bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Analysis of common and rare *VPS13C* variants in late onset Parkinson disease

Uladzislau Rudakou, BSc[1,2], Jennifer A. Ruskey, MSc[2,3], Lynne Krohn, BSc[1,2], Sandra B. Laurent, BSc[2,3], Dan Spiegelman, MSc[2,3], Lior Greenbaum, MD, PhD[4,5,6], Gilad Yahalom, MD[6,7,8], Alex Desautels, MD, PhD[9,10,] Jacques Y. Montplaisir, MD, PhD[9,11], Stanley Fahn, MD[12], Cheryl H. Waters, MD[12], Oren Levy, MD, PhD[12] , Caitlin M. Kehoe, MA[12], Sushma Narayan, MS[12], Yves Dauvilliers, MD, PhD[13], Nicolas Dupré, MD, FRCP[14,15], Sharon Hassin-Baer, MD[5,6,7,8], Roy N. Alcalay, MD, MS[11,16], Guy A. Rouleau, MD, PhD, FRCPC, FRSC[1,2,3], Edward A. Fon, MD, FRCPC[2,3], 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. The Danek Gertner Institute of Human Genetics, Sheba Medical Center, Tel Hashomer, Ramat Gan , Israel
5. The Joseph Sagol Neuroscience Center, Sheba Medical Center, Tel Hashomer, Ramat Gan, Israel
6. Sackler school of medicine, Tel-Aviv University, Tel-Aviv 69978, Israel
7. Department of Neurology, Sheba Medical Center, Tel Hashomer, 52621 Ramat-Gan, Israel
8. Movement Disorders Institute, Sheba Medical Center, Tel Hashomer, 52621 Ramat Gan, Israel
9. Centre d'Études Avancées en Médecine du Sommeil, Hôpital du Sacré-Cœur de Montréal, Montréal, QC, H4J 1C5, Canada
10. Department of Neurosciences, Université de Montréal, Montréal, QC, H3T 1J4, Canada
11. Department of Psychiatry, Université de Montréal, Montréal, QC, H3T 1J4, Canada
12. Department of Neurology, College of Physicians and Surgeons, Columbia University Medical Center, New York, NY 10032, USA
13. National Reference Center for Narcolepsy, Sleep Unit, Department of Neurology, Gui-de-Chauliac Hospital, CHU Montpellier, University of Montpellier, Inserm U1061, Montpellier, France
14. Division of Neurosciences, CHU de Québec, Université Laval, Québec City, QC, G1V 0A6, Canada
15. Department of Medicine, Faculty of Medicine, Université Laval, Québec City, QC, G1V 0A6, Canada;
16. 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

bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Supplemental Data: Table e-1 – e-6

Word count for the paper: 2322
Word count for the abstract: 236
Character count for the title: 96
This paper contains 21 references, 2 tables, and 1 figure

Keywords: Parkinson's disease; genetic variants; VPS13C; association study

Publication history: This manuscript was previously published in bioRxiv: doi: https://doi.org/xxxx

Study funding:

This work was financially supported by the Michael J. Fox Foundation, the Canadian Consortium on Neurodegeneration in Aging (CCNA) and the Canadian Glycomics Network (GlycoNet). This research was also undertaken thanks in part to funding from the Canada First Research Excellence Fund (CFREF), awarded to McGill University for the Healthy Brains for Healthy Lives (HBHL) program. The Columbia University cohort is supported by the Parkinson's Foundation, the National Institutes of Health [K02NS080915, and UL1 TR000040] and the Brookdale Foundation.

Disclosures:

Uladzislau Rudakou – nothing to disclose

Jennifer A. Ruskey - nothing to disclose

Lynne Krohn - nothing to disclose

Sandra B. Laurent - nothing to disclose

Dan Spiegelman - nothing to disclose

Lior Greenbaum - nothing to disclose

Gilad Yahalom - received financial support by Abbvie (unrelated to the current study) and by Allergan and Abbvie for sponsoring training

bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Alex Desautels - Received research grants from Novartis pharma, Flamel Ireland Limited and Biron (all unrelated to the current study) as well as honoraria from speaking engagements from UCB (unrelated to the current study).

Jacques Y. Montplaisir – nothing to disclose

Stanley Fahn – received consulting fees/honoraria for board membership (unrelated to the current study) from Retrophin Inc., Sun Pharma Advanced Research Co., LTD and Kashiv Pharma.

Cheryl Waters - received consulting fees/honoraria (unrelated to the current study) from US World Meds, Acadia, Lundbeck, Cynapsus, Acorda.

Oren Levy – received consulting fees from Schlesinger Associates, and honoraria as board member in Abbvie (all unrelated to the current study).

Caitlin Kehoe – nothing to disclose

Sushma Narayan - nothing to disclose

Yves Dauvilliers - nothing to disclose

Nicolas Dupré – received consultancy fees (unrelated to the current work) from Actelion Pharmaceuticals

Sharon Hassin-Baer – Received consulting fees from Actelion Pharmaceuticals Ltd., Abbvie Israel, Robotico Ltd., Medtronic Israel, Medison Pharma Israel (unrelated to the current study)

Roy N. Alcalay – received consultation fees (unrelated to the current study) from Biogen, Denali, Genzyme/Sanofi and Roche

Guy A. Rouleau - nothing to disclose

Edward A. Fon - Received consulting fees (unrelated to the current study) from Inception Sciences

Ziv Gan-Or – Received consulting fees (unrelated to the current study) from Denali, Inception Sciences, Idorsia, Lysosomal Therapeutics Inc. and Prevail Therapeutics.

bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Abstract

Objective: *VPS13C* is within a risk locus for Parkinson Disease (PD), and bi-allelic *VPS13C* mutations are a rare cause of early onset, rapidly progressive PD. In the current study we aimed to study the role of coding *VPS13C* variants in a large cohort of late-onset PD (LOPD) patients.

Methods: *VPS13C* and its untranslated regions were sequenced using targeted next-generation sequencing in 1,567 PD patients and 1,667 controls from 3 cohorts. Association tests of rare homozygous and compound heterozygous mutations and burden tests for rare heterozygous mutations were performed. Common variants were analyzed using logistic regression adjusted for age and sex in each of the cohorts, followed by a meta-analysis.

Results: No bi-allelic carriers of rare *VPS13C* variants were found among patients and two carriers of compound heterozygous variants were found in two controls. There was no statistically significant burden of rare (MAF<1%) or very rare (MAF<0.1%) coding *VPS13C* variants in PD. A *VPS13C* haplotype including the p.R153H-p.I398I-p.I1132V-p.Q2376Q variants was nominally associated with a reduced risk for PD (meta-analysis of the tagging SNP p.I1132V (OR=0.48, 95%CI=0.28-0.82, p=0.0052). This haplotype was not in linkage disequilibrium (LD) with the known genome-wide association study (GWAS) top hit.

Conclusions: Our results do not support a role for rare heterozygous or bi-allelic *VPS13C* mutations in LOPD. Additional genetic replication and functional studies are needed to examine the role of the haplotype identified here associated with reduced risk for PD.

Keywords: Parkinson's disease; genetic variants; VPS13C; association study

bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Introduction

The vacuolar protein sorting 13C (*VPS13C*) gene is located within a risk locus for Parkinson Disease (PD), reported in large genome-wide association studies (GWAS) of European population [1, 2]. The SNP reported in the GWAS (rs2414739, chr15.hg19:g.61994134G>A) was also studied in several Asian populations and in Iranians, with conflicting results [3-7], possibly due to ethnicity-related differences.

Subsequently, homozygous and compound-heterozygous *VPS13C* mutations were identified as a rare cause of early onset PD (EOPD) characterized by rapid progression and early cognitive dysfunction. It was demonstrated that *VPS13C* is partially localized at the mitochondrial membrane, and its silencing led to mitochondrial dysfunction and increased PINK1/Parkin-dependent mitophagy [8]. A follow-up study in 80 EOPD patients identified an additional patient with compound-heterozygous mutations with similar clinical features to the previously reported patients [9]. In another recent study, a homozygous deletion in *VPS13C* was reported to be the probable cause of an early onset parkinsonism in one patient [10]. Thus far, full sequencing studies of *VPS13C* have not been reported in late onset PD (LOPD).

To further study the potential role of *VPS13C* variants in PD, we sequenced its coding and regulatory regions using targeted next-generation sequencing in three cohorts of PD (predominantly LOPD) and in controls. We examined the association of common, rare and bi-allelic *VPS13C* variants on the risk for PD. We further tested whether any coding variant or variants in the untranslated regions of *VPS13C* are in linkage disequilibrium (LD) with the top *VPS13C*-associated GWAS hit, to determine if any of these variants can explain the GWAS association of this locus.

Methods

Study population

Three cohorts, with a total of 1,567 unrelated PD patients and 1,667 controls, were included in this study, detailed in Table 1. First cohort was composed of French and French-Canadian participants recruited in Quebec (Canada) and in France. The second cohort was recruited in New York - Columbia University,

bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

and was previously described[11]. The majority of participants from New York are of European descent and 38% are Ashkenazi Jewish (AJ). The third cohort was recruited in Israel (Sheba Medical Center) and all participants included in this study from the Israeli cohort are AJ. All samples were consecutively recruited. As detailed below, due to the differences in age and sex (Table 1), statistical analysis was adjusted and included age and sex as co-variates. To account for different ethnicities in the New York cohort, an ethnicity covariate was also introduced in this cohort (GWAS data was not available for this cohort, therefore the reported ethnicity was used and not principal components). All three cohorts were sequenced in the same lab (McGill University), following the same protocol. All PD patients were diagnosed by movement disorder specialists according to the UK brain bank criteria [12], without excluding patients with family history of PD.

Compliance with Ethical Standards

The institutional review board (McGill University Health Center Research Ethics Board - MUHC REB) approved the study protocols (reference number IRB00010120). Informed consent was obtained from all individual participants before entering the study.

DNA extraction and *VPS13C* sequencing

DNA was extracted using a standard salting out protocol. The coding sequence and regulatory regions of *VPS13C* were targeted using molecular inversion probes (MIPs), that were designed as previously described [13]. MIPs were selected based on their predicted coverage, quality and overlap. All MIPs used to sequence *VPS13C* in the present study are included in Table S-1. Targeted DNA capture and amplification was done as previously described [14], and the full protocol is available upon request. The library was sequenced using Illumina HiSeq 2500 platform at the McGill University and Genome Quebec Innovation Centre. Reads were mapped to the human reference genome (hg19) with Burrows-Wheeler Aligner [15]. Genome Analysis Toolkit (GATK, v3.8) was used for post-alignment quality control and variant calling [16] , and ANNOVAR was used for annotation [17]. Data on the frequency of each *VPS13C*

bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

variant were extracted from the public database Genome Aggregation Database (GnomAD) [18]. Validation of the tagging variant p.I1132V was performed using Sanger sequencing, with the following primers: forward 5' – CCGGGAAGGTAATGACAAAA – 3', reverse 5' – CCCCTGATTGAAAAGTCACA– 3'

Quality control

During quality control (QC) filtration using PLINK software v1.9 [19], SNPs with genotyping rate lower than 90% were excluded. Genotyping rate cut-off for individuals was 90%, and individuals with a lower genotyping rate were excluded. SNPs that deviated from Hardy-Weinberg equilibrium set at p=0.001 threshold 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. After these QC steps, cohort composition was as described in Table 1. To be included in the analysis, minimum quality score (GQ) was set to 30. Rare variants (minor allele frequency, MAF<0.01 or 0.001) had to have a minimal coverage of >50x to be included, and common variants had to have a minimal coverage of >15x to be included in the analysis.

Statistical Analysis

The association between common *VPS13C* variants and PD was examined by using 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, and AJ ancestry as an additional covariate in the New York cohort. To analyze rare variants (MAF < 0.01) and very rare variants (MAF < 0.001), an optimised sequence Kernel association test (SKAT-O, R package) was performed [20]. The effects of SNP genotypes on the age at onset (AAO) was tested using analysis of variance (ANOVA; in R software). Meta-analysis of common variants in the three cohorts was performed using Metafor Package in R software [21]. Linkage disequilibrium in our data was examined by PLINK v1.9 and LD between discovered SNPs and the GWAS top hit rs2414739 was tested using LDlink application, selecting all non-Finish Europeans [22].

bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Data availability statement

Anonymized data is available upon request by any qualified investigator.

Results

Rare *VPS13C* mutations and homozygous or compound heterozygous *VPS13C* mutations are not associated with late onset PD.

The average coverage of *VPS13C* with the MIPs was 94% of nucleotides covered at > 10x, and 90% covered at > 50x. There were no differences in coverage between the cohorts and between patients and controls. Table S-2 details the rare variants that are either nonsynonymous or potentially affect a splicing site (n=60) and Table S-3 details the common variants in coding regions (n=18) identified in the three cohorts.

In order to examine whether rare homozygous or compound heterozygous *VPS13C* mutations may cause LOPD, and since patients with *VPS13C* bi-allelic mutations are very rare, we included in this analysis only rare mutations with allele frequency < 0.001. Only two carriers of compound heterozygous mutations were identified, and both were controls, suggesting that bi-allelic mutations are not common in LOPD (Table S-6).

To further study a potential role for rare (allele frequency < 0.01) or very rare (allele frequency < 0.001) *VPS13C* nonsynonymous or splice variants in LOPD, a SKAT-O was performed. There was no association between rare variants (French/French Canadian cohort p=0.44, New York cohort p=0.34, Israel cohort p=0.91) or very rare variants (French/French Canadian cohort p=0.17, New York cohort p=0.85, Israel cohort p=0.89) and PD.

bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

A *VPS13C* haplotype including the p.R153H-p.I398I-p.I1132V-p.Q2376Q coding variants is nominally associated with reduced risk for PD.

To test whether common coding variants in *VPS13C* are associated with LOPD, logistic regression models adjusted for age and sex were performed, and additional adjustment for ethnicity was included in the New York cohort (see methods). A nominal association was observed in four variants (p.R153H [rs12595158, chr15.hg19:g.62316035C>T], p.I398I [rs9635356, chr15.hg19:g.62299603T>G], p.I1132V [rs3784635, chr15.hg19: g.62254989T>C] and p.Q2376Q [rs17238189, chr15.hg19: g.62212781T>C]) with reduced risk for PD in the New York cohort (Table 2). These remained nominally significant with and without including adjustment for ethnicity, and only one of them, p. I1132V (OR 0.28, 95% CI 0.12-0.64, p=0.0025), remained statistically significant after correction for multiple comparisons. In the two other cohorts, these variants showed the same directionality as in the New York cohort, but did not reach statistical significance (Table 2). These four variants, p.R153H, p.I398I, p.I1132V and p.Q2376Q, are in strong or even complete LD (Table S-4). The most significant tagging SNP of this haplotype, p. I1132V, was validated using Sanger sequencing in all three cohorts. Meta-analysis of all the common coding variants showed an association of these four linked SNPs with reduced risk for PD (Figure 1), with similar directionality across all cohorts. To examine whether this haplotype may affect AAO of PD, ANOVA with the status of the p. I1132V variant was performed. This variant was not associated with AAO in all three cohorts (French/French Canadian cohort p=0.65, New York cohort p=0.34, Israel cohort p=0.99).

None of these variants were in LD with the known GWAS top hit, rs2414739, and therefore these associations do not explain the GWAS hit in the *VPS13C* locus. Of all the other common coding variants, only one variant, p.E2008D (rs78071599, chr15.hg19: g.62223303C>G), was in some LD with the GWAS top hit (D'=0.808, r^2=0.006, Table S-5), but this variant was not associated with LOPD, and therefore it also cannot explain the GWAS hit in our populations. Interestingly, an intronic variant, rs78530361 (chr15.hg19: g.62214265C>T), was in strong LD with the top GWAS SNP (D'=1, r^2=0.003). Of note, r^2 is low since this intronic SNP has a much lower allele frequency than the top GWAS hit, but

bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

every time the rs78530361 SNP is found, it is on an allele which harbors the top GWAS hit rs2414739. However, in our cohorts, this variant was not associated with PD (p=0.66), likely due to its very low r^2 with the GWAS hit.

Discussion

Our study, which included full sequencing of *VPS13C* in three cohorts, identified 60 rare *VPS13C* variants (MAF<0.01) that are nonsynonymous or affect splicing, and 18 common variants (MAF>0.01) in coding regions of the gene and splice sites. Our results suggest that rare homozygous and compound heterozygous variants are rare in LOPD and probably have importance mainly in early onset PD, as previously described. We have identified a potentially protective haplotype, which includes four variants, two of which are substitutions of amino acids, p.R153H and p.I1132V. Interestingly, this haplotype is not in LD with the top GWAS SNP, rs2414739, suggesting that this may be a secondary association in the *VPS13C* locus, which was not identified in previous studies. However, this should be considered as preliminary results and needs to be examined in additional cohorts in order to conclude whether this haplotype is indeed associated with reduced risk for PD. Furthermore, the full sequencing analysis did not identify any coding variant that is in LD with the original GWAS hit that can explain the GWAS association in this locus. This may suggest that the variant that has the main effect on the risk for PD in the *VPS13C* is outside of the coding and untranslated regions of the gene, likely being a regulatory element.

Previous studies on the top GWAS hit in the *VPS13C* locus demonstrated conflicting results. While significant associations of the top GWAS hit in this locus (rs2414739) with PD were found in Iranian [5] and East Asian [3] populations, negative results were reported in Taiwanese [4] and Han Chinese [6, 7] populations. Therefore, it is possible that the association of *VPS13C* with PD is population dependent. Of note, that the association in the current study was mainly driven by populations enriched in Ashkenazi Jews, while in the French/French Canadian cohort the differences between patients and controls were

bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

much smaller and not statistically significant (Table 2). This may suggest that this haplotype association is population specific, and additional studies in different populations are required to answer this question.

In our three cohorts, we did not find any very rare homozygous mutations with MAF<0.001 and found only two carriers of compound heterozygous mutations, both of which were controls (Table S-6). We were unable to determine if the mutations were in cis or trans. Previously reported cases of PD in carriers of compound heterozygous or homozygous mutations all shared a specific clinical presentation of PD: early onset, rapid progression and early cognitive dysfunction [8, 9]. In one study, the patient's AAO was 39, and disease progression was moderately severe with psychiatric symptoms and impaired cognition [9]. In another study, the three patients showed severe phenotypes: AAO of 25, 33, and 46 years, severe and early cognitive dysfunction, and became bedridden at 31, 43, and 58 years, respectively. Considering the young age of one of the compound heterozygous *VPS13C* variant carriers in our cohort (30 years), it is still possible, although unlikely that this individual will develop PD in the future. The negative results of the SKAT-O analyses demonstrate that rare heterozygous variants in *VPS13C* do not have an important role in PD in our cohorts.

Our study has several limitations. The differences between PD patients and controls in sex and mainly age are significant in some of our cohorts. To address this limitation, we included age and sex as covariates in the regression models. Therefore, if the association of the protective haplotype was related to age and not to disease status, we would likely not observe an association in the adjusted model. There was no significant difference in the percentage of Ashkenazi Jews between patients and controls in the New York cohort, yet we still performed the regression model with and without ethnicity as a covariate, and in both analyses the results remained significant. Nevertheless, the fact that our populations are enriched with relatively homogeneous populations such as Ashkenazi Jews and French Canadians requires additional studies in other populations.

In conclusion, our results suggest that *VPS13C* variants have a limited role in late-onset PD. The potentially PD protective haplotype located within *VPS13C*, which requires additional replications, may suggest that *VPS13C* could be a future target for PD therapeutic development. If naturally occurring

bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

genetic variants may reduce PD risk, it is conceivable that drugs that can mimic their effects could be developed. Additional genetic and functional studies will be required to determine if *VPS13C* may be a viable target for PD drug development.

bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Acknowledgments

We thank the patients and control subjects for their participation in this study. This work was financially supported by the Michael J. Fox Foundation, the Canadian Consortium on Neurodegeneration in Aging (CCNA) and the Canadian Glycomics Network (GlycoNet). This research was also undertaken thanks in part to funding from the Canada First Research Excellence Fund (CFREF), awarded to McGill University for the Healthy Brains for Healthy Lives (HBHL) program. The Columbia University cohort is supported by the Parkinson's Foundation, the National Institutes of Health [K02NS080915, and UL1 TR000040] and the Brookdale Foundation. 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, 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 Armaghan Alam, Daniel Rochefort, Helene Catoire, Clotilde Degroot and Vessela Zaharieva for their assistance.

bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

References

1. 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.
2. Chang D, Nalls MA, Hallgrimsdottir IB, et al. A meta-analysis of genome-wide association studies identifies 17 new Parkinson's disease risk loci. Nat Genet 2017;49:1511-1516.
3. Zou M, Li R, Wang JY, et al. Association analyses of variants of SIPA1L2, MIR4697, GCH1, VPS13C, and DDRGK1 with Parkinson's disease in East Asians. Neurobiol Aging 2018.
4. Chen CM, Chen YC, Chiang MC, et al. Association of GCH1 and MIR4697, but not SIPA1L2 and VPS13C polymorphisms, with Parkinson's disease in Taiwan. Neurobiol Aging 2016;39:221 e221-225.
5. Safaralizadeh T, Jamshidi J, Esmaili Shandiz E, et al. SIPA1L2, MIR4697, GCH1 and VPS13C loci and risk of Parkinson's diseases in Iranian population: A case-control study. J Neurol Sci 2016;369:1-4.
6. Wang L, Cheng L, Li NN, Yu WJ, Sun XY, Peng R. Association of four new candidate genetic variants with Parkinson's disease in a Han Chinese population. Am J Med Genet B Neuropsychiatr Genet 2016;171B:342-347.
7. Yang X, Zheng J, An R, et al. Polymorphism in MIR4697 but not VPS13C, GCH1, or SIPA1L2 is associated with risk of Parkinson's disease in a Han Chinese population. Neurosci Lett 2017;650:8-11.
8. Lesage S, Drouet V, Majounie E, et al. Loss of VPS13C Function in Autosomal-Recessive Parkinsonism Causes Mitochondrial Dysfunction and Increases PINK1/Parkin-Dependent Mitophagy. Am J Hum Genet 2016;98:500-513.
9. Schormair B, Kemlink D, Mollenhauer B, 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;93:603-612.
10. Darvish H, Bravo P, Tafakhori A, et al. Identification of a large homozygous VPS13C deletion in a patient with early-onset Parkinsonism. Mov Disord 2018;33:1968-1970.
11. Alcalay RN, Levy OA, Wolf P, et al. SCARB2 variants and glucocerebrosidase activity in Parkinson's disease. NPJ Parkinsons Dis 2016;2.
12. 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;55:181-184.
13. O'Roak BJ, Vives L, Fu W, et al. Multiplex targeted sequencing identifies recurrently mutated genes in autism spectrum disorders. Science 2012;338:1619-1622.
14. Ross JP, Dupre N, Dauvilliers Y, et al. Analysis of DNAJC13 mutations in French-Canadian/French cohort of Parkinson's disease. Neurobiol Aging 2016;45:212 e213-212 e217.
15. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 2009;25:1754-1760.
16. McKenna A, Hanna M, Banks E, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res 2010;20:1297-1303.
17. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38:e164.

bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

18. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536:285-291.
19. Purcell S, Neale B, Todd-Brown K, et al. PLINK: a tool set for whole-genome association and population-based linkage analyses. Am J Hum Genet 2007;81:559-575.
20. Lee S, Emond MJ, Bamshad MJ, 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;91:224-237.
21. Viechtbauer W. Conducting Meta-Analyses in R with the metafor Package. Journal of Statistical Software 2010;36.
22. 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;31:3555-3557.

bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Figure legends

Figure 1 Legend: Forest plots – meta-analyses of four *VPS13C* coding variants associated with reduced risk for Parkinson Disease. The forest plots depict the effects of the four *VPS13C* coding variants that create the haplotype associated with reduced PD risk in the three cohorts studied, and their meta-analyzed effect on risk for PD. The results in the random effect model are nearly identical. CI, confidence interval; FE, fixed effect

Table 1: Study population details

Sequenced		Cohort	Min Depth of Coverage	Analyzed						Genotyping call rate %
				Cases			Controls			
N (cases)	N (controls)			N	Mean Age (SD) y	%Male	N	Mean age (SD) y	%Male	
543	866	French/French-Canadian	15x	534	59.7(11.4)	63.7	858	42.0(13.6)	51.6	99.6
			50x	498	59.8(11.2)	63.5	785	42.3(13.5)	52.0	99.2
533	270	New York	15x	520	59.3(11.8)	64.0	262	65.0(9.8)	34.0	99.9
			50x	502	59.3(11.8)	64.1	245	65.4(9.6)	32.7	99.5
491	531	Israel	15x	482	60.6(11.7)	61.8	488	33.9(7.2)	57.8	99.3
			50x	478	60.7(11.7)	61.7	487	33.9(7.2)	57.7	99.3

N, number; Min, minimum; y, year

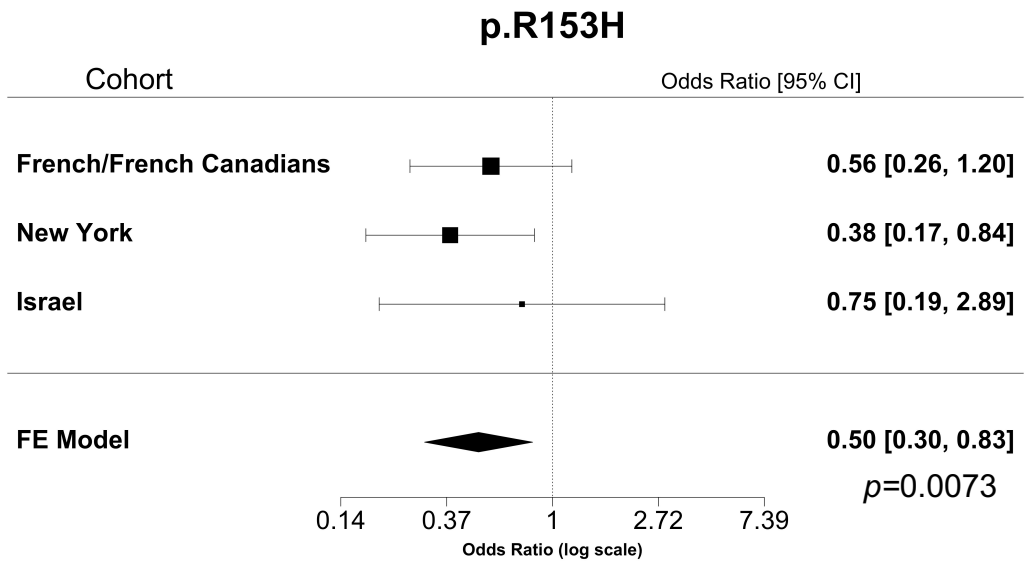
bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

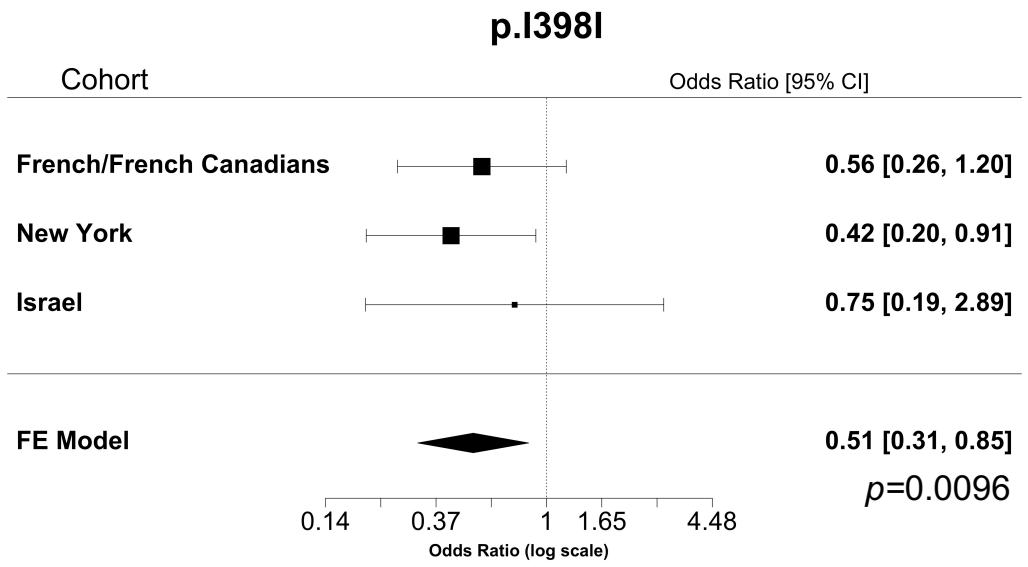
Table 2: Four variants forming the protective haplotype found in *VPS13C* and the results of the logistic regression in three cohorts

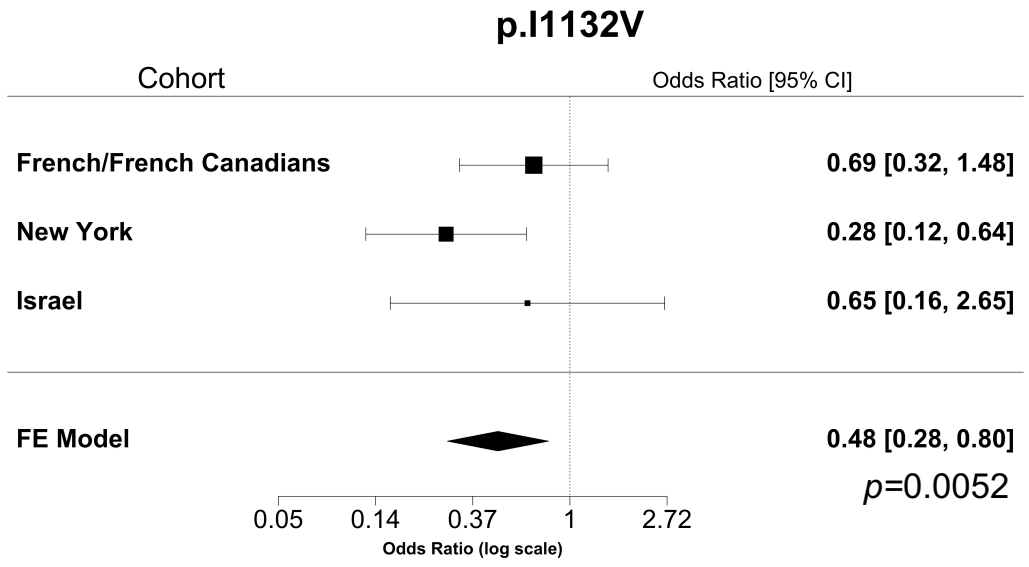
Variant	rs number	nt-substitution	Cohort	N (cases)	N (controls)	Affected MAF	Unaffected MAF	GnomAD MAF†	OR (95%CI)	P
p.R153H	rs12595158	c.G458A	F/FC	15	27	0.0143	0.0157	0.0734	0.56 (0.26-1.20)	0.1358
			NY	14	16	0.0135	0.0324		0.38 (0.17-0.84)	0.0172
			Israel	16	22	0.0166	0.0225	0.0236	0.70 (0.19-2.6)	0.5940
p.I398I	rs9635356	c.A1194C	F/FC	15	28	0.0143	0.0163	0.0737	0.56 (0.26-1.20)	0.1330
			NY	16	16	0.0154	0.0324		0.42 (0.20-0.91)	0.0274
			Israel	15	19	0.0156	0.0195	0.0214	0.75 (0.19-2.89)	0.6752
p.I1132V	rs3784635	c.A3394G	F/FC	15	26	0.0143	0.0152	0.0726	0.69 (0.32-1.48)	0.3404
			NY	11	18	0.0106	0.0363		0.28 (0.12-0.64)	0.0025
			Israel	14	19	0.0145	0.0195	0.0233	0.65 (0.16-2.65)	0.5444
p.Q2376Q	rs17238189	c.A7128G	F/FC	10	24	0.0095	0.0140	0.0698	0.51 (0.21-1.23)	0.1344
			NY	16	17	0.0154	0.0344		0.42 (0.20-0.90)	0.0245
			Israel	16	20	0.0166	0.0215	0.0212	0.47 (0.07-3.10)	0.4350

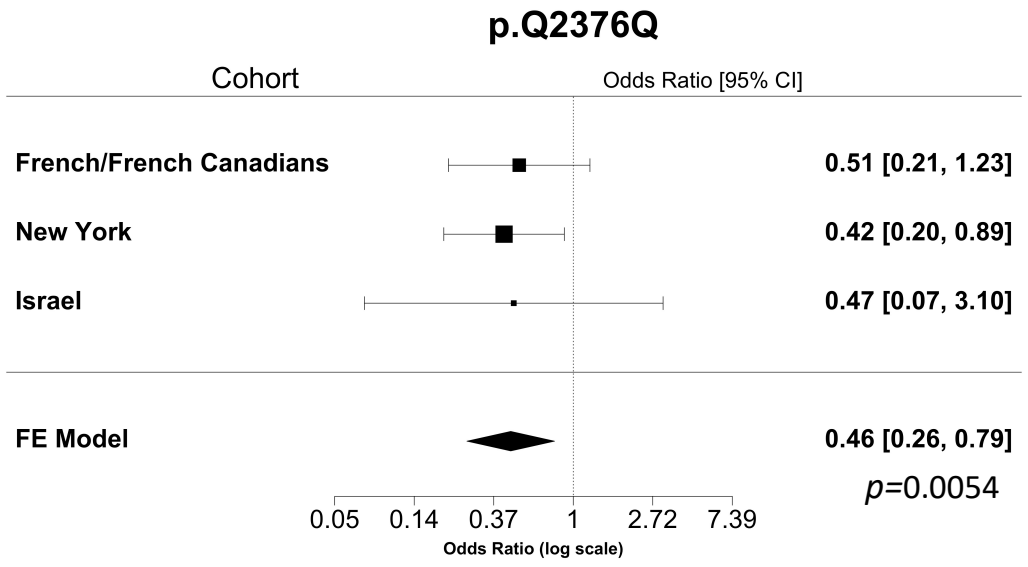
nt: nucleotide; N, number; MAF, minor allele frequency; GnomAD, Genome Aggregation Database; OR, odds ratio; CI, confidence interval; P, p value; F/FC, French/French-Canadian; NY, New York.
†GnomAD_ASJ frequencies were used for the Israeli cohort; all samples are Ashkenazi Jewish.

bioRxiv preprint doi: https://doi.org/10.1101/705533; this version posted July 20, 2019. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.


p.R153H
Cohort
Odds Ratio [95% CI]
French/French Canadians
0.56 [0.26, 1.20]
New York
0.38 [0.17, 0.84]
Israel
0.75 [0.19, 2.89]
FE Model
0.50 [0.30, 0.83]
p=0.0073
0.14
0.37
1
2.72
7.39
Odds Ratio (log scale)


p.I398I
Cohort
Odds Ratio [95% CI]
French/French Canadians
0.56 [0.26, 1.20]
New York
0.42 [0.20, 0.91]
Israel
0.75 [0.19, 2.89]
FE Model
0.51 [0.31, 0.85]
p=0.0096
0.14
0.37
1
1.65
4.48
Odds Ratio (log scale)


p.I1132V
Cohort
Odds Ratio [95% CI]
French/French Canadians
0.69 [0.32, 1.48]
New York
0.28 [0.12, 0.64]
Israel
0.65 [0.16, 2.65]
FE Model
0.48 [0.28, 0.80]
p=0.0052
0.05
0.14
0.37
1
2.72
Odds Ratio (log scale)


p.Q2376Q
Cohort
Odds Ratio [95% CI]
French/French Canadians
0.51 [0.21, 1.23]
New York
0.42 [0.20, 0.89]
Israel
0.47 [0.07, 3.10]
FE Model
0.46 [0.26, 0.79]
p=0.0054
0.05
0.14
0.37
1
2.72
7.39
Odds Ratio (log scale)