Targeted sequencing of Parkinson's disease loci genes highlights *SYT11, FGF20* and other associations

Short title: Parkinson's disease GWAS loci sequencing

Uladzislau Rudakou, BSc[1,2], Eric Yu, BSc[1,2], Lynne Krohn, BSc[1,2], Jennifer A. Ruskey, MSc[2,3], Yves Dauvilliers, MD, PhD[4], Dan Spiegelman, MSc[2,3], Lior Greenbaum, MD, PhD[5,6,7], Stanley Fahn, MD[8], Cheryl H. Waters, MD[8], Nicolas Dupré, MD, FRCP[9,10], Guy A. Rouleau, MD, PhD, FRCPC, FRSC[1,2,3], Sharon Hassin-Baer, MD[7,11,12], Edward A. Fon, MD, FRCPC[2,3], Roy N. Alcalay, MD, MS[8,13], Ziv Gan-Or, MD, PhD[1,2,3].

Affiliations:
1. Department of Human Genetics, McGill University, Montréal, QC, H3A 1A1, Canada
2. Montreal Neurological Institute, McGill University, Montréal, QC, H3A 1A1, Canada
3. Department of Neurology and Neurosurgery, McGill University, Montréal, QC, H3A 1A1, Canada
4. National Reference Center for Narcolepsy, Sleep Unit, Department of Neurology, Gui-de-Chauliac Hospital, CHU Montpellier, University of Montpellier, Inserm U1061, Montpellier, France
5. The Danek Gertner Institute of Human Genetics, Sheba Medical Center, Tel Hashomer, Ramat Gan , Israel
6. The Joseph Sagol Neuroscience Center, Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel
7. Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel
8. Department of Neurology, College of Physicians and Surgeons, Columbia University Medical Center, New York, NY 10032, USA
9. Division of Neurosciences, CHU de Québec, Université Laval, Québec City, QC, G1V 0A6, Canada
10. Department of Medicine, Faculty of Medicine, Université Laval, Québec City, QC, G1V 0A6, Canada;
11. Department of Neurology, Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel
12. Movement Disorders Institute, Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel
13. Taub Institute for Research on Alzheimer's Disease and the Aging Brain, College of Physicians and Surgeons, Columbia University Medical Center, New York, NY 10032, USA;

Corresponding author:

Ziv Gan-Or
Montreal Neurological Institute,
McGill University
1033 Pine Avenue, West,
Ludmer Pavilion, room 312

Montreal, QC, H3A 1A1,
Phone: +1-514-398-5845
e-mail: ziv.gan-or@mcgill.ca

Abstract

Genome-wide association studies (GWAS) have identified numerous loci associated with Parkinson's disease. The specific genes and variants that drive the associations within the vast majority of these loci are unknown. We aimed to perform a comprehensive analysis of selected genes to determine the potential role of rare and common genetic variants within these loci. We fully sequenced 32 genes from 25 loci previously associated with Parkinson's disease in 2,657 patients and 3,647 controls from three cohorts. Capture was done using molecular inversion probes targeting the exons, exon-intron boundaries and untranslated regions (UTRs) of the genes of interest, followed by sequencing. Quality control was performed to include only high-quality variants. We examined the role of rare variants (minor allele frequency < 0.01) using optimized sequence Kernel association tests (SKAT-O). The association of common variants was estimated using regression models adjusted for age, sex and ethnicity as required in each cohort, followed by a meta-analysis. After Bonferroni correction, we identified a burden of rare variants in *SYT11, FGF20* and *GCH1* associated with Parkinson's disease. Nominal associations were identified in 21 additional genes. Previous reports suggested that the *SYT11* GWAS association is driven by variants in the nearby *GBA* gene. However, the association of *SYT11* was mainly driven by a rare 3' UTR variant (rs945006601) and was independent of *GBA* variants (*p*=5.23E-05 after exclusion of all *GBA* variant carriers). The association of *FGF20* was driven by a rare 5' UTR variant (rs1034608171) located in the promoter region. The previously reported association of *GCH1* with Parkinson's Disease is driven by rare nonsynonymous variants, some of which are known to cause dopamine-responsive dystonia. We also identified two *LRRK2* variants, p.Arg793Met and p.Gln1353Lys, in ten and eight controls, respectively, but not in patients. We identified common variants associated with Parkinson's disease in *MAPT, TMEM175, BST1,*

SNCA and *GPNMB* which are all in strong linkage disequilibrium (LD) with known GWAS hits in their respective loci. A common coding *PM20D1* variant, p.Ile149Val, was nominally associated with reduced risk of Parkinson's disease (OR 0.73, 95% CI 0.60-0.89, *p*=1.161E-03). This variant is not in LD with the top GWAS hits within this locus and may represent a novel association. These results further demonstrate the importance of fine mapping of GWAS loci, and suggest that *SYT11*, *FGF20*, and potentially *PM20D1*, *BST1* and *GPNMB* should be considered for future studies as possible Parkinson's disease-related genes.

Keywords (5): Parkinson's disease, genetic fine mapping, risk loci, association study, sequencing

Abbreviations:

AJ: Ashkenazi Jewish

CADD: Combined Annotation Dependent Depletion

CI: confidence interval

CNS: central nervous system

DOPA: dihydroxyphenylalanine

DP: depth of coverage

eQTL: expression quantitative trait loci

GATK: Genome Analysis Toolkit

GnomAD: Genome Aggregation Database

GTEx: Genotype-tissue expression

GWAS: Genome-wide association study

hg19: human genome version 19

LD: linkage disequilibrium

LOF: loss-of-function

MAF: minor allele frequency

MIPs: molecular inversion probes

NS, Nonsyn: non-synonymous

OR: odds ratio

QC: quality control

REM: Rapid Eye Movement

SKAT-O: optimized sequence kernel association test

sQTL: splicing quantitative trait loci

UTR: untranslated region

yo: years of age

Introduction

Genome-wide association studies (GWASs) are an important tool for identifying genetic associations with complex disorders such as Parkinson's Disease (Nalls *et al.*, 2014; Chang *et al.*, 2017; Nalls *et al.*, 2019). GWASs examine the association of multiple common variants across the genome with specific traits. Typically, each associated locus includes multiple genes, and very often GWAS alone cannot identify the specific gene or variant within each locus that drives the association. Furthermore, GWASs generally overlook rare variants, while accumulation of rare variants on a specific common allele may drive some of these associations. Even if not located on a specific common allele, rare variants in GWAS genes may contribute to disease. For example, rare variants in *SNCA* (Polymeropoulos *et al.*, 1997), *GBA* (Sidransky *et al.*, 2009; Ran *et al.*, 2016; Emelyanov *et al.*, 2018; Amaral *et al.*, 2019), *LRRK2* (Paisan-Ruiz *et al.*, 2004; Khan *et al.*, 2005), *VPS13C* (Lesage *et al.*, 2016; Jansen *et al.*, 2017; Darvish *et al.*, 2018; Schormair *et al.*, 2018; Rudakou *et al.*, 2020), *GCH1* (Mencacci *et al.*, 2014; Guella *et al.*, 2015; Lewthwaite *et al.*, 2015; Xu *et al.*, 2017; Yoshino *et al.*, 2018; Rudakou *et al.*, 2019), all GWAS loci genes, also cause or increase the risk of Parkinson's disease.

Recently, a large-scale GWAS has identified 78 loci associated with risk of Parkinson's disease (Nalls *et al.*, 2019), yet only for a few of them we have identified the specific genes and variants that drive the association. In some of these loci, there are multiple hits, suggesting that more than one variant, or perhaps more than one gene within the locus, can be associated with Parkinson's disease. Fine-mapping of GWAS loci may help identifying these genes and variants, as was done for *SNCA* (Soldner *et al.*, 2016; Krohn *et al.*, 2019b), *VPS13C* (Lesage *et al.*, 2016; Jansen *et al.*, 2017; Darvish *et al.*, 2018; Schormair *et al.*, 2018; Rudakou *et al.*, 2020) and *TMEM175* (Krohn *et al.*, 2020). In the *SNCA* region, fine mapping has identified independent 5'

and 3’ variants that drive the associations with different synucleinopathies (Krohn *et al.*, 2019b). Similarly, a potentially protective haplotype in *VPS13C* (Rudakou *et al.*, 2020) and common nonsynonymous variants in *TMEM175* have been associated with Parkinson’s Disease through fine mapping, and in the case of *TMEM175* also by functional studies (Jinn *et al.*, 2019; Krohn *et al.*, 2020).

In the current study, we fully sequenced 32 genes from 25 different Parkinson’s disease GWAS loci in a total of 2,657 cases and 3,647 controls. We examined the association of common and rare variants with Parkinson’s Disease, including the role of rare bi-allelic variants.

Materials and Methods

Study Population

The study population included, consecutively recruited, unrelated patients with Parkinson's Disease (n=2,657) and controls (n=3,647) from three cohorts, collected at McGill University (Quebec, Canada and Montpellier, France), Columbia University (New York, NY, USA), and the Sheba Medical Center (Israel). The McGill cohort was recruited in Quebec, Canada (in part with the assistance of the Quebec Parkinson's Network(Gan-Or *et al.*, 2020)) and in Montpellier, France. It includes 1,026 Parkinson's Disease patients (mean age 61.5±11.3, 62.6% male) and 2,588 controls (mean age 54.5±14.2, 47.4% male), all unrelated of French-Canadian/French ancestry. The Columbia cohort includes 1,026 Parkinson's Disease patients (mean age 59.4±11.7, 64.2% male) and 525 controls (mean age 64.2±10.1, 37.1% male). This cohort has been previously described in detail, and is mainly composed of participants with European descent, but 21.2% of Parkinson's Disease patients and 18.7% controls are of Ashkenazi Jewish (AJ) origin (Alcalay *et al.*, 2016). The Sheba cohort included 605 Parkinson's Disease patients (mean age 60.7±11.9, 62.3% male) and 534 controls (mean age 34.0±7.0, 55.8% male), all of full AJ origin; previously described in more details (Ruskey *et al.*, 2019) . Movement disorder specialists diagnosed Parkinson's Disease with UK brain bank criteria (Hughes *et al.*, 1992) or the MDS clinical diagnostic criteria (Postuma *et al.*, 2015).

As detailed in the Statistical Analyses subsection, due to the differences in age and sex, statistical analyses were adjusted and included age and sex as co-variates. To account for ethnical heterogeneity of the Columbia cohort, an ethnicity covariate was also introduced for the analysis of this cohort (GWAS data was not available, therefore the reported ethnicity was used

and not principal components). All three cohorts were sequenced in the same lab (McGill University), using the same protocol.

Standard Protocol Approvals, Registrations, and Patient Consents

The institutional review boards of McGill University, Columbia University and Sheba Medical Center had approved the study protocols and informed consent was obtained from all individual participants before entering the study.

Target genes

A recent Parkinson's Disease GWAS identified 78 loci, however the current study was designed and performed before its publication and the target genes were taken from earlier literature (Nalls *et al.*, 2014; Chang *et al.*, 2017). Our target genes included: *ACMSD, BST1, CCDC62, DDRGK1, FGF20, GAK, GCH1, GPNMB, HIP1R, INPP5F, ITGA8, LAMP3, LRRK2, MAPT, MCCC1, PM20D1, RAB25, RAB29, RIT2, SCARB2, SETD1A, SIPA1L2, SLC41A1, SNCA, SREBF1, STK39, STX1B, SYT11, TMEM163, TMEM175, USP25* and *VPS13C*. These genes included several known Parkinson's disease related genes (*LRRK2*, *GCH1*, *SNCA*) found in GWAS loci in order to perform full sequencing assessment. Other genes were selected for sequencing based on assessment that included effects of quantitative trait loci, brain expression, potential association with known Parkinson's disease-causing genes and involvement in pathways potentially involved in Parkinson's disease, such the autophagy-lysosomal pathway, mitochondria quality control and endolysosomal recycling.

Sequencing

The coding regions, exon-intron boundaries and the 5' and 3' untranslated regions (UTRs) of the genes listed above were targeted using molecular inversion probes (MIPs), as previously described (O'Roak *et al.*, 2012). All MIPs used to sequence the genes of the study are detailed in Supplementary Table 1. Targeted DNA capture and amplification was done as previously described (Ross *et al.*, 2016), and the full protocol is available upon request. The library was sequenced using Illumina HiSeq 2500/4000 platform at the McGill University and Genome Quebec Innovation Centre.

Data processing and quality control

Reads were mapped to the human reference genome (hg19) with Burrows-Wheeler Aligner (Li and Durbin, 2009). Post-alignment quality control and variant calling was performed using Genome Analysis Toolkit (GATK, v3.8) (McKenna *et al.*, 2010). Variant annotation was done with ANNOVAR (Wang *et al.*, 2010). Data on the frequency of each variant in different populations were extracted from the public database Genome Aggregation Database (GnomAD) (Lek *et al.*, 2016).

The quality control pipeline that was used in this study is available on GitHub (https://github.com/gan-orlab/MIPVar.git). All calls with less than 25% reads with the variant have been removed. The remaining quality control (QC) was performed using GATK and PLINK software v1.9 (Purcell *et al.*, 2007). Genotyping quality (GQ) < 30 was used as a threshold for variant exclusion. Different depths of coverage thresholds were used for common and rare variants. Common variants threshold for inclusion was ≥ 15x, while rare variants (minor allele frequency [MAF] < 0.01) thresholds were more stringent and repeated twice, at ≥ 30x and

≥ 50x. Variants with genotype rate lower than 90% were excluded. Variants that deviated from Hardy-Weinberg equilibrium set at $p \leq 0.001$ were filtered out. Threshold for missingness difference between cases and controls was set at $p = 0.05$ and the filtration script adjusted it with Bonferroni correction. Samples with average genotyping rate of less than 90% were excluded.

Statistical analysis

The burden of rare variants in each of the target genes was tested using the optimized sequence Kernel association test (SKAT-O, R package) (Lee *et al.*, 2012). In addition, we examined the burden of specific variant subgroups including: i) Funct.: Potentially functional included all nonsynonymous mutations, stop gain/loss, frameshift mutations, splicing variants located within two base pairs of exon-intron junctions, and variants that were flagged by ENCODE (explained below). ii) NS: nonsynonymous variants. iii) LOF: loss-of-function variants included stop gain/loss, frameshift, and splicing variants located within two base pairs of exon-intron junctions. iv) CADD: variants with the Combined Annotation Dependent Depletion (CADD) score ≥ 12.37; this threshold indicates the top 2% variants predicted to be the most damaging in the genome (Amendola *et al.*, 2015). v) ENCODE: variants that are predicted to have regulatory activity, such as promoter, enhancer, activator and repressor, based on the chromatin profiling studies (Ernst *et al.*, 2011). To increase power, rare variants were assessed in the three merged cohorts (since rare variants are less affected by ethnicity). Analysis of the individual cohorts was also performed post-hoc. Similarly to exome-wide burden analyses of rare variants (Kenna *et al.*, 2016), Bonferroni correction for statistical significance was performed based on the number of genes that were tested, and the thresholds are $p < 1.85E-03$ for genes with variant coverage of ≥ 50x and $p < 1.56E-03$ at ≥ 30x

The association between common variants and Parkinson's Disease was examined by logistic regression models using PLINK v1.9, with the status (patient or control) as a dependent variable, age and sex as covariates in all cohorts. Ethnicity as an additional covariate in the Columbia University cohort due to the ethnical heterogeneity. Meta-analysis of our three cohorts was conducted using METAL (R package) (Willer *et al.*, 2010) with a fixed effect model. Cochran's Q-test was applied to test for residual heterogeneity. LD of common variants and their respective GWAS top marker was examined using the LDlink online tool LDmatrix (https://ldlink.nci.nih.gov/?tab=ldmatrix) (Machiela and Chanock, 2015). Effects of common and rare variants on expression levels (eQTL) or splicing (sQTL) was evaluated using GTEx Portal online tool (https://www.gtexportal.org/home/).

Data availability

Anonymized data is available upon request by any qualified investigator.

Results

Coverage and variants identified

The average coverage for all 32 genes included in the current analysis was 642X, with a range of 86-1,064 (median 688) across all genes. An average of 97% of the targeted regions was covered at > 15X, 96% at > 30X, and 92% at >50X. Supplementary Table 2 details the average coverage and the percent coverage for each gene. At coverage of > 30X, we found 3,327 rare variants and at > 50X, 1,713 rare were identified (Supplementary Tables 3 and 4). A total of 311 common variants were identified across all genes in all three cohorts (Supplementary Table 5) with a coverage of >15X.

Rare variants in *SYT11, FGF20* and *GCH1* are associated with Parkinson's disease

Table 1 details the SKAT-O results of the analysis in the three combined cohorts, and Supplementary Table 6 details the results per cohort. After Bonferroni correction, *SYT11, FGF20* and *GCH1* were associated with Parkinson's Disease. Nominal associations were identified in 21 genes (Table 1, Supplementary Table 6). When considering all variants with MAF < 0.01 in *SYT11*, there was a strong association with Parkinson's Disease (p = 5.3E-06, Table 1), mainly driven by a *SYT11* 3' UTR variant (rs945006601) with MAF of 0.004 in Parkinson's Disease patients and 0.00015 in controls (OR = 27.0, 95%CI 3.6-202.6, p = 0.0013). As was previously reported (Blauwendraat *et al.*, 2018), LD analysis showed that this variant is in some LD with the known Parkinson's disease-associated variants in *GBA* p.Glu326Lys (p.E326K, r^2 = 1.5E-05, D' = 1) and p.Asn370Ser (p.N370S, r^2 = 4.23E-05, D' = 1). The low r^2 for both variants are due to the different allele frequencies, yet the D' of 1 indicates an LD. To test if the identified association in *SYT11* is driven by these two pathogenic *GBA* variants, we removed the carriers of these variants and reran the *SYT11* SKAT-O analysis (p=5.58E-05). Since we have full sequencing data on *GBA* for all participants, we further excluded carriers of any other *GBA* variant that was ever reported to be potentially pathogenic and the association remained (p=5.23E-05).

When analyzing *FGF20*, a burden of all rare variants was demonstrated (p=0.0002), mainly driven by a rare 5' UTR variant, rs1034608171, which is located in the promoter region of the gene, with MAF of 0.002 in Parkinson's Disease patients and 0.00015 in controls (OR = 13.4, 95% CI 1.7-106.2, p = 0.014). The association of *GCH1* with Parkinson's Disease is driven

by rare nonsynonymous variants (Supplementary Tables 3 and 4), some of which are known to cause dopamine responsive dystonia as we have previously reported(Rudakou *et al.*, 2019).

After removing the *LRRK2* p.Gly2019Ser variant from the analysis, there was no statistically significant burden of other rare *LRRK2* variants in our cohorts. However, we did identify two variants, p.Arg793Met (rs35173587) and p.Gln1353Lys (rs200526782), in ten and eight controls, respectively, but not in Parkinson's Disease. Neither of these two variants had a statistically significant association after Bonferroni correction.

Association of common variants with potential functional effects in known Parkinson's disease loci

To examine the association of common variants in the 32 genes, we performed logistic regression adjusted for age and sex in all three cohorts, and also adjusted for ethnicity in the Columbia University cohort. Additionally, our three cohorts were meta-analyzed. After Bonferroni correction we found 24 variants from five genes associated with risk of Parkinson's Disease (Table 2). Forest plots in Fig.1 show the meta-analysis results for the tagging variants of these six genes. We further examined the LD of these variants with the known Parkinson's Disease-associated variants from the most recent Parkinson's Disease GWAS, and their potential functional effects (detailed in Methods) in ENCODE, GTEx and UCSC genome browser. In the *MAPT* locus, 17 common variants at full LD, all represent the *MAPT* H2 haplotype, were associated with reduced risk of Parkinson's Disease (Table 2). The variants identified in *TMEM175*, *BST1*, *SNCA* and *GPNMB* are all in strong LD with known GWAS hits in their respective loci (Table 2). Two *SNCA* 3' UTR variants were in full LD between themselves, and in strong LD with rs356219, a known 3' GWAS hit in the *SNCA* locus. These two variants are

located within a strong enhancer region and are associated with altered splicing of *SNCA* in the Cortex (Fig. 2). Three *GPNMB* variants were identified, one of which (rs5850) is located at the 3' UTR of *GPNMB* and associated with reduced expression of *GPNMB* in numerous brain regions including the substantia nigra, putamen and cortex (Fig. 3). A coding variant in *PM20D1*, p.Ile149Val (rs1891460), was nominally associated with reduced risk of Parkinson's Disease (OR = 0.73, 95% CI 0.60-0.89, p = 1.161E-03, specific frequencies in each cohort are in Supplementary Table 5). This variant is located within the *PARK16* region which also includes *RAB29* (previously named *RAB7L1*), *SLC41A1* and other genes. However, this variant is not in LD with the top GWAS hits within this locus (r^2 = 0.04, D' = 0.46).

Discussion

In the current study, we examined the role of common and rare variants in 32 genes located within known Parkinson's Disease risk loci, by performing targeted next generation sequencing of the coding and 5' and 3' untranslated regions, followed by association tests. We found a statistically significant burden, after Bonferroni correction, of rare variants in *SYT11, FGF20* and *GCH1*, and nominal associations ($p<0.05$) with other genes and variants (Table 1, Supplementary Tables 3,4,6). In addition, we have identified common coding and putative regulatory variants associated with Parkinson's Disease in *MAPT, TMEM175, BST1, SNCA* and *GPNMB*, including a potential novel association with a nonsynonymous variant in *PM20D1*, p.Ile149Val , which is nominally significant and requires further confirmation.

Three genes, *SYT11*, *FGF20* and *GCH1* were identified with a statistically significant burden of rare variants comparing Parkinson's Disease to controls. *SYT11* is located close to the *GBA* locus, and it was reported that the GWAS signal from *SYT11-GBA* locus in Parkinson's Disease can be attributed to coding *GBA* variants (Blauwendraat *et al.*, 2018). The *SYT11* variant mainly responsible for the association in our data is in weak LD with two coding *GBA* variants, p.Glu326Lys (p.E326K) and p.Asn370Ser (p.N370S). After exclusion of carriers of these variants, and further exclusion of all other *GBA* variants, the association of *SYT11* remained strong. These results suggest that *SYT11* may independently be associated with risk of PD. The Protein encoded by *SYT11*, Synaptotagmin 11, belongs to the synaptotagmin family, known calcium sensors mediating calcium-dependent regulation of membrane trafficking in synaptic transmission (Xie *et al.*, 2017). Synaptotagmin 11 is involved in regulation of endocytosis and vesicle recycling process, which is essential for neurotransmission (Wang *et al.*, 2016; Xie *et al.*, 2017). Overexpression of Synaptotagmin 11 in mice led to impairment of DA transmission

(Wang *et al.*, 2018). Synaptotagmin 11 was also reported to be a physiological substrate for E3 ubiquitin ligase – parkin, playing a critical role in parkin-related neurotoxicity (Wang *et al.*, 2018). Involvement of Synaptotagmin 11 in other mechanisms potentially related to Parkinson's Disease such as autophagy-lysosomal pathway (Bento *et al.*, 2016), phagocytosis of α-synuclein fibrils (Du *et al.*, 2017), and repair of injured astrocytes via lysosome exocytosis regulation (Sreetama *et al.*, 2016), suggests that *SYT11* needs to be considered as a target for further research in PD. Fibroblast growth factor 20 (*FGF20*) encodes a neurotrophic factor preferentially expressed in the substantia nigra pars compacta. It regulates central nervous system development and function (International Parkinson Disease Genomics Consortium, 2011) and plays a role in dopaminergic neurons differentiation and survival (Itoh and Ohta, 2013). There is a number of conflicting reports about FGF20 overexpression increasing α-synuclein levels in dopaminergic neurons (Wang *et al.*, 2008; Wider *et al.*, 2009; Sekiyama *et al.*, 2014; Tarazi *et al.*, 2014). The rare variant (rs1034608171) that drives the association in our data is located in a regulatory region and its potential functional effects would require additional studies. *GCH1* encodes GTP-cyclohydrolase 1, an essential enzyme for the dopamine synthesis pathway in the nigrostriatal cells (Kurian *et al.*, 2011). Rare *GCH1* variants may cause DOPA-responsive dystonia, and the same variants have also been associated with Parkinson's Disease (Mencacci *et al.*, 2014; Rudakou *et al.*, 2019). Since these mutations are rare, they only reached statistical significance in the McGill cohort, as we have previously described (Rudakou *et al.*, 2019). The two *LRRK2* variants, p.Arg793Met and p.Gln1353Lys, that were found only in controls require additional studies to examine their potential protective roles in Parkinson's Disease. *LRRK2* is a well-validated Parkinson's Disease related gene(Bandres-Ciga *et al.*, 2020).

All 17 *MAPT* variants that were found to be significantly associated with Parkinson's Disease are in perfect LD with each other, and represent the H2 haplotype, known to be associated with reduced risk of Parkinson's Disease (Li *et al.*, 2018). The identified coding variants in *MAPT* are located within exons that are likely not expressed in the CNS (Caillet-Boudin *et al.*, 2015), suggesting that the association in this locus may be due to regulatory variants. *TMEM175* is located in the fourth most significant GWAS locus in Parkinson's Disease (Nalls *et al.*, 2019), and encodes a transmembrane potassium channel responsible for K^+ conductance in endosomes and lysosomes (Cang *et al.*, 2015). TMEM175 regulates lysosomal function through maintenance of lysosomal membrane potential and luminal pH stability, as well as its role in autophagosome-lysosome fusion (Lee *et al.*, 2017). In a recent genetic and functional study, two coding *TMEM175* variants, including the variant identified here, were shown to be associated with risk of Parkinson's Disease, risk of REM-sleep behavior disorder, and glucocerebrosidase activity (Krohn *et al.*, 2019a).

The *BST1* and *SNCA* variants associated with risk of Parkinson's Disease in the current study are in strong LD with the known GWAS markers in their respective loci. The *BST1* variant is intronic without known function, and the *SNCA* 3 'UTR variants are located within strong *SNCA* enhancers. Whether these *SNCA* variants drive the association with Parkinson's Disease and their potential functional effects require additional studies. *GPNMB* encodes the glycoprotein nonmetastatic melanoma protein B a type 1, which is a transmembrane glycoprotein shown to have a role in melanoma (Taya and Hammes, 2018). Patients with melanoma have an increased risk of Parkinson's Disease (Liu *et al.*, 2011), and Parkinson's Disease patients have an increased risk for melanoma (Bertoni *et al.*, 2010). GPNMB may have a role in the lysosome (Tomihari *et al.*, 2009) and in regulation of microglial inflammation (Dzamko *et al.*, 2015;

Herrero *et al.*, 2015). While the expression of GPNMB in brain is overall low, the two variants identified in the current study are associated with reduced GPNMB expression in various brain tissues (Fig. 3). One of the variants, rs5850, is located within the 3' UTR of *GPNMB*, but there is no clear regulatory function for this variant.

Peptidase M20 Domain Containing 1 (*PM20D1*) encodes a protein that regulates the production of N-fatty-acyl amino acids which have a function of chemical uncouplers of mitochondrial respiration (Long *et al.*, 2016). Mitochondrial quality control is likely important in Parkinson's Disease, and genes involved in mitophagy are involved in familial forms of Parkinson's Disease (Park *et al.*, 2018). *PM20D1* is located within the *PARK16* locus which also contains *SLC45A3, NUCKS1, RAB29* (previously named *RAB7L1*) and *SLC41A1* (Satake *et al.*, 2009; Simon-Sanchez *et al.*, 2009; Nalls *et al.*, 2019).Several studies have highlighted *RAB29* as the potential gene associated with Parkinson's Disease in this locus (Gan-Or *et al.*, 2012),as it was shown to interact with *LRRK2* (MacLeod *et al.*, 2013) and affect lysosomal function (Eguchi *et al.*, 2018). In the current study we identified a common nonsynonymous variant in *PM20D1*, p.Ile149Val (rs1891460), which is not in LD with the Parkinson's Disease GWAS marker in this locus. Since this variant is not reported in ClinVar, has a low CADD score (1.01), and the statistical association is of borderline significance when correcting by the number of genes, it is possible that this association was found by chance. Replications in other cohorts will be required to determine whether this variant is associated with Parkinson's Disease.

Our study has several limitations. The Parkinson's disease patients and controls are not matched by sex and age, hence we adjusted for these variables in the statistical analysis. We found no significant difference between allele frequencies of young (<40 years old) and old

controls, which allowed us to use all of the controls in our analyses. In addition, the coverage of some of the genes we analyzed was suboptimal (Supplementary Table 2), and we did not target intronic regions and other potentially regulatory regions outside the 5' and 3' UTRs, which may also affect gene function. Whole genome sequencing studies will be required to study these regions and overcome the coverage limitations.

Our results, combined with previous studies, highlight the importance of performing fine mapping of GWAS loci to identify the variants and genes within each locus that drive the reported associations. Thus far, the causative genes within most Parkinson's Disease GWAS loci are unknown, with the exception of *SNCA*, *GBA*, *LRRK2*, and likely *VPS13C*, *MAPT*, *TMEM175* and *GCH1*. The current study suggests that *SYT11*, *FGF20*, and potentially *PM20D1*, *BST1* and *GPNMB* should be considered as possible Parkinson's Disease-related genes and targets for further research.

Acknowledgements

We thank the participants for contributing to the study. GAR holds a Canada Research Chair in Genetics of the Nervous System and the Wilder Penfield Chair in Neurosciences. EAF is supported by a Foundation Grant from the Canadian Institutes of Health Research (FDN grant – 154301). ZGO is supported by the Fonds de recherche du Québec - Santé (FRQS) Chercheurs-boursiers award, in collaboration with Parkinson Quebec, and by the Young Investigator Award by Parkinson Canada. The access to part of the participants for this research has been made possible thanks to the Quebec Parkinson's Network (http://rpq-qpn.ca/en/). We thank Daniel

Rochefort, Helene Catoire, Sandra B. Laurent, Clotilde Degroot and Vessela Zaharieva for their assistance.

Funding

This work was financially supported by grants from the Michael J. Fox Foundation, the Canadian Consortium on Neurodegeneration in Aging (CCNA), the Canada First Research Excellence Fund (CFREF), awarded to McGill University for the Healthy Brains for Healthy Lives initiative (HBHL), and Parkinson Canada. The Columbia University cohort is supported by the Parkinson's Foundation, the National Institutes of Health (K02NS080915, and UL1 TR000040) and the Brookdale Foundation.

Competing interests

Ziv Gan-Or has received consulting fees from Lysosomal Therapeutics Inc., Idorsia, Prevail Therapeutics, Denali, Ono Therapeutics, Deerfield and Inception Sciences (now Ventus). None of these companies were involved in any parts of preparing, drafting and publishing this review.

Cheryl H. Waters disclosed the following: Research support (Biogen, Roche, Sanofi); consulting fees (Acorda, Amneal, Impel, Kyowa, Mitsubishi, and Neurocrine); speaker's honoraria (Acadia, Acorda, Adamas, Amneal, and US WorldMeds)

Figure legends

Figure 1. Forest plots, meta-analyses of common variants in the three cohorts.

Each panel presents the individual odds ratio and 95% confidence interval for each cohort, and the results of the meta-analysis for the tagging variant of six loci identified in our study. Meta-analysis *p*-values: A. *MAPT* (*p*=8.28E-10); B. *TMEM175* (*p*=2.11E-04); C. *BST1* (*p*=3.14E-05); D. *SNCA* (*p*=4.08E-05); E. *GPNMB* (*p*=2.19E-05)

Figure 2. Effect of 3'UTR *SNCA* variants on splicing.

Figure is processed from GTeX (https://www.gtexportal.org), and represent the splicing quantitative trait locus (sQTL) effects of: **A.** rs1045722 in the 3' UTR of *SNCA* and **B.** rs3857053 in the 3' UTR of *SNCA*

Figure 3. The 3'UTR *GPNMB* variant (rs5850) is associated with reduced expression in numerous brain regions.

Figure is processed from GTeX (https://www.gtexportal.org), and shows the expression quantitative trait locus (eQTL) effect of the 3' UTR *GPNMB* variant rs5850 on GPNMB expression in different brain regions.

Supplementary material

Supplementary Table 1 Detailed information on molecular inversion probes (MIPs)

Supplementary Table 2 Coverage statistics

Supplementary Table 3 All identified rare variants with 30x minimum depth of coverage

Supplementary Table 4 All identified rare variants with 50x minimum depth of coverage

Supplementary Table 5 All identified common variants

Supplementary Table 6 SKAT-O results for individual cohorts

References

Alcalay RN, Levy OA, Wolf P, Oliva P, Zhang XK, Waters CH, et al. SCARB2 variants and glucocerebrosidase activity in Parkinson's disease. NPJ Parkinsons Dis 2016; 2.

Amaral CEM, Lopes PF, Ferreira JCC, Alves EAC, Montenegro MVB, Costa ETD, et al. GBA mutations p.N370S and p.L444P are associated with Parkinson's disease in patients from Northern Brazil. Arq Neuropsiquiatr 2019; 77(2): 73-9.

Amendola LM, Dorschner MO, Robertson PD, Salama JS, Hart R, Shirts BH, et al. Actionable exomic incidental findings in 6503 participants: challenges of variant classification. Genome Res 2015; 25(3): 305-15.

Bandres-Ciga S, Diez-Fairen M, Kim JJ, Singleton AB. Genetics of Parkinson's disease: An introspection of its journey towards precision medicine. Neurobiol Dis 2020; 137: 104782.

Bento CF, Ashkenazi A, Jimenez-Sanchez M, Rubinsztein DC. The Parkinson's disease-associated genes ATP13A2 and SYT11 regulate autophagy via a common pathway. Nat Commun 2016; 7: 11803.

Bertoni JM, Arlette JP, Fernandez HH, Fitzer-Attas C, Frei K, Hassan MN, et al. Increased melanoma risk in Parkinson disease: a prospective clinicopathological study. Arch Neurol 2010; 67(3): 347-52.

Blauwendraat C, Bras JM, Nalls MA, Lewis PA, Hernandez DG, Singleton AB, et al. Coding variation in GBA explains the majority of the SYT11-GBA Parkinson's disease GWAS locus. Mov Disord 2018; 33(11): 1821-3.

Caillet-Boudin ML, Buee L, Sergeant N, Lefebvre B. Regulation of human MAPT gene expression. Mol Neurodegener 2015; 10: 28.

Cang C, Aranda K, Seo YJ, Gasnier B, Ren D. TMEM175 Is an Organelle K(+) Channel Regulating Lysosomal Function. Cell 2015; 162(5): 1101-12.

Chang D, Nalls MA, Hallgrimsdottir IB, Hunkapiller J, van der Brug M, Cai F, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. Nat Genet 2017; 49(10): 1511-6.

Darvish H, Bravo P, Tafakhori A, Azcona LJ, Ranji-Burachaloo S, Johari AH, et al. Identification of a large homozygous VPS13C deletion in a patient with early-onset Parkinsonism. Mov Disord 2018; 33(12): 1968-70.

Du C, Wang Y, Zhang F, Yan S, Guan Y, Gong X, et al. Synaptotagmin-11 inhibits cytokine secretion and phagocytosis in microglia. Glia 2017; 65(10): 1656-67.

Dzamko N, Geczy CL, Halliday GM. Inflammation is genetically implicated in Parkinson's disease. Neuroscience 2015; 302: 89-102.

Eguchi T, Kuwahara T, Sakurai M, Komori T, Fujimoto T, Ito G, et al. LRRK2 and its substrate Rab GTPases are sequentially targeted onto stressed lysosomes and maintain their homeostasis. Proc Natl Acad Sci U S A 2018; 115(39): E9115-E24.

Emelyanov AK, Usenko TS, Tesson C, Senkevich KA, Nikolaev MA, Miliukhina IV, et al. Mutation analysis of Parkinson's disease genes in a Russian data set. Neurobiol Aging 2018; 71: 267 e7-e10.

Ernst J, Kheradpour P, Mikkelsen TS, Shoresh N, Ward LD, Epstein CB, et al. Mapping and analysis of chromatin state dynamics in nine human cell types. Nature 2011; 473(7345): 43-9.

Gan-Or Z, Bar-Shira A, Dahary D, Mirelman A, Kedmi M, Gurevich T, et al. Association of sequence alterations in the putative promoter of RAB7L1 with a reduced parkinson disease risk. Arch Neurol 2012; 69(1): 105-10.

Gan-Or Z, Rao T, Leveille E, Degroot C, Chouinard S, Cicchetti F, et al. The Quebec Parkinson Network: A Researcher-Patient Matching Platform and Multimodal Biorepository. J Parkinsons Dis 2020; 10(1): 301-13.

Guella I, Sherman HE, Appel-Cresswell S, Rajput A, Rajput AH, Farrer MJ. Parkinsonism in GTP cyclohydrolase 1 mutation carriers. Brain 2015; 138(Pt 5): e349.

Herrero MT, Estrada C, Maatouk L, Vyas S. Inflammation in Parkinson's disease: role of glucocorticoids. Front Neuroanat 2015; 9: 32.

Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. J Neurol Neurosurg Psychiatry 1992; 55(3): 181-4.

International Parkinson Disease Genomics Consortium (IPDGC). Imputation of sequence variants for identification of genetic risks for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet 2011; 377(9766): 641-9.

Itoh N, Ohta H. Roles of FGF20 in dopaminergic neurons and Parkinson's disease. Front Mol Neurosci 2013; 6: 15.

Jansen IE, Ye H, Heetveld S, Lechler MC, Michels H, Seinstra RI, et al. Discovery and functional prioritization of Parkinson's disease candidate genes from large-scale whole exome sequencing. Genome Biol 2017; 18(1): 22.

Jinn S, Blauwendraat C, Toolan D, Gretzula CA, Drolet RE, Smith S, et al. Functionalization of the TMEM175 p.M393T variant as a risk factor for Parkinson disease. Hum Mol Genet 2019; 28(19): 3244-54.

Kenna KP, van Doormaal PT, Dekker AM, Ticozzi N, Kenna BJ, Diekstra FP, et al. NEK1 variants confer susceptibility to amyotrophic lateral sclerosis. Nat Genet 2016; 48(9): 1037-42.

Khan NL, Jain S, Lynch JM, Pavese N, Abou-Sleiman P, Holton JL, et al. Mutations in the gene LRRK2 encoding dardarin (PARK8) cause familial Parkinson's disease: clinical, pathological, olfactory and functional imaging and genetic data. Brain 2005; 128(Pt 12): 2786-96.

Krohn L, Ozturk TN, Vanderperre B, Ouled Amar Bencheikh B, Ruskey JA, Laurent SB, et al. Genetic, Structural, and Functional Evidence Link TMEM175 to Synucleinopathies. Ann Neurol 2019a.

Krohn L, Ozturk TN, Vanderperre B, Ouled Amar Bencheikh B, Ruskey JA, Laurent SB, et al. Genetic, Structural, and Functional Evidence Link TMEM175 to Synucleinopathies. Ann Neurol 2020; 87(1): 139-53.

Krohn L, Wu RYJ, Heilbron K, Ruskey JA, Laurent SB, Blauwendraat C, et al. Fine-mapping of SNCA in REM sleep behavior disorder and overt synucleinopathies. BioRxiv 2019b: 756528.

Kurian MA, Gissen P, Smith M, Heales S, Jr., Clayton PT. The monoamine neurotransmitter disorders: an expanding range of neurological syndromes. Lancet Neurol 2011; 10(8): 721-33.

Lee C, Guo J, Zeng W, Kim S, She J, Cang C, et al. The lysosomal potassium channel TMEM175 adopts a novel tetrameric architecture. Nature 2017; 547(7664): 472-5.

Lee S, Emond MJ, Bamshad MJ, Barnes KC, Rieder MJ, Nickerson DA, et al. Optimal unified approach for rare-variant association testing with application to small-sample case-control whole-exome sequencing studies. Am J Hum Genet 2012; 91(2): 224-37.

Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, Fennell T, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016; 536(7616): 285-91.

Lesage S, Drouet V, Majounie E, Deramecourt V, Jacoupy M, Nicolas A, et al. Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy. Am J Hum Genet 2016; 98(3): 500-13.

Lewthwaite AJ, Lambert TD, Rolfe EB, Olgiati S, Quadri M, Simons EJ, et al. Novel GCH1 variant in Dopa-responsive dystonia and Parkinson's disease. Parkinsonism Relat Disord 2015; 21(4): 394-7.

Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009; 25(14): 1754-60.

Li J, Ruskey JA, Arnulf I, Dauvilliers Y, Hu MTM, Hogl B, et al. Full sequencing and haplotype analysis of MAPT in Parkinson's disease and rapid eye movement sleep behavior disorder. Mov Disord 2018; 33(6): 1016-20.

Liu R, Gao X, Lu Y, Chen H. Meta-analysis of the relationship between Parkinson disease and melanoma. Neurology 2011; 76(23): 2002-9.

Long JZ, Svensson KJ, Bateman LA, Lin H, Kamenecka T, Lokurkar IA, et al. The Secreted Enzyme PM20D1 Regulates Lipidated Amino Acid Uncouplers of Mitochondria. Cell 2016; 166(2): 424-35.

Machiela MJ, Chanock SJ. LDlink: a web-based application for exploring population-specific haplotype structure and linking correlated alleles of possible functional variants. Bioinformatics 2015; 31(21): 3555-7.

MacLeod DA, Rhinn H, Kuwahara T, Zolin A, Di Paolo G, McCabe BD, et al. RAB7L1 interacts with LRRK2 to modify intraneuronal protein sorting and Parkinson's disease risk. Neuron 2013; 77(3): 425-39.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010; 20(9): 1297-303.

Mencacci NE, Isaias IU, Reich MM, Ganos C, Plagnol V, Polke JM, et al. Parkinson's disease in GTP cyclohydrolase 1 mutation carriers. Brain 2014; 137(Pt 9): 2480-92.

Nalls MA, Blauwendraat C, Vallerga CL, Heilbron K, Bandres-Ciga S, Chang D, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. Lancet Neurol 2019; 18(12): 1091-102.

Nalls MA, Pankratz N, Lill CM, Do CB, Hernandez DG, Saad M, et al. Large-scale meta-analysis of genome-wide association data identifies six new risk loci for Parkinson's disease. Nat Genet 2014; 46(9): 989-93.

O'Roak BJ, Vives L, Fu W, Egertson JD, Stanaway IB, Phelps IG, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science 2012; 338(6114): 1619-22.

Paisan-Ruiz C, Jain S, Evans EW, Gilks WP, Simon J, van der Brug M, et al. Cloning of the gene containing mutations that cause PARK8-linked Parkinson's disease. Neuron 2004; 44(4): 595-600.

Park JS, Davis RL, Sue CM. Mitochondrial Dysfunction in Parkinson's Disease: New Mechanistic Insights and Therapeutic Perspectives. Curr Neurol Neurosci Rep 2018; 18(5): 21.

Polymeropoulos MH, Lavedan C, Leroy E, Ide SE, Dehejia A, Dutra A, et al. Mutation in the alpha-synuclein gene identified in families with Parkinson's disease. Science 1997; 276(5321): 2045-7.

Postuma RB, Berg D, Stern M, Poewe W, Olanow CW, Oertel W, et al. MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord 2015; 30(12): 1591-601.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007; 81(3): 559-75.

Ran C, Brodin L, Forsgren L, Westerlund M, Ramezani M, Gellhaar S, et al. Strong association between glucocerebrosidase mutations and Parkinson's disease in Sweden. Neurobiol Aging 2016; 45: 212 e5- e11.

Ross JP, Dupre N, Dauvilliers Y, Strong S, Ambalavanan A, Spiegelman D, et al. Analysis of DNAJC13 mutations in French-Canadian/French cohort of Parkinson's disease. Neurobiol Aging 2016; 45: 212 e13- e17.

Rudakou U, Ouled Amar Bencheikh B, Ruskey JA, Krohn L, Laurent SB, Spiegelman D, et al. Common and rare GCH1 variants are associated with Parkinson's disease. Neurobiol Aging 2019; 73: 231 e1- e6.

Rudakou U, Ruskey JA, Krohn L, Laurent SB, Spiegelman D, Greenbaum L, et al. Analysis of common and rare VPS13C variants in late-onset Parkinson disease. Neurol Genet 2020; 6(1): 385.

Ruskey JA, Greenbaum L, Ronciere L, Alam A, Spiegelman D, Liong C, et al. Increased yield of full GBA sequencing in Ashkenazi Jews with Parkinson's disease. Eur J Med Genet 2019; 62(1): 65-9.

Satake W, Nakabayashi Y, Mizuta I, Hirota Y, Ito C, Kubo M, et al. Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. Nat Genet 2009; 41(12): 1303-7.

Schormair B, Kemlink D, Mollenhauer B, Fiala O, Machetanz G, Roth J, et al. Diagnostic exome sequencing in early-onset Parkinson's disease confirms VPS13C as a rare cause of autosomal-recessive Parkinson's disease. Clin Genet 2018; 93(3): 603-12.

Sekiyama K, Takamatsu Y, Waragai M, Hashimoto M. Role of genomics in translational research for Parkinson's disease. Biochem Biophys Res Commun 2014; 452(2): 226-35.

Sidransky E, Nalls MA, Aasly JO, Aharon-Peretz J, Annesi G, Barbosa ER, et al. Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. N Engl J Med 2009; 361(17): 1651-61.

Simon-Sanchez J, Schulte C, Bras JM, Sharma M, Gibbs JR, Berg D, et al. Genome-wide association study reveals genetic risk underlying Parkinson's disease. Nat Genet 2009; 41(12): 1308-12.

Soldner F, Stelzer Y, Shivalila CS, Abraham BJ, Latourelle JC, Barrasa MI, et al. Parkinson-associated risk variant in distal enhancer of alpha-synuclein modulates target gene expression. Nature 2016; 533(7601): 95-9.

Sreetama SC, Takano T, Nedergaard M, Simon SM, Jaiswal JK. Injured astrocytes are repaired by Synaptotagmin XI-regulated lysosome exocytosis. Cell Death Differ 2016; 23(4): 596-607.

Tarazi FI, Sahli ZT, Wolny M, Mousa SA. Emerging therapies for Parkinson's disease: from bench to bedside. Pharmacol Ther 2014; 144(2): 123-33.

Taya M, Hammes SR. Glycoprotein Non-Metastatic Melanoma Protein B (GPNMB) and Cancer: A Novel Potential Therapeutic Target. Steroids 2018; 133: 102-7.

Tomihari M, Hwang SH, Chung JS, Cruz PD, Jr., Ariizumi K. Gpnmb is a melanosome-associated glycoprotein that contributes to melanocyte/keratinocyte adhesion in a RGD-dependent fashion. Exp Dermatol 2009; 18(7): 586-95.

Wang C, Kang X, Zhou L, Chai Z, Wu Q, Huang R, et al. Synaptotagmin-11 is a critical mediator of parkin-linked neurotoxicity and Parkinson's disease-like pathology. Nat Commun 2018; 9(1): 81.

Wang C, Wang Y, Hu M, Chai Z, Wu Q, Huang R, et al. Synaptotagmin-11 inhibits clathrin-mediated and bulk endocytosis. EMBO Rep 2016; 17(1): 47-63.

Wang G, van der Walt JM, Mayhew G, Li YJ, Zuchner S, Scott WK, et al. Variation in the miRNA-433 binding site of FGF20 confers risk for Parkinson disease by overexpression of alpha-synuclein. Am J Hum Genet 2008; 82(2): 283-9.

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010; 38(16): e164.
Wider C, Dachsel JC, Soto AI, Heckman MG, Diehl NN, Yue M, et al. FGF20 and Parkinson's disease: no evidence of association or pathogenicity via alpha-synuclein expression. Mov Disord 2009; 24(3): 455-9.
Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010; 26(17): 2190-1.
Xie Z, Long J, Liu J, Chai Z, Kang X, Wang C. Molecular Mechanisms for the Coupling of Endocytosis to Exocytosis in Neurons. Front Mol Neurosci 2017; 10: 47.
Xu Q, Li K, Sun Q, Ding D, Zhao Y, Yang N, et al. Rare GCH1 heterozygous variants contributing to Parkinson's disease. Brain 2017; 140(7): e41.
Yoshino H, Nishioka K, Li Y, Oji Y, Oyama G, Hatano T, et al. GCH1 mutations in dopa-responsive dystonia and Parkinson's disease. J Neurol 2018; 265(8): 1860-70.

Table 1: Results of SKAT-O analysis in the merged cohorts at 30x and 50x minimum depth of coverage

	50x						30x					
Gene	**CADD**	**ENCODE**	**Funct**	**LOF**	**Nonsyn**	**All**	**CADD**	**ENCODE**	**Funct**	**LOF**	**Nonsyn**	**All**
SYT11	0.45(6)	0.89(6)	0.66(14)	NA	0.40(8)	**5.35E-06(34)****	0.31(13)	0.75(32)	0.75(48)	NA	0.22(16)	0.02(86)*
RAB25	0.20(1)	NA	0.20(1)	NA	0.20(1)	0.55(6)	0.34(6)	0.55(6)	0.52(14)	0.48(4)	0.40(4)	0.31(27)
RAB29	0.51(3)	0.04(11)*	0.03(14)*	NA	0.48(4)	0.09(23)	0.90(3)	0.12(19)	0.13(22)	NA	0.91(4)	0.25(41)
SLC41A1	0.13(5)	NA	0.13(5)	NA	0.13(5)	0.25(30)	0.93(18)	0.86(2)	0.97(22)	0.83(4)	0.65(16)	0.56(109)
PM20D1	0.38(9)	0.97(11)	0.22(23)	0.07(5)	0.11(12)	0.73(58)	0.27(15)	0.97(11)	0.13(32)	0.14(6)	0.09(20)	0.54(80)
SIPA1L2	0.24(33)	0.06(5)	0.18(46)	NA	0.25(42)	0.003(121)*	0.56(38)	0.06(5)	0.36(56)	0.06(3)	0.59(49)	0.01(142)*
TMEM163	0.09(4)	NA	0.09(4)	NA	0.09(4)	0.04(28)*	0.12(5)	NA	0.12(5)	NA	0.12(5)	0.03(44)*
ACMSD	0.26(14)	NA	0.40(18)	0.08(2)	0.30(16)	0.15(29)	0.65(19)	NA	0.68(23)	0.04(3)*	0.58(20)	0.32(45)
STK39	0.42(3)	NA	0.11(4)	NA	0.11(4)	0.11(48)	0.30(8)	NA	0.20(9)	NA	0.20(9)	0.31(70)
MCCC1	0.30(19)	0.83(4)	0.42(24)	0.11(4)	0.43(19)	0.20(50)	0.39(29)	0.82(33)	0.58(63)	0.03(6)*	0.53(30)	0.43(99)
LAMP3	0.76(9)	0.56(13)	0.65(21)	0.34(2)	0.84(12)	0.08(28)	0.43(18)	0.65(40)	0.78(55)	0.34(2)	0.63(24)	0.17(64)
GAK	NA	NA	NA	NA	NA	NA	0.79(14)	0.1(5)	0.32(19)	0.70(1)	0.88(16)	0.24(63)
TMEM175	NA	NA	NA	NA	NA	NA	0.66(4)	0.25(4)	0.34(8)	0.34(1)	0.56(3)	0.59(15)
BST1	0.02(12)*	0.33(9)	0.01(23)*	0.76(2)	0.004(14)*	0.06(46)	0.03(14)*	0.21(9)	0.007(25)*	0.76(2)	0.007(16)*	0.11(63)
SCARB2	0.83(8)	0.87(7)	0.05(17)*	0.27(2)	0.05(10)*	0.003(62)*	0.82(10)	0.76(9)	0.08(20)	0.23(3)	0.07(11)	0.01(86)*
SNCA	0.11(2)	0.99(9)	0.79(11)	NA	0.18(3)	0.55(34)	0.11(2)	0.24(29)	0.21(31)	NA	0.18(3)	0.34(57)
GPNMB	0.05(27)*	0.28(26)	0.08(53)	0.97(8)	0.03(27)*	0.07(80)	0.28(30)	0.20(52)	0.12(80)	1(9)	0.09(31)	0.14(118)
FGF20	NA	NA	NA	NA	NA	0.06(8)	0.76(2)	**0.0008(4)****	**0.0005(5)****	NA	0.76(2)	**0.00017(15)****
ITGA8	0.70(26)	0.34(2)	0.85(30)	0.34(2)	0.93(26)	0.26(103)	0.65(39)	0.39(6)	0.27(52)	0.34(2)	0.27(44)	0.06(144)
INPP5F	0.13(5)	0.91(12)	0.59(16)	NA	0.25(6)	0.02(46)*	0.10(27)	0.23(23)	0.10(62)	0.61(5)	0.08(38)	0.003(152)*
LRRK2	0.07(35)	0.07(13)	0.11(52)	0.68(3)	0.08(37)	0.29(106)	0.03(53)*	0.80(17)	0.16(73)	0.83(4)	0.06(55)	0.31(142)
CCDC62	0.12(17)	NA	0.11(25)	0.46(3)	0.11(22)	0.09(59)	0.04(21)*	0.30(20)	0.23(50)	0.17(3)	0.04(28)*	0.17(91)
HIP1R	NA	NA	NA	NA	NA	NA	0.12(5)	0.14(14)	0.14(14)	0.70(1)	0.16(4)	0.08(23)
GCH1	0.02(4)*	NA	0.02(4)*	NA	0.02(4)*	0.64(13)	0.02(8)*	0.63(6)	0.43(14)	NA	0.02(8)*	0.005(76)*
VPS13C	0.38(34)	0.03(5)*	0.17(48)	0.17(3)	0.31(41)	0.32(159)	0.14(33)	0.03(5)*	0.08(42)	0.17(3)	0.20(35)	0.21(180)
SETD1A	0.01(3)*	NA	0.22(4)	NA	0.22(4)	0.11(13)	0.004(15)*	0.73(17)	0.04(32)*	NA	0.03(18)*	0.14(61)
STX1B	NA	NA	NA	NA	NA	NA	0.40(2)	NA	0.20(3)	NA	0.20(3)	0.07(30)
SREBF1	NA	NA	NA	NA	NA	NA	0.78(1)	NA	0.40(3)	NA	0.40(3)	0.46(4)
MAPT	0.60(3)	NA	0.60(3)	0.71(1)	0.65(2)	0.93(7)	0.27(13)	0.23(5)	0.45(24)	0.70(1)	0.07(19)	0.01(106)*
RIT2	0.71(1)	0.21(1)	0.76(2)	NA	0.71(1)	0.51(7)	0.48(4)	0.09(9)	0.11(14)	0.71(1)	0.61(5)	0.15(23)
DDRGK1	0.18(5)	NA	0.18(5)	NA	0.18(5)	0.09(13)	0.16(11)	NA	0.07(13)	0.17(3)	0.21(10)	0.44(53)
USP25	0.47(17)	NA	0.47(17)	0.76(2)	0.44(15)	0.37(63)	0.75(33)	0.22(15)	0.31(46)	0.12(6)	0.81(28)	0.45(128)

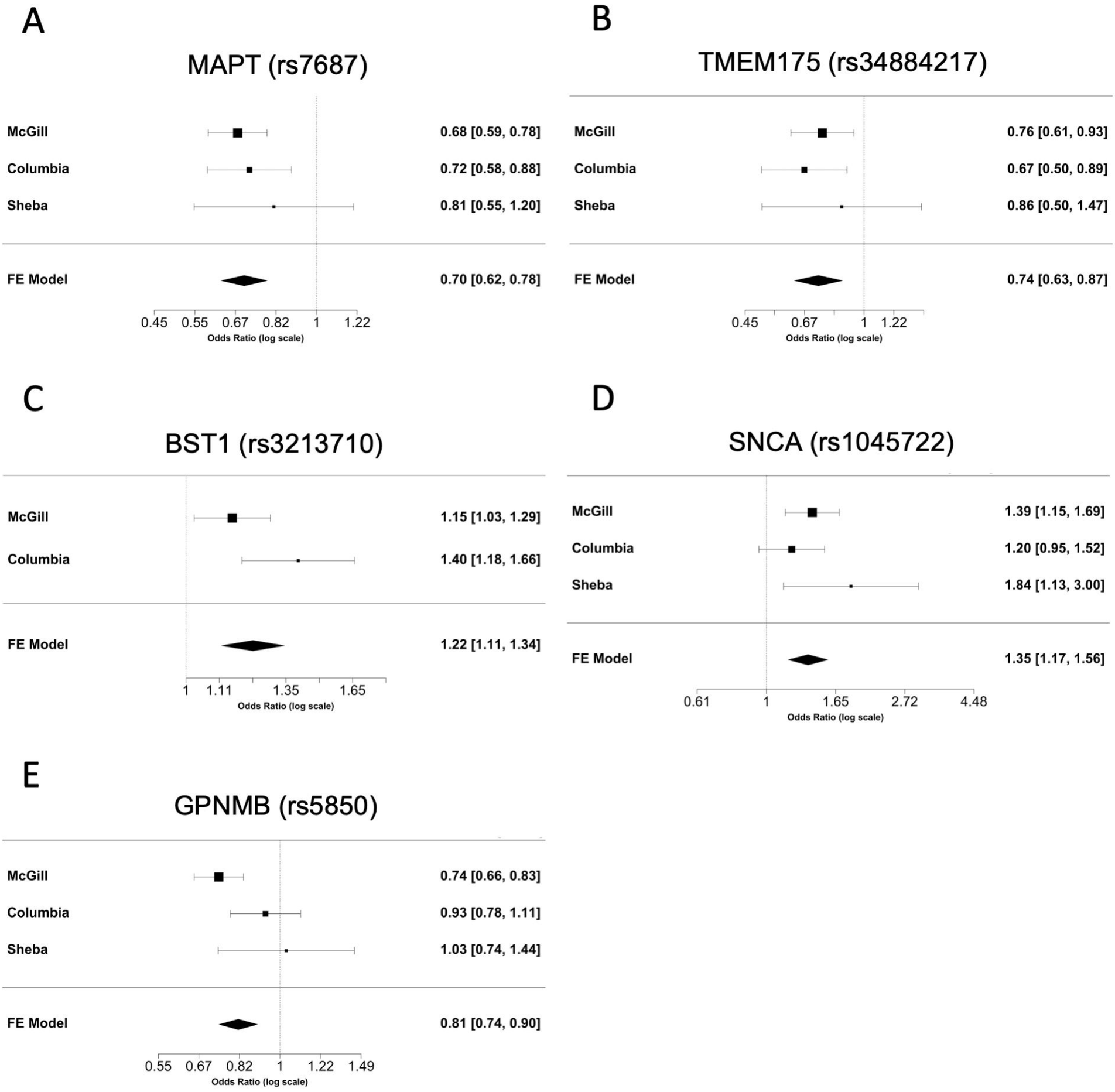
The table shows *p*-values of the SKAT-O analysis in the specified subgroup; the number of tested variants is indicated in the parentheses; *result with nominal significance ($p<0.05$); **in bold - result significant after Bonferroni correction for the number of genes
CADD subgroup: CADD score ≥ 12.37; ENCODE subgroup: variants in regions with predicted regulatory function; Funct subgroup: potentially functional variants (Methods); LOF subgroup: loss-of-function variants; NA: not applicable - no variants tested; Nonsyn subgroup: non-synonymous; SKAT-O: Sequence Optimized Kernel Association Test

Table 2 The common variants with statistically significant association with Parkinson's Disease in the meta-analysis of three cohorts

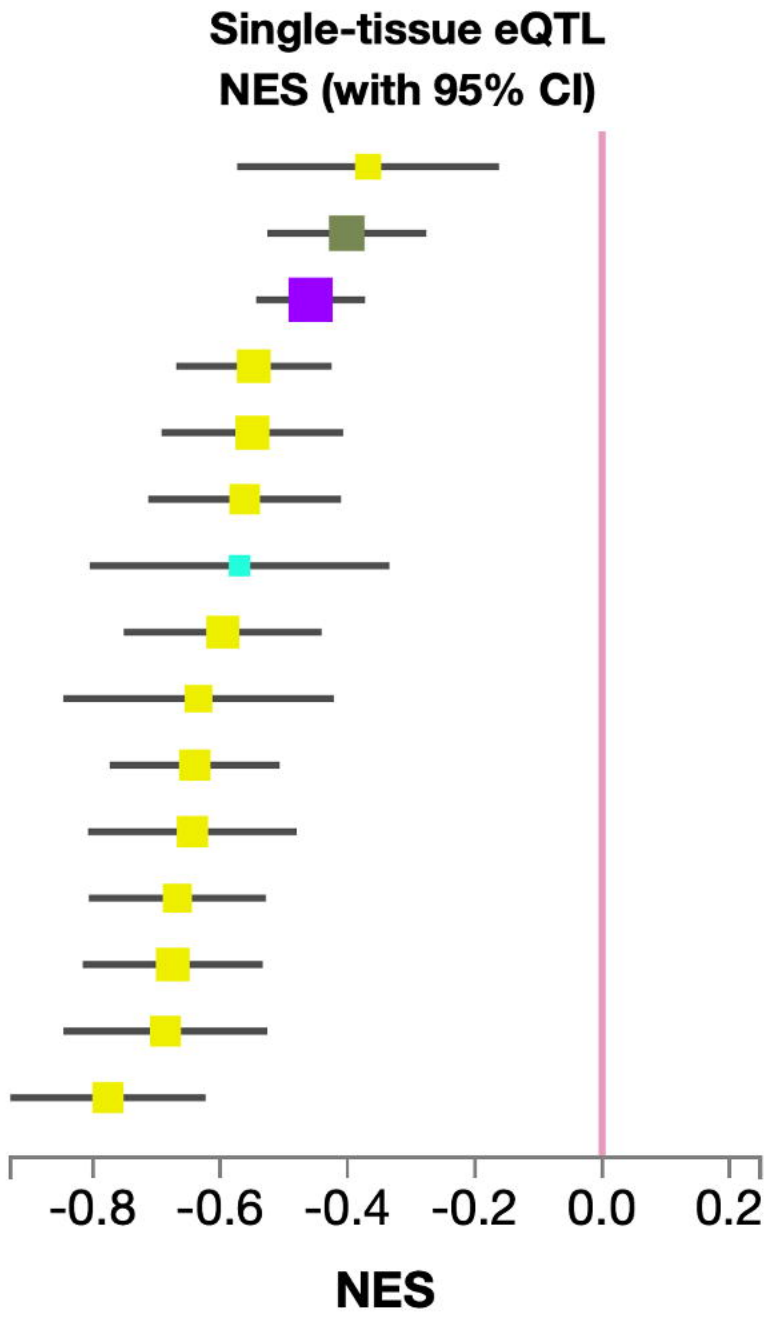
Gene	Chr	Position	Ref Allele	Mutant Allele	SNP	D'	Rsq	CADD	Detailed annotation of the variant	Meta OR(95%CI)	Meta *p*-value
MAPT	17	44049329	C	T	rs75242405	0.99	0.98		NM_016835:intronic	0.72(0.64-0.81)	1.05E-07
MAPT	17	44051846	A	G	rs1800547	0.99	0.98		NM_016835:intronic	0.74(0.62-0.89)	1.24E-03
MAPT	17	44061023	G	A	rs62063786	0.99	0.98	7.7	NM_016835:exon6:c.G853A:p.D285N	0.71(0.63-0.79)	2.42E-09
MAPT	17	44061036	T	C	rs62063787	0.99	0.98	0.001	NM_016835:exon6:c.T866C:p.V289A	0.71(0.64-0.80)	2.66E-09
MAPT	17	44061278	C	T	rs17651549	0.99	0.98	34	NM_016835:exon6:c.C1108T:p.R370W	0.71(0.64-0.80)	3.35E-09
MAPT	17	44067400	T	C	rs10445337	0.99	0.98	9.93	NM_016835:exon8:c.T1339C:p.S447P	0.74(0.62-0.89)	1.04E-03
MAPT	17	44067508	A	G	rs79447161	0.99	0.98		NM_016835:intronic	0.71(0.64-0.80)	6.01E-09
MAPT	17	44071294	T	C	rs62063845	0.99	0.98		NM_016835:intronic	0.72(0.64-0.80)	6.82E-09
MAPT	17	44101563	T	C	rs9468	0.99	0.98		NM_016835:UTR3:c.*26T>C	0.72(0.64-0.81)	8.35E-09
MAPT	17	44102604	T	C	rs1052587	0.99	0.98		NM_016835:UTR3:c.*1067T>C	0.72(0.65-0.81)	9.98E-09
MAPT	17	44102638	A	G	rs1052590	0.99	0.98		NM_016835:UTR3:c.*1101A>G	0.72(0.64-0.81)	9.51E-09
MAPT	17	44102865	A	C	rs17574040	0.99	0.98		NM_016835:UTR3:c.*1328A>C	0.72(0.64-0.80)	4.79E-09
MAPT*	17	44103296	T	C	rs7687	0.99	0.98		NM_016835:UTR3:c.*1759T>C	0.70(0.62-0.78)	8.28E-10
MAPT	17	44103616	C	T	rs17652748	0.99	0.98		NM_016835:UTR3:c.*2079C>T	0.71(0.64-0.80)	3.81E-09
MAPT	17	44103825	T	C	rs75010486	0.99	0.98		NM_016835:UTR3:c.*2288T>C	0.71(0.64-0.80)	4.02E-09
MAPT	17	44104343	A	C	rs2158257	0.99	0.98		NM_016835:UTR3:c.*2806A>C	0.73(0.65-0.81)	2.36E-08
MAPT	17	44104410	TCTC	T	rs568475466	0.99	0.98		NM_016835:UTR3:c.*2874_*2876delCTC	0.72(0.64-0.80)	7.73E-09
TMEM175	4	944210	A	C	rs34884217	1	0.01	0.2	NM_032326:exon4:c.A194C:p.Q65P	0.74(0.63-0.87)	2.11E-04
BST1	4	15717321	G	A	rs3213710	0.94	0.79		NM_004334:intronic	1.22(1.11-1.34)	3.14E-05
SNCA*	4	90645671	T	A	rs1045722	1	0.12		NM_000345:UTR3:c.*2108A>T	1.35(1.17-1.56)	4.08E-05
SNCA	4	90645674	C	T	rs3857053	1	0.12		NM_000345:UTR3:c.*2105G>A	1.35(1.17-1.56)	4.56E-05
GPNMB	7	23300049	A	C	rs199351	0.95	0.89		NM_002510:intronic	0.82(0.75-0.90)	2.70E-05
GPNMB	7	23307634	ATGGG	A	rs2307783	0.95	0.89		NM_002510:intronic	0.83(0.76-0.91)	7.55E-05
GPNMB*	7	23314547	C	T	rs5850	0.95	0.89		NM_002510:UTR3:c.*704C>T	0.81(0.74-0.90)	2.19E-05

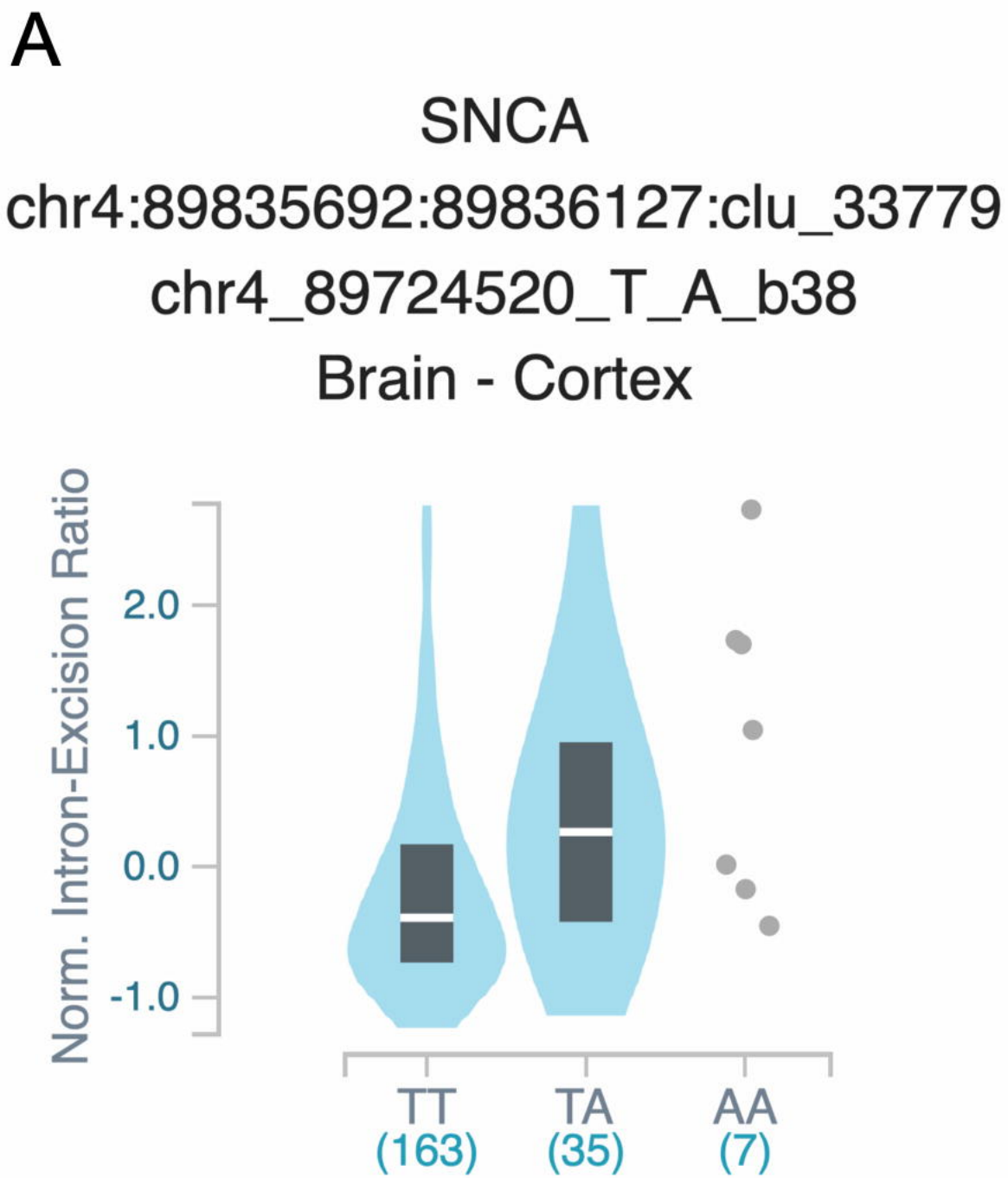
*tagging SNP for the haplotype

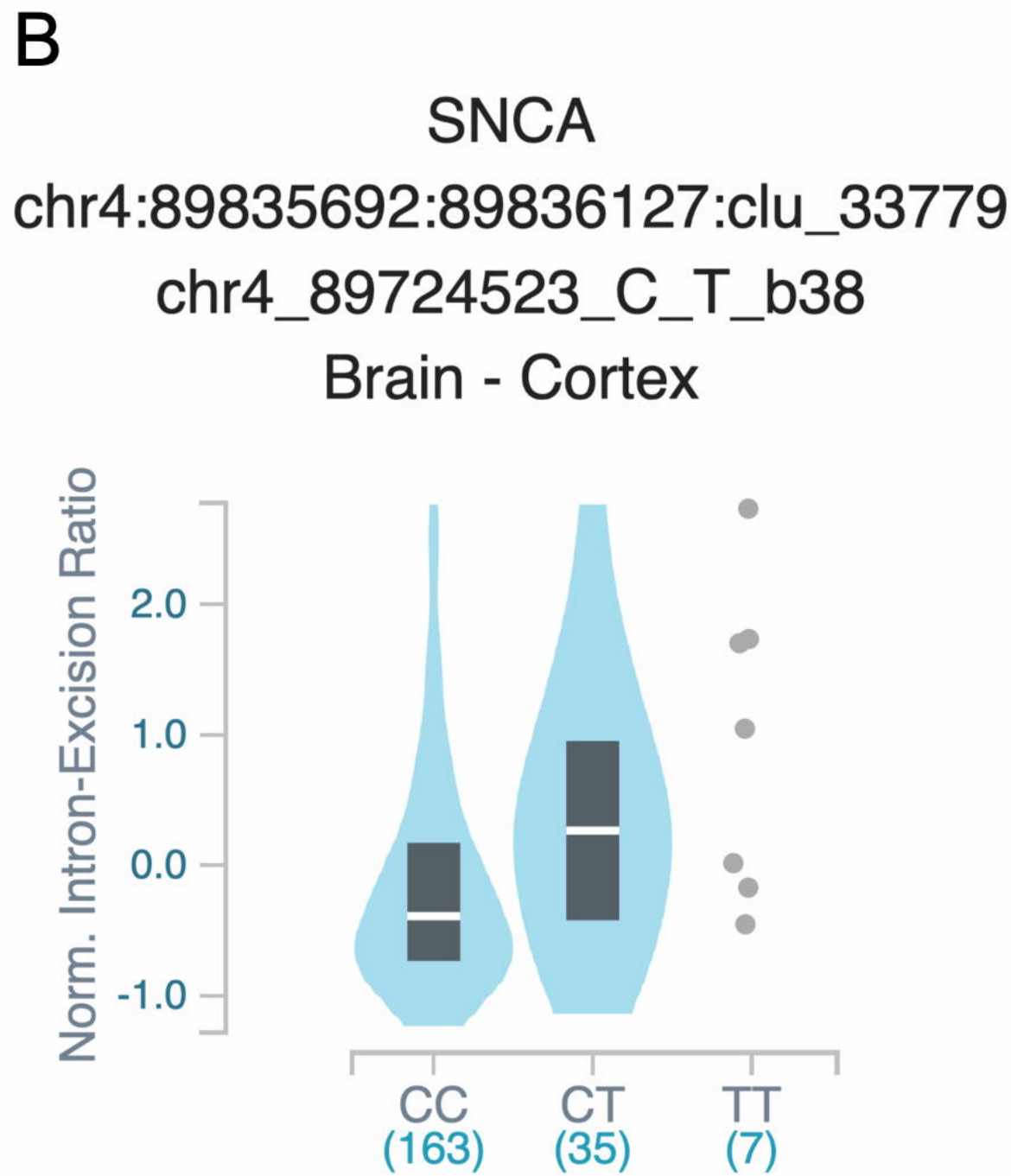
Chr: chromosome; CI: confidence interval; OR: odds ratio; Ref: reference; Rsq: r-square; SNP: single nucleotide polymorphism


A
MAPT (rs7687)
McGill
0.68 [0.59, 0.78]
Columbia
0.72 [0.58, 0.88]
Sheba
0.81 [0.55, 1.20]
FE Model
0.70 [0.62, 0.78]
0.45 0.55 0.67 0.82 1 1.22
Odds Ratio (log scale)
B
TMEM175 (rs34884217)
McGill
0.76 [0.61, 0.93]
Columbia
0.67 [0.50, 0.89]
Sheba
0.86 [0.50, 1.47]
FE Model
0.74 [0.63, 0.87]
0.45 0.67 1 1.22
Odds Ratio (log scale)
C
BST1 (rs3213710)
McGill
1.15 [1.03, 1.29]
Columbia
1.40 [1.18, 1.66]
FE Model
1.22 [1.11, 1.34]
1 1.11 1.35 1.65
Odds Ratio (log scale)
D
SNCA (rs1045722)
McGill
1.39 [1.15, 1.69]
Columbia
1.20 [0.95, 1.52]
Sheba
1.84 [1.13, 3.00]
FE Model
1.35 [1.17, 1.56]
0.61 1 1.65 2.72 4.48
Odds Ratio (log scale)
E
GPNMB (rs5850)
McGill
0.74 [0.66, 0.83]
Columbia
0.93 [0.78, 1.11]
Sheba
1.03 [0.74, 1.44]
FE Model
0.81 [0.74, 0.90]
0.55 0.67 0.82 1 1.22 1.49
Odds Ratio (log scale)

Tissue	Samples	NES	p-value	m-value
Brain - Substantia nigra	114	-0.368	7.1e-4	1.00
Spleen	227	-0.401	2.1e-9	1.00
Heart - Atrial Appendage	372	-0.458	4.1e-22	1.00
Brain - Cortex	205	-0.547	1.9e-15	1.00
Brain - Cerebellum	209	-0.549	2.5e-12	1.00
Brain - Hippocampus	165	-0.562	3.3e-11	1.00
Kidney - Cortex	73	-0.570	1.9e-5	1.00
Brain - Caudate (basal ganglia)	194	-0.596	4.3e-12	1.00
Brain - Amygdala	129	-0.634	5.7e-8	1.00
Brain - Frontal Cortex (BA9)	175	-0.640	1.9e-16	1.00
Brain - Cerebellar Hemisphere	175	-0.644	2.7e-12	1.00
Brain - Anterior cingulate cortex (BA24)	147	-0.667	4.0e-16	1.00
Brain - Nucleus accumbens (basal ganglia)	202	-0.675	6.5e-17	1.00
Brain - Hypothalamus	170	-0.686	6.8e-14	1.00
Brain - Putamen (basal ganglia)	170	-0.776	1.3e-17	1.00




A
SNCA
chr4:89835692:89836127:clu_33779
chr4_89724520_T_A_b38
Brain - Cortex
Norm. Intron-Excision Ratio
2.0
1.0
0.0
-1.0
TT
(163)
TA
(35)
AA
(7)


B
SNCA
chr4:89835692:89836127:clu_33779
chr4_89724523_C_T_b38
Brain - Cortex
Norm. Intron-Excision Ratio
2.0
1.0
0.0
-1.0
CC
(163)
CT
(35)
TT
(7)