medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Analysis of heterozygous *PRKN* variants and copy number variations in Parkinson's disease

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

Authors affiliations:
1. Department of Human Genetics, McGill University, Montréal, Quebec, Canada
2. Montreal Neurological Institute and Hospital, McGill University, Montréal, Quebec, Canada
3. Department of Neurology and Neurosurgery, McGill University, Montréal, Quebec, Canada
4. Department of Neurology, College of Physicians and Surgeons, Columbia University Medical Center, New York, NY 10032, USA
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 school of medicine, Tel-Aviv University, Tel-Aviv, Israel
8. UC Gardner Neuroscience Institute and Gardner Family Center for Parkinson's Disease and Movement Disorders, Cincinnati, OH, USA.
9. National Reference Center for Narcolepsy, Sleep Unit, Department of Neurology, Gui-de-Chauliac Hospital, CHU Montpellier, University of Montpellier, Inserm U1061, Montpellier, France
10. Division of Neurosciences, CHU de Québec, Université Laval, Québec City, Quebec, Canada
11. Department of Medicine, Faculty of Medicine, Université Laval, Québec City, Quebec, Canada
12. Movement Disorders Institute, Department of Neurology, 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, USA

Correspondence to:
Ziv Gan-Or
Department of Neurology and Neurosurgery
McGill University
1033 Pine Avenue, West, Ludmer Pavilion, room 312
Montreal, QC, H3A 1A1, Canada
Email: ziv.gan-or@mcgill.ca
Telephone: +1-514-398-5845
Word Count: 2,877 words.
Running title: Heterozygous *PRKN* variants in PD

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Key Words:

Parkinson's disease, copy number variation, association study, genetics, neurodegeneration.

Funding agencies:

This study 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.

Relevant conflicts of interest/financial disclosures:

SF received consulting fees/honoraria for board membership from Retrophin Inc., Sun Pharma Advanced Research Co., LTD and Kashiv Pharma. CHW received research support from Sanofi, Biogen, Roche, consulting fees/honoraria from Amneal, Adamas, Impel, Kyowa, Mitsubishi, Neurocrine, US World Meds, Acadia, Acorda. AJE received grant support from the NIH and the Michael J Fox Foundation; personal compensation as a consultant/scientific advisory board member for Abbvie, Adamas, Acadia, Acorda, Neuroderm, Neurocrine, Impax/Amneal, Sunovion, Lundbeck, Osmotica Pharmaceutical, and US World Meds; publishing royalties from Lippincott Williams & Wilkins, Cambridge University Press, and Springer; and honoraria from US World Meds, Lundbeck, Acadia, Sunovion, the American Academy of Neurology, and the Movement Disorders Society. ND received consultancy fees from Actelion Pharmaceuticals. SHB received consulting fees from Actelion Pharmaceuticals Ltd., Abbvie Israel, Robotico Ltd., Medtronic Israel, Medison Pharma Israel. EAF received consulting fees from Inception Sciences. RNA received consultation fees from Biogen, Denali, Genzyme/Sanofi and Roche. ZGO received consultancy fees from Lysosomal Therapeutics Inc. (LTI), Idorsia, Prevail Therapeutics, Inceptions Sciences (now Ventus), Ono Therapeutics, Denali and Deerfield. Rest of the authors have nothing to report.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Abstract

Background: Biallelic *PRKN* mutation carriers with Parkinson's disease (PD) typically have an earlier disease onset, slow disease progression and, often, different neuropathology compared to sporadic PD patients. However, the role of heterozygous *PRKN* variants in the risk of PD is controversial.

Objectives: We aimed to examine the association between heterozygous *PRKN* variants, including single nucleotide variants and copy-number variations, and PD.

Methods: We fully sequenced *PRKN* in 2,809 PD patients and 3,629 healthy controls, including 1,965 late onset (63.97±7.79 years, 63% men) and 553 early onset PD patients (43.33±6.59 years, 68% men). *PRKN* was sequenced using targeted next-generation sequencing with molecular inversion probes. Copy-number variations were identified using a combination of multiplex ligation-dependent probe amplification and ExomeDepth. To examine whether rare heterozygous single nucleotide variants and copy-number variations in *PRKN* are associated with PD risk and onset, we used optimized sequence kernel association tests and regression models.

Results: We did not find any associations between all types of *PRKN* variants and risk of PD. Pathogenic and likely-pathogenic heterozygous single nucleotide variants and copy-number variations were less common among PD patients (1.52%) than among controls (1.8%, false discovery rate-corrected p=0.55). No associations with age at onset and in stratified analyses were found.

Conclusions: Heterozygous single nucleotide variants and copy-number variations in *PRKN* are not associated with Parkinson's disease. Molecular inversion probes allow for rapid and cost-

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

effective detection of all types of *PRKN* variants, which may be useful for pre-trial screening and for clinical and basic science studies specifically targeting *PRKN* patients.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Introduction

Parkinson's disease (PD) is a common neurodegenerative disorder with a typical age at onset (AAO) ranging between 60-70 years.[1] However, a subgroup of patients has early onset PD (EOPD), typically defined as AAO < 50 years.[2] The most common genetic cause of EOPD are homozygous or compound heterozygous variants in the *PRKN* gene, found in 6.0-12.4% of individuals who present with PD symptoms before the age of 50.[3-5] *PRKN* has a high rate of single nucleotide variants (SNV) and copy number variations (CNVs), since it is located in a genomic region prone to rearrangements.[6, 7] *PRKN* encodes Parkin, an E3 ubiquitin protein ligase important in mitophagy.[8]

Neuropathological studies have demonstrated that individuals with biallelic *PRKN* variants diagnosed with PD do not have the typical PD neuropathology, as Lewy bodies are absent in most cases, and the neurodegenerative process is limited to the substantia nigra.[9, 10] It is therefore possible that patients with biallelic *PRKN* variants represent a distinct subgroup, or arguably a distinct disease with similar clinical features.[10] Since we are moving towards therapies targeting specific genetic defects in PD (such as *GBA* and *LRRK2*-targeting therapies), or α-synuclein accumulation[11] (which is mostly absent in *PRKN*-related patients),[9] it is crucial to properly identify these patients. However, the role of rare heterozygous *PRKN* SNVs and CNVs in PD has not been clearly established by association studies,[12] and it is currently controversial. For example, a previous study with 159 patients and 170 controls showed significant difference in heterozygous *PRKN* SNVs and CNVs between PD patients and controls, while larger studies suggested a lack of association.[13-15] Additional studies have also shown contradictory results in familial PD, EOPD and late onset PD (LOPD) using SNVs and/or CNVs.[13-34] Therefore, the role of heterozygous *PRKN* variants remains controversial. Towards future clinical trials targeting

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

PRKN, it will be crucial to determine whether heterozygous *PRKN* variants are associated with PD.

To investigate the potential effect of rare heterozygous SNVs and CNVs in PD, we applied a simple, fast and cost-effective method to detect both types of variants. Using targeted next generation sequencing and bioinformatic approaches, we fully sequenced *PRKN* to identify both SNVs and CNVs in a large cohort of PD, including LOPD and EOPD.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Methods

Study Population

A total of 2,809 unrelated and consecutively recruited PD patients and 3,629 controls from three cohorts were sequenced, including 1,965 LOPD patients (mean [SD], 63.97±7.79 years, 1,231 men [63%]) and 553 EOPD patients (mean [SD], 43.33±6.59, 374 men [68%]). Age and sex were not available for 291 patients, 88 controls and 22 patients, 4 controls, respectively. After excluding low sequencing quality samples and biallelic *PRKN* carriers, we performed statistical analysis on 6,090 individuals: 2,627 patients and 3,463 controls. The three cohorts are detailed in Table 1 and include: a) a cohort of European ancestry, confirmed by principal component analysis, collected at McGill University, including French-Canadian (mostly recruited through the Quebec Parkinson Network)[35] and French participants recruited in Quebec, Canada and Montpellier, France b) a cohort recruited at Columbia University, New York, as previously described,[36] primarily composed of individuals of self-reported European origin and Ashkenazi Jews, and c) a cohort collected at the Sheba Medical Center, Israel, of self-reported Ashkenazi Jewish ancestry, as previously described.[37] PD was diagnosed by movement disorder specialists according to the UK Brain Bank Criteria, without excluding patients with positive family history [38] or the Movement Disorders Society Criteria.[39] Study protocols were approved by the relevant Institutional Review Boards and all patients signed informed consent before participating in the study.

Genetic analysis

PRKN sequencing

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

All samples were sequenced at McGill University, Canada using the same method. A total of 50 genes were captured using molecular inversion probes (MIPs) and sequenced as previously described.[40] In brief, probes that specifically target the coding sequences of the genes of interest were designed, followed by capture and PCR amplification of the targeted regions. After adding barcodes, samples were pooled and sequenced at the McGill University and Génome Québec Innovation Centre with Illumina HiSeq 2500/4000. The full protocol is available upon request. Alignment (GRCh37/hg19), quality control and variant calls were done using the Burrows-Wheeler Aligner (BWA),[41] Genome Analysis Toolkit (GATK v3.8),[42] and ANNOVAR [43] as previously described.[44] Only rare variants (minor allele frequency, MAF, < 0.01) according to the public database Genome Aggregation Database (GnomAD) [45] with a minimum coverage of 30x were included in the analysis. Samples with more than 10% missingness were excluded. The script for these analyses can be found at https://github.com/gan-orlab/MIPVar. We examined all rare exonic variants using the Integrative Genomics Viewer (IGV v 2.7).[46] All variants were classified using Varsome [47] according to the American College of Medical Genetics and Genomics (ACMG) standards and guidelines into five categories: pathogenic, likely pathogenic, uncertain significance, likely benign and benign.

Detection and validation of copy number variations

There are four general types of methods to infer CNVs from next-generation sequencing.[48] Because MIPs target only a small portion of the genome, most CNV breakpoints will not be sequenced. Therefore, only read-depth based methods can be applied for MIPs since other types of methods utilize reads that span breakpoints. In order to detect CNVs, we examined two methods based on read depth for the MIP data, ExomeDepth v1.1.10 [49] and panelcn.MOPS v1.4.0 in R.[50] When using ExomeDepth, each test sample is compared to the best set of reference

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

samples out of 3,629 controls, chosen by the software according to the correlation of the coverage for each probe between the test sample and the reference samples. A filter for samples with correlation above 0.97 per the suggestion of the developer was applied to remove false positives. Panelcn.MOPS also selects the best set of reference samples according to correlation and includes several quality control (QC) steps, such as a minimum user defined depth of coverage per probe. Probes are marked as low quality if their read count shows high variance across the test sample and selected references. To validate CNVs, we performed multiplex ligation-dependent probe amplification (MLPA) using the SALSA MLPA P051-D2 Parkinson probemix 1 kit according to the manufacturer instructions (MRC Holland), which is the gold standard for *PRKN* CNV detection.

Quality Control of MIPs for CNV detection

The highest performing parameters were achieved by excluding probes from genes in our library where the average coverage was below 100X in more than 15% of the coding and untranslated regions of the genes. Probes with average coverage below 100X, and samples with average coverage across all genes less than 50X were also excluded. Figure 1 details the numbers of patients and controls in each cohort after different stages of quality control.

Statistical Analysis

The associations between rare heterozygous SNVs (MAF < 0.01), heterozygous CNVs and PD were tested using optimized sequence kernel association tests (SKAT-O v1.3.2 in R)[51] in all cohorts separately, adjusted for age, sex and ancestry as needed. The initial analysis was performed after excluding biallelic carriers of pathogenic and likely pathogenic mutations and adjusting for age, sex, ethnicity and the presence of *GBA* and *LRRK2* variants (Figure 1, yellow).

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Rare variants were grouped by: a) CADD score (CADD>12.37), which represent the top 2% of variants predicted to be deleterious, b) functional variants, which include stop gain, nonsynonymous, splice-site and frameshift variants, c) nonsynonymous variants, and d) loss-of-function variants, which include frameshift, splice-site and stop gain variants. A meta-analysis of the results from the three cohorts was performed using MetaSKAT (MetaSKAT v0.80, R)[52] for heterozygous SNVs, CNVs, and both combined, according to the five ACMG categories (pathogenic, likely pathogenic, uncertain significance, likely benign and benign). Since the age- and sex-adjusted model removes samples without available data on age and sex, we also performed an unadjusted model to avoid this exclusion (Figure 1, blue). We have also repeated all analyses after several additional filtering and adjusting stages, including: adjusting for all *GBA* and *LRRK2* p.Gly2019Ser variant carriers (Supplementary Table 1), excluding these *GBA* and *LRRK2* variant carriers, analyzing only samples with early onset PD (defined as AAO < 50 years), and excluding samples with CNVs in which phasing was not possible (n=8, for example, in a sample with a reported deletion of exons 3-4 the deletion could be on the same allele, or each exon can be deleted on a different allele, Figure 1, grey). The association between heterozygous SNVs, CNVs and AAO of Parkinson's disease was also calculated using linear regression adjusted for sex and ancestry as needed in all cohorts separately. Here too, patients carrying *GBA* variants or the *LRRK2* p.Gly2019Ser variant (Supplementary Table 1) were excluded and all analyses were repeated. METAL [53] was used to performed fixed-effect meta-analysis on all cohorts in the AAO analysis. Since we have performed multiple interdependent analyses, we used a false discovery rate (FDR) correction for multiple comparisons with a FDR-corrected q<0.05 considered as statistically significant.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Results

Identification of *PRKN* SNVs and CNVs

The average coverage of *PRKN* (NM_004562) across all samples was 988X, with 98% of nucleotides covered at >30X, and 94% covered at >100X. We identified 199 rare SNVs in 237 patients and 300 controls in the main analysis (Table 2), including nonsynonymous, frameshift deletions and splice site variants in *PRKN* across all cohorts (the specific variants are detailed in Supplementary Table 2).

To identify CNVs, we first aimed to examine which calling method is best suited to properly call CNVs from our MIP targeted sequencing panel. For this purpose, we screened for CNVs in 510 samples using MLPA, the gold standard for CNV detection in *PRKN*. We specifically enriched these samples with EOPD patients to increase the chances to detect CNVs. Out of the 510 samples, 46 carried CNVs in *PRKN* (32 patients and 14 controls). The 32 patients included four homozygous *PRKN* deletion carriers, 17 heterozygous deletion carriers and 11 duplication carriers. Subsequently, we have examined which method (ExomeDepth or panel.cnMOPS) has the highest performance. Except for one deletion for which the MIPs data did not pass QC due to low coverage call rate, deletions and duplications in *PRKN* were identified with 97% sensitivity and 95% specificity using ExomeDepth. In contrast, using the best parameters, panel.cnMOPS had 98% sensitivity but only 54% specificity using samples that passed QC when compared to MLPA. The parameters and CNV call rates for each method are detailed in Supplementary Table 3. Due to its superior performance, we applied ExomeDepth on all cohorts, and identified a total of 62 carriers of CNVs in patients and controls. Supplementary Table 4 details all carriers of CNVs, including heterozygous and bi-allelic carriers of other CNVs or other SNVs.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Heterozygous *PRKN* SNVs and CNVs are not associated with Parkinson's disease

To examine the association of rare (MAF < 0.01) heterozygous SNVs and CNVs on risk of PD, we took two approaches. First, we performed a SKAT-O in each cohort to determine whether there is a burden of heterozygous *PRKN* variants of different types. "Pathogenic" variants included pathogenic and likely pathogenic variants, while "non-benign" variants included pathogenic, likely pathogenic and variants of uncertain significance. All CNVs were considered as pathogenic loss-of-function variations. No statistically significant associations were found in any of the SKAT-O analyses (Table 2). Second, we performed a series of meta-analyses by collapsing in each cohort SNVs alone, CNVs alone, and combined. In these analyses too, adjusted for age, sex and ethnicity, no association between heterozygous carriage of *PRKN* mutations and PD was found (Table 2). Pathogenic and likely pathogenic variants were less frequent in patients (1.52%) than in controls (1.8%, p = 0.55, , Table 2), suggesting lack of association with risk of PD. In order to avoid the possibility that the exclusion of samples without available data on age and sex had biased the results, we have also performed an unadjusted analysis including all samples. Additional analyses with and without *GBA* and *LRRK2* p.Gly2019Ser variants, with and without CNVs of unknown phasing, and including only samples patients with AAO < 50 have also been performed. In these analyses too, there were no statistically significant differences between patients and controls (Supplementary Table 5-6).

Heterozygous *PRKN* SNVs and CNVs are not associated with AAO of Parkinson's disease

The association between rare heterozygous SNVs and CNVs on AAO of PD was examined using linear regression in each cohort alone on the same groups of mutations mentioned in the previous

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

association study. After adjusting for sex, ancestry, and the presence of *GBA* and *LRRK2* variants, we found no association in any analyses. We also performed meta-analysis by collapsing each cohort which yielded no statistically significant results (Table 3). When examining CNVs, the meta-analysis shows an earlier AAO in heterozygous *PRKN* CNV carriers (3.6 years younger compared to non-carriers), but the association was not statistically significant after correction for multiple comparisons. This difference in AAO was mainly driven by an effect of CNVs in the Columbia cohort, which was almost 8 years younger in carriers of CNVs (average AAO of 51.85 years) compared to non-carriers of CNVs (59.44 years). This difference was not statistically significant after correction for multiple comparisons as well. Larger studies for AAO of heterozygous *PRKN* carriers are needed to further study these findings. Association analyses between different types of heterozygous *PRKN* variants and AAO of PD, including with and without *LRRK2* and *GBA* variant carriers, with and without ambiguous phasing (see methods), and in AAO < 50 can be found in Supplementary Tables 7-8. In all analyses, there were no statistically significant associations.

Identification of *PRKN*-associated parkinsonism patients

Overall, we were able to identify 9 patients with pathogenic or likely pathogenic homozygous and compound heterozygous *PRKN* SNVs and/or CNVs (Table 4). The most common pathogenic SNV in our cohort was p.Gln34ArgfsTer5 mutation, found in 3 (33%) *PRKN* patients, and the most common CNV was heterozygous deletion of exon 3, found in 7 (77%) *PRKN* patients. The average AAO of PD in biallelic *PRKN* SNV/CNV carriers was 28.0 ±7.82 years old.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Discussion

In the current study, we found that the frequencies of heterozygous SNVs and CNVs in *PRKN* are similar in PD patients and controls. These results do not support a role for heterozygous *PRKN* variants in the risk of PD or its AAO. Of note, in one cohort (Columbia), the average AAO of CNV carriers was about 8 years younger compared to non-carriers (Table 3), yet in the other cohorts there was no difference between CNV carriers and non-carriers. Additional studies on AAO in heterozygous *PRKN* carriers are required to conclusively determine whether or not they are associated with earlier AAO. Since the *PRKN* region is prone to genetic variance,[6] including multiple SNVs and CNVs, properly genotyping all types of *PRKN* variants could be challenging. Using a simple, fast and cost-effective method, we were able to successfully detect all CNVs, SNVs and indels. With MIPs, deep coverage can be achieved, and the probes always target the exact same region, as opposed to whole-exome or whole-genome sequencing where there is no full overlap between all the reads. When the coverage is high, it provides an advantage that allows for more accurate calls of CNVs as well as SNVs and indels. Using this approach, we have identified 199 rare *PRKN* variants and 62 participants with *PRKN* CNVs, with very high sensitivity and specificity (97% and 95%, respectively, when compared to the gold standard MLPA method). Our approach can therefore be used for large-scale screening of PD cohorts, with only validation of detected *PRKN* CNVs with MLPA, instead of fully screening all patients with MLPA. Of note, we identified 9 patients with pathogenic and likely pathogenic biallelic *PRKN* variants. This number of patients is lower than previously reported in EOPD. It is possible that in Ashkenazi Jewish Parkinson's disease patients (comprising the entire Sheba cohort and a large portion of the Columbia cohort), the frequency of *PRKN* variants is lower, as evident by the lack of such patients in the Sheba cohort. This is also supported by the Columbia

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

cohort, in which all biallelic *PRKN* patients are of European ancestry and none among the Ashkenazi Jewish origin.

There have been multiple studies analyzing the role of heterozygous *PRKN* mutations with conflicting results, shown in Supplementary Table 9. These conflicts may arise from different screening approaches. Some studies first sequenced all patients for rare SNVs and/or CNVs, then sequenced only for selected variants in controls. This approach will create a bias, as the controls may carry other pathogenic *PRKN* variants. Other studies sequenced all patients and controls for heterozygous SNVs and/or CNVs more systematically, and the majority of them were negative. Systematic analysis, as was done in the current study, will avoid misrepresenting the genetic landscape of the study population. Our results do not support an association between heterozygous SNVs and CNVs in *PRKN* and PD, which is supported by other systematic studies of *PRKN* as shown in Supplementary Table 9.[14-17, 24, 33] These results also emphasize the need for determining the pathogenicity of different *PRKN* variants, as many variants are currently defined as variants of unknown significance. Having a reliable assay for Parkin activity, as previously suggested, would provide an experimental way to assess pathogenicity of *PRKN* variants.[54]

To further study the potential effect of heterozygous *PRKN* variants, previous studies have compared the rate of 18F-dopa uptake in biallelic *PRKN* patients, asymptomatic heterozygous *PRKN* mutation carriers and healthy controls.[55, 56] These studies have suggested that some *PRKN* heterozygous carriers may have reduced uptake of ^{18}F - dopa, especially in the caudate and putamen. A follow-up longitudinal study by one of these groups, however, suggested that this reduction is subclinical, and that the rate of progression is very slow and unlikely to lead to clinical parkinsonism manifestations.[57]

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

In recent years, treatments that target specific genes and proteins implicated by human genetic studies, such as *SNCA* (α-synuclein), *GBA* and *LRRK2*, are being tested in clinical trials.[58] Therefore, identifying patients that may benefit from these trials, or conversely, patients that are less likely to benefit, is crucial. Neuropathological studies on brains of patients with *PRKN*-associated parkinsonism have demonstrated that the vast majority of patients with biallelic *PRKN* mutations do not have accumulation of α-synuclein and the typical Lewy bodies that are seen in PD.[59] Since α-synuclein does not accumulate, it is likely that treatment targeting α-synuclein will not be efficient for these patients, who should therefore be excluded from these clinical trials. Furthermore, the neurodegenerative process in *PRKN*-associated Parkinsonism is limited to the substantia nigra and locus coeruleus, and does not spread to other brain regions.[60] Since we did not detect an association between heterozygous *PRKN* variants and PD, we recommend that heterozygous carriers of *PRKN* variants should not be excluded from such trials, as it is likely that the presence of heterozygous *PRKN* variants in PD patients is due to chance. Clinically, patients with *PRKN*-associated Parkinsonism are also different, as they have early onset disease, slowly progressing and typically without or with very limited non-motor symptoms.[59] Therefore, it is important to identify these patients, and our method for rapid and cost-effective detection of *PRKN* variants would be useful for pre-trial screening and for clinical and basic science studies specifically targeting *PRKN* patients.

Although this study examined heterozygous mutations systematically, there are several limitations. The error rate of ExomeDepth CNV detection could affect the results of the association study because not all samples were analysed using MLPA. Furthermore, potentially pathogenic intronic variants have not been examined since intronic regions were not sequenced. In addition, our cohorts were not matched for age and sex. Our controls are on average younger

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

and our patients are predominantly composed of men, yet age and sex were adjusted for when possible. The missing age at onset of patients underpowers our AAO study, however, because data were missing at random, its effect on our results is likely minimal. Another limitation is that in a case-control set-up, phasing cannot be performed, and patients with two variants are considered as compound heterozygous carriers. Since all patients with two mutations had AAO<50, it is likely that indeed they are all compound heterozygous, but we cannot rule out that they carry two variants on the same allele. In addition, individuals with CNVs in consecutive exons are considered as heterozygous carriers, while in fact they can have separate deletions of each exon in different alleles. To examine whether inclusion of these patients affected the results, we repeated the analysis after excluding them, which did not substantially change the results (Supplementary Tables 5-6). An additional limitation of our study is that it includes predominantly individuals of European and Ashkenazi Jewish ancestries. While we adjusted for ancestry in the analysis, studies in additional ancestries are required to determine if heterozygous *PRKN* variants may have a role in PD in other populations.

To conclude, our findings do not support a role for heterozygous *PRKN* variants in PD, and additional large-scale studies are required for a definite conclusion. Our study and the methods we have used provide a framework and a cost-effective method for rapidly screening for all types of *PRKN* variants, which will be useful in future genetic and clinical studies, and for stratification or patient selection for clinical trials.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Acknowledgment

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, Hélène Catoire, Clotilde Degroot and Vessela Zaharieva for their assistance.

Authors' Roles

1) Research project: A. Conception (EY, ZGO), B. Organization (EY, LK, JAR, FA, MS, SF, CHW, LG, AJE, YD, ND, GAR, SH-B, EAF, RNA, ZGO), C. Execution (EY, UR, LK, KM, JAR, FA, DS).

2) Data Generation: A. Experimental (EY, UR, LK, KM, JAR, FA, MAE, DS, MS, SF, CHW, LG, AJE, YD, ND, GAR, SH-B, EAF, RNA, ZGO).

3) Statistical Analysis: A. Design (EY, ZGO), B. Execution (EY).

4) Manuscript Preparation: A. Writing of the first draft (EY, ZGO), B. Review and Critique (EY, UR, LK, KM, JAR, FA, MAE, DS, MS, SF, CHW, LG, AJE, YD, ND, GAR, SH-B, EAF, RNA, ZGO).

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

References

1. Pagano G, Ferrara N, Brooks DJ, Pavese N. Age at onset and Parkinson disease phenotype. Neurology 2016;86(15):1400-1407.
2. Schrag A, Quinn N, Ben-Shlomo Y. Heterogeneity of Parkinson's disease. Journal of Neurology, Neurosurgery & Psychiatry 2006;77(2):275-276.
3. Bonifati V. Autosomal recessive parkinsonism. Parkinsonism Relat Disord 2012;18:S4-S6.
4. Kilarski LL, Pearson JP, Newsway V, et al. Systematic review and UK‐based study of PARK2 (parkin), PINK1, PARK7 (DJ‐1) and LRRK2 in early‐onset Parkinson's disease. Mov Disord 2012;27(12):1522-1529.
5. Alcalay RN, Caccappolo E, Mejia-Santana H, et al. Frequency of known mutations in early-onset Parkinson disease: implication for genetic counseling: the consortium on risk for early onset Parkinson disease study. Arch Neurol 2010;67(9):1116-1122.
6. Ambroziak W, Koziorowski D, Duszyc K, et al. Genomic instability in the PARK2 locus is associated with Parkinson's disease. J Appl Genet 2015;56(4):451-461.
7. La Cognata V, Morello G, D'Agata V, Cavallaro S. Copy number variability in Parkinson's disease: assembling the puzzle through a systems biology approach. Hum Genet 2017;136(1):13-37.
8. Brüggemann N, Klein C. Parkin type of early-onset Parkinson disease. GeneReviews®[Internet]: University of Washington, Seattle, 2013.
9. Mata IF, Lockhart PJ, Farrer MJ. Parkin genetics: one model for Parkinson's disease. Hum Mol Genet 2004;13(suppl_1):R127-R133.
10. Schneider SA, Alcalay RN. Neuropathology of genetic synucleinopathies with parkinsonism: Review of the literature. Mov Disord 2017;32(11):1504-1523.
11. Sardi SP, Cedarbaum JM, Brundin P. Targeted therapies for Parkinson's disease: from genetics to the clinic. Mov Disord 2018;33(5):684-696.
12. Nalls MA, Blauwendraat C, Vallerga CL, et al. Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. The Lancet Neurology 2019;18(12):1091-1102.
13. Lesage S, Lohmann E, Tison F, et al. Rare heterozygous parkin variants in French early-onset Parkinson disease patients and controls. J Med Genet 2008;45(1):43-46.
14. Kay D, Stevens C, Hamza T, et al. A comprehensive analysis of deletions, multiplications, and copy number variations in PARK2. Neurology 2010;75(13):1189-1194.
15. Kay DM, Moran D, Moses L, et al. Heterozygous parkin point mutations are as common in control subjects as in Parkinson's patients. Ann Neurol 2007;61(1):47-54.
16. Bandrés-Ciga S, Price TR, Barrero FJ, et al. Genome-wide assessment of Parkinson's disease in a Southern Spanish population. Neurobiol Aging 2016;45:213. e213-213. e219.
17. Benitez BA, Davis AA, Jin SC, et al. Resequencing analysis of five Mendelian genes and the top genes from genome-wide association studies in Parkinson's Disease. Mol Neurodegener 2016;11(1):29.
18. Bras J, Guerreiro R, Ribeiro M, et al. Analysis of Parkinson disease patients from Portugal for mutations in SNCA, PRKN, PINK1 and LRRK2. BMC Neurol 2008;8(1):1.
19. Brooks J, Ding J, Simon-Sanchez J, Paisan-Ruiz C, Singleton A, Scholz S. Parkin and PINK1 mutations in early-onset Parkinson's disease: comprehensive screening in publicly available cases and control. J Med Genet 2009;46(6):375-381.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

20. Camargos ST, Dornas LO, Momeni P, et al. Familial Parkinsonism and early onset Parkinson's disease in a Brazilian movement disorders clinic: Phenotypic characterization and frequency of SNCA, PRKN, PINK1, and LRRK2 mutations. Movement Disorders 2009;24(5):662-666.
21. Clark LN, Afridi S, Karlins E, et al. Case-control study of the parkin gene in early-onset Parkinson disease. Arch Neurol 2006;63(4):548-552.
22. Fiala O, Zahorakova D, Pospisilova L, et al. Parkin (PARK 2) mutations are rare in Czech patients with early-onset Parkinson's disease. PLoS One 2014;9(9).
23. Klein C, Djarmati A, Hedrich K, et al. PINK1, Parkin, and DJ-1 mutations in Italian patients with early-onset parkinsonism. European journal of human genetics 2005;13(9):1086-1093.
24. Lincoln SJ, Maraganore DM, Lesnick TG, et al. Parkin variants in North American Parkinson's disease: cases and controls. Movement disorders: official journal of the Movement Disorder Society 2003;18(11):1306-1311.
25. Macedo MG, Verbaan D, Fang Y, et al. Genotypic and phenotypic characteristics of Dutch patients with early onset Parkinson's disease. Movement disorders: official journal of the Movement Disorder Society 2009;24(2):196-203.
26. Moura KCV, Campos Junior M, de Rosso ALZ, et al. Genetic analysis of PARK2 and PINK1 genes in Brazilian patients with early-onset Parkinson's disease. Dis Markers 2013;35(3):181-185.
27. Sironi F, Primignani P, Zini M, et al. Parkin analysis in early onset Parkinson's disease. Parkinsonism Relat Disord 2008;14(4):326-333.
28. Spataro N, Roca‐Umbert A, Cervera‐Carles L, et al. Detection of genomic rearrangements from targeted resequencing data in Parkinson's disease patients. Movement Disorders 2017;32(1):165-169.
29. Wang Y, Clark LN, Louis ED, et al. Risk of Parkinson disease in carriers of parkin mutations: estimation using the kin-cohort method. Arch Neurol 2008;65(4):467-474.
30. Camacho JLG, Jaramillo NM, Gómez PY, et al. High frequency of Parkin exon rearrangements in Mexican‐mestizo patients with early‐onset Parkinson's disease. Movement disorders 2012;27(8):1047-1051.
31. Huttenlocher J, Stefansson H, Steinberg S, et al. Heterozygote carriers for CNVs in PARK2 are at increased risk of Parkinson's disease. Hum Mol Genet 2015;24(19):5637-5643.
32. Pankratz N, Dumitriu A, Hetrick KN, et al. Copy number variation in familial Parkinson disease. PLoS One 2011;6(8).
33. Simon‐Sanchez J, Scholz S, del Mar Matarin M, et al. Genomewide SNP assay reveals mutations underlying Parkinson disease. Hum Mutat 2008;29(2):315-322.
34. Hopfner F, Mueller SH, Szymczak S, et al. Private variants in PRKN are associated with late-onset Parkinson's disease. Parkinsonism Relat Disord 2020;75:24-26.
35. Gan-Or Z, Rao T, Leveille E, et al. The Quebec Parkinson Network: A Researcher-Patient Matching Platform and Multimodal Biorepository. J Parkinsons Dis 2020;10:301-313.
36. Alcalay RN, Levy OA, Waters CC, et al. Glucocerebrosidase activity in Parkinson's disease with and without GBA mutations. Brain 2015;138(Pt 9):2648-2658.
37. Ruskey JA, Greenbaum L, Ronciere L, et al. Increased yield of full GBA sequencing in Ashkenazi Jews with Parkinson's disease. Eur J Med Genet 2019;62(1):65-69.
38. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. Journal of Neurology, Neurosurgery & Psychiatry 1992;55(3):181-184.
39. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. Mov Disord 2015;30(12):1591-1601.
40. 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.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

41. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 2009;25(14):1754-1760.
42. 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(9):1297-1303.
43. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 2010;38(16):e164-e164.
44. Rudakou U, Ruskey JA, Krohn L, et al. Analysis of common and rare VPS13C variants in late-onset Parkinson disease. Neurol Genet 2020;6(1):385.
45. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature 2016;536(7616):285-291.
46. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. Nat Biotechnol 2011;29(1):24.
47. Kopanos C, Tsiolkas V, Kouris A, et al. VarSome: the human genomic variant search engine. Bioinformatics 2018;35(11):1978-1980.
48. Zhao M, Wang Q, Wang Q, Jia P, Zhao Z. Computational tools for copy number variation (CNV) detection using next-generation sequencing data: features and perspectives. BMC Bioinformatics 2013;14(11):S1.
49. Plagnol V, Curtis J, Epstein M, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics 2012;28(21):2747-2754.
50. Povysil G, Tzika A, Vogt J, et al. panelcn. MOPS: Copy‐number detection in targeted NGS panel data for clinical diagnostics. Hum Mutat 2017;38(7):889-897.
51. 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(2):224-237.
52. Lee S, Teslovich TM, Boehnke M, Lin X. General framework for meta-analysis of rare variants in sequencing association studies. Am J Hum Genet 2013;93(1):42-53.
53. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. Bioinformatics 2010;26(17):2190-2191.
54. Yi W, MacDougall EJ, Tang MY, et al. The landscape of Parkin variants reveals pathogenic mechanisms and therapeutic targets in Parkinson's disease. Hum Mol Genet 2019;28(17):2811-2825.
55. Pavese N, Moore RY, Scherfler C, et al. In vivo assessment of brain monoamine systems in parkin gene carriers: a PET study. Exp Neurol 2010;222(1):120-124.
56. Khan NL, Scherfler C, Graham E, et al. Dopaminergic dysfunction in unrelated, asymptomatic carriers of a single parkin mutation. Neurology 2005;64(1):134-136.
57. Pavese N, Khan NL, Scherfler C, et al. Nigrostriatal dysfunction in homozygous and heterozygous parkin gene carriers: An 18F‐dopa PET progression study. Movement disorders: official journal of the Movement Disorder Society 2009;24(15):2260-2266.
58. Sardi SP, Simuni T. New Era in disease modification in Parkinson's disease: Review of genetically targeted therapeutics. Parkinsonism Relat Disord 2019;59:32-38.
59. Schneider SA, Alcalay RN. Neuropathology of genetic synucleinopathies with parkinsonism: review of the literature. Mov Disord 2017;32(11):1504-1523.
60. Dawson TM, Dawson VL. The role of parkin in familial and sporadic Parkinson's disease. Mov Disord 2010;25(S1):S32-S39.

medRxiv preprint doi: https://doi.org/10.1101/2020.05.07.20072728; this version posted August 20, 2020. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY 4.0 International license .

Figure legeneds

Figure 1. Flow chart of different analysis phases. The flow chart detail the total numbers of patients and controls included in different phases of the analysis. In red, the total number of samples sequenced. In green, the total numbers of samples which passed the quality control phase. In blue, the total numbers of samples after exclusion of 9 patients with biallelic pathogenic and likely pathogenic mutations in *PRKN*. In yellow, the total number of samples included in the analysis aadjusted for age, sex, ethnicity and the presence of *GBA* and *LRRK2* variants. In grey, the total number of samples included in the analysis after excluding additional samples with potentially pathoigenic biallelic copy number variations that could not be phased, i.e. samples with deletions of consecutive exons, for which we could not determine if they occur on the same allele or if they are biallelic.

	McGil	Columbia	Sheba	Total
Sequenced	Total: 3675 Case: 1126 Control: 2549	Total: 1522 Case: 1007 Control: 515	Total: 1241 Case: 676 Control: 565	Total: 6438 Case: 2809 Control: 3629
Standard QC	Total: 3490 Case: 1039 Control: 2451	Total: 1434 Case: 943 Control: 491	Total: 1175 Case: 654 Control: 521	Total: 6099 Case: 2636 Control: 3463
Exclude Biallelic	Total: 3485 Case: 1034 Control: 2451	Total: 1430 Case: 939 Control: 491	Total: 1175 Case: 654 Control: 521	Total: 6090 Case: 2636 Control: 3463
Statistical analysis exluding missing age, sex	Total: 2964 Case: 745 Control: 2219	Total: 1420 Case: 929 Control: 491	Total: 1139 Case: 632 Control: 507	Total: 5523 Case: 2306 Control: 3217
Statistical analysis exluding unknown phasing	Total: 2960 Case: 743 Control: 2217	Total: 1415 Case: 926 Control: 489	Total: 1137 Case: 631 Control: 506	Total: 5512 Case: 2300 Control: 3212