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

Eric Yu, BSc[1,2], Uladzislau Rudakov, 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], 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[13,14] 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. Department of Psychiatry, Université de Montréal, Montréal, Quebec, Canada
14. 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

Corresponding author:
Ziv Gan-Or
Department of Neurology and Neurosurgery
McGill University
1033 Pine Avenue, West, Ludmer Pavilion, room 312
Montreal, QC, H3A 1A1,
Phone: +1-514-398-5845
Email: ziv.gan-or@mcgill.ca
Word Count: 2,730 words.

Key Points

Question: Are rare heterozygous *PRKN* single nucleotide variants and copy number variations associated with Parkinson's disease?

Findings: In this case-controls study including 2,807 patients and 3,627 controls, we found no significant association between rare heterozygous *PRKN* variants and risk of Parkinson's disease.

Meaning: *PRKN*-associated parkinsonism is an autosomal recessive Mendelian disease, and based on our data, heterozygous carriers are not likely to be at increased risk of Parkinson's disease.

Abstract

Importance: 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 patients. However, the role of heterozygous *PRKN* variants in the risk of PD remains unclear.

Objective: To examine the association between heterozygous *PRKN* variants, including single nucleotide variants and copy number variations (CNVs), and PD status.

Design: This case-control study was conducted in Canada using data from 4 centers in Canada, France, the United States and Israel, collected between 2013 and 2020.

Setting: Multicenter, population matched, unrelated participants.

Participants: PD patients were compared with healthy controls. Participants were excluded if genetic data quality was poor, *PRKN* variants were biallelic, or relatedness was discovered. In total, 2,807 patients and 3,627 controls were recruited for the study. Data analyses were performed from January 2019 to March 2020.

Exposure: Inheritance of heterozygous *PRKN* variants.

Main Outcome(s) and Measure(s): PD was diagnosed by movement disorder specialists according to the UK Brain Bank Criteria or the Movement Disorder Society criteria. Targeted next-generation sequencing with molecular inversion probes and multiplex ligation-dependent probe amplification were used to detect rare variants and CNVs. These variants were examined with optimized sequence kernel association tests after accounting for potential pathogenicity. *P*

values of optimized sequence kernel association tests between heterozygous *PRKN* variants and PD were measured.

Results: DNA of all participants were sequenced, including 1,903 late onset (mean [SD], 64.02±7.81 years, 1,196 men [63%]) and 542 early onset patients (mean [SD], 43.30±6.60, 368 men [68%]). Age at onset was not available for 349 patients. Carriers of two *PRKN* variants were excluded from the analysis. No associations were found between heterozygous variants and risk of PD. Pathogenic and likely pathogenic heterozygous variants and CNVs were less common among PD patients (1.0%) than among controls (1.3%).

Conclusion and Relevance: This study suggests that heterozygous variants and CNVs in *PRKN* are not associated with PD. Molecular inversion probes allow for rapid and cost-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.

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 have early onset PD (EOPD), typically defined as AAO < 50 years.[2] The most common known 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 neuropathology of PD, 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 have not been clearly established by association studies.[12] Prior studies have shown contradictory results in familial PD, EOPD and late onset PD (LOPD) using SNVs and/or CNVs.[13-33]

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.

Materials and Methods

Study Population

A total of 2,807 unrelated and consecutively recruited PD patients and 3,627 controls from three cohorts were sequenced, including 1,903 LOPD patients (mean [SD], 64.02±7.81 years, 1,196 men [63%]) and 542 EOPD patients (mean [SD], 43.30±6.60, 368 men [68%]). AAO was not available for 349 patients. After excluding low quality samples, we performed statistical analysis on 6,121 individuals. The three cohorts are detailed in Table 1 and include: a) a cohort collected at McGill University, including French-Canadian (mostly recruited through the Quebec Parkinson Network)[34] and French participants recruited in Quebec, Canada and Montpellier, France, b) a cohort recruited at Columbia University, New York, as previously described[35], primarily composed of individuals of European origin and Ashkenazi Jews, and c) a cohort collected at the Sheba Medical Center, Israel, of Ashkenazi Jewish ancestry, as previously described.[36] PD was diagnosed by movement disorder specialists according to the UK Brain Bank Criteria, without excluding patients with positive family history[37] or the Movement Disorders Society Criteria.[38] 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

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.[39] 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),[40] Genome Analysis Toolkit (GATK v3.8),[41] and ANNOVAR[42] as previously described.[43] Only rare variants (minor allele frequency, MAF, < 0.01) according to the public database Genome Aggregation Database (GnomAD)[44] 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/.We examined all rare exonic variants using the Integrative Genomics Viewer (IGV v 2.7).[45] All variants were classified using Varsome[46] 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.[47] 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[48] and panelcn.MOPS v1.4.0 in R.[49] When using ExomeDepth, each test sample is compared to the best set of reference samples out of 3,627 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 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.

Statistical Analysis

The association between rare heterozygous variants (MAF<0.01), heterozygous CNV and PD were tested using an optimized sequence kernel association test (SKAT-O v1.3.2 in R)[50] in all cohorts separately. 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 was performed using MetaSKAT (MetaSKAT v0.80, R) [51] on all cohorts for heterozygous SNVs, CNVs, and both combined, according to the five ACMG categories (pathogenic, likely pathogenic, uncertain significance, likely benign and benign). The association between heterozygous SNVs, CNVs and AAO of PD was also calculated using linear regression

adjusted for sex and ancestry as needed in all cohorts separately. Patients carrying *GBA* variants or the *LRRK2* p.Gly2019Ser variant were excluded. METAL[52] was used to performed fixed-effect meta-analysis on all cohorts in the AAO analysis.

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 variants including nonsynonymous, frameshift deletions and splice site variants in *PRKN* across all cohorts (eTable 1 in the Supplement).

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 96% 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 eTable 2 in the Supplement. Due to its superior performance, we applied ExomeDepth on all cohorts, and identified a total of 53 carriers of CNVs in patients and controls (eTable 3 in the Supplement).

Heterozygous *PRKN* SNVs and CNVs are not associated with PD

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 this analysis too, adjusted for age, sex and ethnicity, no association between heterozygous carriage of *PRKN* mutations and PD was found (Table 2). The strongest association was found in the meta-analysis for pathogenic and likely pathogenic variants which were less frequent in patients (1%) than in controls (1.3%, $P = .006$, not statistically significant after Bonferroni correction, Table 2), ruling out association with risk of PD. Additional association tests between different types of heterozygous *PRKN* variants and risk for PD, all negative, can be found in eTable 4 in the Supplement.

Heterozygous *PRKN* SNVs and CNVs are not associated with AAO of PD

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 association study. After adjusting for sex and ancestry, 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 pathogenic SNVs and CNVs combined, the meta-analysis shows an AAO earlier in heterozygous *PRKN* carriers by 4.09 years, but the association was not statistically significant after correction for multiple comparisons. This was mainly driven by an effect of CNVs in the Columbia cohort, which was also not statistically significant on its

own after correction for multiple comparisons. Larger studies for AAO of heterozygous *PRKN* carriers are needed to validate these findings. Association between different types of heterozygous *PRKN* variants and AAO for PD can be found in eTable 5 in the Supplement.

Identification of *PRKN*-associated parkinsonism patients

Overall, we were able to identify 11 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 (23%) *PRKN* patients, and the most common CNV was heterozygous deletion of exon 3, found in 8 (62%) *PRKN* patients. The average AAO of PD in biallelic *PRKN* SNV/CNV carriers was 34.27 [±14.67] years old.

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. 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 53 participants with *PRKN* CNVs, with very high sensitivity and specificity (97% and 96%, 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 11 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 PD 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 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 eTable 6 in the Supplement. 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 eTable 6 in the Supplement.[13,14,20,23,31,33] 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.[53]

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.[54] 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.[55] 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.[56] Since we did not detect an association between heterozygous *PRKN* variants and PD, we recommend that only

biallelic *PRKN* variant carriers will be excluded from such trials, as it is likely that the presence of heterozygous *PRKN* variants in PD patients is by 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.[55] 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 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.

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.

Table 1. Study populations included in the analysis

	McGill University		Columbia University		Sheba Medical Center	
	Patients (n=1,046)	Controls (n=2,456)	Patients (n=943)	Controls (n=491)	Patients (n=661)	Controls (n=524)
Age, y[a,b]	58.9 (10.7)	53.4 (14.1)	59.2 (11.8)	64.6 (9.9)	60.4 (11.9)	33.7 (7.8)
Male, No. (%)	663 (64)	1162 (47)	613 (65)	177 (36)	413 (62)	307 (59)

[a] Data are presented as mean (SD).
[b] Data for age are missing for 324 (12.2%) patients. Difference of age and sex between patients and controls were adjusted in our analysis.

Table 2. Rare *PRKN* heterozygous SNV and CNV analysis for risk of PD using SKAT-O

	Carriers in patients, No. (%)	Carriers in controls, No. (%)	*P* value[a]
McGill University (n=3,417)			
SNV	100 (9.9)	213 (8.8)	.32
CNV	2 (0.19)	8 (0.33)	.90
SNV & CNV	102 (10.1)	221 (9.2)	.33
Patho SNV	12 (1.2)	21 (0.87)	.64
Patho SNV & CNV	14 (1.4)	29 (1.2)	.22
No Benign SNV	84 (8.3)	159 (6.6)	.73
No Benign SNV & CNV	86 (8.5)	167 (6.9)	.22
Columbia University (n=1,396)			
SNV	85 (9.2)	37 (7.7)	.20
CNV	7 (0.76)	8 (1.6)	.07
SNV & CNV	92 (10.0)	45 (9.3)	.16
Patho SNV	5 (0.54)	5 (1.0)	.12
Patho SNV & CNV	12 (1.3)	13 (2.7)	.13
No Benign SNV	69 (7.5)	26 (5.4)	.008
No Benign SNV & CNV	76 (8.3)	34 (7.1)	.11
Sheba Medical Center (n=1,172)			
SNV	45 (6.9)	23 (4.4)	.02
CNV	1 (0.15)	1 (0.19)	.61
SNV & CNV	46 (7.0)	24 (4.6)	.03
Patho SNV	0 (0.0)	1 (0.19)	.61
Patho SNV & CNV	1 (0.15)	2 (0.39)	.02
No Benign SNV	42 (6.4)	21 (4.0)	.69
No Benign SNV & CNV	43 (6.6)	22 (4.2)	.03
Meta-Analysis (n=5,985)			
SNV	230 (8.9)	273 (8.0)	.03
CNV	10 (0.39)	17 (0.50)	.18
SNV & CNV	240 (9.3)	290 (8.5)	.04
Patho SNV	17 (0.66)	27 (0.79)	.57
Patho SNV & CNV	27 (1.0)	44 (1.3)	.006
No Benign SNV	195 (7.6)	206 (6.0)	.21
No Benign SNV & CNV	205 (8.0)	223 (6.5)	.01

Abbreviations: SNV—Single Nucleotide Variant; CNV—Copy Number Variation; PD—Parkinson's Disease; Patho—pathogenic and likely pathogenic variants; No Benign—analysis excluding benign and likely benign variants.

[a] After Bonferroni correction, alpha = 0.0018 (28 tests)

Table 3. Rare *PRKN* heterozygous SNV and CNV analysis for Age at Onset of PD

	AAO of carriers, y mean (SD)	AAO of non-carriers, y mean (SD)	Coeff (95% CI)	*P* values[a]
McGill University (n=664)				
SNV	61.00 (11.17)	58.88 (10.55)	2.13 (-0.50, 4.76)	.11
CNV	57.00 (22.63)	59.11 (10.61)	-2.73 (-17.55, 12.09)	.72
SNV & CNV	60.89 (11.36)	58.89 (10.53)	1.99 (-0.61, 4.59)	.13
Patho SNV	56.38 (15.16)	59.14 (10.57)	-2.68 (-10.11, 4.74)	.48
Patho SNV & CNV	56.50 (15.35)	59.15 (10.55)	2.01 (-0.85, 4.88)	.47
No Benign SNV	60.95 (10.80)	58.93 (10.60)	-2.70 (-0.85, 4.88)	.42
No Benign SNV & CNV	60.82 (11.04)	58.94 (10.58)	1.85 (-0.97, 4.67)	.20
Columbia University (n=733)				
SNV	58.39 (12.94)	59.44 (11.55)	-1.04 (-3.91, 1.82)	.47
CNV	46.50 (9.77)	59.44 (11.64)	-12.94 (-22.31, -3.56)	.007
SNV & CNV	57.47 (13.07)	59.56 (11.50)	-2.09 (-4.85, 0.67)	.14
Patho SNV	60.40 (18.86)	59.33 (11.64)	1.04 (-9.29, 11.36)	.84
Patho SNV & CNV	52.82 (15.59)	59.44 (11.60)	1.08 (-2.11, 4.27)	.51
No Benign SNV	60.34 (12.03)	59.25 (11.66)	-6.63 (-13.59, 0.33)	.06
No Benign SNV & CNV	59.00 (12.46)	59.37 (11.62)	-0.37 (-3.42, 2.68)	.81
Sheba Medical Center (n=442)				
SNV	57.22 (13.38)	61.82 (12.22)	-4.71 (-9.50, 0.07)	.05
CNV	69 (NA)	61.52 (12.34)	8.18 (-16.07, 32.43)	.51
SNV & CNV	57.64 (13.32)	61.80 (12.23)	-4.24 (-8.96, 0.47)	.08
Patho SNV	NA	61.54 (12.33)	NA	NA
Patho SNV & CNV	69 (NA)	61.52 (12.34)	-4.74 (-9.71, 0.23)	.06
No Benign SNV	57.16 (13.68)	61.80 (12.21)	8.18 (-16.07, 32.43)	.51

No Benign SNV & CNV	57.62 (13.61)	61.78 (12.22)	-4.23 (-9.11, 0.65)	.09
Meta-Analysis (n=1,839)				
SNV	59.29 (12.33)	59.83 (11.43)	-0.08 (-1.88, 1.71)	.93
CNV	51.33 (13.72)	59.82 (11.49)	-8.26 (-15.78, -0.75)	.03
SNV & CNV	58.89 (12.48)	59.88 (11.41)	-0.53 (-2.28, 1.23)	.56
Patho SNV	57.93 (12.48)	59.88 (11.48)	-1.42 (-7.43, 4.60)	.64
Patho SNV & CNV	55.23 (15.15)	59.84 (11.46)	-4.09 (-8.80, 0.62)	.09
No Benign SNV	60.02 (11.85)	59.76 (11.49)	0.61 (-1.35, 2.57)	.54
No Benign SNV & CNV	59.49 (12.10)	59.81 (11.46)	0.05 (-1.85, 1.96)	.96

Abbreviations: SNV—Single Nucleotide Variant; CNV—Copy Number Variation; PD—Parkinson's Disease; AAO—Age At Onset; Coeff—Regression Coefficient; CI—Confidence Interval; Patho—pathogenic and likely pathogenic variants; No Benign—analysis excluding benign and likely benign variants. NA—Not Applicable.

[a] After Bonferroni correction, alpha = 0.0018 (28 tests)

Table 4: PD patients with biallelic pathogenic and likely pathogenic *PRKN* SNVs and CNVs.

Sample	SNV	CNV	Sex	AAS	AAO
McGill cohort					
S29243	p.Gly430Asp	Exon 9 duplication	M	38	28
S21069	p.Gln34ArgfsTer5, p.Arg42Pro		F	73	29
S35319		Homozygous exon 3 deletion	M	35	29
S06128	p.Gln34ArgfsTer5	Exon 3, 4 deletion	M	44	39
S29254		Homozygous exon 3 deletion	M	57	NA
S29263	p.Arg275Trp	Exon 3 deletion	F	41	NA
Columbia cohort					
S23082		Homozygous exon 3, 4 deletion	F	54	16
S22919	p.Gln34ArgfsTer5	Exon 3, 4 deletion	M	52	19
S23324	p.Arg275Trp	Exon 3 deletion	F	70	28
S33552	p.Arg366Trp, p.Met458Leu	Exon 2 deletion	M	36	34
S33346		Homozygous exon 3, 4 deletion	M	37	37

Sample are sorted by age at onset with one SNV or CNV per locus at each line.
Abbreviations: SNV—Single nucleotide Variants; CNV—Copy Number Variation; Zyg—Zygosity of mutation; AAS—Age At Sample; AAO—Age At Onset; Het—Heterozygote; Hom—Homozygote; M—Male; F—Female; NA—Not Available.

Acknowledgments

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, Clotilde Degroot and Vessela Zaharieva for their assistance.

Author Contributions:

Eric Yu and Ziv Gan-Or have full access to all the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis.

Study concept and design: Yu, Gan-Or.

Acquisition, analysis, or interpretation of data: All authors.

Drafting of the manuscript: Yu, Gan-Or.

Critical revision of the manuscript for important intellectual content: All authors.

Statistical analysis: Yu.

Obtained funding: Alcalay, Gan-Or

Administrative, technical, or material support: Ruskey, Asayesh, Spiegelman, Gan-Or.

Study supervision: Gan-Or

Conflict of Interest Disclosures:

Dr. Fahn received consulting fees/honoraria for board membership from Retrophin Inc., Sun Pharma Advanced Research Co., LTD and Kashiv Pharma. Dr. Waters received rsearch support from Sanofi, Biogen, Roche, consulting fees/honoraria from Amneal, Adamas, Impel, Kyowa, Mitsubishi, Neurocrine, US World Meds, Acadia, Acorda. Dr. Espay 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. Dr. Dupré received consultancy fees from Actelion Pharmaceuticals. Dr. Hassin-Baer received consulting fees from Actelion Pharmaceuticals Ltd., Abbvie Israel, Robotico Ltd., Medtronic Israel, Medison Pharma Israel. Dr. Fon received consulting fees from Inception Sciences. Dr. Alcalay received consultation fees from Biogen, Denali, Genzyme/Sanofi and Roche. Dr. Gan-Or received consultancy fees from Lysosomal Therapeutics Inc. (LTI), Idorsia, Prevail Therapeutics, Inceptions Sciences (now Ventus), Ono Therapeutics, Denali and Deerfield. No other disclosures were reported.

Funding/Support

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.

Role of the Funder/Sponsor

The funders had no role in any part of the study design, analysis, interpretation and drafting the manuscript.

Data Sharing Statement

The data that support the findings of this study are available from the corresponding author upon reasonable request.

Reference

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. In: *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. 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.
14. 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.
15. 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.
16. Brooks J, Ding J, Simon-Sanchez J, et al. 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.
17. 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.
18. 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.
19. 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).
20. 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.

21. 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.
22. 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.
23. 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.
24. 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.
25. 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.
26. Sironi F, Primignani P, Zini M, et al. Parkin analysis in early onset Parkinson's disease. *Parkinsonism Relat Disord.* 2008;14(4):326-333.
27. 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.
28. 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.
29. 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.
30. 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.
31. 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.
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. 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.
35. 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.
36. 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.
37. 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.
38. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord.* 2015;30(12):1591-1601.
39. 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.
40. Li H, Durbin R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics.* 2009;25(14):1754-1760.
41. 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.

42. 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.
43. 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.
44. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature.* 2016;536(7616):285-291.
45. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. *Nat Biotechnol.* 2011;29(1):24.
46. Kopanos C, Tsiolkas V, Kouris A, et al. VarSome: the human genomic variant search engine. *Bioinformatics.* 2018;35(11):1978-1980.
47. 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.
48. 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.
49. 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.
50. 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.
51. 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.
52. Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics.* 2010;26(17):2190-2191.
53. 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.
54. 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.
55. Schneider SA, Alcalay RN. Neuropathology of genetic synucleinopathies with parkinsonism: review of the literature. *Mov Disord.* 2017;32(11):1504-1523.
56. Dawson TM, Dawson VL. The role of parkin in familial and sporadic Parkinson's disease. *Mov Disord.* 2010;25(S1):S32-S39.