealed by a transcriptome analysis using a custom microarray. Autophagy. 2011; 7:760–770. [PubMed: 21490427]

63. Kanehisa M, Goto S. KEGG: Kyoto Encyclopedia of Genes and Genomes. Nucleic Acids Res. 2000; 28:27–30. [PubMed: 10592173]

64. Gene Ontology Consortium. Gene Ontology Consortium: going forward. Nucleic Acids Res. 2015; 43:D1049–D1056. [PubMed: 25428369]

65. Chang CC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience. 2015; 4:7. [PubMed: 25722852]

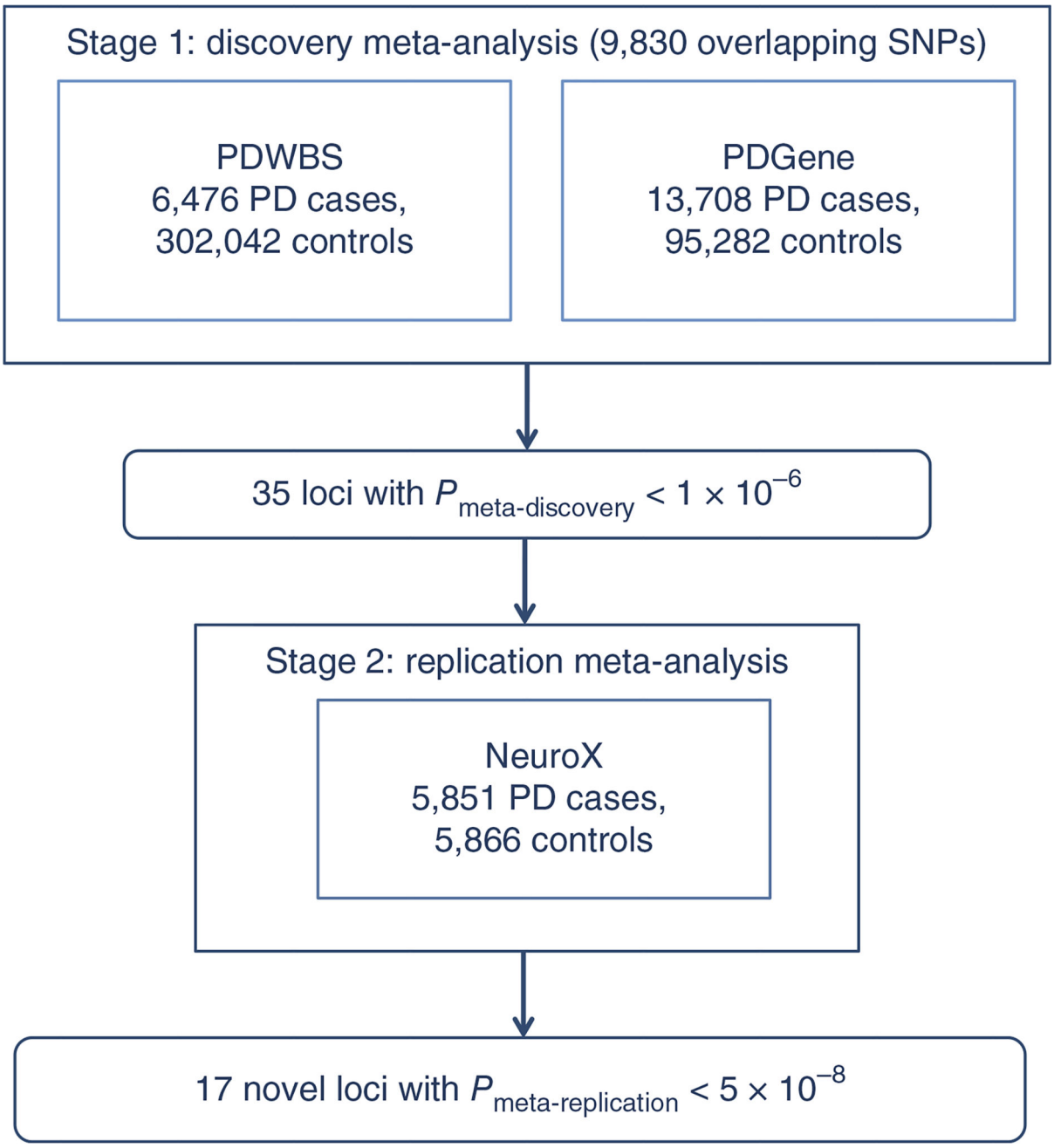

Figure 1.

A flow chart of the two-stage meta-analysis design. In stage 1, we carried out a meta-analysis of 9,830 SNPs between the PDWBS and PDGene studies. Thirty-five loci with $P < 1 \times 10^{-6}$ were carried forward into the replication-phase meta-analysis. In stage 2, we carried out a meta-analysis between the two discovery-phase studies and the NeuroX study for these 35 loci. Of these loci, 16 of the 29 available in NeuroX and 1 locus without replication data were carried forward for downstream analyses (see the main text for further details).

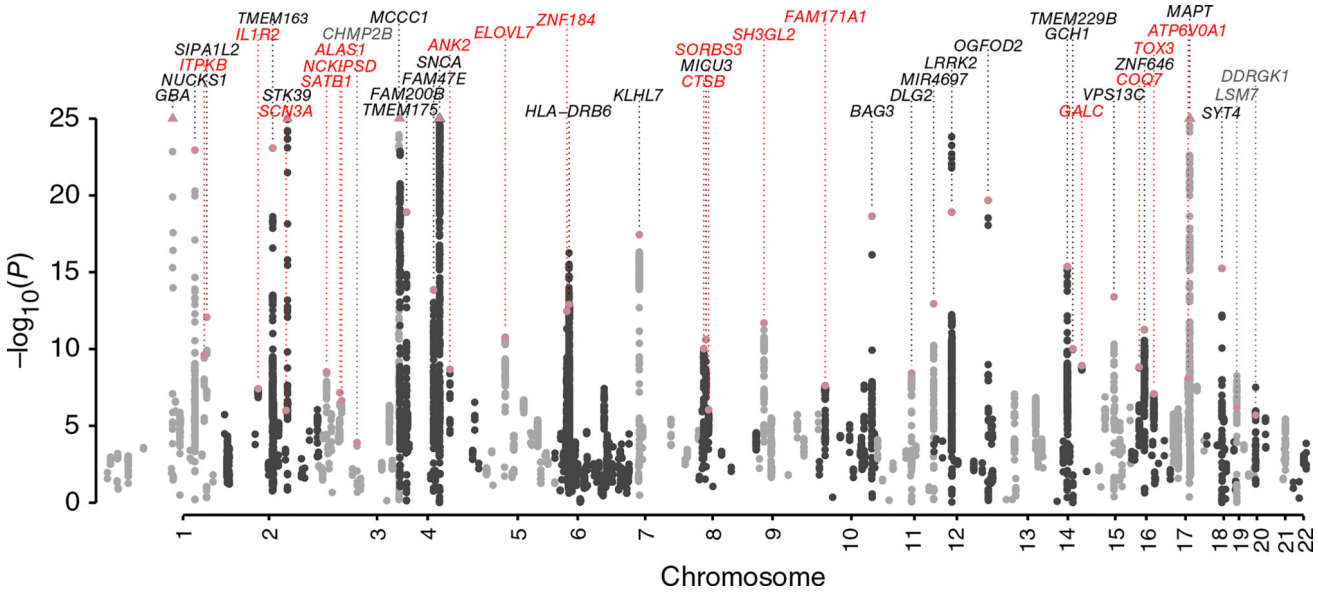

Figure 2.

Results of the Parkinson's disease discovery-phase meta-analysis. The top SNPs in associated regions are indicated by pink symbols. Candidate genes for previously associated loci are labeled in black ($P < 5 \times 10^{-8}$ in the discovery phase) or gray text ($P > 5 \times 10^{-8}$ in the discovery phase); candidate genes for newly identified loci are labeled in red. The *y*-axis shows the two-sided unadjusted $-\log_{10}(P)$ values for association with PD. SNPs with $P < 1 \times 10^{-25}$ are indicated by triangles.

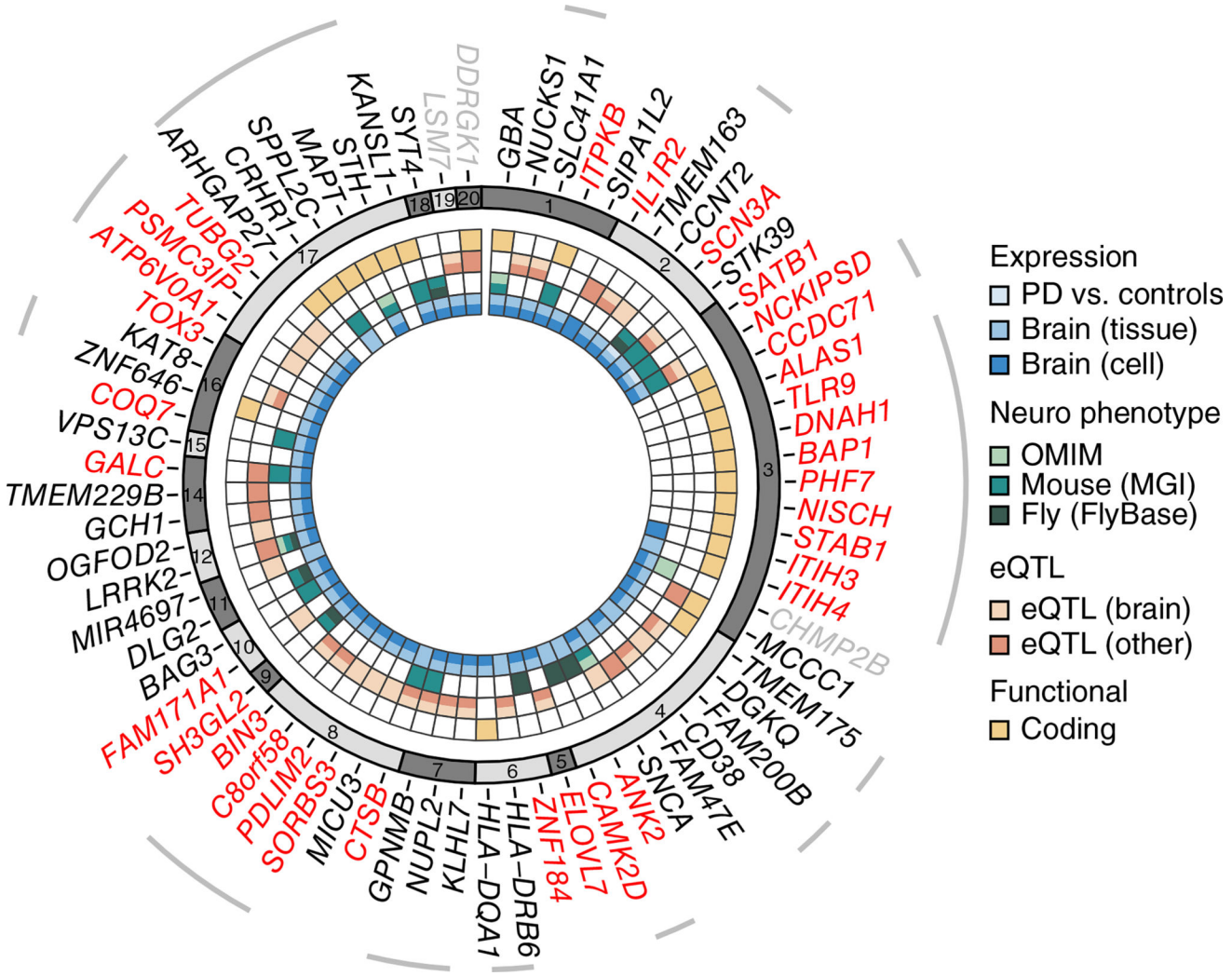

Figure 3.

The candidate genes for regions associated with Parkinson's disease. The most likely candidate gene is annotated for each region that was significantly associated with PD in the final joint analysis. Black or gray text indicates previously reported loci that had *P* values less than or greater than 5×10^{-8} in the discovery phase, respectively. Red text indicates newly identified loci that were significantly associated with PD in the final joint analysis. Gray lines at the outer edge spanning multiple genes indicate candidate genes within a single locus. Chromosome numbers are shown in the gray shaded ring, and support for candidate genes is indicated by color-coding in the inner rings. The innermost ring indicates expression of the gene in brain cell types (in a mouse expression data set) or in human brain regions (in GTEx), or differential expression between PD brains and healthy control brains.

Table 1

Parkinson's disease risk loci previously reported at genome-wide significance levels

CHR:BP[a]	SNP	Candidate gene[b]	Effect allele/ alternate allele	EAF in 1000 Genomes	$EAF_{cases/controls}$[c]	P_{PDGene}[d]	OR_{PDGene}	P_{PDWBS}[e]	OR_{PDWBS}	$P_{discovery}$	$OR_{discovery}$	$OR_{discovery}$ 95% CI
1:155135036	rs35749011	*GBA*	G/A	0.976	0.979/0.988	6.10×10^{-23}	0.57	5.33×10^{-14}	0.59	2.59×10^{-35}	0.58	0.53–0.63
1:205723572	rs823118	*NUCKS1, SLC41A1*	C/T	0.467	0.419/0.443	1.96×10^{-16}	0.89	8.78×10^{-9}	0.90	1.12×10^{-23}	0.89	0.87–0.91
1:232664611	rs10797576	*SIPA1L2*	T/C	0.137	0.145/0.135	1.76×10^{-10}	1.13	7.4×810^{-4}	1.10	8.41×10^{-13}	1.12	1.09–1.15
2:135539967	rs6430538	*TMEM163,CCNT2*	T/C	0.488	0.426/0.450	3.35×10^{-19}	0.88	1.5×410^{-6}	0.91	8.24×10^{-24}	0.89	0.87–0.91
2:169110394	rs1474055	*STK39*	C/T	0.881	0.855/0.874	7.11×10^{-16}	0.82	1.11×10^{-11}	0.83	5.68×10^{-26}	0.83	0.80–0.86
3:87520857[f]	**rs115185635**	*CHMP2B*	**C/G**	**0.036**	**0.040/0.039**	**2.2×10^{-8}**	**1.79**	**0.182**	**1.08**	**1.22×10^{-4}**	**1.21**	**1.10–1.33**
3:182762437	rs12637471	*MCCC1*	A/G	0.219	0.175/0.198	5.38×10^{-22}	0.84	4.27×10^{-10}	0.86	2.11×10^{-30}	0.85	0.82–0.87
4:951947	rs34311866	*TMEM175,DGKQ*	C/T	0.199	0.212/0.184	6.00×10^{-41}	1.26	2.48×10^{-12}	1.18	1.47×10^{-50}	1.23	1.20–1.27
4:15737101	rs11724635	*FAM200B, CD38*	C/A	0.437	0.437/0.452	4.26×10^{-17}	0.89	1.0×410^{-4}	0.93	1.22×10^{-19}	0.90	0.88–0.92
4:77198986	rs6812193[g]	*FAM47E*	T/C	0.398	0.351/0.370	1.85×10^{-11}	0.91	1.24×10^{-4}	0.93	1.43×10^{-14}	0.92	0.90–0.94
4:90626111	rs356182	*SNCA*	G/A	0.375	0.406/0.349	1.85×10^{-82}	1.34	1.44×10^{-42}	1.31	5.21×10^{-123}	1.33	1.30–1.36
6:32666660	rs9275326	*HLA-DRB6, HLA-DQA1*	T/C	0.114	0.099/0.105	5.81×10^{-13}	0.80	1.04×10^{-3}	0.90	1.26×10^{-13}	0.85	0.82–0.89
7:23293746	rs199347	*KLHL7, NUPL2, GPNMB*	G/A	0.368	0.389/0.412	5.62×10^{-14}	0.90	8.66×10^{-6}	0.92	3.51×10^{-18}	0.91	0.89–0.93
8:16697091	rs591323	*MICU3*	A/G	0.293	0.258/0.274	3.17×10^{-8}	0.91	1.61×10^{-4}	0.92	2.38×10^{-11}	0.91	0.89–0.94
10:121536327	rs117896735	*BAG3*	A/G	0.012	0.021/0.015	1.21×10^{-11}	1.77	1.75×10^{-9}	1.57	2.23×10^{-19}	1.65	1.48–1.85
11:83544472	rs3793947	*DLG2*	A/G	0.463	0.431/0.442	2.59×10^{-8}	0.91	8.92×10^{-3}	0.95	3.72×10^{-9}	0.93	0.91–0.95
11:133765367	rs329648	*MIR4697*	T/C	0.327	0.369/0.351	8.05×10^{-12}	1.11	9.16×10^{-4}	1.07	1.11×10^{-13}	1.09	1.07–1.12
12:40614434	rs76904798[h]	*LRRK2*	T/C	0.132	0.152/0.137	4.86×10^{-14}	1.16	4.10×10^{-7}	1.14	1.21×10^{-19}	1.15	1.12–1.19
12:123303586	rs11060180	*OGFOD2*	G/A	0.45	0.423/0.449	3.08×10^{-11}	0.91	4.95×10^{-11}	0.88	2.05×10^{-20}	0.90	0.88–0.92
14:55348869	rs11158026	*GCH1*	T/C	0.307	0.309/0.331	2.88×10^{-10}	0.91	2.65×10^{-7}	0.90	4.30×10^{-16}	0.91	0.89–0.93
14:67984370	rs1555399	*TMEM229B*	T/A	0.544	0.518/0.514	5.70×10^{-16}	1.15	0.453	1.01	9.61×10^{-11}	1.09	1.06–1.11
15:61994134	rs2414739	*VPS13C*	G/A	0.292	0.250/0.266	3.59×10^{-12}	0.90	1.1×10^{-3}	0.93	3.94×10^{-14}	0.91	0.89–0.93
16:31121793	rs14235	*ZNF646, KAT8*	A/G	0.397	0.388/0.378	3.63×10^{-12}	1.10	0.0339	1.04	5.44×10^{-12}	1.08	1.06–1.10
17:43994648	rs17649553	*ARHGAP27, CRHR1, SPPL2C, MAPT, STH, KANSL1*	T/C	0.232	0.187/0.221	6.11×10^{-49}	0.77	9.24×10^{-22}	0.80	1.26×10^{-68}	0.78	0.76–0.80
18:40673380	rs12456492	*SYT4*	G/A	0.332	0.336/0.315	2.15×10^{-11}	1.10	5.13×10^{-6}	1.10	5.56×10^{-16}	1.10	1.07–1.12

CHR:BP[a]	SNP	Candidate gene[b]	Effect allele/ alternate allele	EAF in 1000 Genomes	$EAF_{cases/controls}$[c]	P_{PDGene}[d]	OR_{PDGene}	P_{PDWBS}[e]	OR_{PDWBS}	$P_{discovery}$	$OR_{discovery}$	$OR_{discovery}$ 95% CI
19:2363319[f]	**rs62120679**	***LSM7***	**t/C**	**0.324**	**0.314/0.310**	**2.52×10^{-9}**	**1.14**	**0.24O**	**1.03**	**6.64×10^{-7}**	**1.08**	**1.05–1.11**
20:3168166[f]	**rs8118008**	***DDRGK1***	**A/G**	**0.596**	**0.615/0.609**	**2.32×10^{-8}**	**1.11**	**0.283**	**1.02**	**1.99×10^{-6}**	**1.07**	**1.04–1.09**

Rows in bold text refer to loci that did not pass the genome-wide significance threshold (5×10^{-8}) in the discovery-phase meta-analysis.

[a] Chromosome and physical position according to Hg19.

[b] Details regarding the assignment of candidate genes are provided in the Online Methods.

[c] Effect allele frequency (EAF) measured in PDWBS controls or cases.

[d] *P* value for SNP in the publicly available PDGene data (13,708 cases, 95,282 controls). Publicly available data for the following SNPs include an additional 5,450 cases and 5,798 controls genotyped on NeuroX: rs115185635, rs35749011, rs117896735, rs62120679, rs9275326, rs3793947, rs1555399, rs1474055, and rs8118008.

[e] *P* value for SNP in PDWBS (6,476 cases, 302,042 controls).

[f] The alternate SNP is genome-wide significant (rs12651582; $P = 3.51 \times 10^{-8}$).

[g] The alternate SNP is genome-wide significant (rs76904798; $P = 4.45 \times 10^{-75}$).

Table 2

Seventeen novel regions associated with Parkinson's disease at genome-wide significance levels

CHR:BP[a]	SNP	Candidate gene[b]	Effect allele/ alternate allele	EAF in 1000 Genomes	$P_{discovery}$	$OR_{discovery}$	P_{NeuroX}	OR_{NeuroX}	P_{joint}	OR_{Joint}	OR_{Joint} (95% CI)
1:226916078	rs4653767	*ITPKB*	C/T	0.315	2.40×10^{-10}	0.92	0.017	0.93	1.63×10^{-11}	0.92	0.90–0.94
2:102413116	rs34043159	*IL1R2*	C/T	0.352	3.83×10^{-8}	1.07	1.91×10^{-4}	1.11	5.48×10^{-11}	1.08	1.06–1.10
2:166133632	rs353116	*SCN3A*	T/C	0.385	9.73×10^{-7}	0.94	8.98×10^{-3}	0.93	2.98×10^{-8}	0.94	0.92–0.96
3:18277488	rs4073221	*SATB1*	G/T	0.132	3.02×10^{-9}	1.11	0.583	1.02	1.57×10^{-8}	1.10	1.06–1.13
3:48748989	rs12497850	*NCKIPSD, CDC71*	G/T	0.347	6.80×10^{-8}	0.93	0.040	0.94	9.16×10^{-9}	0.93	0.91–0.96
3:52816840	rs143918452	*ALAS1, TLR9, DNAH1, BAP1, PHF7, NISCH, STAB1, ITIH3, ITIH4*	G/A	0.996	2.25×10^{-7}	0.68	0.095	0.73	3.20×10^{-8}	0.68	0.60–0.78
4:114360372	rs78738012	*ANK2, CAMK2D*	C/T	0.106	2.11×10^{-9}	1.14	7.5×10^{-3}	1.12	4.78×10^{-11}	1.13	1.09–1.17
5:60273923	rs2694528	*ELOVL7*	C/A	0.115	1.69×10^{-11}	1.15	6.25×10^{-5}	1.19	4.84×10^{-15}	1.15	1.11–1.20
6:27681215	rs9468199	*ZNF184*	A/G	0.172	3.44×10^{-13}	1.12	0.302	1.04	1.46×10^{-12}	1.11	1.08–1.14
8:11707174	rs2740594[c]	*CTSB*	A/G	0.753	9.54×10^{-11}	1.10	7.95×10^{-3}	1.08	5.91×10^{-12}	1.09	1.07–1.12
8:22525980	rs2280104	*SORBS3, PDLIM2, C8orf58, BIN3*	T/C	0.367	9.06×10^{-7}	1.06	7.87×10^{-3}	1.08	2.53×10^{-8}	1.07	1.04–1.09
9:17579690	rs13294100	*SH3GL2*	T/G	0.371	1.99×10^{-12}	0.91	0.037	0.94	4.84×10^{-13}	0.92	0.89–0.94
10:15569598	rs10906923	*FAM171A1*	C/A	0.306	2.37×10^{-8}	0.93	0.133	0.96	1.35×10^{-8}	0.93	0.91–0.96
14:88472612	rs8005172	*GALC*	T/C	0.424	1.20×10^{-9}	1.08	0.022	1.06	8.77×10^{-11}	1.08	1.05–1.10
16:19279464	rs11343	*COQ7*	T/G	0.454	1.46×10^{-9}	1.07	0.019	1.06	9.13×10^{-11}	1.07	1.05–1.10
16:52599188	rs4784227	*TOX3*	T/C	0.265	8.29×10^{-8}	1.08	1.47×10^{-4}	1.12	9.75×10^{-11}	1.09	1.06–1.12
17:40698158	rs601999	*ATP6V0A1, PSMC3IP, TUBG2*	C/T	0.699	8.03×10^{-9}	0.93	NA	NA	NA	NA	NA

Summary statistics are shown for the discovery cohort (PDWBS and PDGene), NeuroX (5,851 cases, 5,866 controls), and the joint meta-analysis of the discovery and NeuroX data. Additional summary statistics for NeuroX and the joint meta-analysis are available in Supplementary Table 5. EAF, effect allele frequency.

[a] Chromosome and physical position according to Hg19.

[b] Details regarding the assignment of candidate genes are provided in the Online Methods.

[c] NeuroX and joint statistics are shown for proxy SNP rs1293298.