Assessment of private variants in *PRKN, PARK7* and *PINK1* in Parkinson's disease

Jing Hu1,2, Cheryl H. Waters3, Dan Spiegelman4, Edward A. Fon4,5, Eric Yu4,6, Farnaz Asayesh4,5, Lynne Krohn4,6, Prabhjyot Saini4,6, Roy N. Alcalay3,7, Sharon Hassan-Baer4,8,9, Ziv Gan-Or4,5,6, Dimitri Krainc2, BaoRong Zhang1*, Bernabe I. Bustos2*, Steven J. Lubbe2*, t

1 Department of Neurology, The Second Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou, Zhejiang, 310009, China.
2 Ken and Ruth Davee Department of Neurology and Simpson Querrey Center for Neurogenetics, Northwestern University Feinberg School of Medicine, Chicago, Illinois, USA.
3 Department of Neurology, Columbia University Irving Medical Center
4 The Neuro (Montreal Neurological Institute-Hospital), Montreal, Quebec, Canada.
5 The Department of Neurology & Neurosurgery, McGill University, Montreal, Quebec, Canada.
6 The Department of Human Genetics, McGill University, Montreal, Quebec, Canada.
7 Center of Movement Disorders, Tel Aviv Sourasky Medical Center, Tel Aviv, Israel.
8 The Movement Disorders Institute, Department of Neurology, Chaim Sheba Medical Center, Tel-Hashomer, Israel.
9 The Sackler Faculty of Medicine, Tel Aviv University, Tel Aviv, Israel.

# these authors contributed equally

Correspondence to: Bernabe I. Bustos (bernabe.bustos@northwestern.edu) and BaoRong Zhang (brzhang@zju.edu.cn).

Abstract

Recessive mutations in *PRKN, PARK7* and *PINK1* are established causes of early-onset Parkinson’s disease (EOPD). Previous studies have interrogated the role of heterozygous variants in these genes but mainly focused on rare (minor allele frequency [MAF] <1%) damaging variants or established mutations. Here, we assessed heterozygous private *PRKN, PARK7* and *PINK1* variants in PD risk in four large-scale PD case-control datasets by performing gene-wise burden analyses using sequencing data totaling 5,831 PD cases and 7,221 controls, and summary allele counts from 9,506 PD cases and 48,207 controls. Results showed no significant burden in all three genes after meta-analyses. Burden in EOPD (age at onset <50 years) and late-onset PD (≥50 years) remained nonsignificant. In summary, we found no evidence to support the association of the excess burden
of heterozygous private variants in PRKN, PARK7, and PINK1 with PD risk in the European population. Larger, more diverse cohorts are needed to accurately determine their role in PD.

Keywords: private variants, Parkinson’s disease, PRKN, PARK7, PINK1, genetics

Introduction

Biallelic variants in PRKN (Parkin), PARK7 (DJ-1) and PINK1 cause early-onset Parkinson’s disease (EOPD, defined here as age at onset [AAO] <50 years) (Bonifati 2012; Kilarski et al. 2012; Kasten et al. 2018). The role of heterozygous mutations is more controversial. Several recent large-scale studies did not uncover any association for heterozygous mutations in these genes with PD (Yu et al. 2021; Krohn et al. 2020; Lubbe et al. 2021; Zhu et al. 2021). These studies focused on rare (minor allele frequency [MAF] <1%) damaging variants or established mutations, and largely ignored the specific role of private variants that are only seen in single individuals within a cohort and are absent or very rare in public population databases such as gnomAD (https://gnomad.broadinstitute.org; (Karczewski et al. 2020)). An increased burden of rare (MAF <1%) and private heterozygous PRKN variants in late-onset PD (LOPD, AAO ≥50 years) was recently reported (Hopfner et al. 2020). Here, we analyzed heterozygous private variants in PRKN, PARK7 and PINK1 using four large PD cohorts to assess their role in PD risk.

Methods

Sequencing data were obtained from three independent PD case-control cohorts totaling 5,831 PD cases and 7,221 controls, and summary count data from 9,506 PD cases and 48,207 controls was extracted from the PD Variant Browser (Kim et al. 2021) and two of the sequencing cohorts (International Parkinson’s disease Genomics Consortium (IPDGC)-whole exome sequencing (WES) and McGill). Cohort information is described in Supplementary Table 1. Heterozygous private PRKN, PARK7 and PINK1 variants were prioritized using, as a guideline, a MAF<1% threshold in >141,000 population-based individuals (Karczewski et al. 2020). Damaging private
variants were identified as described in Supplementary Methods. To avoid overlap between the sequencing cohorts and those used to extract allele counts, different analytical methods were used. R package MetaSKAT (Lee et al. 2013) and Metafor (Viechtbauer 2010) were utilized for meta-analysis of genotyping data and allele counts data, respectively. Bonferroni correction based on the number of genes was applied for multiple testing. See Supplementary Methods for more details.

Results

Using available sequencing data, we first performed a gene-based Optimized Sequence-kernel-association test (SKAT-O) (Lee et al. 2012) to assess the burden of all heterozygous private and private damaging variants that yielded no significant association for all genes in individual cohort analysis, nor after meta-analysis (Table 1). Similarly, after stratifying PD cases into EOPD and LOPD, no significant enrichment was observed (Supplementary Tables 2 and 3).

To further assess the contribution of PRKN, PARK7 and PINK1 heterozygous private variants, we next extracted allele counts and performed an allelic gene-based Fisher's exact burden test. Again, no significant associations at the individual cohort level were observed for both private and private damaging variants (Supplementary Table 4). After meta-analysis, no significant results were found for private variants (Supplementary Figure 1) and private potentially damaging variants (Figure 1) in the three genes. After stratifying PD cases into EOPD and LOPD, private and private potentially damaging PRKN variants demonstrated some level of enrichment in the McGill EOPD New York (NY) cohort (Supplementary Table 4), however, these findings no longer remained after meta-analysis (Supplementary Figures 2-5).

Discussion

Here we assessed the contribution of heterozygous private variants within PRKN, PARK7 and PINK1 to PD risk in several large European ancestry cohorts. Overall, no enrichment across all genes was observed for individual cohorts and upon meta-analysis for both the sequencing and allele count data, and also not after stratifying by early and late onset PD cases. Together this
suggests that private \textit{PRKN}, \textit{PARK7} and \textit{PINK1} variants do not influence PD risk in the European population. While in agreement with previous studies assessing heterozygous mutations (Yu et al. 2021; Krohn et al. 2020; Lubbe et al. 2021; Zhu et al. 2021), our findings are not consistent with those of Hopfner et al. (2020). Differences in genotyping methodologies (pooled DNA sequencing and individual genotype simulation compared to whole-exome, genome, or targeted sequencing in our case) could explain this discordance (see Supplementary Discussion). Additionally, cohort characteristics and failure to assess all possible private variants, which can influence \textit{PRKN} burden assessments (Yu et al. 2021; Krohn et al. 2020; Lubbe et al. 2021; Zhu et al. 2021), may also play a role. Despite our study being well powered to detect significant gene-based variant enrichment (power=85.7\%; \alpha=0.05; based on 3.4-fold enrichment and gene-based carriership of 0.03\% in controls (Hopfner et al. 2020) and global disease prevalence of 0.2\% (Keller et al. 2012)), we failed to uncover evidence that private \textit{PRKN}, \textit{PARK7} and \textit{PINK1} variants influence PD risk in the European population. Additional studies of private variants in more diverse populations are needed to further understand their role in PD.
Figure 1. Meta-analysis of carriership burden of private damaging variants in PRKN, PARK7 and PINK1 genes in Parkinson’s disease (PD) using allele count data.

Forest plots assessing private, potentially damaging variants (CADD >20) in (A) PRKN (Parkin), (B) PARK7 (DJ-1) and (C) PINK1 genes. Allele counts for each gene were extracted from the PD Variant Browser, International Parkinson’s disease Genomics Consortium (IPDGC) Whole Exome Sequencing Project cohort and individual McGill cohorts separately. Meta-analyses were performed using Metafor. Panels show Odds ratio, confidence intervals (CI), corrected Fisher’s exact test gene burden p-value per cohort (right side of each plot), and the overall meta-analysis p-value per gene (bottom left of each plot).

Key: CADD=Combined Annotation Dependent Depletion phred-scaled score; GENOME=PD Whole Genome Project cohort; RESEQ=IPDGC Resequencing Project cohort; UKB=UK Biobank cohort; EXOME=IPDGC Whole Exome Sequencing Project cohort; ISR=McGill Israel cohort; FC=McGill French/French-Canadian cohort; NY=McGill Columbia University Spot Study; FE=fixed effects; I²=I-square test for heterogeneity.
Table 1. Assessment of private variants in PRKN, PARK7 and PINK1 with Parkinson’s disease (PD) in cohorts with available sequencing data.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Variant Group</th>
<th>SKAT-O (p-values)</th>
<th></th>
<th></th>
<th>MetaSKAT</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>AMP-PD</td>
<td>IPDGC-WES</td>
<td>ISR</td>
<td>FC</td>
</tr>
<tr>
<td>PRKN</td>
<td>Private</td>
<td>0.41</td>
<td>0.51</td>
<td>0.65</td>
<td>0.81</td>
</tr>
<tr>
<td></td>
<td>Private damaging</td>
<td>0.57</td>
<td>0.36</td>
<td>0.65</td>
<td>0.42</td>
</tr>
<tr>
<td>PINK1</td>
<td>Private</td>
<td>0.13</td>
<td>0.24</td>
<td></td>
<td>0.23</td>
</tr>
<tr>
<td></td>
<td>Private damaging</td>
<td>0.21</td>
<td>0.24</td>
<td></td>
<td>0.23</td>
</tr>
<tr>
<td>PARK7</td>
<td>Private</td>
<td>0.53</td>
<td>0.26</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Private damaging</td>
<td>0.49</td>
<td>0.37</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sequence-kernel-association test (SKAT-O) was used to assess variant burden for individual cohort analysis and MetaSKAT was used to perform the meta-analysis. Uncorrected P-values for the SKAT-O tests are indicated. Corrected P-values for the meta-analyses are indicated in parentheses.

Key: AMP-PD=Accelerating Medicine Partnership in Parkinson’s disease cohort; IPDGC-WES=International Parkinson’s disease Genomics Consortium whole exome sequencing cohort; ISR=McGill Israel cohort; FC=McGill French-Canadian cohort; NY=McGill
Columbia University Spot Study.
Reference