



Heterozygous *PRKN* mutations are common but do not increase the risk of Parkinson's disease

William Zhu,¹ Xiaoping Huang,¹ Esther Yoon,² Sara Bandres-Ciga,³
Cornelis Blauwendraat,^{4,5} Kimberly J. Billingsley,³ Joshua H. Cade,¹ Beverly P. Wu,¹
Victoria H. Williams,¹ Alice B. Schindler,⁶ Janet Brooks,³ J. Raphael Gibbs,³
Dena G. Hernandez,³ Debra Ehrlich,² Andrew B. Singleton^{3,5} and Derek P. Narendra¹

PRKN mutations are the most common recessive cause of Parkinson's disease and are a promising target for gene and cell replacement therapies. Identification of biallelic *PRKN* patients at the population scale, however, remains a challenge, as roughly half are copy number variants and many single nucleotide polymorphisms are of unclear significance. Additionally, the true prevalence and disease risk associated with heterozygous *PRKN* mutations is unclear, as a comprehensive assessment of *PRKN* mutations has not been performed at a population scale.

To address these challenges, we evaluated *PRKN* mutations in two cohorts with near complete genotyping of both single nucleotide polymorphisms and copy number variants: the NIH-PD + AMP-PD cohort, the largest Parkinson's disease case-control cohort with whole genome sequencing data from 4094 participants, and the UK Biobank, the largest cohort study with whole exome sequencing and genotyping array data from 200 606 participants. Using the NIH-PD participants, who were genotyped using whole genome sequencing, genotyping array, and multi-plex ligation-dependent probe amplification, we validated genotyping array for the detection of copy number variants. Additionally, in the NIH-PD cohort, functional assays of patient fibroblasts resolved variants of unclear significance in biallelic carriers and suggested that cryptic loss of function variants in monoallelic carriers are not a substantial confounder for association studies.

In the UK Biobank, we identified 2692 *PRKN* copy number variants from genotyping array data from nearly half a million participants (the largest collection to date). Deletions or duplications involving exon 2 accounted for roughly half of all copy number variants and the vast majority (88%) involved exons 2, 3, or 4. In the UK Biobank, we found a pathogenic *PRKN* mutation in 1.8% of participants and two mutations in ~1/7800 participants. Those with one *PRKN* pathogenic variant were as likely as non-carriers to have Parkinson's disease [odds ratio = 0.91 (0.58–1.38), *P*-value 0.76] or a parent with Parkinson's disease [odds ratio = 1.12 (0.94–1.31), *P*-value = 0.19]. Similarly, those in the NIH-PD + AMP + PD cohort with one *PRKN* pathogenic variant were as likely as non-carriers to have Parkinson's disease [odds ratio = 1.29 (0.74–2.38), *P*-value = 0.43].

Together our results demonstrate that heterozygous pathogenic *PRKN* mutations are common in the population but do not increase the risk of Parkinson's disease.

1 Inherited Disorders Unit, Neurogenetics Branch, Division of Intramural Research, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892-3705, USA

2 Parkinson's Disease Clinic, Office of the Clinical Director, National Institute of Neurological, Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892-3705, USA

3 Molecular Genetics Section, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892-3705, USA

4 Integrative Neurogenomics Unit, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of

Health, Bethesda, MD 20892-3705, USA

5 Center for Alzheimer's and Related Dementias, National Institutes of Health, Bethesda, MD 20892-3705, USA

6 National Institute of Neurological Disorders and Stroke, Neurogenetics Branch, National Institutes of Health, Bethesda, MD 20892-3705, USA

Correspondence to: Dr Derek P. Narendra

35 Convent Drive

Building 35 Room 2A-209

Bethesda, MD 20892-3705 USA

E-mail: derek.narendra@nih.gov

Keywords: PARK2; parkin; mitophagy; early onset Parkinson's disease; young onset Parkinson's disease

Abbreviations: CNV = copy number variant; SNP = single nucleotide polymorphism; UPDRS = Unified Parkinson's Disease Rating Scale; UPSIT = University of Pennsylvania smell identification test; VUS = variant of unclear significance; WES = whole-exome sequencing

Introduction

Parkinson's disease is the second most common neurodegenerative disorder with a prevalence of about 0.5% in individuals ≥ 45 years of age.¹ Five to ten per cent of cases are caused by mutation(s) in a single gene.² Mutations in *PRKN* are the most common recessive form of Parkinson's disease (*PRKN*-PD), and *PRKN*-PD, in particular, is a promising target for gene and cell replacement therapies.^{3–5} Patients with *PRKN*-PD typically have disease onset before the age of 40 years.⁶

For gene-targeted trials, the target gene needs to be identified in a large group of individuals, ideally early in the disease course. Cost-effective genotyping platforms, notably genotyping arrays, have facilitated screening for mutation carriers for clinical trials, sometimes through partnership with consumer-based genotyping companies.⁷ Genotyping arrays are particularly effective for mutations such as *APOE* E4 or *LRRK2* p.G2019S, in which the locus can be genotyped from one or a few single nucleotide polymorphisms (SNPs).

Identification of individuals with recessive disorders like *PRKN*-PD, however, presents additional challenges. Recessive disorders are typically caused by loss of function variants that may be scattered throughout the gene, two pathogenic variants must be identified to establish causality, and both SNPs and copy number variants (CNVs) may lead to loss of function, necessitating analytic methods that can detect both. The challenges of identifying *PRKN* mutations at the population scale have several implications. Because large scale studies do not typically capture all *PRKN* SNPs and/or all CNVs, the true prevalence of *PRKN* mutations in the general population is not known, with estimates ranging from 0.17–3.7%.⁸ Additionally, a large cohort study has suggested that single *PRKN* mutation carriers have an increased risk of developing idiopathic Parkinson's disease.⁹ This study, however, may be confounded by cases with a missed second *PRKN* mutation. Notably, other smaller studies with a case-control design have failed to find this association, and a recent meta-analysis suggested that the association may depend on a missed mutation in biallelic variant carriers.¹⁰ The case-control design has some limitations relative to a cohort design, as an appropriate control group must be matched to the disease group. Improving genotyping of *PRKN* mutations at the population scale may help to clarify the prevalence of *PRKN* in the general population and the risk of idiopathic Parkinson's disease conferred by a single *PRKN* mutation, using a cohort study design with near complete genotyping of *PRKN*.

To help address these challenges, we assessed the frequency and risk conferred by pathogenic *PRKN* mutations in two large cohorts in which the *PRKN* gene was nearly fully genotyped: the NIH-PD + AMP-PD cohort, which is the largest Parkinson's disease whole genome sequencing (WGS) study, and the UK Biobank, which is the largest cohort study. Although smaller and with a case-control design, the NIH subjects in the NIH-PD + AMP-PD cohort allowed validation of *PRKN* CNV genotyping by array and assessment of potential missing second mutations through functional assessment of patient fibroblasts in single *PRKN* mutation carriers. We then assessed *PRKN* mutations in the UK Biobank, a large population study with genotyping array data available from nearly half a million participants and whole exome sequencing (WES) data from ~200 000 participants.¹¹ This allowed us to assess for the first time the frequency of *PRKN* mutations in the population and the risk conferred by a single heterozygous *PRKN* mutation with near complete genotyping of *PRKN*, using a cohort study design. Together our results show that there is a high prevalence of heterozygous pathogenic *PRKN* mutations in the general population and that they do not increase Parkinson's disease risk.

Materials and methods

NIH-PD and AMP-PD cohorts

To establish the NIH-PD cohort, study participants were recruited to the Parkinson's Disease Clinic of the National Institute of Neurological Disorders and Stroke (NINDS) and the National Institutes of Health (NIH) Clinical Centre between the years 2006 and 2019. All participants gave written informed consent according to the Declaration of Helsinki to protocols approved by the Institutional Review Board of NINDS before undergoing research procedures. All patients were evaluated by a board-certified neurologist with specialized training in movement disorders, using the full Unified Parkinson's Disease Rating Scale (UPDRS). Many were also tested for olfaction using the University of Pennsylvania smell identification test (UPSIT) and screened for cognitive deficits using the Montreal cognitive assessment. Diagnosis of Parkinson's disease was based on the UK Parkinson's Disease Society Brain Bank Diagnostic Criteria. Patients with an identified known pathogenic mutation were invited back to the NIH for genetic counselling and CLIA certified testing. The AMP-PD cohort consisted of those individuals with publicly available WGS sequencing

data made available through the AMP PD v1_release and included participants from several multi-centre studies, including BioFIND, Parkinson's Disease Biomarkers Program, Parkinson's Progression Markers Initiative, and the Harvard Biomarker Study, as described previously.¹² Patient characteristics of the NIH-PD + AMP-PD cohorts are shown in (Supplementary Table 1).

Functional analysis of patient fibroblasts

Patient fibroblast lines were established from 3-mm punch biopsies taken from the forearm. Cell lines were assayed at passage 11 or before. Where indicated, additional lines indicated were obtained from the NINDS Human Cell and Data Repository (<https://stemcells.ninds.genetics.org/>). For analysis of MFN1 and pS65-Ub, cells treated with DMSO or valinomycin 10 mM overnight (>16 h) were washed in PBS and pelleted. Cell pellets were lysed in RIPA buffer on ice for 30 min and then cleared by centrifugation at 21 130g. Protein concentration was determined using the BCA assay. Twenty micrograms of protein were loaded for each sample. Samples were separated on a 7.5% Criterion TGX precast midi protein gel (Biorad) and proteins were transferred to nitrocellulose membranes. Membranes were blotted with MFN1 using an antibody from Proteintech (cat no. 13798-1-AP). Some of the same lysates were blotted for pS65-ubiquitin (Sigma, cat no. ABS1513). Indicated patient fibroblast lines were transduced with mt-Keima and analysed by flow cytometry as previously reported.¹³ For examination of PRKN exon 1–exon 2 splicing, RNA was isolated from the indicated fibroblast lines using the Direct-zol RNA miniprep kit (RPI research products, cat no. ZR2051) and reverse transcribed using a kit from ThermoFisher (cat no. 4368814). The PRKN exon 1–3 product was amplified with MyTaq DNA polymerase (Bioline) from 60 ng of cDNA using the following primers: 5'-aggatttaaccaggagagc-3' and 5'-aatgctctctgatccaggt-3'. Beta-actin was amplified using primers from ThermoFisher (cat no. Hs01060665_g1).

Genotyping for NIH-PD and combined NIH-PD + AMP-PD cohorts

Genotyping of the NIH-PD cohort was performed using a genome-wide coverage genotyping array (NeuroX or Neuro Consortium Array, Illumina, Inc.) and/or WGS. The NeuroX array is based on the Illumina Human Exome array v1.1 and the NeuroChip array is based on the Infinium HumanCore-24 c1.0 array.^{14,15} Both have additional custom content covering neurodegenerative disease-related variants. To identify SNPs from the genotyping array, Illumina GenomeStudio (v.2.0) was used to cluster genotypes. Quality-control measures included limiting to samples with call rates of >95%, excluding samples with excess heterogeneity (F statistic $\geq \pm 0.25$), and excluding samples whose genotyped sex did not match the sample demographics. CNVs were identified by manual inspection of the B allele frequency and Log R ratio for the PRKN gene region (Chr6: 161770811–163140694, hg19), using the gglot2 visualization package for R (<https://www.r-project.org/>), as described previously.¹⁶

For WGS, 1 mg of total genomic DNA was sheared to a target size of 450 base pairs (bp) by ultrasonication. The library was prepped with the TruSeq DNA PCR-free high throughput library prep kit and IDT for Illumina TruSeq DNA UD Indexes (96 indexes, 96 samples). Samples were sequenced on an Illumina HiSeq X system using 150 bp paired end reads. Called SNPs were annotated using ANNOVAR.¹⁷ In the initial analysis of the NIH-PD cohort, CNVs in the PRKN region were detected using the Manta structural variant

caller (Illumina).¹⁸ For cases with available DNA, detected PRKN CNVs were verified by multiplex ligation dependent probe amplification with SALSA MLPA Probemixes P051 and P052 (MRC Holland).

For analysis of the NIH-PD + AMP-PD cohorts, CNVs were called from a harmonized WGS dataset that included participants from NIH-PD and AMP-PD with available WGS data, using the gatk-sv caller (<https://github.com/broadinstitute/gatk-sv>).¹⁹ SNPs were called as described previously.¹² Quality-control measures included excluding samples whose genotyped sex did not match the sample demographics. One participant (NIHPD_B3 in Supplementary Table 3), who came from the NIH-PD cohort, was identified as having three pathogenic mutations (two deletions and R275W variant). Comparison to multiplex ligation dependent probe amplification and microarray results suggested that an additional exon 3 deletion had been miscalled within the exon 3–4 deletion. This participant's genotype was corrected prior to analysis. Ancestry of participants was determined by clustering of genetic SNPs with HapMap3 populations, using principal component analysis (PCA).

UK Biobank data

To identify CNVs in the UK Biobank data,²⁰ the UK Biobank B allele frequency and Log R ratio data were downloaded from UK Biobank (v2), containing 488 377 participants. B allele frequency and Log R ratio data were extracted for the PRKN gene region (Chr6: 161770811–163140694, hg19). Potential CNVs were detected using three algorithms. The first, PRKN deletion finder 1, averaged the Log R ratio for SNPs within or immediately flanking each exon of PRKN. It flagged an exon as possibly deleted if the exon average was 1.5 SD less than the average of all the exons or possibly duplicated if it was greater than 2 SD of the average of all the exons. The second algorithm, PRKN duplicate finder, divided the PRKN locus into 12 regions flanking each PRKN exon, and then counted the number of SNPs with a B allele frequency in the ranges of 0.125–0.375 and 0.625–0.875. Samples were flagged if ≥ 2 positive SNPs were identified in any region. Finally, the third algorithm, PRKN deletion finder 2, counted the number of SNPs with low Log R ratio intensities in these same regions flanking each exon. Samples were flagged if there were ≥ 3 positive SNPs in any region. All flagged samples were then assessed visually for CNVs by plotting the B allele frequency and Log R ratio values with the gglot2 visualization package for R (<https://www.r-project.org/>). The following phenotypic data were obtained from the UK Biobank: ICD10 codes (field code: 41270), Parkinson's disease (field code: 131023), parkinsonism (field code: 42031), illnesses of father and mother (field codes: 2017 and 20110), genetic ethnic grouping (field code: 22006), year of birth (field code: 34), and age of recruitment (field code: 21022). Assessment of proband Parkinson's disease status was based on indication of Parkinson's disease or parkinsonism in field codes 131023, 42031, and 41270. Thus, the assessment that resulted in a diagnosis of Parkinson's disease may have been established by a health care professional other than a neurologist. UK Biobank exome sequencing data (FE dataset, field code: 23156) were downloaded from the UK Biobank. Variants were annotated using ANNOVAR.¹⁷ Pathogenicity of variants was determined using their ClinVar annotation (<https://www.ncbi.nlm.nih.gov/clinvar/>). Likely pathogenic variants were grouped with pathogenic and likely benign variants were grouped with benign. We inspected the evidence for variants annotated as 'conflicting interpretations of pathogenicity' and categorized as pathogenic if most reports listed the variant as pathogenic or likely pathogenic and benign if most

reports listed it as benign or likely benign. The primary association test for Parkinson's disease risk due to PRKN variants was testing whether having one pathogenic PRKN variant identified by WES or genotyping array increases risk of Parkinson's disease relative to having no pathogenic PRKN variants. Additional tests were performed to assess sensitivity of the analysis including testing whether one PRKN variant increases the risk of having a parent with Parkinson's disease, testing different classes of variants separately, and testing detected variants in the larger sample with just genotyping array data. As these secondary sensitivity analyses were considered exploratory, reported *P*-values were not adjusted for multiple comparisons. No assumptions about Hardy-Weinberg equilibrium were made. No methods were used to infer genotypes or haplotypes. No specific methods were used to address population stratification in the primary analysis, although subgroup analysis tested by groups split by ancestry. No method was used to address relatedness among participants.

Statistical analyses

In analyses of patient fibroblasts, statistical significance was determined using one-way ANOVA followed by Tukey's multiple comparisons test implemented in Prism 9 (GraphPad). Odds ratios and *P*-values for contingency tables were calculated in R using Fisher's exact test (<https://www.r-project.org/>). For calculating the frequency and odds ratio for each mutation, samples with missing data were omitted.

Data and code availability

UK Biobank data is publicly available upon application at the UK Biobank website (<https://www.ukbiobank.ac.uk/>). AMP-PD data is publicly available upon application at the AMP-PD website (<https://amp-pd.org/>). gnomAD v2.1.1 summary statistics are available from (<https://gnomad.broadinstitute.org/>). MDS gene database summary statistics are available from (<http://www.mdsgene.org/>). Code used for analysis is available on our GitHub repository (<https://github.com/NarendraLab/Parkin/>).

Results

Identification of biallelic PRKN patients in NIH-PD cohort by WGS

To validate genotyping methods used in both the NIH-PD + AMP-PD and the UK Biobank cohorts, we first assessed PRKN mutations among the NIH-PD participants (the NIH-PD cohort) (Fig. 1A). This helped establish the ground truth for analysis in both cohorts. The NIH-PD participants had been genotyped both by WGS and by genotyping array and, in most cases, DNA was available to confirm CNVs by MLPA, allowing validation of WGS and genotyping array for CNV calls. Additionally, phenotypic and functional assessment was possible for many of the NIH-PD participants to verify variant pathogenicity and rule out missed pathogenic variants as a substantial confounder. This analysis was not possible for the AMP-PD and UK Biobank participants for whom only coded data were available.

In the NIH-PD cohort, WGS data were available for 742 patients seen consecutively at the NIH from 2006–2019 and for whom DNA was available (hereafter, the NIH-PD cohort). A total of 17 known pathogenic SNPs and 13 exon-spanning deletions in PRKN were identified (Fig. 1B, in black, and Table 1). Multiplex ligation-dependent probe amplification verified the deletions in 12/12 samples that were available for testing.

In addition to the known pathogenic SNPs, we identified a novel intronic variant, c.7+5G>A, in a patient with early-onset Parkinson's disease, who also carried an exon 3–4 deletion (Fig. 1B, in red, and 1C). This variant was absent in MDSGene v3.5.95, gnomAD 2.1.1, and ClinVar.^{6,21} On further inspection, we determined that this variant lies in the predicted splice acceptor of intron 1 and would disrupt base pairing with the U1 small nuclear RNA (Fig. 1C, left).²² To phase the variants, we tested DNA from the proband's unaffected sister and detected only the exon 3–4 deletion, demonstrating that the proband's variants are *in trans* (Fig. 1C, right). Consistent with c.7+5G>A interfering with RNA splicing, a PCR product was not detected from the patient's cDNA using primers that spanned the predicted splice site (Fig. 1C, right). Based on these findings and additional functional studies described below, we considered the c.7+5G>A variant to be pathogenic. Altogether, 31 pathogenic variants were found in a total of nine biallelic PRKN patients and 13 monoallelic PRKN patients by WGS.

Validating PRKN CNV detection by genotyping array

We next assessed the accuracy of a genotyping array for identifying PRKN biallelic patients, using the WGS genotyping as the reference. Genotyping array data (from the NeuroX or Neuro Consortium arrays^{14,15}) were available for 732/742 patients in the NIH cohort, including all but one patient with an identified PRKN mutation. Genotyping array identified 60% of the pathogenic variants, including all PRKN deletions (12/12) and a third of pathogenic SNPs (5/17) (Fig. 1D, right, and Table 1). Notably, the missense variant p.R275W and deletions involving exon 2, exon 3, and/or exon 4 accounted for the majority (51.5%) of pathogenic PRKN mutations. All SNPs that were probed on the genotyping array, namely p.R275W, were concordant with WGS calls. As expected, all SNPs without probes were missed. These included the frameshift mutations p.N34fs, p.N52fs, and p.113fs; a stop gain mutation, p.E409X; and the splice site mutation, c.7+5G>A. Deletion span was likewise concordant between the genotyping array and WGS, as visualized in the Log R ratio and B allele frequency plots of the genotyping array data with the WGS calls overlaid (Fig. 1D, left). Altogether at least one pathogenic mutation was identified by genotyping array in 7/8 biallelic PRKN patients and 5/13 monoallelic PRKN patients.

Phenotypic differences between Parkinson's disease patients with one and two PRKN mutations

Next, we compared the phenotypes of monoallelic and biallelic PRKN patients. We reviewed charts for the monoallelic and biallelic PRKN groups identified by WGS, LRRK2 p.G2019S (*n* = 16) and GBA p.N409S (*n* = 14) patient groups identified from genotyping array or WGS data, and the idiopathic Parkinson's disease patients with an available chart enrolled before each genetic Parkinson's disease patient. We also reviewed charts for a second group of six biallelic PRKN patients seen at the NIH (Supplementary Table 2).

We found that age-at-onset was significantly younger and UPSIT was significantly higher for both biallelic PRKN groups compared to the idiopathic Parkinson's disease and monoallelic PRKN groups (Table 2), similar to what has been reported in other cohorts previously.^{23,24} In contrast, Montreal cognitive assessment and UPDRS subscales were not significantly different among the groups (Table 2). Notably, all PRKN biallelic patients in our cohort had an age-at-onset ≤38 years and an UPSIT ≥20.

UPSIT and age-at-onset scores were separately available for a subset (334/739) of the WGS sequenced group. This allowed us to

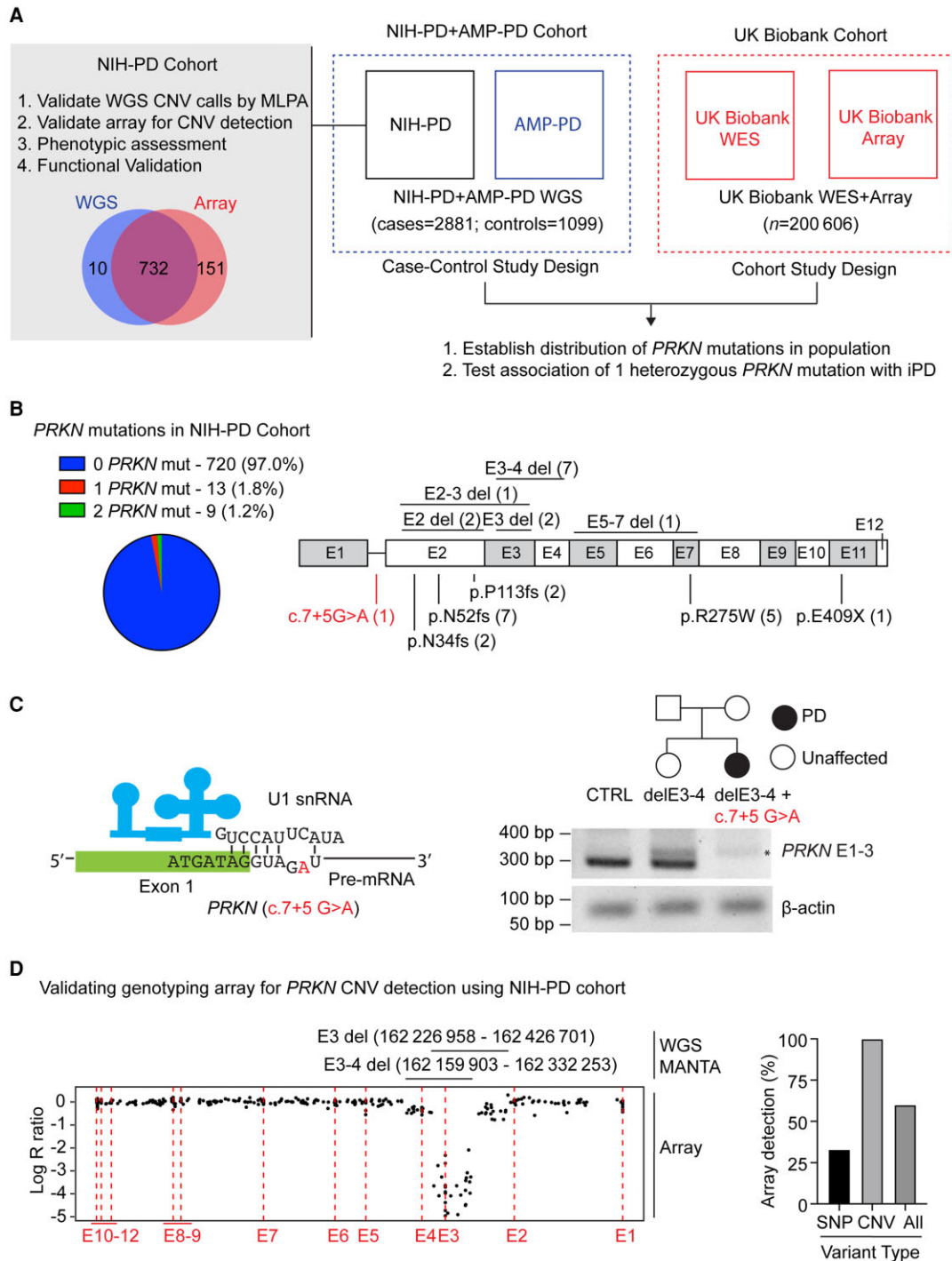


Figure 1 Study design and investigation of PRKN variants in the NIH-PD cohort. (A) PRKN variants were investigated in two large cohorts, the US based NIH-PD + AMP-PD cohort and the UK Biobank cohort. More extensive investigation was undertaken for NIH-PD participants, which cross-validated genotyping methods used in the two cohorts and provided context for the overall study through phenotypic and functional analysis of PRKN mutation carriers. (B) Proportion of PRKN variants detected in the NIH-PD cohort by whole genome sequencing is shown in the pie chart (left). The position of pathogenic PRKN variants (black) and a novel variant of unclear significance (VUS) (c.7+5G>A) detected by whole genome sequencing are shown in schematic of the PRKN gene (right). CNVs are shown on the top and SNPs on the bottom. (C) Schematic of the exon 1–intron 1 junction is shown with the position of the intronic substitution c.7+5G>A in red (left). This is predicted to disrupt a base pairing with the U1 small nuclear RNA of the spliceosome. Pedigree of family with this mutation appears on the top right. RT-PCR of RNA isolated from fibroblast cell lines from a control (CTRL), the unaffected sister of the proband with a single exon 3–4 deletion, and the proband with an exon 3–4 deletion and c.7+5G>A variants is depicted on the bottom right. The c.7+5G>A mutation disrupted the PCR product from primers in exons 1 and 3. An aberrant product (asterisk) was seen in both exon 3–4 deletion carriers. β -Actin served as a loading control. (D) Correspondence between CNV calls by WGS and by genotyping array is shown for a representative case with compound heterozygous PRKN deletions (left). The x-axis represents the relative genomic position. The position of the exons is indicated by the dotted lines. Log R ratio is shown on the y-axis. A value of -0.5 corresponds to a single deletion and a lower value is expected for two deletions. The relative position of the deletions called from WGS is shown above. Log R ratio is around -0.5 for probes in the non-overlapping regions of the deletions and <-2 for probes in the overlapping region. The graph on the right represents the proportion of PRKN variants (either SNPs or CNVs) detected by genotyping array.

Table 1 Pathogenic PRKN variants detected in NIH-PD cohort by WGS

	RSID	Consequence	Position start	Position end	Deletion size	MLPA	Array
Biallelic 1	Mut 1	E3 del	162 226 958	162 426 701	199 743	E3 del	E3 del
	Mut 2	E3-4 del	162 159 903	162 332 253	172 350	E3-4 del	E3-4 del
Biallelic 2	Mut 1	E3-4 del	162 090 797	162 375 645	284 848	E3-4 del	E3-4 del
	Mut 2	E3-4 del	162 090 797	162 375 645	284 848	E3-4 del	E3-4 del
Biallelic 3	Mut 1	E3-4 del	162 199 998	162 278 845	78 847	E3-4 del	E3-4 del
	Mut 2	rs34424986	p.R275W				p.R275W
Biallelic 4	Mut 1	E3 del	162 215 325	162 405 338	190 013	E3 del	E3 del
	Mut 2	rs34424986	p.R275W				p.R275W
Biallelic 5	Mut 1	E3-4 del	162 121 439	162 381 034	259 595	E3-4 del	E3-4 del
	Mut 2	rs34424986	p.R275W				p.R275W
Biallelic 6	Mut 1	E5-7 del	161 748 422	162 084 070	335 648	E5-7 del	E5-7 del
	Mut 2	rs754809877	p.N52fs				NT
Biallelic 7	Mut 1	E3-4 del	162 155 184	162 290 716	135 532	E3-4 del	E3-4 del
	Mut 2		c.7+5G>A				NT
Biallelic 8	Mut 1	rs754809877	p.N52fs				NT
	Mut 2	rs754809877	p.N52fs				NT
Biallelic 9	Mut 1	E3-4 del	162 090 797	162 375 645	284 848	NT	NT
	Mut 2	rs754809877	p.N52fs				NT
Monoallelic 1	Mut 1	rs1554252213	p.E409X				NT
Monoallelic 2	Mut 1	rs771529549	p.P113fs				NT
Monoallelic 3	Mut 1	rs754809877	p.N52fs				NT
Monoallelic 4	Mut 1	rs34424986	p.R275W				p.R275W
Monoallelic 5	Mut 1	rs771529549	p.P113fs				NT
Monoallelic 6	Mut 1		E2 del	162 439 838	162 454 794	14 956	E2 del
Monoallelic 7	Mut 1	rs55777503	p.Q34fs				NT
Monoallelic 8	Mut 1		E2 del	162 334 681	162 444 403	109 722	E2 del
Monoallelic 9	Mut 1	rs34424986	p.R275W				p.R275W
Monoallelic 10	Mut 1	rs754809877	p.N52fs				NT
Monoallelic 11	Mut 1		E2-3 del	162 227 931	162 466 456	238 525	E2-3 del
Monoallelic 12	Mut 1	rs55777503	p.N34fs				NT
Monoallelic 13	Mut 1	rs754809877	p.N52fs				NT

del = deletion; E = exon; MLPA = multiplex ligation dependent probe amplification; Mut = mutation; NT = not tested; RSID = reference SNP cluster ID.

Table 2 Clinical characteristics of NIH-PD cohort

Clinical	Idiopathic Parkinson's disease	GBA	LRRK2	PRKN 1 mut	PRKN 2 mut (group 1)	PRKN 2 mut (group 2)	P-value (unadjusted)	P-value (corrected)
Age-at-onset (years)	52.7 ± 11.8	58.1 ± 6.9	53.8 ± 8.8	51.8 ± 15.2	26.5 ± 8.3***	30.7 ± 5.5**	<0.0001	<0.0001
UPSIT	20.3 ± 7.9	20.7 ± 8.2	23.5 ± 7.1	20.8 ± 6.7	31.9 ± 6.5**	33.7 ± 4.5**	0.0002	0.0014
MoCA	24.8 ± 4.5	26.9 ± 3.3	26.7 ± 2.3	24.2 ± 5.2	24.6 ± 2.9	26.8 ± 3.1	0.4733	3.3131
UPDRSI	2.9 ± 2.9	3.8 ± 1.8	3.3 ± 2.3	3.9 ± 3.0	3.0 ± 3.0	1.8 ± 1.7	0.7627	5.3389
UPDRSII	11.3 ± 8.2	13.5 ± 5.4	11.0 ± 5.1	15.7 ± 8.9	16.0 ± 5.7	11.0 ± 8.0	0.642	4.494
UPDRSIII	26.8 ± 12.8	27.9 ± 10.0	23.4 ± 12.9	32.7 ± 15.6	31.1 ± 8.8	22.4 ± 9.4	0.9547	6.6829
UPDRSIV	4.9 ± 3.3	2.1 ± 1.5	4.1 ± 4.2	6.6 ± 3.8	4.0 ± 3.0	4.3 ± 5.8	0.3235	2.2645

MoCA = Montreal cognitive assessment. **<0.01 and ***<0.001 for Dunnett's multiple comparison test.

P-value (corrected) is a Bonferroni correction of the one-way ANOVA P-value.

estimate how many patients with early onset and relatively preserved olfaction in our study have biallelic PRKN mutations. In total, 21 (6.3%) patients had age-at-onset ≤ 38 years and UPSIT ≥ 20 . Of these, three (14.2%) were biallelic for PRKN mutations. Comparing the age-at-onset and UPSIT for all available PRKN biallelic and monoallelic patients, biallelic PRKN patients formed a tight cluster, whereas monoallelic PRKN patients were distributed in a similar pattern as idiopathic Parkinson's disease patients (Fig. 2A). Of note, five patients with variants that were novel or of uncertain significance clustered with the other biallelic PRKN patients, providing additional support for the variants' pathogenicity.

Together these data suggested that age-at-onset and UPSIT scores can distinguish biallelic from monoallelic PRKN patients.

Assessing loss of PARKIN function in patient fibroblasts

Three patients in our study clustered phenotypically with biallelic PRKN patients but had only one detected PRKN mutation (Fig. 2A). This raised the question of whether they may have a second undetected PRKN variant that was missed in our initial analysis of WGS data, such as a deep intronic variant or an exon inversion.²⁵

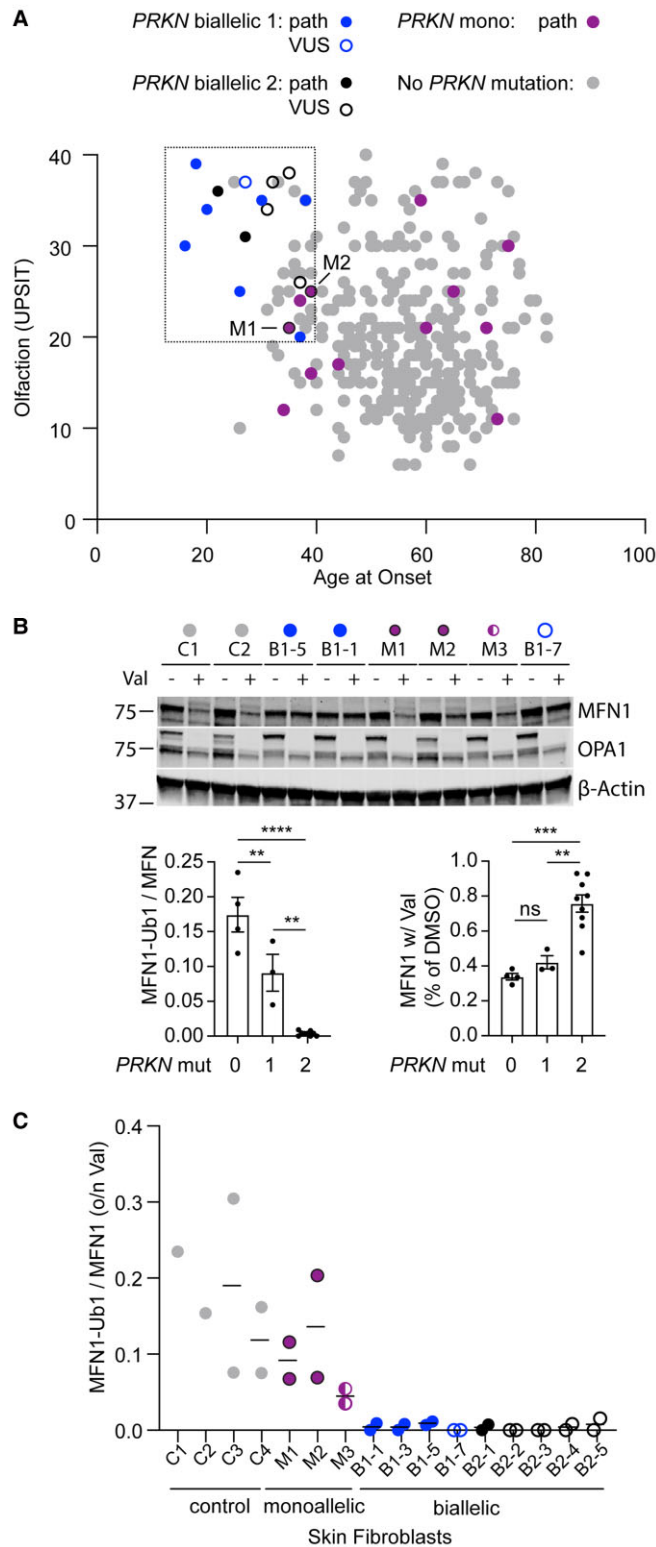


Figure 2 Biallelic and monoallelic PRKN mutation carriers can be distinguished by their phenotype and functional assays in their fibroblasts. (A) Scatterplot depicts scores from smell identification test (UPSIT, y-axis) and age-at-onset (x-axis) for genotyped Parkinson disease patients in the NIH-PD cohort with no PRKN mutations (grey), one PRKN mutation (magenta), or two PRKN mutations (blue). A second group of patients with two PRKN mutations (black) was identified outside of the consecutive series. Those with a novel variant or VUS are shown as open symbols. Two patients with one PRKN mutation and whose fibroblasts were evaluated in C have a black border. (B and C) Patient fibroblasts were treated with vehicle (DMSO) or 10 mM of valinomycin (val) overnight (o/n), separated on SDS-PAGE gels, and immunoblotted for the PARKIN substrate MFN1, as shown in representative blot. The ratio of MFN-Ub1 (corresponding to the slower migrating band) to MFN1 was calculated for each cell line treated with val. Each cell line was assayed in two biological replicates with exception of lines C1 and C2, which were assayed in one replicate each. All cell lines from PRKN carriers were from individuals with Parkinson’s disease, except for M3, who was the unaffected sibling of B1-7. Graphs of the summary data are shown on the bottom. (C) Individual data for experiment described in B.

Additionally, five patients had novel variants or variants annotated as of uncertain significance (Fig. 2A, C and Supplementary Fig. 1), raising the question of whether the variants (namely, c.7+5G>A, p.V56G, and p.G429D) are pathogenic. To resolve definitively which patients had complete loss of PRKN function, we tested available fibroblast lines for ubiquitination of PRKN substrate Mitofusin-1 (MFN1) following PRKN activation by mitochondrial membrane depolarization.^{26–30} This assay has been used previously to differentiate biallelic PRKN carriers from healthy controls and PRKN carriers within a family, but to our knowledge has not been previously used to differentiate unrelated monoallelic and biallelic PRKN carriers.^{30,31} Altogether we tested fibroblast lines from nine biallelic PRKN patients (including five with variants that were novel or of unclear significance), one patient with the novel c.7+5G>A PRKN mutation (*in trans* with an exon 3–4 deletion, as discussed above and in Fig. 1C), three monoallelic PRKN patients (including two that phenotypically clustered with biallelic PRKN patients), and four healthy controls (Fig. 2B, C and Table 3). Monitoring the change in MFN1-Ub1/MFN1 ratio following depolarization with valinomycin distinguished all healthy controls and monoallelic PRKN patients from all biallelic PRKN patients (Fig. 2B, C and Table 3). Additionally, as a group, monoallelic PRKN carriers had an average MFN1-Ub1/MFN1 ratio that was about half that of the controls, consistent with their possessing intermediate PRKN activity (Fig. 2B). Total MFN1 levels were also significantly decreased biallelic in PRKN cell lines compared to the control and monoallelic cell lines, consistent with the PRKN-dependent degradation of MFN1 (Fig. 2B). To further validate the assay, we tested three publicly available monoallelic PRKN fibroblast lines and four publicly available biallelic PRKN lines, obtaining similar results (Supplementary Fig. 1A and B). Two other measures of PRKN activity, levels of phospho-ubiquitin (Ser65) and mitophagy using the mt-Keima assay, failed to distinguish individual biallelic and monoallelic carriers, suggesting that these assays may not be sensitive enough in fibroblasts with low endogenous PRKN levels (Supplementary Fig. 1C–E).

Table 3 PRKN mutations in fibroblast lines

Sample	Mutation	Variant	ClinVar
Monoallelic 1	Mut 1	p.Q34RfxX5	Pathogenic
Monoallelic 2	Mut 1	exon 2 deletion	
Monoallelic 3	Mut 1	exon 3–4 deletion	
Biallelic 2-1	Mut 1	p.T415N	Likely pathogenic
	Mut 2	p.T415N	Likely pathogenic
Biallelic 2-2	Mut 1	exon 11 deletion	
	Mut 2	p.G429D	Not reported
Biallelic 2-3	Mut 1	exon 11 deletion	
	Mut 2	p.G429D	Not reported
Biallelic 2-4	Mut 1	exon 7 deletion	
	Mut 2	p.V56G	VUS
Biallelic 2-5	Mut 1	exon 7 deletion	
	Mut 2	p.V56G	VUS
Biallelic 1-1	Mut 1	exon 3 deletion	
	Mut 2	exon 3–4 deletion	
Biallelic 1-3	Mut 1	exon 2–3 deletion	
	Mut 2	p.R275W	Pathogenic
Biallelic 1-5	Mut 1	exon 3–4 deletion	
	Mut 2	p.R275W	Pathogenic
Biallelic 1-7	Mut 1	exon 3–4 deletion	
	Mut 2	c.7+5G>T	Not reported

Variants present in each of the fibroblast lines tested in Fig. 2B and C. Mut = mutation.

In summary, functional assessment of patient fibroblasts resolved three VUSs (c.7+5G>A, p.V56G, and p.G429D) as pathogenic in five of our biallelic patients. Additionally, it ruled out a missed PRKN mutation in two early onset patients with one detected PRKN variant, validating our discovery cohort. Together these results suggested that ‘cryptic’ second PRKN mutations (such as from deep intronic mutations or exon inversions) are likely not a substantial source of missed second mutations.

Identifying PRKN copy number variants in the UK Biobank cohort

We next used our high-confidence discovery cohort to develop sensitive screening algorithms for identifying PRKN CNVs in genotyping array data (see Methods). These optimized algorithms were then applied to genotyping array data from 488 264 subjects in the UK Biobank. Collectively, they flagged 25 906 subjects in the UK Biobank as potentially carrying a CNV, of which 2687 were confirmed to have a CNV on visual inspection of the Log R ratio and B allele frequency plots. From the whole cohort a deletion was detected in 0.30% of samples, and a duplication in 0.25% of samples. This was close to the percentage of deletions (0.34%) and duplications (0.26%) that we detected by visual inspection of every 100th sample in the UK Biobank. Altogether, we estimated that ~92% of detectable CNVs were identified in the UK Biobank.

To visualize the CNVs, we generated heatmaps for all single deletions and duplications, as well as the average values for each class of CNV (Fig. 3). Exon 2 deletions and exon 2 duplications were the most common classes of deletion and duplication, respectively, together accounting for 48.9% of all CNVs. Interestingly, while the lengths of exon 2 deletions were highly variable, the lengths of exon 2 duplications were more uniform. Although genotyping array does not provide the resolution to define exact breakpoints, this suggests that exon 2 duplications may be composed of fewer distinct CNVs compared to exon 2 deletions.

To provide additional support for this observation, we examined PRKN exon 2 duplications and exon 2 deletions called from a harmonized WGS dataset that included both participants in the NIH-PD cohort and participants in the US based AMP-PD cohort (the NIH-PD + AMP-PD cohort). In this cohort we identified two pathogenic PRKN mutations in 20 individuals, all of whom were cases, and one pathogenic PRKN mutation in 78 individuals (57 cases and 17 controls; Supplementary Tables 3 and 4). Although age-at-onset was not available in the harmonized dataset, the age at last evaluation was significantly younger for cases with two PRKN mutations than those without PRKN mutations, as expected (Supplementary Fig. 2A). In contrast, the age of those with one PRKN did not significantly differ from those with no mutations (Supplementary Fig. 2A). Assessment of the participants' ancestry by PCA analysis of their genetic data showed the cohort was of predominantly European ancestry (Supplementary Fig. 2B). We also examined PRKN exon 2 duplications and exon 2 deletions in the gnomAD structural variants v2.1 database.¹⁹

Consistent with results from the UK Biobank, a common exon 2 duplication in the European population (Chr6:162296324–162494975) accounted for 40–75% of all exon 2 duplications, whereas the most frequent exon 2 deletion accounted for only 13–17% of exon 2 deletion in both gnomAD structural variants v2.1 and the NIH-PD + AMP-PD cohort. This confirmed that there is greater diversity among exon 2 deletions than exon 2 duplications (Fig. 3). The common exon 2 duplication identified in the European population was absent in all non-European populations in gnomAD structural

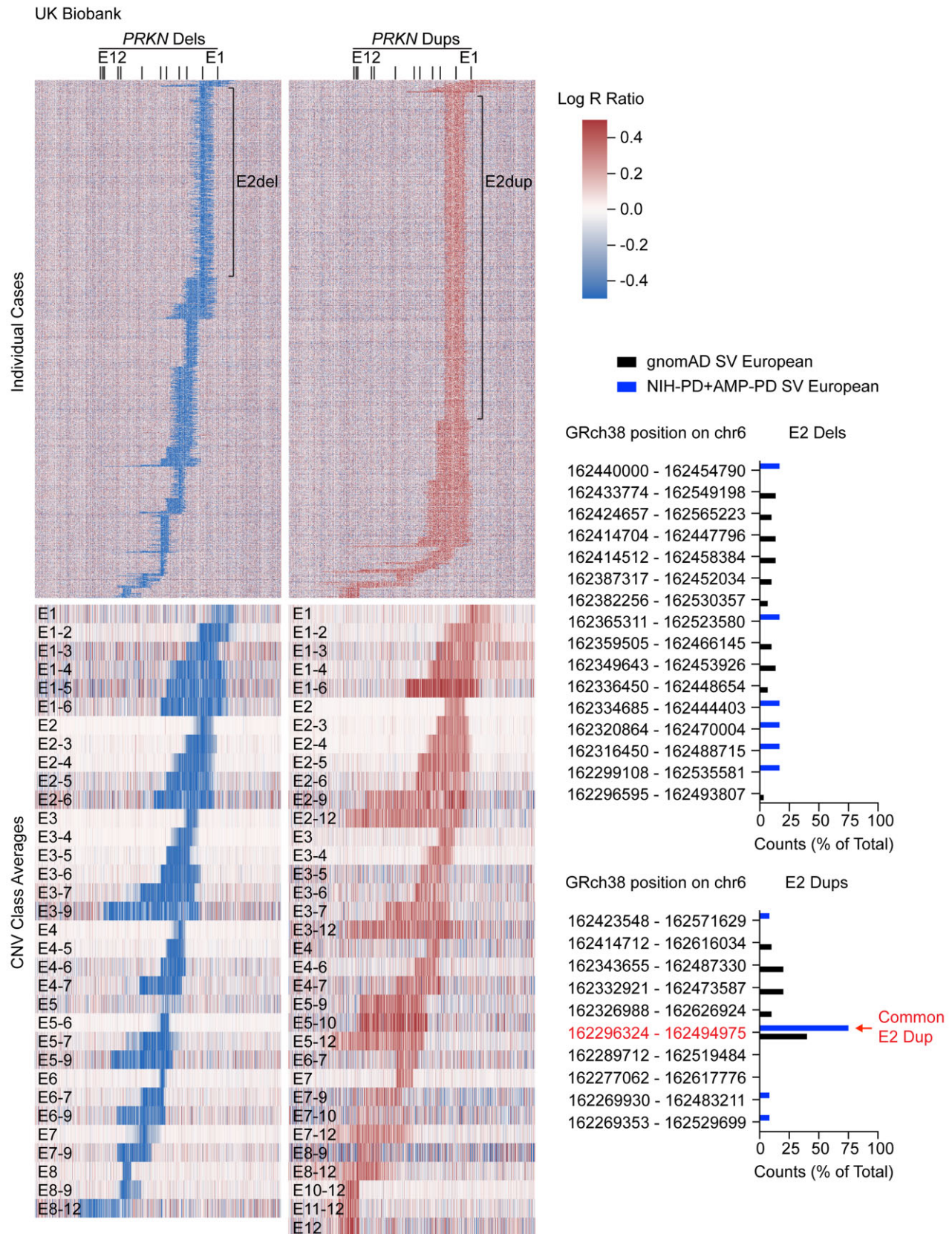


Figure 3 Identification of PRKN CNVs in UK Biobank from genotyping array data. Heat maps represent Log R ratio across the PRKN locus (on the x-axis). Deletions (dels) are shown on the left and duplications (dups) on the right. The top heat maps depict all individual cases (on the y-axis) in which a single CNV was detected. The locations of exon 2 deletions and exon 2 duplications are shown. Heat maps (bottom) represent the average Log R ratios for each class of CNV. Graphs (right) show the distributions of exon 2 deletions (top) and exon 2 duplications (bottom) for the non-Finnish European population in the gnomAD structural variant database and the NIH-PD + AMP-PD cohort. A common exon 2 duplication detected in European populations is indicated by the arrow.

Table 4 Odds of Parkinson's disease or parent with Parkinson's disease in PRKN heterozygous mutation carriers in UK Biobank

Mutation	Mut_Cases	Mut_controls	NoMut_Cases	NoMut_controls	Odds	CI	P-value
For single PRKN mutation carriers							
Odds of Parkinson's disease in UK Biobank participants with WES and array (n = 200 606)							
Frameshift_SNP	2	683	1372	198 549	0.42	0.05–1.54	0.35
AnySNP	15	2517	1359	196 715	0.86	0.48–1.43	0.71
AnyMut	23	3648	1351	195 584	0.91	0.58–1.38	0.76
Odds of parent with Parkinson's disease in UK Biobank participants with WES and array (n = 200 606)							
Frameshift_SNP	29	656	7465	192 456	1.14	0.76–1.65	0.48
AnySNP	104	2428	7390	190 684	1.11	0.90–1.35	0.32
AnyMut	152	3519	7342	189 593	1.12	0.94–1.31	0.19
Odds of Parkinson's disease in NIH-PD + AMP-PD (n = 2862 cases versus 1099 controls)							
Frameshift_SNP	8	1	2853	1098	3.08	0.41–136.70	0.46
AnySNP	27	7	2834	1092	1.49	0.63–4.05	0.44
AnyCNV	30	10	2831	1089	1.15	0.55–2.66	0.86
AnyMut	57	17	2804	1082	1.29	0.74–2.38	0.43
For double PRKN mutation carriers							
Odds of Parkinson's disease in UK Biobank Participants with WES and microarray (n = 200 606)							
AnyMut	2	22	1351	195 584	13.16	1.50–53.57	0.012
Odds of Parkinson's disease in NIH-PD + AMP-PD (n = 2862 cases versus 1099 controls)							
AnyMut	20	0	2804	1082	Infinity	1.90–infinity	0.0020

variants v2.1, suggesting that its prevalence may be due to a founder effect in the European population.

We next compared the distribution of the 2693 CNVs (identified in 2687 carriers) in the UK Biobank to NIH-PD + AMP-PD and three other studies/databases with >100 CNVs each: the deCode study of the Icelandic population, which detected CNVs from genotyping array (993 CNVs)⁹; the MDS gene database (435 CNVs), a curated database of the published literature⁶; and the gnomAD structural variants v2.1 database,¹⁹ which called CNVs from WGS data (171 CNVs) (Fig. 4A and B and Supplementary Table 5). In all studies, most deletions and duplications involved exon 2, exon 3, or exon 4 (88.2% in the UK Biobank, 96.4% in deCode, 65.7% in the MDS gene database, and 94.7% in gnomAD, and 93.3% in the NIH-PD + AMP-PD cohort). Notably, the two most common classes of deletions in UK Biobank, exon 2 and exon 3–4 deletions, were among the top four in MDSGene and the top three in deCode, NIH-PD + AMP-PD, and gnomAD. An exon 6–9 deletion that was particularly common in deCode (21% of deletions) was uncommon in UK Biobank (1.2% of deletions) and absent in MDSGene and gnomAD. This suggests that the high exon 6–9 deletion frequency may be due to a founder effect in the Icelandic population.

Among duplications, the exon 2 duplication was the most common duplication in the UK Biobank, the deCode study, and gnomAD, although it was rare in the MDS gene database. Interestingly, an exon 3 duplication accounted for 18.8% of duplications in gnomAD and 17.8% of duplications in MDSGene but was absent in the UK Biobank. In gnomAD all nine exon 3 duplications had the same breakpoints (Chr6:162638588–162730211) and the allelic frequency was highest for the African/African-American population, suggesting that it may reflect a founder effect.

Overall, duplications were less common in the literature as represented in the MDS gene database than in the population-based studies (10% of CNVs in MDSGene versus 28% in gnomAD, 29.7% in NIH-PD + AMP-PD, 48% in UK Biobank, and 70% in deCode). This could reflect a bias in their detection or reporting. Alternatively, one or more dup affecting an exon may not result in complete loss of function. Notably, while dups represented 20% of mutations

in single PRKN mutation carriers in NIH-PD + AMP-PD, only one (5% of total) was found among the two PRKN mutation carriers (Supplementary Fig. 2C).

Overall, the distribution in CNVs was similar in the UK Biobank as in other studies and databases with PRKN CNV data with most CNVs affecting exons 2, 3, or 4. Founder effects may account for the increased prevalence of exon 2 duplications in the UK and Icelandic populations relative to others. Similarly, an exon 6–9 deletion and an exon 3 duplication may be due to founder effects in the Icelandic and African-American/African populations, respectively.

Estimating pathogenic PRKN variant frequency in the UK Biobank

To obtain a more complete view of the prevalence of pathogenic PRKN variants in the UK population, we next examined the frequency of pathogenic PRKN missense variants in the UK Biobank. The UK Biobank genotyping array contains probes for four known pathogenic PRKN missense variants: p.R275W, p.C253Y, p.T240M, and p.K211N. To validate these probes, we benchmarked their performance against WES data from 192 490 participants genotyped on both platforms. Genotyping array calls agreed with WES calls for p.R275W and p.T240M variants in most samples (98.5% and 100%, respectively), but were less reliable for p.C253Y and p.K211N variants (4.4% and 70%, respectively). Thus, we restricted further analysis of genotyping array data to p.R275W and p.T240M variants.

Assessed in the entire UK Biobank from genotyping array data, p.R275W and p.T240M mutations had allelic frequencies of 0.0039 and 0.00022, respectively (Supplementary Table 6). These were similar to their allelic frequencies in the subset of samples with WES data (0.0038 and 0.00019, respectively), as well as their frequency in the non-Finnish European population in gnomAD v2.1.1 database (0.0033 and 0.00023, respectively). In the UK Biobank, p.R275W was more common in Europeans, whereas p.T240M was more common in non-Europeans. The latter finding may reflect the higher allelic frequency of p.T240M in the South

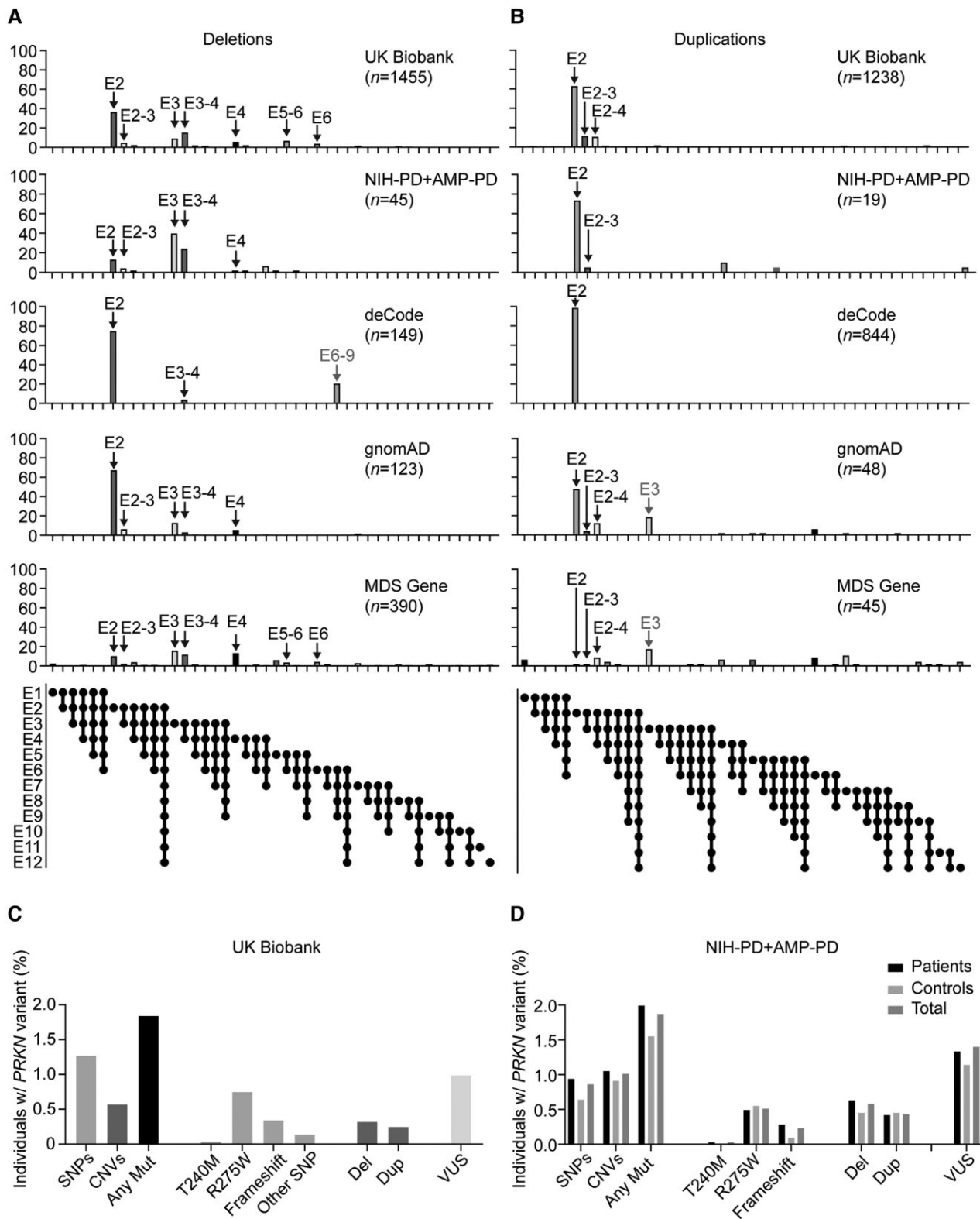


Figure 4 Distribution of PRKN CNVs in UK Biobank and other studies. (A and B) Graphs depict distributions of PRKN deletions (A) and duplications (B) reported here for the UK Biobank and NIH-PD + AMP-PD cohorts versus those reported in the deCode study, the gnomAD structural variant database, and the MDS gene database. Major peaks identified in the UK Biobank are indicated by black arrows. Grey arrows signify major peaks (>15% of deletions or duplications, respectively) that were detected in one of the other databases and were rare in the UK Biobank (<1%). (C and D) Graphs represent PRKN variants detected in the UK Biobank for participants with both WES and genotyping array data (C) and PRKN variants detected in the NIH-PD + AMP-PD cohort by WGS (D).

Asian population, as also seen in gnomAD v2.1.1 (0.0017 versus 0.00023). A total of 217 rare PRKN missense variants that were annotated as VUSs or not annotated were present in an additional 1% of samples (Supplementary Table 7). Combining WES and genotyping array estimates, we found a pathogenic PRKN mutation in 1.84% of samples. Twenty-four (~1/8400) carried biallelic pathogenic variants. The frequency of pathogenic PRKN variants was similar in the US based NIH-PD + AMP-PD cohort (Fig. 4D).

Risk of Parkinson's disease in single mutation carriers

Altogether, we detected single pathogenic mutations in 6612 of samples in all UK Biobank samples and 3671 in samples with WES and array data, more than has been reported in any previous study to date.¹⁰ Given its size, near complete genotyping of PRKN, and its cohort study design, we reasoned the UK Biobank dataset may be particularly valuable for testing whether heterozygous PRKN variants increase the risk of developing Parkinson's disease. We also tested whether PRKN carriers were more likely to have a parent with Parkinson's disease, given one of the parents is an obligate PRKN carrier in most cases. Additionally, while the median age of UK Biobank participants at enrollment (i.e. 58 years) is less than the average age of Parkinson's disease onset, their parents' age is likely greater than the average age of onset. Thus, Parkinson's disease prevalence is likely higher in the parents than the probands.

Among the 200 606 fully genotyped participants with zero or one detected heterozygous PRKN mutations, those with one detected PRKN mutation were as likely as those with no mutation to have Parkinson's disease (odds ratio = 0.91, 95%CI = 0.58–1.38, P-value = 0.76) or a parent with Parkinson's disease (odds ratio = 1.12, 95%CI = 0.94–1.31, P-value = 0.19; Table 4). As expected, those with two detected PRKN mutations were more likely to have Parkinson's disease (odds ratio = 13.16, 95%CI = 1.50–53.57, P-value = 0.012; Table 4). Not all participants with two detected PRKN mutations were reported to have Parkinson's disease, most likely representing incomplete ascertainment of disease status. Similarly, among all 488 341 participants with array data in the UK Biobank those carrying one detected heterozygous PRKN mutation were as likely as those without a mutation to have Parkinson's disease (odds ratio = 1.18, 95%CI = 0.88–1.54, P-value = 0.24) or a parent with Parkinson's disease (odds ratio = 1.06, 95%CI = 0.93–1.20, P-value = 0.39; Supplementary Table 8). We obtained similar findings testing each class of mutation individually.

We additionally assessed the risk of Parkinson's disease in the NIH-PD + AMP-PD cohort. Although smaller than the UK Biobank and a case-control design, it was the largest publicly available WGS Parkinson's disease dataset. It had more complete ascertainment of Parkinson's disease status, as evidenced by the fact that all participants with two PRKN mutations were among the cases (Table 4). Additionally, it had about twice as many Parkinson's disease patients overall and represented a different (albeit predominantly European population) drawn from the US rather than the UK. Consistent with results from the UK Biobank, those with one detected PRKN mutation were as likely as those with no mutation to have Parkinson's disease (odds ratio = 1.29, 95%CI = 0.74–2.38, P-value = 0.43; Table 4). Similar results were obtained when limiting to participants with European ancestry or to only subjects in the AMP-PD cohort (Supplementary Table 8). As expected, simulating incomplete genotyping by considering only SNPs or only CNVs led to higher odds ratios (2.03 and 1.50, respectively), confirming that missed second mutations can falsely inflate the odds ratio.

Together our findings show that monoallelic heterozygous PRKN mutations do not significantly increase the risk of having Parkinson's disease or a parent with Parkinson's disease.

Discussion

In this study, we assessed the frequency and risk conferred by pathogenic PRKN mutations in two large cohorts in which the PRKN gene was nearly fully genotyped, including the NIH-PD + AMP-PD cohort and the UK Biobank. The NIH cohort allowed us to validate that both WGS and genotyping array capture the vast majority of CNVs. Additionally, through functional studies of patient fibroblasts, it allowed us to establish that 'cryptic' PRKN mutations (such as deep intronic mutations or exon inversions) are likely not a substantial source of missed second mutations. Finally, sampling nearly half million individuals in the UK Biobank, we found that PRKN mutations were common in the general population, and that carrying a single heterozygous PRKN mutation did not significantly increase the risk of having Parkinson's disease or a parent with Parkinson's disease. These results were validated in the independent US based NIH-PD + AMP-PD cohort.

Notably, genotyping array detected the majority of PRKN mutations in both the NIH-PD cohort and the UK Biobank. In the NIH-PD cohort, genotyping array identified all deletions that were also called by WGS, as well as a third of PRKN SNPs. Genotyping array performed similarly in the UK Biobank, identifying an estimated 74.1% of pathogenic variants, including most pathogenic SNPs. Notably, similar results were achieved in these two cohorts even though different genotyping platforms were used (Illumina in NIH-PD versus Affymetrix in the UK Biobank).

Genotyping array performed well compared to next-generation sequencing, identifying at least one PRKN mutation in 87.5% of PRKN-PD patients. The strong performance of genotyping arrays in our cohorts was due to good coverage of the PRKN locus by the arrays used and the high frequency of the missense variant p.R275W and CNVs involving exon 2, exon 3, and/or exon 4 in our cohorts. Together these accounted for 51.5% of mutations in the NIH-PD cohort and ~72% of mutations in the UK Biobank. Similar results are likely achievable with other genotyping arrays in populations with high European ancestry, provided they genotype the p.R275W variant and they have adequate coverage of exon 2, exon 3, and exon 4 to determine their copy number (typically at least four probes per exon). For arrays that currently do not have this coverage, the addition of ~13 probes would allow for screening of most PRKN-PD patients in populations of predominantly European-ancestry for PRKN-PD targeted trials.

In non-European populations, the p.R275W variant is less frequent, and, therefore, genotyping arrays may not capture the majority of PRKN variants. This highlights the need for PRKN sequencing in populations of non-European ancestry, and inclusion of PRKN SNPs from these populations on genotyping arrays. The novel NeuroBooster array, used as part of GP2, the Global Parkinson's Genetics Program, will help address these needs.³² Nonetheless, genotyping arrays are still likely to be effective for identifying a substantial portion of PRKN CNVs in non-European populations. We found CNVs were present in >0.5% of participants in the UK Biobank, regardless of ancestry. This likely reflects the diversity of CNVs detected. The most common classes of deletions, for instance, were composed of several CNVs of distinct length, suggesting that they were generated by several independent recombination events. This diversity makes it more likely that CNVs are

evenly distributed among populations. The frequency and diversity of PRKN CNVs is likely related to the large size of the gene (1.38 Mb), making recombination events more likely.

This is the first comprehensive population scale study to estimate the frequency of pathogenic PRKN mutations using methods that capture both most SNPs (by WES) and most CNVs (by genotyping array). Altogether we detected a pathogenic PRKN mutation in 1.8% of participants in the UK Biobank. We found a similar frequency among cases and controls in the US based NIH-PD + AMP-PD cohort. Although higher than some estimates based on SNP frequency alone, our estimate is generally in line with those from smaller studies that have employed methods to capture most SNPs and CNVs. Yu *et al.*,³³ for instance, recently found that 1.8% of control subjects carried a PRKN mutation. Along these lines, we found biallelic PRKN mutations in ~1/8400 UK Biobank participants, which would correspond to ~7900 biallelic PRKN carriers in the UK and close to a million world-wide. Together these findings suggest that pathogenic PRKN variants are common in the general population.

Strikingly, the clinical significance of many PRKN variants remains unknown. Although individually rare, 217 PRKN missense variants of unclear significance or without annotation in ClinVar were found in 1% of UK Biobank samples. This represents about a 1/3 of variants that affect an exon and are not annotated as benign. In the NIH-PD cohort we found that assaying PRKN function in PRKN fibroblasts can differentiate unrelated samples with one or two heterozygous pathogenic mutations.

One of our key findings was that functional assays of fibroblasts from unrelated patients can resolve PRKN VUS and rule out missing variants in monoallelic carriers. This allowed us to resolve the pathogenicity of three variants that were novel or annotated as VUSs in the NIH-PD cohort. This also allowed us to rule out cryptic pathogenic variants in two patients with single pathogenic PRKN variants and a PRKN-PD phenotype (i.e. early age-at-onset and preserved olfaction), suggesting that these are unlikely to be a substantial confounder of association studies. Functional studies may help to clarify which of the 217 missense VUSs in the UK Biobank are loss of function and therefore likely pathogenic.

Whether heterozygous PRKN variants increase the risk of Parkinson's disease has been unclear. A large cohort study of the Icelandic population involving 105 749 genotyping array samples found that heterozygous PRKN CNVs increased the odds of Parkinson's disease (with an odds ratio of 1.69), however, this study was not able to fully assess PRKN SNPs as it was limited to genotyping array data.⁹ A recent study nominally replicated this association in a case-control study and meta-analysis but found that the association is lost if missed mutations in biallelic PRKN carriers are taken into account.¹⁰ Here, in the first population scale cohort study with near complete genotyping of PRKN, we did not find an association between heterozygous PRKN mutations and Parkinson's disease risk. Similarly, heterozygous PRKN mutations did not increase the odds of having a parent with Parkinson's disease. This is notable as one parent of most PRKN carriers is an obligate carrier, and most parents are beyond the age-at-onset for most Parkinson's disease cases. The number of parents with Parkinson's disease in the UK Biobank is much larger than the number of Parkinson's disease cases (17 675 versus 3465), which increases the power to detect an association. In contrast to the lack of association between heterozygous PRKN variants and Parkinson's disease in the UK Biobank, association of common variants with Parkinson's disease risk has been replicated in the UK Biobank previously.³⁴ We additionally replicated this finding in a large case-control study using the NIH-PD + AMP-PD cohort.

Possible reasons for the lack of association between single PRKN mutations in our study versus the positive association in previous studies include: (i) the cohort design of the UK Biobank, which avoids potential confounding effects inherent to matching cases and controls; (ii) near complete genotyping of PRKN at the population scale through the combined use of WES and microarray; and (iii) not limiting analysis to predominately early onset cases. While a recent meta-analysis by Lubbe *et al.*¹⁰ found a positive association between single PRKN mutations and Parkinson's disease, they found this association was no longer significant if they excluded studies with two of these confounding factors, namely, incomplete genotyping and inclusion of predominately early onset Parkinson's disease cases. Consistent with incomplete genotyping potentially leading to a false association, in the NIH-PD + AMP-PD cohort, we likewise found that simulating only CNV detection or only SNP detection inflated the odds ratio for association between single PRKN mutations and Parkinson's disease.

Although our data suggest that single pathogenic PRKN mutations do not increase the risk of Parkinson's disease, single PRKN mutations may still have a subclinical effect on dopamine neurons. Single PRKN mutation carriers have a slight dopamine deficit in the striatum compared to controls, when measured by PET.^{35–40} Progression of this dopamine deficit, however, is slow, with an estimated annual decrease of 0.62% in the most affected region (compared to 2% in biallelic PRKN mutation carriers and 9–12% in idiopathic Parkinson's disease).³⁸ The gradual nature of the striatal dopamine loss and compensatory changes in cortical regions contributing to motor control have been credited with preventing overt motor symptoms in single PRKN mutation carriers.^{38,41,42} Thus, PRKN-PD may be similar to other recessive disorders such as cystic fibrosis and sickle cell anemia, in which haploinsufficiency has a subtle physiological effect but does not predispose to overt disease.

In summary, through comprehensive analysis of PRKN mutations in a large Parkinson's disease cohort and a population scale cohort, we validated genotyping array screening for the detection of biallelic PRKN patients. Additionally, we demonstrated that functional assays in patient cell can resolve PRKN VUSs in biallelic patients and can rule out second cryptic variants in patients with one heterozygous pathogenic mutation. Finally, we demonstrate that heterozygous PRKN variants are common in the population but do not increase the risk of Parkinson's disease.

Study limitations

We detected ~92% of CNVs in the UK Biobank that can be identified by visual inspection. However, small CNVs that were not well covered by probes on the genotyping array may have been missed, particularly those involving single exons 1, 5, 9, 10, and 12. These likely make up a small proportion of PRKN CNVs, however; they account for only 0.6% of all CNVs in gnomAD (all are absent except exon 5 duplications; [Supplementary Table 3](#)). These would tend to underestimate the prevalence of PRKN mutations in the UK population but are not predicted to affect the association between single pathogenic PRKN variants and Parkinson's disease risk, as they are as likely to occur in cases as controls. Ascertainment of Parkinson's disease in the UK Biobank is based on medical codes placed in the health record and may be incomplete and/or may have been obtained by a medical professional other than a movement disorders specialist. As even specialists following established criteria may incorrectly diagnose Parkinson's disease in as many as 18% of cases, some cases may be misdiagnosed as Parkinson's disease and subtle parkinsonian signs may have been missed or not reported in participants classified as

controls.⁴³ However, as misclassification of Parkinson's disease status is as likely to occur in carriers of no or one *PRKN* mutation, it should not affect the association between single pathogenic *PRKN* variants and Parkinson's disease risk, apart from decreasing statistical power. Additionally, as the controls in the NIH-PD + AMP-PD cohort may not have received the same level of assessment as the cases, there is the potential for bias that would decrease the odds ratio due to misclassification of a case as a control. However, because Parkinson's disease is likely uncommon among the controls, this bias is unlikely to have a meaningful impact on the results. A limitation of this study is analysis of participants of predominately European ancestry. It will be important to replicate these in non-European cohorts as data becomes available from efforts such as GP2.³²

Acknowledgements

This research has been conducted using the UK Biobank Resource under Application Number 33601. This work utilized the computational resources of the NIH HPC Biowulf cluster (<http://hpc.nih.gov>). Data used in the preparation of this article were obtained from the AMP PD Knowledge Platform. For up-to-date information on the study, <https://www.amp-pd.org>.

AMP PD—a public-private partnership—is managed by the FNIH and funded by Celgene, GSK, the Michael J. Fox Foundation for Parkinson's Research, the National Institute of Neurological Disorders and Stroke, Pfizer, Sanofi, and Verily. Clinical data and biosamples used in preparation of this article were obtained from the Fox Investigation for New Discovery of Biomarkers (BioFIND), the Harvard Biomarker Study (HBS), the Parkinson's Progression Markers Initiative (PPMI), and the Parkinson's Disease Biomarkers Program (PDBP). BioFIND is sponsored by the Michael J. Fox Foundation for Parkinson's Research with support from NINDS. The BioFIND Investigators have not participated in reviewing the data analysis or content of the article. For up-to-date information on the study, visit michaeljfox.org/biofind. The Harvard Biomarkers Study (HBS) is a collaboration of HBS investigators (full list of HBS investigator found at <https://www.bwhparkinsoncenter.org/biobank/>) and funded through philanthropy and NIH and non-NIH funding sources. The HBS investigators have not participated in reviewing the data analysis or content of the article. PPMI, a public-private partnership, is funded by the Michael J. Fox Foundation for Parkinson's Research and funding partners (the full names of all the PPMI funding partners can be found at www.ppmi-info.org/fundingpartners). The PPMI investigators have not participated in reviewing the data analysis or content of the article. For up-to-date information on the study, visit www.ppmi-info.org. The PDBP consortium is supported by the NINDS at the NIH. A full list of PDBP investigators can be found at <https://pdbp.ninds.nih.gov/policy>. The PDBP investigators have not participated in reviewing the data analysis or content of the article.

Funding

This work was supported in part by the Intramural Research Program of the NINDS, NIH: project number 1ZIA-NS003169. This work was supported in part by the Intramural Research Programs of the National Institute on Aging (NIA): project numbers 1ZIA-NS003154, Z01-AG000949-02, and Z01-ES10198.

Competing interests

The authors declare no competing interest.

Supplementary material

Supplementary material is available at *Brain* online.

References

- Marras C, Beck JC, Bower JH, et al. Prevalence of Parkinson's disease across North America. *NPJ Parkinsons Dis.* 2018;4:21.
- Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. *Lancet Neurol.* 2020;19(2):170–178.
- Kitada T, Asakawa S, Hattori N, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature.* 1998;392(6676):605–608.
- Mochizuki H. Parkin gene therapy. *Parkinsonism Relat Disord.* 2009;15:S43–S45.
- Kunath T, Natalwala A, Chan C, et al. Are PARKIN patients ideal candidates for dopaminergic cell replacement therapies? *Eur J Neurosci.* 2019;49(4):453–462.
- Kasten M, Hartmann C, Hampf J, et al. Genotype-phenotype relations for the Parkinson's disease genes Parkin, PINK1, DJ1: MDSGene systematic review. *Mov Disord.* 2018;33(5):730–741.
- Lopez Lopez C, Tariot PN, Caputo A, et al. The Alzheimer's prevention initiative generation program: Study design of two randomized controlled trials for individuals at risk for clinical onset of Alzheimer's disease. *Alzheimers Dement.* 2019;5:216–227.
- Brüggemann N, Klein C. Parkin type of early-onset Parkinson disease. In: Adam MP, Ardinger HH, Pagon RA, et al., eds. *GeneReviews*®. University of Washington; 1993. Accessed July 29, 2021. <http://www.ncbi.nlm.nih.gov/books/NBK1478/>
- 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.
- Lubbe SJ, Bustos BI, Hu J, et al. Assessing the relationship between monoallelic *PRKN* mutations and Parkinson's risk. *Hum Mol Genet.* 2021;30(1):78–86.
- Sudlow C, Gallacher J, Allen N, et al. UK biobank: An open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* 2015;12(3):e1001779.
- Iwaki H, Leonard HL, Makarios MB, et al. Accelerating medicines partnership: Parkinson's disease. Genetic resource. *Mov Disord.* 2021;36(8):1795–1804.
- Liu Y-T, Sliter DA, Shammass MK, et al. Mt-Keima detects PINK1-PRKN mitophagy in vivo with greater sensitivity than mito-QC. *Autophagy.* 17;2021:3753–3762.
- Nalls MA, Bras J, Hernandez DG, et al. NeuroX, a fast and efficient genotyping platform for investigation of neurodegenerative diseases. *Neurobiol Aging.* 2015;36(3):1605.e7–1605.e12.
- Blauwendraat C, Faghri F, Pihlstrom L, et al. NeuroChip, an updated version of the NeuroX genotyping platform to rapidly screen for variants associated with neurological diseases. *Neurobiol Aging.* 2017;57:247.e9–247.e13.
- Bandres-Ciga S, Ahmed S, Sabir MS, et al. The genetic architecture of Parkinson disease in Spain: Characterizing population-specific risk, differential haplotype structures, and providing etiologic insight. *Mov Disord.* 2019;34(12):1851–1863.
- Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res.* 2010;38(16):e164.

18. Chen X, Schulz-Trieglaff O, Shaw R, et al. Manta: rapid detection of structural variants and indels for germline and cancer sequencing applications. *Bioinformatics*. 2016;32(8):1220–1222.
19. Collins RL, Brand H, Karczewski KJ, et al. A structural variation reference for medical and population genetics. *Nature*. 2020; 581(7809):444–451.
20. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203–209.
21. Karczewski KJ, Francioli LC, Tiao G, et al. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv*. [Preprint] <https://doi:10.1101/531210>
22. Zhuang Y, Weiner AM. A compensatory base change in U1 snRNA suppresses a 5' splice site mutation. *Cell*. 1986;46(6): 827–835.
23. Khan NL, Katzenschlager R, Watt H, et al. Olfaction differentiates parkin disease from early-onset parkinsonism and Parkinson disease. *Neurology*. 2004;62(7):1224–1226.
24. Alcalay RN, Siderowf A, Ottman R, et al. Olfaction in Parkin heterozygotes and compound heterozygotes: The CORE-PD study. *Neurology*. 2011;76(4):319–326.
25. Mor-Shaked H, Paz-Ebstein E, Basal A, et al. Levodopa-responsive dystonia caused by biallelic PRKN exon inversion invisible to exome sequencing. *Brain Commun*. 2021;3(3):fcab197.
26. Ziviani E, Tao RN, Whitworth AJ. Drosophila parkin requires PINK1 for mitochondrial translocation and ubiquitinates mitofusins. *Proc Natl Acad Sci USA*. 2010;107(11):5018–5023.
27. Poole AC, Thomas RE, Yu S, Vincow ES, Pallanck L. The mitochondrial fusion-promoting factor mitofusins is a substrate of the PINK1/parkin pathway. *PLoS One*. 2010;5(4):e10054.
28. Gegg ME, Cooper JM, Chau K-Y, Rojo M, Schapira AHV, Taanman J-W. Mitofusins 1 and mitofusins 2 are ubiquitinated in a PINK1/parkin-dependent manner upon induction of mitophagy. *Hum Mol Genet*. 2010;19(24):4861–4870.
29. Tanaka A, Cleland MM, Xu S, et al. Proteasome and p97 mediate mitophagy and degradation of mitofusins induced by Parkin. *J Cell Biol*. 2010;191(7):1367–1380.
30. Rakovic A, Grünewald A, Kottwitz J, et al. Mutations in PINK1 and Parkin impair ubiquitination of Mitofusins in human fibroblasts. *PLoS One*. 2011;6(3):e16746.
31. Koentjoro B, Park J-S, Ha AD, Sue CM. Phenotypic variability of parkin mutations in single kindred. *Mov Disord*. 2012;27(10):1299–1303.
32. Global Parkinson's Genetics Program. GP2: The global Parkinson's genetics program. *Mov Disord*. 2021;36(4): 842–851.
33. Yu E, Rudakou U, Krohn L, et al. Analysis of heterozygous PRKN variants and copy-number variations in Parkinson's disease. *Mov Disord*. 2021;36(1):178–187.
34. Nalls MA, Blauwendraat C, Vallergera 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. *Lancet Neurol*. 2019;18(12):1091–1102.
35. Guo J-F, Wang L, He D, et al. Clinical features and [11C]-CFT PET analysis of PARK2, PARK6, PARK7-linked autosomal recessive early onset Parkinsonism. *Neurol Sci*. 2011;32(1):35–40.
36. Hilker R, Klein C, Ghaemi M, et al. Positron emission tomographic analysis of the nigrostriatal dopaminergic system in familial parkinsonism associated with mutations in the parkin gene. *Ann Neurol*. 2001;49(3):367–376.
37. Hilker R, Klein C, Hedrich K, et al. The striatal dopaminergic deficit is dependent on the number of mutant alleles in a family with mutations in the parkin gene: evidence for enzymatic parkin function in humans. *Neurosci Lett*. 2002; 323(1):50–54.
38. Pavese N, Khan NL, Scherfler C, et al. Nigrostriatal dysfunction in homozygous and heterozygous parkin gene carriers: an 18F-dopa PET progression study. *Mov Disord*. 2009;24(15):2260–2266.
39. Khan NL, Brooks DJ, Pavese N, et al. Progression of nigrostriatal dysfunction in a parkin kindred: an [18F]dopa PET and clinical study. *Brain*. 2002;125(Pt 10):2248–2256.
40. 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.
41. Anders S, Sack B, Pohl A, et al. Compensatory premotor activity during affective face processing in subclinical carriers of a single mutant Parkin allele. *Brain*. 2012;135(Pt 4):1128–1140.
42. Buhmann C, Binkofski F, Klein C, et al. Motor reorganization in asymptomatic carriers of a single mutant Parkin allele: a human model for presymptomatic parkinsonism. *Brain*. 2005; 128(Pt 10):2281–2290.
43. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry*. 1992;55(3):181–184.