HHS Public Access

Author manuscript

Ann Neurol. Author manuscript; available in PMC 2021 November 18.

Published in final edited form as:

Ann Neurol. 2021 July ; 90(1): 22–34. doi:10.1002/ana.26051.

Common X-Chromosome Variants Are Associated with Parkinson Disease Risk

Yann Le Guen, PhD[1,†], **Valerio Napolioni, PhD**[2,†], **Michael E. Belloy, PhD**[1], **Eric Yu, BSc**[3,4], **Lynne Krohn, MSc**[3,4], **Jennifer A. Ruskey, MSc**[4,5], **Ziv Gan-Or, MD, PhD**[3,4,5], **Gabriel Kennedy, BSc**[1], **Sarah J. Eger, BA**[1], **Michael D. Greicius, MD, MPH**[1]

[1]Department of Neurology and Neurological Sciences, Stanford University, Stanford, CA, USA

[2]School of Biosciences and Veterinary Medicine, University of Camerino, Camerino, Italy

[3]Department of Human Genetics, McGill University, Montreal, QC, Canada

[4]Montreal Neurological Institute and Hospital, McGill University, Montreal, QC, Canada

[5]Department of Neurology and Neurosurgery, McGill University, Montreal, QC, Canada

Abstract

Objective: The objective of this study was to identify genetic variants on the X-chromosome associated with Parkinson disease (PD) risk.

Methods: We performed an X-chromosome–wide association study (XWAS) of PD risk by meta-analyzing results from sex-stratified analyses. To avoid spurious associations, we designed a specific harmonization pipeline for the X-chromosome and focused on a European ancestry sample. We included 11,142 cases, 280,164 controls, and 5,379 proxy cases, based on parental history of PD. Additionally, we tested the association of significant variants with (1) PD risk in an independent replication with 1,561 cases and 2,465 controls and (2) putamen volume in 33,360 individuals from the UK Biobank.

Results: In the discovery meta-analysis, we identified rs7066890 (odds ratio [OR] = 1.10, 95% confidence interval [CI] = 1.06–1.14, $p = 2.2 \times 10^{-9}$), intron of *GPM6B,* and rs28602900 (OR = 1.10, 95% CI = 1.07–1.14, $p = 1.6 \times 10^{-8}$) in a high gene density region including *RPL10, ATP6A1, FAM50A*, and *PLXNA3*. The rs28602900 association with PD was replicated (OR = 1.16, 95% CI = 1.03–1.30, $p = 0.016$) and shown to colocalize with a significant expression quantitative locus (eQTL) regulating *RPL10* expression in the putamen and other brain tissues in the Genotype-Tissue Expression Project. Additionally, the rs28602900 locus was found to be

Address correspondence to Dr Le Guen, Department of Neurology and Neurological Sciences–Greicius Lab, Stanford University, 290 Jane Stanford Way, E265, CA 94305-5090. yleguen@stanford.edu.

[†]Y.L.G. and V.N. contributed equally to this work.

Author Contributions

Y.L.G., V.N., and M.D.G. contributed to the conception and design of the study; Y.L.G., V.N., M.E.B., E.Y., L.K., J.A.R., Z.G.O., S.J.E., G.K, M.D.G., and IPDGC contributed to the acquisition and analysis of data; Y.L.G., V.N., and M.D.G. contributed to drafting the text and preparing the figures. Y.L.G and V.N. contributed equally to this work. A full list of IPDGC members is listed in Table S8.

Potential Conflicts of Interest

Nothing to report.

Additional supporting information can be found in the online version of this article.

associated with reduced brain putamen volume. No results reached genome-wide significance in the sex-stratified analyses.

Interpretation: We report the first XWAS of PD and identify 2 genome-wide significant loci. The rs28602900 association was replicated in an independent PD dataset and showed concordant effects in its association with putamen volume. Critically, rs26802900 is a significant eQTL of *RPL10*. These results support a role for ribosomal proteins in PD pathogenesis and show that the X-chromosome contributes to PD genetic risk.

Parkinson disease (PD), a neurodegenerative disorder characterized by a broad range of motor and non-motor symptoms, is thought to be caused by a combination of aging, genetics, and environmental factors. PD affects the nervous system at multiple levels from the enteric nervous system to the cortex, but prominent motor symptoms associated with PD are linked to the substantia nigra and striatum. PD heritability has been estimated at 34% in a twin study[1] and 16 to 36% in genome-wide association studies (GWASs).[2] The latter estimates were derived from autosomal variants and did not include the genetic contribution of sex chromosomes. Similarly, the latest PD GWAS meta-analysis, which identified 90 common genetic variants associated with PD risk, focused solely on autosomes.[2] The effect of common sex chromosome variants on PD risk remains unexplored, despite the relative risk for PD being 1.5 times greater in males than in females.[3–6] Various mechanisms have been proposed to account for this sex difference in PD risk, including sex hormone levels and other environmental and genetic factors influenced by biological sex.[7,8] To date, a few studies have identified variants or loci on the X-chromosome linked to PD, notably, a linkage analysis in PD multiplex families that identified the *PARK12* locus.[9] However, this locus is quite large and has not yet been linked to a gene.[10] Whole-exome sequencing studies in pedigree and single case reports have linked loss of function mutations on *RAB39B* with Waisman syndrome, an X-linked, early onset, α-synucleinopathy with Parkinsonism and intellectual disability.[11–14]

Although the X-chromosome is 155Mb and accounts for 5% of the human genome, it remains excluded from the vast majority of GWASs, because it requires different quality control than autosomal chromosomes[15] due to the way it is inherited. Females inherit one X-chromosome from each parent, while males only inherit a maternal copy, and outside of pseudoautosomal regions, male X-chromosomes do not undergo any recombination during meiosis. This causes the X-chromosome to have reduced genetic diversity. Notably, the X-chromosome structure is more sensitive to historical events such as population bottleneck and sex-biased demographic events, and has a different mutation rate from autosomes.[16] Additionally, until recently, genotyping arrays were poorly designed for the X-chromosome.[15]

Our work addresses this gap by conducting the first PD X-chromosome–wide association study (XWAS). Specifically, we performed meta-analyses including all publicly available PD cohorts with X-chromosome data, as well as the UK Biobank cohort, in which we used both diagnosed PD individuals and proxy cases based on parental history of PD. The use of this PD proxy phenotype was recently shown to provide increased power for the discovery of PD genetic risk factors.[2] We first meta-analyzed males and females separately to identify

potential sex-specific effects. Then, to increase statistical power, we meta-analyzed across the sexes to identify variants influencing PD risk in a sex-independent fashion. Variants that reached genome-wide significance were tested for replication in an independent PD meta-analysis and tested for association with putamen volume in the UK Biobank, and functionally annotated in gene expression datasets through colocalization analyses, leading to the identification of a putatively causal gene for PD that may provide novel targets for future therapeutic development.

Materials and Methods

Samples

For the discovery PD meta-analysis, we used 2 types of samples: PD cohorts published in previous studies and the UK Biobank, using both directly diagnosed PD and a proxy phenotype based on parental history of PD. The PD cohorts include the International PD Genomics Consortium (IPDGC) NeuroX dataset,[17] the National Institute of Neurological Disorders and Stroke (NINDS) PD dataset,[18] the Autopsy-Confirmed PD GWAS Consortium (APDGC) dataset genotyped by the Center for Inherited Disease Research, the NeuroGenetics Research Consortium (NGRC) dataset,[19] the UK PD Consortium, and the Wellcome Trust Case Control Consortium 2,[20] referred to here as EBI MERGE. The whole-genome sequencing (WGS) data were obtained from the Accelerating Medicines Partnership–Parkinson's Disease (AMP PD), composed of the following cohorts: New Discovery of Biomarkers (BioFIND), the Harvard Biomarker Study, the Parkinson's Progression Markers Initiative,[21] and the Parkinson's Disease Biomarkers Program. Table 1 shows a detailed description of the cohorts including genotyping platform, number of X-chromosome single nucleotide polymorphisms (SNPs), and demographics.

In the replication, we meta-analyzed 2 PD datasets (Table 2): the first composed of the Parkinson's Disease Cognitive Genetics Consortium (PDCGC)[22] and healthy individuals from the Adult Change in Thought (ACT) longitudinal cohort,[23] referred to here as PDCGC & ACT; and the second from McGill University (McGill), including samples from the Quebec Parkinson Network.[24]

The putamen volume analysis consisted of individuals with magnetic resonance imaging (MRI) scans that were successfully processed by the UK Biobank FSL (FMRIB Software Library, where FMRIB stands for "Oxford Centre for Functional Magnetic Resonance Imaging of the Brain) pipeline[25] as of January 2020. The putamen volume used in our analysis corresponds to the average of the left and right hemispheres. UK Biobank details regarding the MRI scan parameters, and image quality control are described elsewhere.[26,27]

The current study protocol was granted an exemption by the Stanford University Institutional Review Board because the analyses were carried out on deidentified, off-the-shelf data; therefore, further informed consent was not required.

Quality Control, Ancestry Determination, and Imputation

To perform the first PD XWAS, we introduced a quality control and harmonization pipeline of X-chromosome SNP array data from homogenous ancestry samples prior

to X imputation. Autosomal variants were extracted from the SNP array genetic data and processed through our harmonization pipeline in several stages (Table S1). First, we removed multiallelic SNPs, SNPs located on structural variants, and duplicated or monomorphic SNPs. The list of multiallelic SNPs or SNPs located on common structural variants was created using Tri-Typer[28] and gnomAD.[29] Additionally, we removed variants located on potential probe polymorphisms identified from gnomAD. These are defined as SNPs for which the probe may have variable affinity due to the presence of other SNP(s) within 20bp and with minor allele frequency (MAF) > 1%. SNPs were checked for consistency with the Haplotype Reference Consortium (HRC) panel[30] using a checking script developed by the McCarthy group (https://www.well.ox.ac.uk/~wrayner/tools/). This script enabled us to flip SNPs reported on the incorrect strand and excluded palindromic SNPs and SNPs with >10% MAF difference from the reference panel.

Using these quality control SNPs as inputs, we further excluded SNPs with a genotyping rate < 95% and Hardy–Weinberg equilibrium (HWE) $p < 10^{-5}$ in controls and also removed individuals with >5% genotype missingness. Next, the individual ancestry was determined with SNP weights v.2.1[31] using reference populations from the 1000 Genomes Project.[32] By applying an ancestry percentage cutoff of >75%, the samples were stratified into the 5 superpopulations, South-Asians, East-Asians, Americans, Africans, and Europeans. Because most of the samples belonged to the European population, we also determined their percent ancestry to 3 major ethnicities, Northwestern, Southeastern, and Ashkenazi Jewish, using reference populations available from SNP weights v.2.1. European subjects were stratified into the abovementioned ethnicities when their ancestry percentage was >50%. Further analyses were focused on the Northwestern European (NWE) subsample, which represents the vast majority of subjects and provided a highly homogenous sample (Table S2). Two main reasons explain the more extreme population structure on the X-chromosome compared to autosomes: (1) the X-chromosome has a smaller effective population size, and thus the rate of genetic drift of X-linked loci is amplified; and (2) local adaptation will lead to a higher level of differentiation between geographically isolated populations.[33] As such, to better control our population structure, we decided to be stringent and restrict our analyses to NWE subjects.

Additionally, we merged all genotype quality control autosomes and performed an identity-by-descent on the whole NWE sample to identify duplicates and first-degree relatives. If the diagnosis was incongruent between cohorts for a duplicate subject, we chose the cohort with a PD diagnosis. Otherwise, we kept the subject's data in the cohort with the following order of priority based on sequencing coverage of the X-chromosome: AMP AD WGS, APDGC, EBI MERGE, NINDS, NGRC, and IPDGC. When removing first-degree relatives, we prioritized the inclusion of a male relative over a female one, given that one copy of X-chromosome will be randomly inactivated in females, making the X-chromosome genotype–phenotype association noisier in females. Additionally, we kept cases over controls, used the same order of priority for cohort selection when relatives had the same diagnosis, and chose the older of the two relatives if they were controls in the same cohort.

The X-chromosome variants underwent a similar harmonization pipeline as the autosomes (Tables S3 and S4). We excluded multiallelic SNPs, SNPs within structural variations,

and potential probe polymorphism SNPs. Additionally, our analysis excluded the pseudoautosomal regions of the X-chromosome and used only the NWE subsample as derived above. Several steps were performed to avoid spurious findings: (1) variants with <95% genotyping rate were excluded; (2) individuals with >5% genotype missingness were excluded; (3) reported sex was checked using PLINK1.9[34] *-check-sex* flag, with 0.2 maximum value for females and 0.94 minimum value for men, and all individuals with a discordant sex label were excluded; (4) heterozygous SNPs in males were set as missing in males; (5) SNPs with differential missingness between PD cases and controls were removed ($p < 10^{-5}$); (6) HWE was tested in female controls, and SNPs with $p < 10^{-5}$ were removed; (7) any monomorphic SNPs that remained were removed; (8) differential missingness and differential MAF between males and females were both tested, and SNPs with $p < 10^{-5}$, for either one of the tests, were excluded; and finally, as for the autosomes, (9) remaining SNPs were checked for consistency with the HRC panel, flipping palindromic SNPs and excluding SNPs with >10% MAF difference from the reference panel.

The EBI MERGE dataset, which includes 2 SNP arrays, had these quality control steps performed separately on each array, and the remaining SNPs were merged and reprocessed through the whole harmonization pipeline. Similarly, PDCGC & ACT autosomes were first harmonized separately, then merged and reharmonized together before ancestry determination and PCA computation. By contrast, the X-chromosome data were imputed separately and merged after imputation, due to a low X coverage of the NeuroX array used in PDCGC that resulted in only 111 overlapping ACT-PDCGC variants after harmonization.

The remaining X-chromosome quality-controlled SNPs were imputed on the Michigan Imputation Server (v1.2.4),[35] which uses minimac 4 for imputation and improved X-chromosome support. The following parameters were selected: reference panel HRC r1.1 2016 (hg19) with the European subpanel, phasing Eagle v2.4, r-square imputation score cutoff of 0.3. Before the XWAS with PD status, SNPs with MAF < 0.01 or genotyping rate < 95% were removed in WGS and imputed data.

Genetic principal components (PCs) were computed within the NWE subsample for each cohort to account for population stratification in the downstream analysis.[36] PCs were inspected for each cohort and indicated a homogenous sample.

The UK Biobank genotyping, imputation, and quality control data are described in detail elsewhere.[37] We considered the 33,360 British individuals with MRI scan, whose genetic data successfully passed the UK Biobank quality control and were not diagnosed with PD, nor had a parent with PD.

Study-Level Analyses and Meta-Analysis

To identify sex-specific effects, we first performed sex-stratified meta-analyses separating males and females. We then performed a combined meta-analysis across all sex-stratified cohorts, to identify sex-independent effects with increased statistical power.

The male meta-analysis included 5,745 cases and 4,875 controls of NWE ancestry from PD cohorts, 706 PD cases, and 128,275 controls from the UK Biobank consisting of unrelated

British males, and 3,551 proxy male cases and 143,316 controls using the father's PD status in unrelated British females. The rationale is that the female dosage, in this case, corresponds to the probability of the father carrying the variant and testing it against his PD status. Put differently, if a female has 2 copies of the effect allele, then her father is a carrier; if a female has one copy, then there is a 50% chance that her father is a carrier; if a female has no copies, then her father is a noncarrier. The female meta-analysis included 3,104 cases and 4,925 controls of NWE ancestry from PD cohorts, 396 PD cases, and 146,867 controls from the UK Biobank consisting of unrelated British females, and 1,828 proxy female cases and 126,447 controls using the mother's PD status in unrelated British males. In this scenario, the mother must carry at least one copy for a male carrier and zero or no more than one copy for a male noncarrier. We excluded directly diagnosed PD subjects from the proxy phenotype analysis.

The associations were estimated with PLINKv2.0[34] using the *-glm* flag, which performs a standard logistic regression for case/control phenotype. For each cohort, we covaried by the age provided in each study and the 10 PCs computed within the NWE sample. In each separate study, cases and controls were roughly age matched, with cases slightly older than controls (Table S5). This small age difference per cohort resulted in a small positive effect of age on PD risk that was adjusted for,[38] except in the EBI MERGE cohort, where age was unavailable. The age provided varied across studies between age at onset, age at clinical diagnosis or age at death for cases, and age at last examination or age at death for controls. To address this heterogeneity, we ran the analysis with and without age as a covariate and the loci passing the genome-wide significance threshold did not change. For the UK Biobank, the covariates used were the first 10 PCs provided by the UK Biobank and the age at last visit of the participant. Finally, we meta-analyzed all XWAS summary statistics using fixed effect meta-analysis as implemented in GWAMA.[39] Additionally, GWAMA testing was done for sex-heterogenous effects by combining male and female effects into a chi-squared distribution with 1 degree of freedom (assuming the same allelic effect in males and females) and testing the null hypothesis that the association is the same in males and females. The genome-wide significance threshold was set at $p < 5 \times 10^{-8}$.

We queried the expression quantitative loci (eQTLs) among our XWAS loci in brain tissues of the Genotype-Tissue Expression Project (GTEx) v8.[40] GTEx (https://gtexportal.org/) includes bulk RNASeq data from 12 brain tissues types: amygdala (n = 129), anterior cingulate cortex (n = 147), caudate (n = 194), cerebellar hemisphere (n = 175), cerebellum (n = 209), cortex (n = 205), frontal cortex (n = 175), hippocampus (n = 165), hypothalamus (n = 170), nucleus accumbens (n = 202), putamen (n = 170), and substantia nigra (n = 114). Additionally, we queried the Braineac database[41] (http://braineac.org/), which includes n = 134 individuals with up to 12 different brain regions sequenced with RNA exon arrays. Determination of the eQTLs significance and false discovery rate (FDR) correction[42] were performed by GTEx[40] and Braineac.[41] Colocalization was performed using the R package *coloc*,[43] and we report the posterior probability of colocalization (PP4) between PD and QTL associations. Colocalization between X-chromosome–wide association and eQTL results were inspected with *locuscompareR*[44] for all FDR significant results from GTEx.

We tested the association of the significant loci with putamen volume using a Bayesian linear mixed model as implemented in LMM-BOLT.[45] In this second validation analysis, males and females were analyzed together coding males as 0/2 and females as 0/1/2. We adjusted the analysis by brain volume, age at MRI scan, genotyping array, sex, and 10 PCs accounting for population stratification.

Results

PD X-Chromosome Discovery Meta-Analyses

We meta-analyzed all XWAS summary results in males, females, and both sexes combined (Fig 1). In visually inspecting each cohort's QQ plots and genomic control lambdas, we observed no inflation ($\boldsymbol{\lambda}$ = 0.983–1.042). As a sensitivity analysis, we also conducted the analysis without accounting for any covariates. The top associations remained unchanged, and lambdas remained within the normal range ($\boldsymbol{\lambda}$ = 0.971–1.069).

In the combined sex meta-analysis ($\boldsymbol{\lambda}$ = 1.041), 2 loci reached genome-wide significance with top SNPs: rs7066890 near *GPM6B* (odds ratio [OR] = 1.10, 95% confidence interval [CI] = 1.07–1.14, MAF = 0.16, $p = 2.4 \times 10^{-9}$) and rs28602900 near *RPL10* (OR = 1.10, 95% CI = 1.06–1.14, MAF = 0.12, $p = 1.8 \times 10^{-8}$; Fig 2, Table S6). For each SNP, the minor allele was associated with increased risk, and in all analyses described below the rsID will refer to the minor allele. These 2 SNPs had a concordant direction of effect across cohorts (Fig 3). The first SNP is an intron of *GPM6B*, and the second SNP is located in a region with a high gene density (39 genes within ±400kb).

For the sex-stratified analyses ($\boldsymbol{\lambda}_{male}$ = 1.004, $\boldsymbol{\lambda}_{female}$ = 0.980) there were no genome-wide significant results. The 2 genome-wide significant loci from the combined sex meta-analysis had a nondifferentiated effect across sex (Table 3).

Independent PD Replication Meta-Analysis

In 2 independent datasets, we tested the association with PD of the 2 lead SNPs from the genome-wide significant loci. The rs28602900 association with PD significantly replicated (OR = 1.16, 95% CI = 1.03–1.30, $p = 0.016$), whereas rs7066890 was imputed in only 1 of the 2 datasets and showed a discordant direction of effect (OR = 0.96, 95% CI = 0.83–1.11, $p = 0.54$). Thus, the rs7066890 association did not replicate but remained significant genome-wide after meta-analysis with the replication ($p = 1.2 \times 10^{-8}$; see Table 3).

Functional Annotation via eQTL Analysis

The identified loci lie in regions with multiple genes, so we sought to identify the potentially causal genes through eQTL analysis. To this aim, we queried the genome-wide significant SNPs in the GTEx v8[40] and Braineac[41] databases, which both contain gene expression data from brain tissues and include eQTL analyses of X-chromosome SNPs.

rs28602900 was significantly associated with decreased *RPL10* expression in 11 of 12 queried brain tissues, including putamen and substantia nigra. rs28602900 was also significantly associated with decreased *PLXNA3* expression in 6 of 12 brain tissues, including putamen but not substantia nigra. For 3 other genes, rs28602900 was a significant

eQTL for 4 or fewer brain tissues, none of which included putamen or substantia nigra. In Braineac, rs28602900 was significantly associated with decreased *RPL10* and *PLXNA3* average expression across brain tissues ($p < 0.05$, FDR-corrected). No FDR-significant eQTL associations were observed in brain tissues for rs7066890 (*GPM6B* locus) in GTEx or Braineac.

The *RPL10* eQTL colocalized across 11 brain tissues, with the locus associated with PD risk (PP4 between 0.90 and 0.97; Table S7). The colocalization is illustrated for the putamen (PP4 = 0.97; Fig 4A). On the other hand, the *PLXNA3* eQTL association is driven by a nearby stronger *PLXNA3* eQTL, which is in low linkage disequilibrium with the locus associated with PD risk. This is illustrated for the putamen (PP4 = 0; see Fig 4B) and occurred similarly in the 5 other brain tissues with a significant eQTL for *PLXNA3* (PP4 ≈ 0).

Association with Putamen Volume in the UK Biobank

To further assess and validate the role of the 2 genome-wide significant loci, we tested their association in UK Biobank with the putamen volume, an established PD biomarker.[46] In a sample composed of 32,896 British individuals, excluding PD cases and PD proxy cases, the lead SNP in the *RPL10* locus, rs28602900, was significantly associated with decreased putamen volume ($\beta = -11.01$, $\sigma = 4.35$, $p = 0.011$). The lead SNP in the *GPM6B* locus, rs7066890, was not significantly associated with putamen volume.

Discussion

Our study demonstrates a genome-wide significant association of PD with 2 X-chromosome loci and paves the way for further XWASs to decipher the role of common X-chromosome variants in neurodegenerative diseases. This analysis expands our understanding of the genes involved in PD. These new loci will help refine polygenic risk scores to determine an individual's genetic susceptibility to PD. We have also leveraged eQTL data to focus attention on a gene new to PD that may have important pathogenic implications.

The *RPL10* locus has the strongest support across our analyses and was replicated in an independent PD meta-analysis. rs28602900, the top SNP in the locus, was significantly associated with increased risk in our combined sex meta-analysis and with reduced putamen volume in the UK Biobank analysis. Although this locus is in a high gene density region, our colocalization analysis showed that the signal associated with PD risk colocalizes with the *RPL10* eQTL in all GTEx brain tissues. In both GTEx and Braineac, *RPL10* expression was lower in carriers of the rs28602900 minor allele. *RPL10*, ribosomal protein L10, has been associated with intellectual disability,[47] cerebellar hypoplasia,[48] microcephaly,[49] and autism spectrum disorder.[50,51] More generally, in a gene and protein network analysis of PD, Alzheimer disease, and amyotrophic lateral sclerosis, Monti et al emphasized that 6 of 25 PD-specific proteins were ribosomal.[52] Additionally, ribosomal proteins interact with *LRRK2* and elevate kinase activity.[53] Specifically, the ribosomal protein, *RPS15*, was shown to be a key pathogenic *LRRK2* substrate mediating dopamine neuron degeneration.[53] The reduced expression of several ribosomal proteins in PD brain tissues is associated with alpha-synuclein aggregation.[54]

The second locus, near *GPM6B*, did not replicate in an independent PD meta-analysis and did not have any eQTL signals in brain tissue that survived FDR correction. The top SNP at the locus is intronic in *GPM6B*, whose associated protein is highly expressed in brain. The locus is also situated near a number of other brain-expressed proteins such as *OFD1*, *TRAPPC2*, and *GEMIN8*, but for the time being, the gene associated with PD risk at this locus remains uncertain.

Interestingly, a linkage analysis in multiplex PD families identified a large PD X-linked region, *PARK12*, spanning Xq21-q25.[9] Our analysis highlights a well-defined signal within this region, with a lead variant missense of *RTL9* (OR = 1.15, 95% CI = 1.09–1.21, $p = 9.0 \times 10^{-9}$; see Table S6). Linkage studies are well powered to discover genetic loci associated with disease and reduce the risk of false positives due to population structure, but this approach often cannot fine-map the causal gene. Our complementary XWAS approach allowed us to refine this original signal and emphasized the most likely polymorphism increasing disease risk at this locus.

The main limitation of our study is that although we have 3 suggestive loci and 2 loci meeting the genome-wide significance threshold—including one replicating in an independent PD meta-analysis and strongly supported by eQTL and imaging analyses—we do not have access to additional replication datasets to validate more of these loci. Two large datasets used in the recent PD GWAS[2] could not, unfortunately, be used here because access to the data is restricted by the investigative teams. When additional datasets become available, future studies will allow further validation and extension of the findings presented here. In addition, the inclusion of UK Biobank proxy phenotypes that rely on participant reports of family history likely injects some diagnostic uncertainty into those data, but we believe this is outweighed by the gain in statistical power from being able to include this large cohort in our analyses.

In conclusion, our work provides compelling evidence for 2 novel loci on the X-chromosome that increase risk for PD and that can be used to refine polygenic risk scores of PD. Most critically, these results focus attention on a new-to-PD gene, *RPL10*, that should further increase enthusiasm for the role of ribosomal proteins in PD pathogenesis.[52,53]

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

Funding for this study was provided by the Iqbal Farrukh & Asad Jamal Fund, and the NIH-National Institute on Aging (NIA) (AG060747 and AG047366). Z.G.-O. is supported by the Fonds de recherche du Québec–Santé Chercheurs-boursiers award, in collaboration with Parkinson Quebec, and by the Young Investigator Award by Parkinson Canada.

We thank all of the subjects who donated their time and biological samples to be a part of this study.

Acknowledgments for the Use of PD Datasets

Some of the data used in the preparation of this article were obtained from the IPDGC, NeuroX dataset provided through dbGaP accession phs000918.v1.p1. Funding sources for IPDGC included the NIH NINDS

and National Institute on Aging, Bethesda, Maryland, and US Army Medical Research Acquisition Activity, Fort Detrick, Maryland/Department of Defense, Washington, District of Columbia. PD data also included the NINDS Parkinson's Disease dataset provided through dbGaP accession phs001172.v1.p2. PD data also included the APDGC dataset provided through dbGaP accession phs000394.v1.p1. PD data also included the Genome-Wide Association Study of Parkinson Disease: Genes and Environment dataset provided through dbGaP accession phs000196.v3.p1 and the genotyping center Johns Hopkins University Center for Inherited Disease Research. PD data also included the PDCGC Stage I, NeuroX Dataset (phs001664.v1.p1), for which we acknowledge the Pacific Northwest Udall Center. Healthy individuals were included from the ACT study, which is a longitudinal prospective cohort study that began in 1994. PD data also included the dataset generated by the Wellcome Trust Case–Control Consortium 2 from UK patients with PD and UK control individuals from the 1958 Birth Cohort and National Blood Service (available through the European Bioinformatics Institute, accession EGAD00000000022 and EGAD00000000057). This study makes use of data generated by the Wellcome Trust Case–Control Consortium. A full list of the investigators who contributed to the generation of the data is available from https://www.wtccc.org.uk/. Some of the data used in the preparation of this article were obtained from the AMP PD Knowledge Platform. For up-to-date information on the study, visit https://www.amp-pd.org. AMP PD—a public–private partnership—is managed by the Foundation for the National Institutes of Health and funded by Celgene, GSK, the Michael J. Fox Foundation for Parkinson's Research, the NINDS, Pfizer, Sanofi, and Verily. The AMP PD data included clinical data and bio-samples obtained from the Fox Investigation for New Discovery of Biomarkers (BioFIND), the Harvard Biomarker Study (HBS), the Parkinson's Progression Markers Initiative (PPMI), and the Parkinson's Disease Biomarker Program (PDBP). BioFIND is sponsored by the Michael J. Fox Foundation for Parkinson's Research with support from the NINDS. The BioFIND Investigators have not participated in reviewing the data analysis or content of the article. For up-to-date information on the study, visit https://biofind.loni.usc.edu/. The Harvard NeuroDiscovery Biomarker Study (HBS) is a collaboration of HBS investigators funded through philanthropy and NIH and non-NIH funding sources. A full list of HBS investigator can found at https://neurodiscovery.harvard.edu/. PPMI is a public–private partnership. For up-to-date information on the study, visit http://www.ppmi-info.org/. The PDBP consortium is supported by the NINDS. A full list of PDBP investigators can be found at https://pdbp.ninds.nih.gov/policy. The HBS, PPMI, and PDBP investigators have not participated in reviewing the data analysis or content of the article. The McGill cohort was financially supported by grants from the Michael J. Fox Foundation, the Canadian Consortium on Neurodegeneration in Aging, the Canada First Research Excellence Fund, awarded to McGill University for the Healthy Brains for Healthy Lives initiative, and Parkinson Canada. Access to some of the participants for this research has been made possible thanks to the Quebec Parkinson's Network (http://rpq-qpn.ca/en/).

Acknowledgments for the Use of the UK Biobank Data

This research has been conducted using the UK Biobank Resource under application number 45420.

References

1. Wirdefeldt K, Gatz M, Reynolds CA, et al. Heritability of Parkinson disease in Swedish twins: a longitudinal study. Neurobiol Aging 2011;32:1923.e1–1923.e8.
2. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol 2019;18:1091–1102. [PubMed: 31701892]
3. Van Den Eeden SK, Tanner CM, Bernstein AL, et al. Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. Am J Epidemiol 2003;157:1015–1022. [PubMed: 12777365]
4. Wooten GF, Currie LJ, Bovbjerg VE, et al. Are men at greater risk for Parkinson's disease than women? J Neurol Neurosurg Psychiatry 2004;75:637–639. [PubMed: 15026515]
5. Taylor KSM, Cook JA, Counsell CE. Heterogeneity in male to female risk for Parkinson's disease. J Neurol Neurosurg Psychiatry 2007;78: 905–906. [PubMed: 17635983]
6. Miller IN, Cronin-Golomb A. Gender differences in Parkinson's disease: clinical characteristics and cognition. Mov Disord 2010;25: 2695–2703. [PubMed: 20925068]
7. Gillies GE, Pienaar IS, Vohra S, Qamhawi Z. Sex differences in Parkinson's disease. Front Neuroendocrinol 2014;35:370–384. [PubMed: 24607323]
8. Jurado-Coronel JC, Cabezas R, Ávila Rodríguez MF, et al. Sex differences in Parkinson's disease: features on clinical symptoms, treatment outcome, sexual hormones and genetics. Front Neuroendocrinol 2018;50:18–30. [PubMed: 28974386]
9. Pankratz N, Nichols WC, Uniacke SK, et al. Genome-wide linkage analysis and evidence of gene-by-gene interactions in a sample of 362 multiplex Parkinson disease families. Hum Mol Genet 2003;12: 2599–2608. [PubMed: 12925570]

10. Klein C, Westenberger A. Genetics of Parkinson's disease. Cold Spring Harb Perspect Med 2012;2:a008888. [PubMed: 22315721]
11. Wilson GR, Sim JCH, McLean C, et al. Mutations in RAB39B cause X-linked intellectual disability and early-onset Parkinson disease with α-synuclein pathology. Am J Hum Genet 2014;95:729–735. [PubMed: 25434005]
12. Lesage S, Bras J, Cormier-Dequaire F, et al. Loss-of-function mutations in RAB39B are associated with typical early-onset Parkinson disease. Neurol Genet 2015;1:1–3. Available at: https://ng.neurology.org/content/1/1/e9. Accessed December 10, 2020.
13. Mata IF, Jang Y, Kim C-H, et al. The RAB39B p.G192R mutation causes X-linked dominant Parkinson's disease. Mol Neurodegener 2015; 10:50. [PubMed: 26399558]
14. Ciammola A, Carrera P, Fonzo AD, et al. X-linked parkinsonism with intellectual disability caused by novel mutations and somatic mosaicism in RAB39B gene. Parkinsonism Relat Disord 2017;44:142–146. [PubMed: 28851564]
15. Wise AL, Gyi L, Manolio TA. eXclusion: toward integrating the X chromosome in genome-wide association analyses. Am J Hum Genet 2013;92:643–647. [PubMed: 23643377]
16. Gottipati S, Arbiza L, Siepel A, et al. Analyses of X-linked and autosomal genetic variation in population-scale whole genome sequencing. Nat Genet 2011;43:741–743. [PubMed: 21775991]
17. Nalls MA, Pankratz N, Lill CM, 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]
18. Fung HC, Scholz S, Matarin M, et al. Genome-wide genotyping in Parkinson's disease and neurologically normal controls: first stage analysis and public release of data. Lancet Neurol 2006;5:911–916. [PubMed: 17052657]
19. McCulloch CC, Kay DM, Factor SA, et al. Exploring gene-environment interactions in Parkinson's disease. Hum Genet 2008; 123:257–265. [PubMed: 18210157]
20. Spencer CCA, Plagnol V, Strange A, et al. Dissection of the genetics of Parkinson's disease identifies an additional association 50 of SNCA and multiple associated haplotypes at 17q21. Hum Mol Genet 2011; 20:345–353. [PubMed: 21044948]
21. Nalls MA, Keller MF, Hernandez DG, et al. Baseline genetic associations in the Parkinson's Progression Markers Initiative (PPMI). Mov Disord 2016;31:79–85. [PubMed: 26268663]
22. Mata IF, Johnson CO, Leverenz JB, et al. Large-scale exploratory genetic analysis of cognitive impairment in Parkinson's disease. Neurobiol Aging 2017;56:211.e1–211.e7.
23. Montine TJ, Sonnen JA, Montine KS, et al. Adult changes in thought study: dementia is an individually varying convergent syndrome with prevalent clinically silent diseases that may be modified by some commonly used therapeutics. Curr Alzheimer Res 2012;9:718–723. [PubMed: 22471861]
24. Gan-Or Z, Rao T, Leveille E, et al. The Quebec Parkinson network: a researcher-patient matching platform and multimodal biorepository. J Parkinsons Dis 2020;10:301–313. [PubMed: 31868683]
25. Alfaro-Almagro F, Jenkinson M, Bangerter NK, et al. Image processing and quality control for the first 10,000 brain imaging datasets from UK Biobank. Neuroimage 2018;166:400–424. [PubMed: 29079522]
26. Miller KL, Alfaro-Almagro F, Bangerter NK, et al. Multimodal population brain imaging in the UK Biobank prospective epidemiological study. Nat Neurosci 2016;19:1523–1536. [PubMed: 27643430]
27. Le Guen Y, Leroy F, Philippe C, et al. Enhancer locus in ch14q23.1 modulates brain asymmetric temporal regions involved in language processing. Cereb Cortex 2020;30:5322–5332. [PubMed: 32432689]
28. Franke L, de Kovel CGF, Aulchenko YS, et al. Detection, imputation, and association analysis of small deletions and null alleles on oligonucleotide arrays. Am J Hum Genet 2008;82:1316–1333. [PubMed: 18519066]
29. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. Nature. 2020;581:434–443. 10.1038/s41586-020-2308-7. [PubMed: 32461654]

30. McCarthy S, Das S, Kretzschmar W, et al. A reference panel of 64,976 haplotypes for genotype imputation. Nat Genet 2016;48: 1279–1283. [PubMed: 27548312]
31. Chen CY, Pollack S, Hunter DJ, et al. Improved ancestry inference using weights from external reference panels. Bioinformatics 2013; 29:1399–1406. [PubMed: 23539302]
32. Auton A, Abecasis GR, Altshuler DM, et al. A global reference for human genetic variation. Nature 2015;526:68–74. [PubMed: 26432245]
33. Lambert CA, Connelly CF, Madeoy J, et al. Highly punctuated patterns of population structure on the X chromosome and implications for African evolutionary history. Am J Hum Genet 2010;86:34–44. [PubMed: 20085712]
34. Chang CC, Chow CC, Tellier LC, et al. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 2015; 4:7. [PubMed: 25722852]
35. Das S, Forer L, Schönherr S, et al. Next-generation genotype imputation service and methods. Nat Genet 2016;48:1284–1287. [PubMed: 27571263]
36. Price AL, Patterson NJ, Plenge RM, et al. Principal components analysis corrects for stratification in genome-wide association studies. Nat Genet 2006;38:904–909. [PubMed: 16862161]
37. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 2018;562: 203–209. [PubMed: 30305743]
38. Le Guen Y, Belloy ME, Napolioni V, et al. A novel age-informed approach for genetic association analysis in Alzheimer's disease. medRxiv 2021;21249292:2021.01.05.21249292. 10.1101/2021.01.05.21249292v1.
39. Mägi R, Morris AP. GWAMA: software for genome-wide association meta-analysis. BMC Bioinformatics 2010;11:288. [PubMed: 20509871]
40. Consortium GTE. Genetic effects on gene expression across human tissues. Nature 2017;550:204–213. [PubMed: 29022597]
41. Ramasamy A, Trabzuni D, Guelfi S, et al. Genetic variability in the regulation of gene expression in ten regions of the human brain. Nat Neurosci 2014;17:1418–1428. [PubMed: 25174004]
42. Storey JD, Tibshirani R. Statistical significance for genomewide studies. Proc Natl Acad Sci U S A 2003;100:9440–9445. [PubMed: 12883005]
43. Giambartolomei C, Vukcevic D, Schadt EE, et al. Bayesian test for Colocalisation between pairs of genetic association studies using summary statistics. PLoS Genet 2014;10:e1004383. [PubMed: 24830394]
44. Liu B, Gloudemans MJ, Rao AS, et al. Abundant associations with gene expression complicate GWAS follow-up. Nat Genet 2019;51: 768–769. [PubMed: 31043754]
45. Loh P-R, Tucker G, Bulik-Sullivan BK, et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. Nat Genet 2015;47:284–290. [PubMed: 25642633]
46. Pitcher TL, Melzer TR, MacAskill MR, et al. Reduced striatal volumes in Parkinson's disease: a magnetic resonance imaging study. Transl Neurodegener 2012;1:17. [PubMed: 23210661]
47. Thevenon J, Michot C, Bole C, et al. RPL10 mutation segregating in a family with X-linked syndromic intellectual disability. Am J Med Genet A 2015;167:1908–1912.
48. Zanni G, Kalscheuer VM, Friedrich A, et al. A novel mutation in RPL10 (ribosomal protein L10) causes X-linked intellectual disability, cerebellar hypoplasia, and Spondylo-epiphyseal dysplasia. Hum Mutat 2015;36:1155–1158. [PubMed: 26290468]
49. Brooks SS, Wall AL, Golzio C, et al. A novel ribosomopathy caused by dysfunction of RPL10 disrupts neurodevelopment and causes X-linked microcephaly in humans. Genetics 2014;198:723–733. [PubMed: 25316788]
50. Chiocchetti A, Pakalapati G, Duketis E, et al. Mutation and expression analyses of the ribosomal protein gene RPL10 in an extended German sample of patients with autism spectrum disorder. Am J Med Genet A 2011;155:1472–1475.
51. Gong X, Delorme R, Fauchereau F, et al. An investigation of ribosomal protein L10 gene in autism spectrum disorders. BMC Med Genet 2009;10:1–5. [PubMed: 19133158]
52. Monti C, Colugnat I, Lopiano L, et al. Network analysis identifies disease-specific pathways for Parkinson's disease. Mol Neurobiol 2018;55:370–381. [PubMed: 28004338]

53. Martin I, Kim JW, Lee BD, et al. Ribosomal protein s15 phosphorylation mediates LRRK2 neurodegeneration in Parkinson's disease. Cell 2014;157:472–485. [PubMed: 24725412]

54. Garcia-Esparcia P, Hernández-Ortega K, Koneti A, et al. Altered machinery of protein synthesis is region- and stage-dependent and is associated with α-synuclein oligomers in Parkinson's disease. Acta Neuropathol Commun 2015;3:1–25. [PubMed: 25627031]

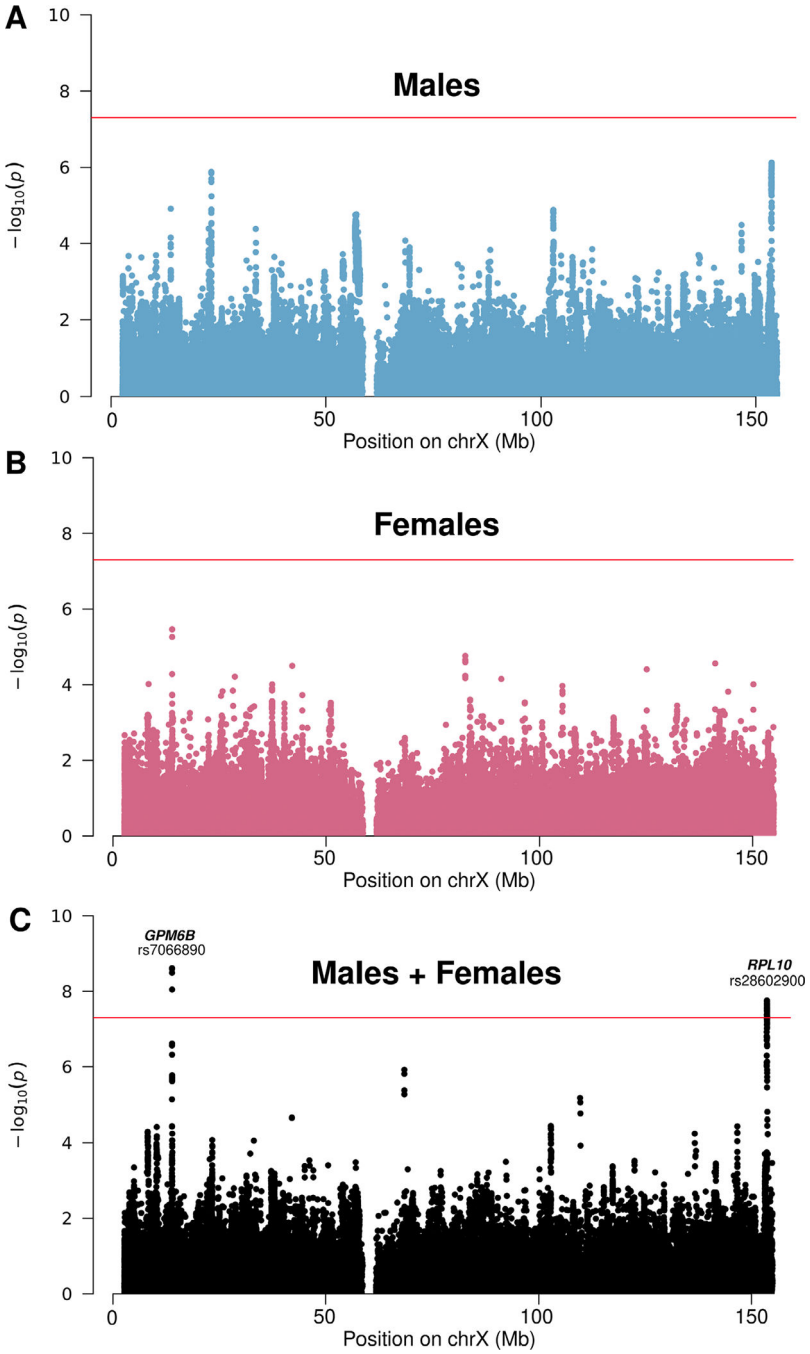

FIGURE 1:

X-chromosome wide analyses reveal 2 genome-wide significant loci in the combined sex meta-analysis. X-chromosome–wide sex-stratified discovery analyses in males (A) and in females (B) and combined sex meta-analysis (C) are shown. The most significantly regulated gene in brain tissues (*RPL10*) or the nearest gene (*GPM6B*) is annotated on top of the lead single nucleotide polymorphism at each locus. The horizontal line indicates the genome-wide significance threshold ($p < 5 \times 10^{-8}$).

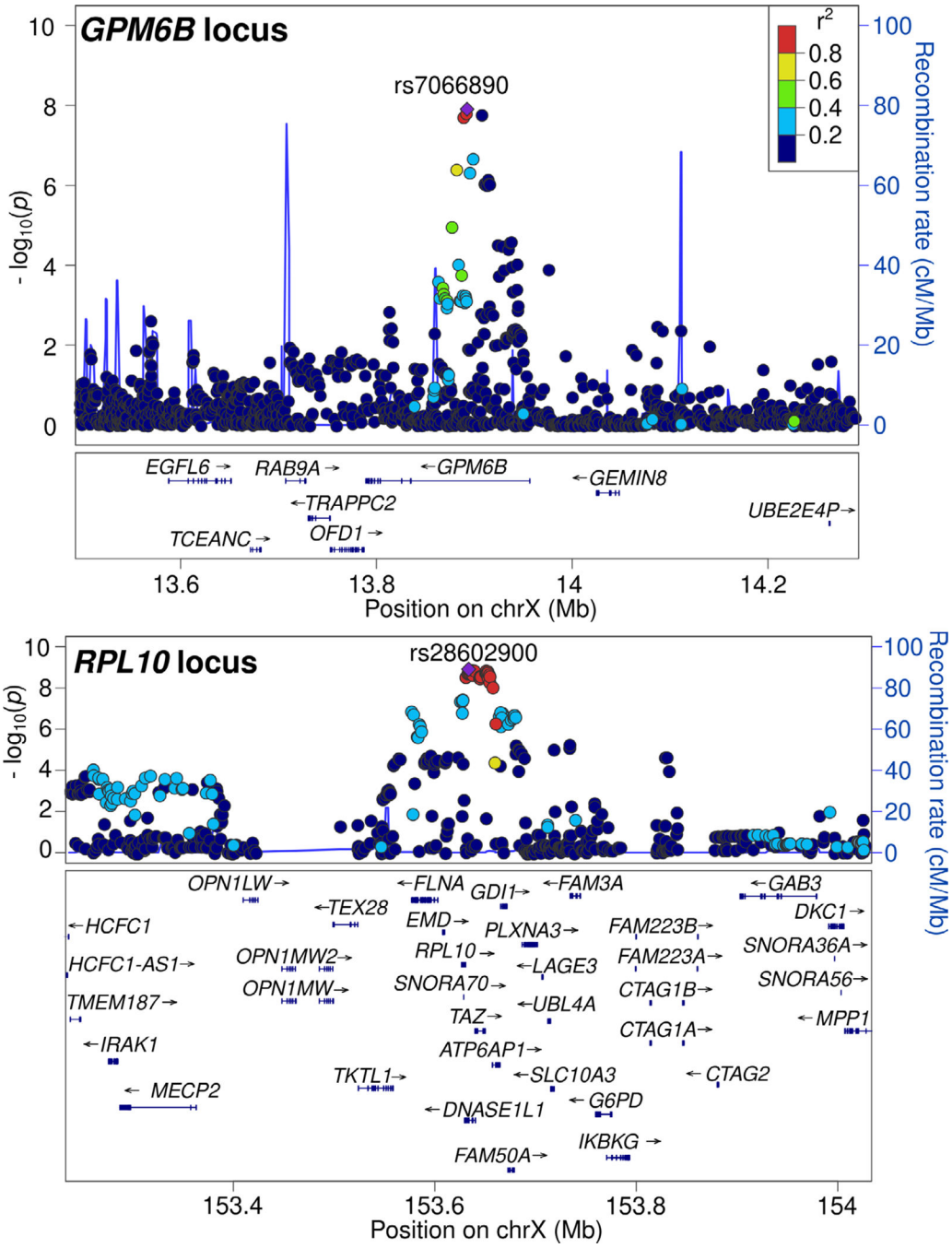

FIGURE 2:
LocusZoom plots of the genome-wide significant loci in the combined sex meta-analysis.

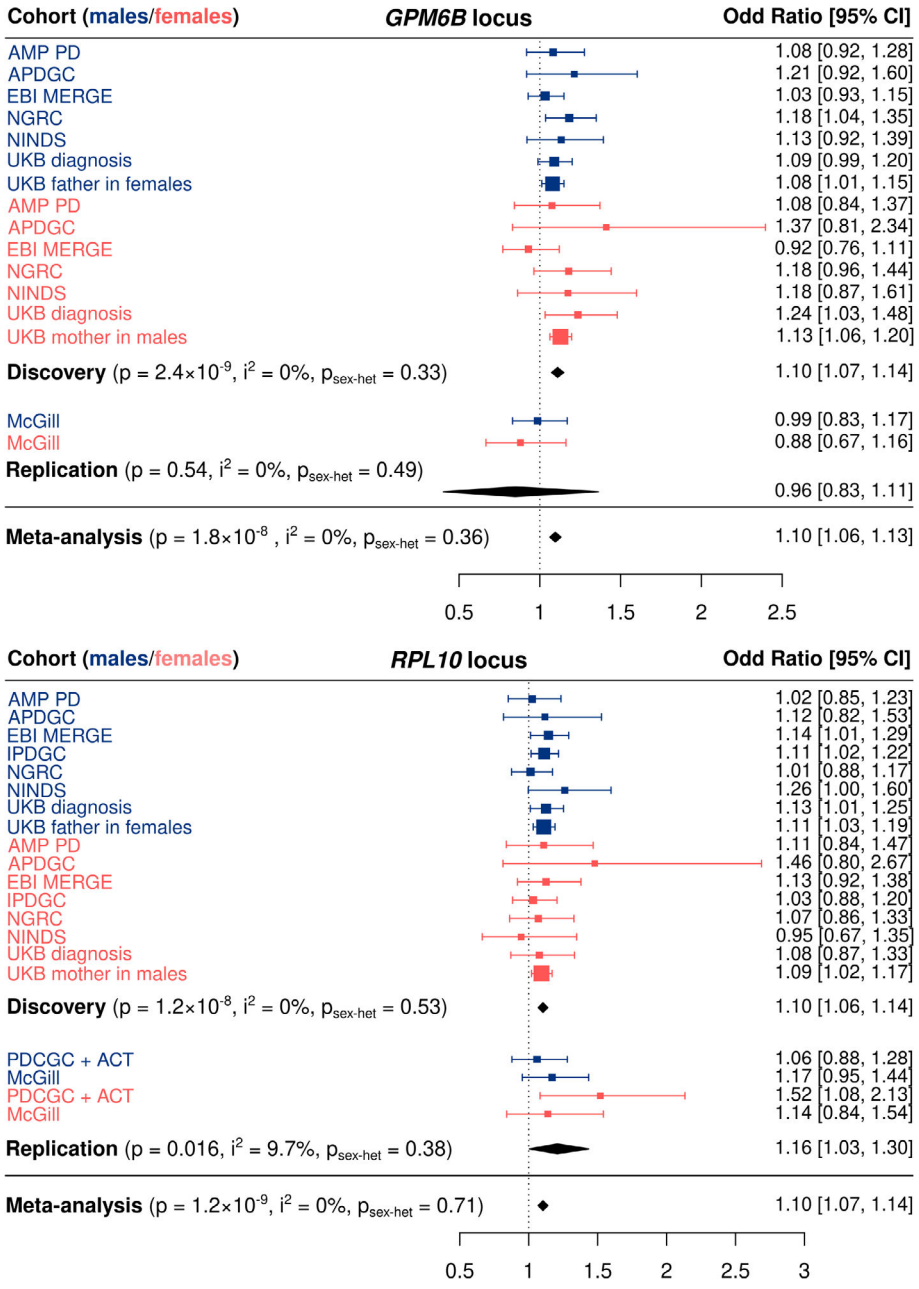

FIGURE 3:

Forest plots of the genome-wide significant loci in the combined sex meta-analysis, emphasizing the consistent direction of effect of the respective lead variants across datasets. ACT = Adult Change in Thought; AMP PD = Accelerating Medicines Partnership–Parkinson's Disease; APDGC = Autopsy-Confirmed PD GWAS Consortium; CI = confidence interval; EBI = European Bioinformatics Institute; IPDGC = International PD Genomics Consortium; NGRC = NeuroGenetics Research Consortium; NINDS = National Institute of Neurological Disorders and Stroke; PDCGC = Parkinson's Disease Cognitive Genetics Consortium; $p_{sex\text{-}het}$ = *p* value from the sex-heterogeneity test; UKB = UK Biobank.

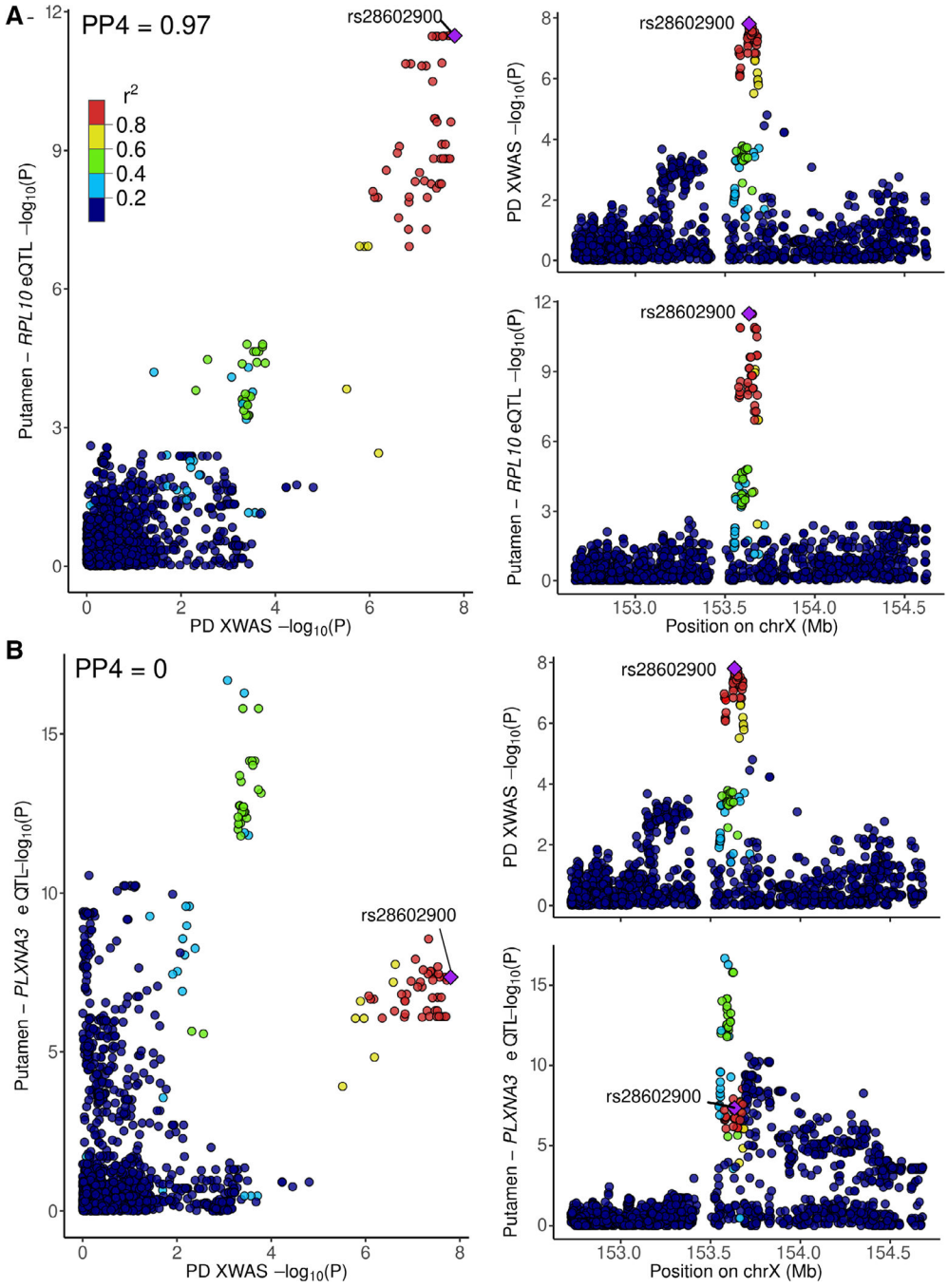

FIGURE 4:

Colocalization analyses between association with Parkinson disease (PD) risk and association with gene expression in brain tissues. (A) The locus, on Xq28, associated with PD risk colocalized with the *RPL10* expression quantitative locus (eQTL) in the putamen (and in the 10 other Genotype-Tissue Expression Project brain tissues with false discovery rate [FDR]-significant *RPL10* eQTLs; data not shown). (B) The locus, on Xq28, associated with PD risk did not colocalize with the *PLXNA3* eQTL in the putamen (or in other brain tissues with FDR-significant *PLXNA3* eQTLs). Rather, the eQTL signal is driven by a

nearby eQTL locus in low linkage disequilibrium with the PD risk locus. PP4 = posterior probability of colocalization; XWAS = X-chromosome–wide association study.

TABLE 1.

Overview of the Cohorts Included in the PD Discovery Meta-Analysis

Cohort	Genotyping Platform	SNPs Pre-IMP, n[a]	SNPs Post-IMP, n[b]	NWE, n	Cases, %
IPDGC	Illumina NeuroX Array	1,664	26,276		
Females			3,071	40.2	
Males			4,551	51.7	
NINDS	Illumina HumanHap550	19,788	223,218		
Females			715	44.9	
Males			720	62.2	
APDGC	Illumina HumanOmni1-Quad	17,496	224,713		
Females			261	52.9	
Males			497	74.0	
NGRC	Illumina HumanOmni1-Quad	11,732	225,792		
Females			1,692	33.9	
Males			1,920	62.5	
EBI MERGE	Illumina 610 k Quad & Illumina 1.2 M	5,727	217,464		
Females			2,034	32.5	
Males			2,563	43.2	
AMP PD WGS	WGS	2,466,931	221,601		
Females			1,005	58.4	
Males			1,412	73.4	
UK Biobank PD diagnosis	Affymetrix Axiom Array	cf Bycroft et al[37]	261,450		
Females			147,263	0.27	
Males			128,981	0.55	
UK Biobank proxy PD	Affymetrix Axiom Array	cf Bycroft et al[37]	261,450		
Mother PD in males			128,275	1.4	
Father PD in females			146,867	2.4	

[a]Number of SNPs passing quality control prior to IMP, or number of SNPs in the WGS.

[b]Number of SNPs with IMP quality $r^2 > 0.3$ (n), and after filtering for minor allele frequency > 0.01 and genotyping rate > 95%. r^2 represents imputation quality.

AMP PD = Accelerating Medicines Partnership–Parkinson's Disease; APDGC = Autopsy-Confirmed PD GWAS Consortium; EBI = European Bioinformatics Institute; IMP = imputation; IPDGC = International PD Genomics Consortium; NGRC = NeuroGenetics Research Consortium; NINDS = National Institute of Neurological Disorders and Stroke; NWE = Northwestern Europeans; PD = Parkinson disease; SNP = single nucleotide polymorphism; WGS = whole-genome sequencing.

TABLE 2.

Overview of the Cohorts Included in the Parkinson Disease Replication Meta-Analysis

Cohort	Genotyping Platform	SNPs Pre-IMP, n[a]	SNPs Post-IMP, n[b]	NWE, n	Cases, %
PDCGC & ACT	Illumina NeuroX & Illumina 660 W	1,322 in PDCGC, 8,908 in ACT	29,283 in both		
Females				1,094	20.0
Males				1,141	36.5
McGill	Illumina Infinium OmniExpress-24	15,912	227,144		
Females				758	41.7
Males				1,033	59.1

[a]Number of SNPs passing quality control prior to IMP, or number of SNPs in the whole-genome sequencing.

[b]Number of SNPs with IMP quality $r^2 > 0.3$ (n), and after filtering for minor allele frequency > 0.01 and genotyping rate > 95%.

ACT = Adult Change in Thought; IMP = imputation; NWE = Northwestern Europeans; PDCGC = Parkinson's Disease Cognitive Genetics Consortium; SNP = single nucleotide polymorphism.

TABLE 3.

Lead SNPs in Genome-Wide Significant Loci Passing in the Combined Sex Meta-Analysis

SNP	Pos (chrX/hg19)	Alleles, m/M	MAF	Nearest Gene	Sex	OR	95% CI	*p*	*p*-het
rs7066890	13892582	T/C	0.155	*GPM6B*	Combined sex	1.10	1.06–1.13	1.2×10^{-8}	0.36
	Xp22.2				Females	1.12	1.06–1.17	1.6×10^{-5}	
					Males	1.08	1.04–1.14	1.3×10^{-4}	
rs28602900	153633533	A/G	0.115	*RPL10*	Combined sex	1.10	1.07–1.14	1.2×10^{-9}	0.71
	Xq28				Females	1.09	1.04–1.15	7.2×10^{-4}	
					Males	1.11	1.07–1.15	4.0×10^{-7}	

ORs are calculated with respect to the minor allele. These values are provided for the overall meta-analysis; values in the discovery and replication are provided in Table S5.

CI = confidence interval; m/M = minor allele/major allele; MAF = minor allele frequency; OR = odds ratio; *p*-het = *p* value from the sex-heterogeneity test; SNP = single nucleotide polymorphism.