



RESEARCH ARTICLE

Compound Heterozygous Structural Variants in Cases with Unsolved *PRKN*-Associated Parkinson's Disease

Agata Fant, MSc,¹ Sara Trova, PhD,¹  Edoardo Monfrini, MD, PhD,^{2,3}  Gaia Treves, MSc,¹ Francesco Musacchia, PhD,^{1,4} Fabio Landuzzi, PhD,⁵ Paola Mandich, MD, PhD,⁶ Antonio Amoroso, MD,¹ Remo Sanges, PhD,^{4,7} Luca Pandolfini, PhD,⁴  Francesco Cavallieri, MD, PhD,⁸  Franco Valzania, MD,⁹ Valentina Fioravanti, MD, PhD,¹⁰ Giulia Di Rauso, MD,^{8,9}  Gloria Brescia, PhD,¹⁰ Enza Maria Valente, MD, PhD,^{11,12}  Valeria Tiranti, MSc,¹³ Luigi Michele Romito, MD, PhD,¹⁴  Chiara Reale, MSc,¹³  Barbara Garavaglia, PhD,¹³ Antonio Emanuele Elia, MD, PhD,¹⁴ Andrea Cavalli, PhD,^{5,15,16} Alessio Di Fonzo, MD, PhD,^{3*}  Manuela Vecchi, PhD,^{1*}  and Stefano Gustincich, PhD^{1,4*} 

ABSTRACT: Background: Biallelic mutations in the *PRKN* gene are a common cause of early-onset Parkinson's disease (EOPD). In addition to single nucleotide variants, structural variants contribute substantially to the mutational profile of *PRKN*. A significant portion of patients with EOPD remains genetically unsolved.

Objectives: By using short-read whole genome sequencing (sr-WGS), we aimed to uncover complex genetic alterations at the *PRKN* locus in EOPD cases which tested negative for mutations in Mendelian PD genes with clinical exome sequencing (CES) and multiplex ligation-dependent probe amplification (MLPA).

¹Non-Coding RNAs and RNA-Based Therapeutics, Italian Institute of Technology (IIT), CMP3VdA, Aosta, Italy; ²Dino Ferrari Center, Neuroscience Section, Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy; ³Foundation IRCCS Ca' Granda Ospedale Maggiore Policlinico, Neurology Unit, Milan, Italy; ⁴Center for Human Technologies, Non-Coding RNAs and RNA-Based Therapeutics, Italian Institute of Technology (IIT), Genoa, Italy; ⁵Computational and Chemical Biology, Italian Institute of Technology (IIT), CMP3VdA, Aosta, Italy; ⁶Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics and Maternal and Child Health, University of Genoa, Genoa, Italy; ⁷Area of Neuroscience, International School for Advanced Studies (SISSA), Trieste, Italy; ⁸Neurology Unit, Neuromotor & Rehabilitation Department, Azienda USL-IRCCS di Reggio Emilia, Reggio Emilia, Italy; ⁹Clinical and Experimental Medicine PhD Program, University of Modena and Reggio Emilia, Modena, Italy; ¹⁰Department of Pathophysiology and Transplantation, University of Milan, Milan, Italy; ¹¹Department of Molecular Medicine, University of Pavia, Pavia, Italy; ¹²Neurogenetics Research Centre, IRCCS Mondino Foundation, Pavia, Italy; ¹³Medical Genetics and Neurogenetics Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy; ¹⁴Department of Clinical Neurosciences, Parkinson and Movement Disorders Unit, Fondazione IRCCS Istituto Neurologico Carlo Besta, Milan, Italy; ¹⁵Center for Human Technologies, Computational and Chemical Biology, Italian Institute of Technology (IIT), Genoa, Italy; ¹⁶Centre Européen de Calcul Atomique et Moléculaire (CECAM), Ecole Polytechnique Fédérale de Lausanne, Lausanne, Switzerland

This is an open access article under the terms of the [Creative Commons Attribution](https://creativecommons.org/licenses/by/4.0/) License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

*Correspondence to: Stefano Gustincich, CMP³VdA, 11100 Aosta, Italy; E-mail: stefano.gustincich@iit.it; Manuela Vecchi, Italian Institute of Technology (IIT), CMP³VdA, 11100 Aosta, Italy; E-mail: manuela.vecchi@iit.it; Alessio Di Fonzo, Foundation IRCCS Ca' Granda Ospedale Maggiore Policlinico, 20122, Milan, Italy; E-mail: alessio.difonzo@policlinico.mi.it

Agata Fant and Sara Trova contributed equally to this article.

Stefano Gustincich, Manuela Vecchi, and Alessio Di Fonzo are co-last authors and co-corresponding authors.

Relevant conflicts of interest/financial disclosures: There are no financial disclosures or conflicts of interest reported by the authors related to this article.

Funding agencies: This work was funded by the Project 5000genomi@VdA, co-funded by "Fondo Europeo di Sviluppo Regionale" (FESR CUPB68H19005520007) to A.A., A.C., and S.G.; S.T., F.M., M.V., and F.L. were supported by "Fondo Europeo di Sviluppo Regionale (FESR CUPB68H19005520007); A.F. and G.T. were recipient of a fellowship from Programma Investimenti per la crescita e l'occupazione 2014/20" (European Social Fund, ESF CUPB65F19001200009), and G.S., G.T., and A.F. continued their activity with a fellowship from Programma Regionale Fondo sociale europeo Plus 2021/2027 (European Social Fund, ESF CUPJ51B24000170002). A.D.F. acknowledges grant support by the EJP RD Joint Transnational Call 2022) and "Fondazione Regionale per la Ricerca Biomedica" awarded to the project PreDYT (PREdictive biomarkers in DYsTonia, grant agreement number: 825575). V.T., L.M.R., C.R., B.G., and A.E.E. were supported by "The Mariani Foundation and the Italian Ministry of Health". The Parkinson's Progression Markers Initiative (PPMI) – a public–private partnership – is funded by The Michael J. Fox Foundation for Parkinson's Research and funding partners, including 4D Pharma, AbbVie, AcureX, Allergan, Amathus Therapeutics, Aligning Science Across Parkinson's, AskBio, Avid Radiopharmaceuticals, BIAL, BioArctic, Biogen, Biohaven, BioLegend, BlueRock Therapeutics, Bristol-Myers Squibb, Calico Labs, Capsida Biotherapeutics, Celgene, Cerevel Therapeutics, Coave Therapeutics, DaCapo Brainscience, Denali, Edmond J. Safra Foundation, Eli Lilly, Gain Therapeutics, GE HealthCare, Genentech, GSK, Golub Capital, Handl Therapeutics, Insiteo, Jazz Pharmaceuticals, Johnson & Johnson Innovative Medicine, Lundbeck, Merck, Meso Scale Discovery, Mission Therapeutics, Neurocrine Biosciences, Neuron23, Neupore, Pfizer, Piramal, Prevail Therapeutics, Roche, Sanofi, Servier, Sun Pharma Advanced Research Company, Takeda, Teva, UCB, Vanqua Bio, Verily, Voyager v. 25MAR2024 Therapeutics, the Weston Family Foundation, and Yumanity Therapeutics.

Received: 3 April 2025; **Revised:** 4 August 2025; **Accepted:** 12 August 2025

Published online in Wiley Online Library ([wileyonlinelibrary.com](https://www.wileyonlinelibrary.com)). DOI: 10.1002/mds.70027

Methods: We evaluated 498 unrelated EOPD patients, who tested negative using gold-standard diagnostic methods, using sr-WGS. In selected cases, long-read whole genome sequencing (lr-WGS) with Oxford Nanopore technology was employed for an in-depth analysis and validation. The Parkinson's Progression Markers Initiative (PPMI) dataset was interrogated to assess the prevalence of any newly identified elusive pathogenic genetic configurations.

Results: sr-WGS revealed elusive compound heterozygous structural variations, consisting of partially overlapping deletions and duplications within the *PRKN* gene in three unrelated EOPD cases (two familial, one sporadic). In familial cases, biallelic *PRKN* structural variants co-segregated with the disease. The exact structure of

each variant was resolved using lr-WGS. Similar variants were absent in the large PPMI database, suggesting that they are a rare occurrence.

Conclusions: In this article we describe a rare configuration of compound heterozygous structural variations involving partially overlapping chromosomal regions at the *PRKN* locus, which are difficult to detect through standard diagnostic genetic technologies. This study highlights the importance of integrating WGS into clinical practice. © 2025 The Author(s). *Movement Disorders* published by Wiley Periodicals LLC on behalf of International Parkinson and Movement Disorder Society.

Key Words: *PRKN*; Parkinson's disease; WGS; structural variants; diagnostic

Parkinson's disease (PD) is the second most common neurodegenerative disorder, affecting 1% of the population over the age of 60 years, with prevalence increasing to 4% by age 80 years.¹ It is considered a multifactorial disorder, as both genetic and environmental factors can contribute to its pathology.² While most cases are sporadic, up to 20% present as familial.^{3,4} Variants in the *PRKN* gene are the most common cause of autosomal recessive PD, accounting for between 2.6% and 14.9% of early-onset Parkinson's disease (EOPD) cases.⁵⁻⁷ Patients with *PRKN* mutations typically exhibit a median age of onset at 31 years, with a predominantly motor-related presentation that responds well to dopaminergic therapy, progresses slowly, and shows little to no cognitive decline.⁸

The *PRKN* gene spans 1.3 Mb and contains 12 exons encoding for the 465 amino acid Parkin protein. *PRKN* is located within FRA6E, a common fragile site prone to genomic instability. Its core region, spanning exons 3 to 8, represents a known mutation hotspot, making it particularly susceptible to rearrangements.⁹ Parkin plays a crucial role in the physiology of dopaminergic neurons of the mesencephalon, the cell type that degenerates in PD. It is an E3 ubiquitin ligase and mediates mitophagy, which is the selective degradation of mitochondria, a function essential for mitochondrial homeostasis.¹⁰ Given its neuroprotective role, *PRKN* mutations compromise the ability of this protein to regulate mitochondrial function leading to dopaminergic cells degeneration and PD.¹¹

The up-to-date typical diagnostic approaches in PD involve multiplex ligation-dependent probe amplification (MLPA) and targeted next-generation sequencing (NGS) gene panels or clinical/whole exome sequencing (CES/WES). However, these tests exhibit low sensitivity for complex structural variations. By taking advantage of these and other technologies, various types and combinations of *PRKN* pathogenic variants have been found to be causally associated to PD, including single nucleotide variants (SNVs) and

structural variants (SVs).⁸ Nevertheless, an important portion of typical EOPD cases remain genetically unsolved.

In this study, we demonstrate that whole genome sequencing (WGS) can successfully solve some of these typical EOPD cases by identifying elusive combinations of SVs in the *PRKN* gene. Prior gold-standard diagnostic methods had failed to detect these instances of compound heterozygosity, leaving these cases undiagnosed for years despite extensive genetic investigations.

Methods

Study Participants

Patients with PD, diagnosed according to International Parkinson and Movement Disorder Society (MDS) clinical diagnostic criteria,¹² were enrolled in a prospective observational study, entitled NeurOmics (Prot. NeurOmics: 209/2024 - DB id 13 896) designed to investigate omics alterations for the identification of pathogenic mechanisms and biomarkers in neurodegenerative disorders. Written informed consent for publication of pseudonymized genetic and clinical details was obtained from all the subjects involved. For this study cohort, 498 unrelated EOPD were selected by neurologists with expertise in movement disorders and neurogenetics to be negative or with inconclusive results to previous gold-standard diagnostic assays (CES and MLPA) for genes known to be associated with Mendelian forms of PD (namely *SNCA*, *LRRK2*, *PRKN*, *PINK1*, *PARK7*, *DNAJC6*, *SYNJ1*, *VPS13C*, *PLA2G6*, *ATP13A2*, *FBXO7*, *RAB39B*, *PTRHD1*, and *GRN*). For diagnostic genetic testing in the clinical setting, CES was performed using the Sure Select CD Clinical Focused Exome kit (Agilent Technologies) that enables analysis of only disease-associated targets (5629 targets in total) on the Illumina NextSeq2000 platform, and CNVs in *PRKN*, *PARK7*, *PINK1*, *SNCA*, and

ATP13A2 were analyzed by MLPA with the SALSA MLPA Parkinson probemix P051 (MRC Holland).

This cohort includes patients carrying monoallelic pathogenic variants in *PRKN* ($n = 32$) and *PINK1* ($n = 2$). Concerning the index cases of Families A and B, we enrolled additional family members, specifically $n = 2$ affected siblings and $n = 4$ first-degree healthy relatives from Family A, and $n = 1$ affected and $n = 1$ healthy sibling from Family B. Clinical features of the whole cohort are reported in Table S1.

DNA Extraction and Quality Assessment

Genomic DNA was extracted from 400 μ L of fresh whole blood using the Qiacube Automated Nucleic Acid Extraction Kit on the Qiacube Connect instrument (Qiagen), following the manufacturer's instructions.

The extracted genomic DNA (gDNA) was quantified using a Nanodrop OneC spectrophotometer (Life Technologies, Thermo Fisher Scientific) to assess purity ratios, and a Qubit Fluorometer 4.0 (Thermo Fisher Scientific) to determine nucleic acid concentration (ng/ μ L) using the QubitTM dsDNA High Sensitivity Assay Kit (InvitrogenTM, Thermo Fisher Scientific). The quality of gDNA was also assessed for size distribution using the 4200 TapeStation system (Agilent Technologies) with the Genomic DNA ScreenTape. For all samples, the DNA integrity number (DIN) was >7 as required for Nanopore sequencing.

Short-Read Whole Genome Sequencing

sr-WGS of gDNA extracted from blood samples was performed with the Illumina NovaSeq 6000 platform (Illumina Inc.) to produce paired 151 base pair reads. The library preparation and sequencing of gDNA from the blood samples utilized barcode adapters. Specifically, 350 ng of gDNA was prepared using the Illumina DNA PCR-Free Prep, Tagmentation library preparation kit (Cat. No. 20041795), and IDT[®] for Illumina[®] DNA/RNA Unique Dual Indexes Set A and B, Tagmentation (96 indexes, Cat. Nos 20027213 and 20027214). The library preparation was carried out with a T100 Thermal Cycler instrument (Bio-Rad) and standard laboratory equipment. Libraries were pooled by volume and adjusted for the index correction factor according to Illumina's instructions to ensure balanced sample coverage across each run. An average coverage of 39.1X (± 8.7) was achieved for Families A and B and the single case (Table S2).

Alignment on GRCh38 and variant calling for SNVs was performed with the GPU optimized version of the Genome Analysis Toolkit (GATK) developed by NVIDIA ClaraTM Parabricks[®]. For SV detection, to decrease false-positive calls, we use four different tools (Manta v1.6.0,¹³ Lumpy v0.2.13,¹⁴ CNVnator v0.3.3,¹⁵ and BreakDancer v1.4.5¹⁶) making them a consensus

using Survivor (v1.0.7)¹⁷ and selecting only SVs detected by at least two tools. SVs were annotated first with SVAfotate¹⁸ to obtain population level allele frequency information and then with AnnotSV v3.2.3.¹⁹

Long-Read Whole Genome Sequencing

Selected gDNA samples used for sr-WGS were also analyzed by long-read whole genome sequencing (lr-WGS). The gDNA was further fragmented by pipetting it through a p20 tip (Gilson) and fragments below 10 kb were removed using the Short Read Eliminator XS (SRE XS, PacBio[®] #102-208-200). Library preparation was performed using the DNA Ligation Sequencing Kit (SQK-LSK114) according to the manufacturer's instructions. Long-read sequencing was performed on a FLO-PRO00114M flow cell (R10.4.1) using the PromethION 24 sequencer (Oxford Nanopore Technologies, ONT). For all samples, the starting material was >1500 ng and a total of >500 ng was obtained for each library. Initially, 30–50fM were directly loaded on the flow cell and after 36 hr a second loading with 30–50fM was required, following a flow cell nuclease wash (according to ONT instructions). An average coverage of 23.6X (± 6.9) and average N50 of 22.5 kb (± 5.9) was achieved for Families A and B and the single case (Table S3).

Basecalling and concurrent genome alignment were performed with dorado (ONT; v0.6.0). For variant calling and phasing, the ONT wf-human-variation pipeline v.2.2.4 was used, with options “–sv –snv –cnv –phase”. Sniffles2 options were modified with “–long-del-length 520000 –long-dup-length 520000” to increase the SV size limits of detection.

Validation of Compound Heterozygous SVs in *PRKN* Using PCR

The same gDNA used for the WGS analysis was used for the validation with polymerase chain reaction (PCR). Additional gDNA for PCR validation was available for I-2, II-1, II-2, II-3 in Family A, for II-1 and II-2 in Family B, and for the single case (SC).

The primers for each SV were designed with the PRIMER BLAST tool²⁰ from the National Center for Biotechnology Information (NCBI) website to target the regions surrounding the breakpoints of the deletion and duplication (Fig. S1), after checking for non-specific regions of annealing with the same online tool. The PCR reaction was carried out with TaKaRa Ex Taq DNA Polymerase according to manufacturer's instructions in a T100 Thermal Cycler (Biorad) under the following conditions: 30 cycles of 98°C for 10 s, 60°C for 30 s, 72°C for the time needed for the elongation (60 s each 1000 bp amplified). Around 50 ng of DNA were used as input for each sample. All samples were run in a 2% agarose gel made with 1X TAE buffer and SYBR Safe DNA Gel Stain (Invitrogen), and GeneRuler

100 bp Plus DNA (Invitrogen) was used as ladder. Commercial primers for glyceraldehyde 3-phosphate dehydrogenase (GAPDH) were used as the positive control (Hs02786624_g1, Thermo Fisher). All the primers were newly designed (Table S4).

Analysis of WGS Data of the PPMI Cohort

Data used for this article were obtained on November 15, 2020 from the Parkinson's Progression Markers Initiative (PPMI) database (www.ppmi-info.org/access-dataspecimens/download-data), RRID:SCR_006431. Clinical information for the reanalyzed cases was obtained from the curated data provided by the PPMI data team in the latest PPMI release (21-03-2025) and is summarized in Table S5. A total of 851 subjects were included, all lacking pathogenic variants in *LRKK2*, *PINK1*, *PRKN*, or *SNCA* according to the curated PPMI data. Information about pathogenic variants was obtained from Participant_Status and iu_genetic_consensus tables. This cohort comprised 144 prodromal individuals, 63 SWEDD cases (clinically diagnosed with PD but with normal dopamine transporter single photon emission computed tomography [DAT SPECT]), 195 healthy individuals, and 449 PD patients (87 EOPD, 354 late-onset Parkinson's disease [LOPD], and 8 without available onset information). SV detection was performed using the same pipeline as previously described.

Results

Study Cohort

Our study cohort comprised 498 unrelated typical EOPD patients from Italy. In addition, three affected and five healthy relatives from two families were enrolled. The larger family, designated as Family A, consists of two healthy parents and five siblings (three affected), while the other family, Family B, is composed of three siblings (two affected). Clinical and demographic features of the analyzed cohort are reported in Table S1. All EOPD cases were previously analyzed using gold-standard diagnostic assays (CES and MLPA) yielding negative or inconclusive results for genes known to be associated with Mendelian forms of PD (see Methods).

Identification of Biallelic SVs in *PRKN* by WGS

sr-WGS was employed to reconstruct the *PRKN* gene locus. To analyze WGS data, we implemented an ad hoc GPU-based pipeline for SNV variant calling and a consensus approach of four different tools (Manta v1.6.0,¹³ Lumpy v0.2.13,¹⁴ CNVnator v0.3.3,¹⁵ and BreakDancer v1.4.5¹⁶) using Survivor (v1.0.7)¹⁷ to detect SVs (as described in Methods). We initially analyzed the two unrelated probands of Family A and Family B and identified a peculiar configuration of

compound heterozygous SVs of *PRKN* involving partially overlapping regions. For these two familial cases, we extended the analysis to available family members (Fig. 1A,D).

Regarding Family A, genomic DNA was available for all the members of the family, namely two healthy parents, two healthy sisters, and three affected siblings. Bioinformatic analysis identified in the three affected individuals (II-3, II-4, and II-5) a unique combination of SVs in the *PRKN* gene, consisting of a partially overlapping duplication (of ~509 kb) and a deletion (of ~405 kb) in *trans*, both spanning exons 2–3–4 of *PRKN* (Fig. 1A,B, Table 1). The two healthy sisters (II-1, II-2) carried only the deletion, inherited from the father (I-1), while the duplication present in the three affected siblings was inherited from the mother (I-2) (Fig. 1A,B).

In Family B, genomic DNA was available for two affected and one healthy brother. Both affected siblings (II-2, II-3) carried a ~15 kb duplication and a ~133 kb deletion both spanning exon 3 of *PRKN* while the healthy brother (II-1) presented the deletion only (Fig. 1C,D; Table 1).

Collectively, we demonstrated the biallelic configuration of the identified *PRKN* SVs (Fig. 1).

We then extended the sr-WGS analysis to the entire EOPD cohort. An additional sporadic EOPD case (SC) was detected carrying a *PRKN* deletion involving exons 3 and 4 on one allele, and a duplication of exon 3 on the other allele (Fig. 1E,F; Table 1). This configuration was not captured by MLPA, as only a dosage alteration of exon 4 was detected (Fig. S2).

WGS did not identify any relevant variants in other PD genes in these carriers (see Methods).

All the identified SVs are classified as pathogenic (https://www.mdsgene.org/d/1/g/4?fc=0&mu=0&_mu=1&_country=1) and their association in *trans* indicated that *PRKN* variants were the genetic cause of EOPD in these cases. The presence of overlapping deletions and duplications with breakpoints in introns in the two different alleles of *PRKN* explains the missed diagnosis through gold-standard diagnostic methods, as MLPA could not reveal exon dosage abnormalities and CES did not cover these regions (Fig. S2). Overall, three unrelated EOPD cases (0.6%, 3/498), two familial and one sporadic, were found to carry a peculiar configuration of compound heterozygous SVs of *PRKN* involving partially overlapping regions. The clinical phenotype of carriers was highly suggestive of *PRKN*-related PD (Table 2).

For a precise characterization of the compound heterozygous *PRKN* SVs identified by sr-WGS, lr-WGS was also performed in all the probands and available family members, taking advantage of the Oxford Nanopore Technologies platform. lr-WGS confirmed the presence of biallelic *PRKN* SVs (Fig. 2; Fig. S3) in

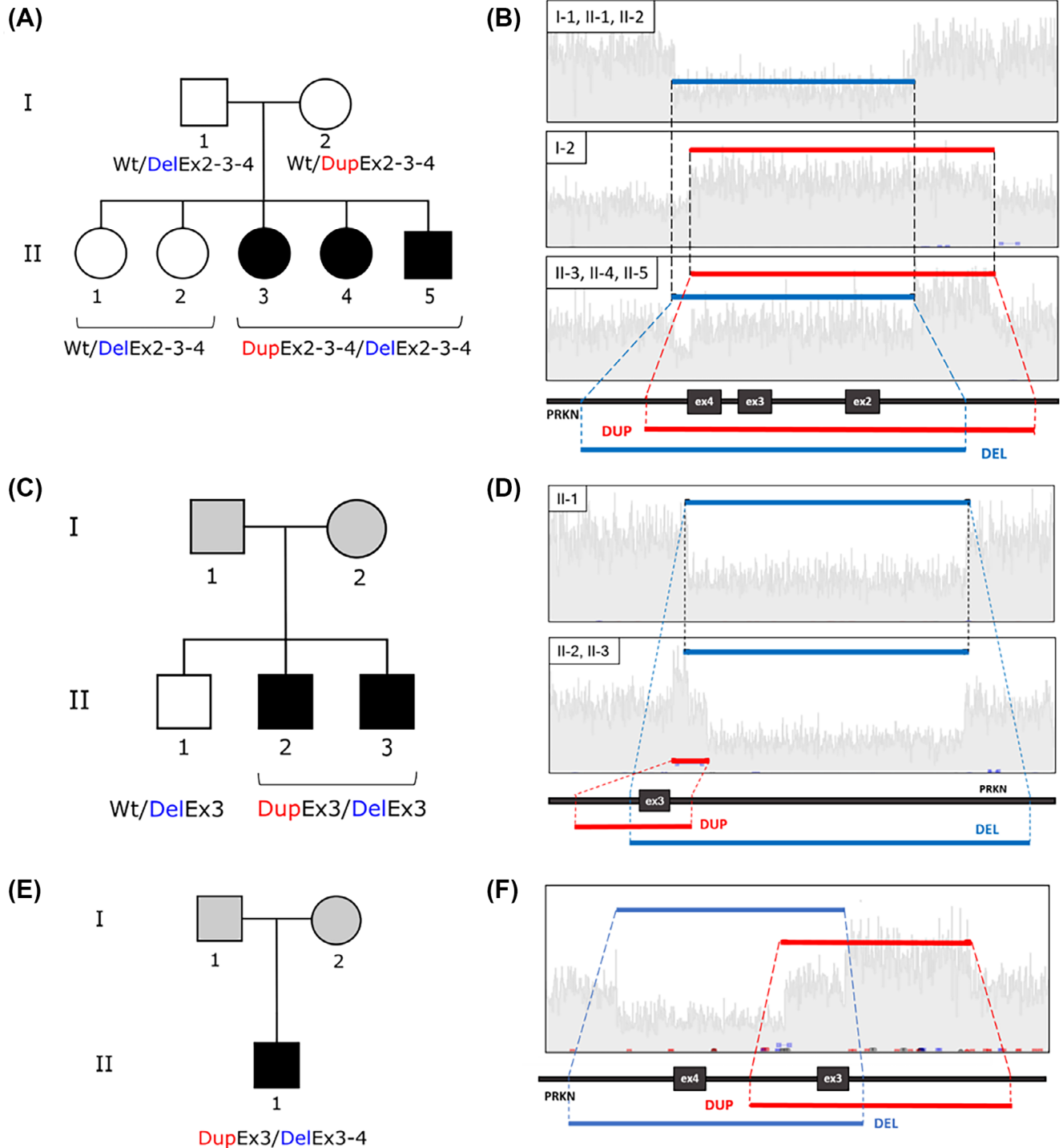


FIG. 1. Configuration of compound heterozygous structural variations of *PRKN* involving partially overlapped regions. (A, C, E) Pedigrees of Family A, Family B, and the single case, respectively, showing affected individuals (black symbols) and subjects not sequenced (grey symbols). (B, D, F) Samplot images depicting the deletions and duplications in *PRKN* identified in Family A, Family B, and the sporadic case, respectively.

PD probands and monoallelic *PRKN*-SVs in healthy relatives of the two analyzed families. Additionally, IrWGS allowed us to precisely determine all the SVs breakpoints, which were confirmed through genomic PCR amplification (Table 1; Table S4; Fig. S1).

Analysis of Compound *PRKN* SVs in the PPMI Cohort

The analysis was extended to the WGS data extracted from PPMI dataset by applying the same tailored bioinformatic pipeline to detect SVs. We analyzed a total of

Table 1 Location of the breakpoints of all the *PRKN* structural variants identified in the two families and the sporadic Parkinson's disease case with short-read whole genome sequencing, aligned on hg38.

Family	Subjects	Location	Start	End	SV type	Size (bp)	Exons involved
A	I-1, II-1, II-2, II-3, II-4, II-5	6q26	162,154,938	162,560,715	Del	405,964	2,3,4
	I-2, II-3, II-4, II-5		162,185,434	162,695,422	Dup	508,975	2,3,4
B	II-1, II-2, II-3	6q26	162,260,786	162,393,577	Del	133,194	3
	II-2, II-3		162,253,473	162,269,839	Dup	15,125	3
Single case	SC	6q26	162,149,134	162,265,728	Del	116,594	3,4
			162,234,822	162,328,795	Dup	93,973	3

Abbreviations: SV, structural variant; bp, base pair; del, deletion; dup, duplication.

851 individuals without pathogenic variants in *LRRK2*, *PINK1*, *PRKN*, or *SNCA*, including 449 PD patients (87 EOPD, 354 LOPD, and 8 with unknown age at onset), 144 prodromal subjects, 63 SWEDD cases (clinically diagnosed with PD but exhibiting normal DAT SPECT imaging), and 195 healthy controls. The clinical characteristics of these groups are detailed in Table S5.

Initially, we compared the sequencing quality metrics between the PPMI and our Italian cohort and found no significant technical differences. Both cohorts were sequenced using Illumina technology with a PCR-free library preparation kit, achieving comparable average coverage: 34.79X (± 4.96) in the PPMI cohort versus 34.02X (± 7.13) in this Italian cohort. When focusing specifically on the EOPD subsets, 87 samples from PPMI and 498 from our cohort, the average coverage remained similar at 35.48X (± 5.29) versus 34.02X (± 7.1), respectively. No substantial technical or clinical differences were observed between the two cohorts (Table S1; Table S5), although the PPMI's EOPD subset was notably smaller than the cohort of the present study. Overall, 13 heterozygous SVs (either a deletion or a duplication comprising exons) were detected in the *PRKN* gene of all PPMI cohorts (Table S6). Nonetheless, no patient showed either compound SVs heterozygosity involving both a deletion and a duplication spanning overlapping genomic regions within the *PRKN* locus or compound heterozygosity with a pathogenic SV and a single nucleotide variant in *PRKN*.

Discussion

Current technologies used in clinical practice are precious tools to detect pathogenic variants in PD-related genes with high confidence and reproducibility.^{21,22} However, they present important limitations. Accumulating evidence indicates that SNVs or SVs could remain undetected due to challenges in accurately mapping these variants within analyzed sequences and the inherent difficulties in identifying SVs without

introducing artifacts.^{23,24} This limitation is particularly relevant when clinical phenotypic traits suggest the involvement of a specific disease-causing gene. The co-existence of inheritance, early-onset, and specific symptoms is typical of *PRKN*-PD patients.²⁵ Nevertheless, in several cases with a typical *PRKN*-PD clinical presentation, no pathogenic *PRKN* variants are identified, even after years of molecular diagnostic testing using traditional approaches.²⁶

In this study, we demonstrate that both short-read and long-read WGS approaches can accurately detect a rare configuration of compound heterozygous SVs in the *PRKN* gene (Figs. 1 and 2; Fig. S3). This configuration was identified in six Italian EOPD individuals, including five familial cases from two unrelated families and one sporadic patient, all of whom had previously undergone CES and MLPA analyses that were either negative or inconclusive (Fig. 1). All the identified variants involve partially overlapping chromosomal regions comprising exon 3, consistent with its location within FRA6E, a common fragile site known to be prone to genomic instability.⁹ This compound heterozygous SVs configuration differs from previously reported SV patterns, such as two deletions or two duplications in *trans*²⁷ or a duplication on one allele entirely overlapped by a deletion on the other, as recently reported in one European family and in one German EOPD case^{28,29}.

Our data corroborate recent findings reporting biallelic complex alterations in the *PRKN* gene in unsolved cases carrying a single heterozygous *PRKN* mutation and further expand the genetic heterogeneity of this specific locus in EOPD patients³⁰. Additionally, we unveiled partially overlapping SVs in *PRKN* also in patients in which conventional methods did not show any pathogenic *PRKN* variant. These findings from the PD cohort enrolled in Italy underscores the broader relevance of complex *PRKN* SVs in EOPD across diverse ancestral backgrounds. Finally, we calculated the prevalence of this distinctive type of genetic alteration, occurring in 0.6% of EOPD patients who had previously tested negative using standard diagnostic approaches.

Table 2 Clinical information of the affected Parkinson's disease cases with compound heterozygous PRKN structural variants.

Parameter	Family 1			Family 2			Proband 3	
	M	F	F	M	M	M	M	M
Age at onset (years)	32	40	34	38	37			40
Disease duration (years)	19	14	6	14	13			22
Non-motor symptoms	Sleeplessness, RLS	Anxiety, constipation	None	Constipation	Anxiety, depression, constipation			Depression, ICD (hypersexuality with dopamine-agonist)
Motor symptoms at onset	Right-hand rest tremor	Bilateral lower limbs bradykinesia	Right leg rest tremor	Left foot dystonia	Rigidity			Right upper limb bradykinesia
Other motor symptoms	Bradykinesia, rigidity.	Bradykinesia, rigidity, mild for limbs and gait ataxia.	Bradykinesia, rigidity	Bradykinesia, rigidity	Bradykinesia			Tremor
Clinical improvement of motor symptoms with levodopa	Yes	Yes	Yes	Yes	Yes			Yes
Current LEDD (mg)	105	790	105	660	300			900
Deep brain stimulation	No	No	No	No	No			Yes, favorable response (after 13 years of disease duration)
Other advanced therapies	No	No	No	No	No			No

Abbreviations: M, male; F, female; RLS, restless legs syndrome; ICD, impulse control disorder; LEDD, levodopa equivalent daily dose.

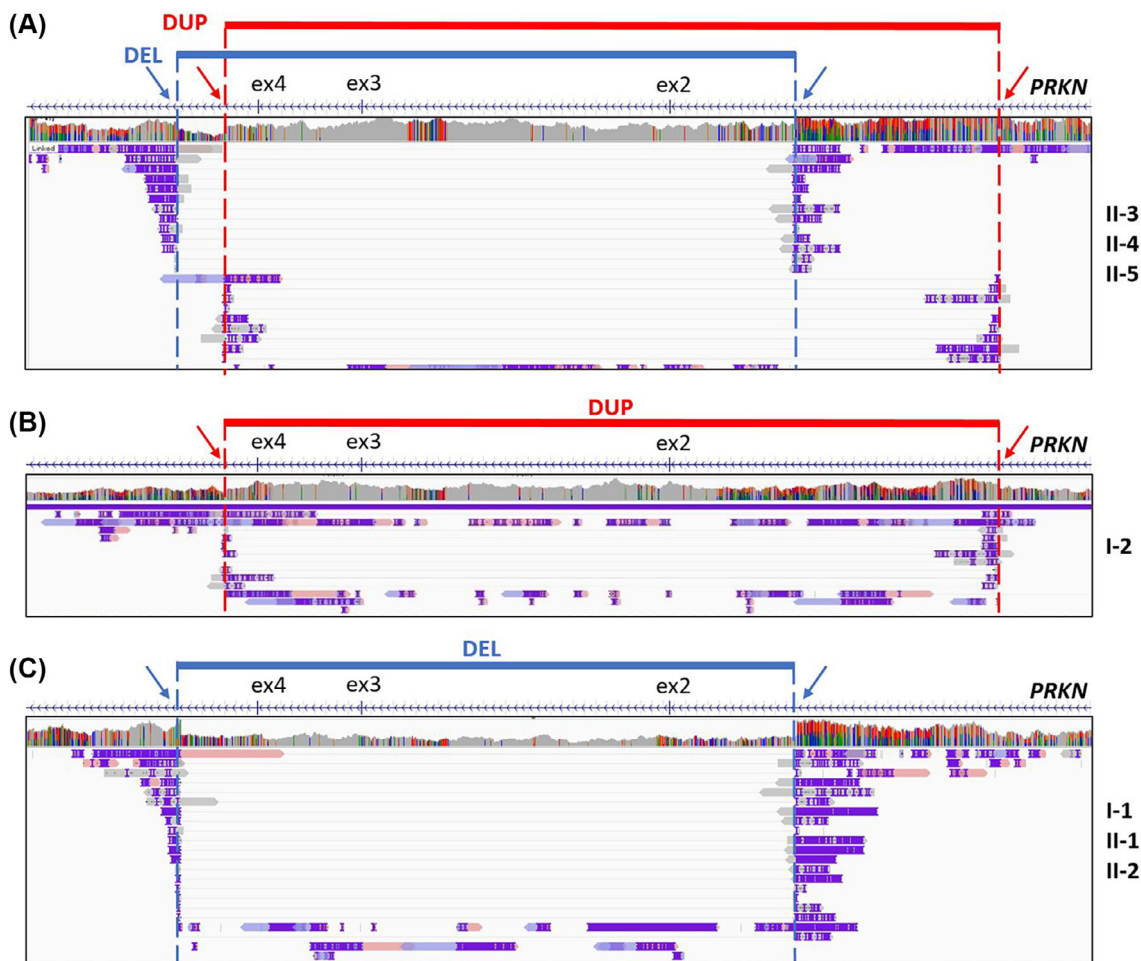


FIG. 2. Analysis of the identified biallelic *PRKN* structural variants by long-read sequencing in Family A. Representative images from the Integrative Genomics Viewer (IGV) of long-read sequencing split reads mapped on the *PRKN* gene of Family A, displaying the deletion of exons 2–3–4 (blue) and the duplication of exons 2–3–4 (red) in all the three affected siblings (A), the duplication in the mother (B), and the deletion in the father and unaffected siblings (C). Blue and red arrows indicate deletion and duplication breakpoints, respectively. Gray reads indicate the default state for reads that match the reference genome. Pink and violet reads are supplementary alignments.

A comprehensive analysis of the PPMI study dataset did not reveal any elusive biallelic *PRKN* SVs event in 87 EOPD cases, suggesting that these events occur at low frequency. As the sample size of the PPMI EOPD subset is limited (approximately one-sixth of the cohort presented in this study) we cannot exclude that the PPMI cohort lacked sufficient power to detect these rare SVs in *PRKN*. Of note, the PPMI cohort represents a more genetically and geographically diverse population than the relatively homogeneous Italian cohort included in the present study. The diversity in ethnic background may contribute to the variability in the detection of rare, population-specific SVs in *PRKN*.

These elusive SVs were confirmed using a third-generation sequencing platform, Oxford Nanopore Technologies (Fig. 2; Fig. S3). Nonetheless, sr-WGS was sufficient to accurately detect these compound heterozygous SVs without the need of more advanced technologies (Fig. 1; Table 1). sr-WGS effectively identifies the elusive biallelic SV configurations when analyzed

with split read-based methods. Since breakpoints are located within intronic regions, area typically not covered by CES and/or WES, these SVs had previously gone undetected. In contrast, WGS, with its uniform coverage of both exonic and intronic regions and the availability of specialized bioinformatics tools, significantly improves the detection of variants that escape conventional genetic approaches. Furthermore, these genomic variants would likely be missed by tools that rely exclusively on read depth or by quantitative methods such as MLPA. Although the adoption of WGS as a first-line diagnostic approach still requires the implementation of appropriate technological and bioinformatics infrastructure in healthcare systems, its improved diagnostic yield and steadily decreasing cost support its use in selected patient populations.

Based on our findings, we propose the use of WGS to diagnose both familial and sporadic EOPD cases at large and, in particular, in those individuals that present clinical features suggestive of *PRKN*-PD and have

tested negative or inconclusive by conventional methods, including targeted gene panels, CES/WES, and MLPA. Rather than limiting this analysis to individuals carrying monoallelic mutation in autosomal recessive genes (eg, *PRKN*, *PINK1*, etc.), WGS should be considered for a broader group of unresolved PD cases.

The choice between sr-WGS versus lr-WGS as second-line diagnostic test largely depends on the available computational infrastructure, laboratory equipment, and bioinformatic expertise within hospitals and specialized centers. sr-WGS can efficiently detect these elusive compound heterozygous SVs and is less technically demanding than lr-WGS. However, it may be less accurate for identifying large or complex rearrangements.^{31,32} Therefore, a sequential strategy, starting with sr-WGS and reserving lr-WGS for cases that remain unresolved, may optimize diagnostic yield. For familial or sporadic EOPD cases that remain undiagnosed after MLPA/CES/WES and sr-WGS, lr-WGS may provide a definitive diagnosis. By integrating both sequencing approaches, the likelihood of uncovering the genetic cause in undiagnosed patients is significantly enhanced. Identifying a definitive genetic diagnosis can also inform clinical decision-making and potential therapeutic strategies. As WGS costs decline, its application is expected to expand, particularly among patients with strong indications of a monogenic etiology. While WGS, both sr-WGS and lr-WGS, demonstrates improved diagnostic yield in PD, a formal cost-effectiveness analysis compared with WES or traditional testing is warranted prior to its adoption as a first-line diagnostic tool.

For the remaining unsolved EOPD cases in our cohort, we are currently performing an in-depth analysis of WGS data by exploring repeat expansions and non-coding regions to identify and characterize variants in key regulatory elements, such as enhancers and promoters, and to assess their potential impact on three-dimensional genome organization. These investigations may require the integration of WGS with complementary datasets, including chromatin interactions maps (eg, Hi-C) and epigenetic profiles (eg, histone modifications). Such integrated multi-omics approaches offer a comprehensive framework for unraveling complex disease mechanisms, identifying potential biomarkers for diagnosis and prognosis, and advancing precision medicine in PD. Finally, given emerging evidence suggesting a potential role of somatic mutations in neurodegenerative diseases,³³⁻³⁶ comparative genomic analysis between post-mortem neuronal tissue and peripheral blood may yield novel insights into the molecular basis of unresolved cases and uncover pathogenetic mechanisms relevant for the development of new therapeutic strategies.

In conclusion, as genomic technologies become increasingly accessible, our findings strongly support

the clinical implementation of sr-WGS to uncover previously undetectable SVs. This approach promises to improve diagnostic accuracy and advance the application of precision medicine in PD. ■

Author Roles: (1) Research Project: A. Conception, B. Organization, C. Execution; (2) Data Analysis: A. Design, B. Execution, C. Collection, D. Review and Critique; (3) Manuscript Preparation: A. Writing of the First Draft, B. Review and Critique; (4) Clinical Activity: A. Supervision and Organization, B. Data and Sample Collection.

A.F.: 1B, 1C, 2B, 2C, 2D, 3A, 3B.

S.T.: 1B, 1C, 2B, 2C, 2D, 3A, 3B.

E.M.: 1B, 1C, 2C, 2D, 3A, 4B.

G.T.: 1C, 2B, 2C, 3A, 3B.

F.M.: 2A, 2C, 3B.

F.L.: 2C, 3B.

P.M.: 2C, 2D, 3A, 3B, 4A, 4B.

A.A.: 3A, 3B.

R.S.: 2D, 3B.

L.P.: 2A, 2D, 3B.

F.C.: 3B, 4B.

F.V.: 3B, 4B.

V.F.: 3B, 4B.

G.D.R.: 3B, 4B.

G.B.: 3B, 4B.

E.M.V.: 3B, 4B.

V.T.: 3B, 4B.

L.M.R.: 3B, 4B.

C.R.: 3B, 4B.

B.G.: 3B, 4B.

A.E.E.: 3B, 4B.

A.C.: 2D, 3B.

A.D.F.: 1A, 1B, 2D, 3A, 3B, 4A.

M.V.: 1A, 1B, 1C, 2A, 2C, 2D, 3A, 3B.

S.G.: 1A, 1B, 2A, 2D, 3A, 3B.

Acknowledgments: We are grateful to the subjects and families who participated in this study and donated their samples for research under the NeurOmics project. We thank also “Cell Line and DNA Bank of Genetic Movement Disorders and Mitochondrial Diseases” of the Telethon Network of Genetic Biobanks (GTB18001) who supplied biological specimens of the Fondazione Carlo Besta. We are indebted to all the members of the CMP³VdA laboratory and of the RNA Flagship@IIT for thought-provoking discussions. We are grateful to the technical IT support, technicians, the project manager Alessandra Gastaldi, Arianna De Arcangelis for regulatory documents, and all the administrative staff of CMP³VdA, Aosta. Special thanks to Emilie Jacquemod and Eva Ferri (IIT). Open access publishing facilitated by Istituto Italiano di Tecnologia, as part of the Wiley - CRUI-CARE agreement.

Full Financial Disclosure for the Previous 12 Months: A.D.F. received honoraria from Sanofi for expert meeting participation. None of the other authors report any conflicts of interest.

Data Availability Statement

Anonymized patient data will be made available on reasonable request. Data used in this study are available from the PPMI database (<https://www.ppmi-info.org/access-data-specimens/download-data>).

References

1. Lin MK, Farrer MJ. Genetics and genomics of Parkinson's disease. *Genome Med* 2014;6(6):1–16.
2. Jankovic J, Tan EK. Parkinson's disease: etiopathogenesis and treatment. *J Neurol Neurosurg Psychiatry* 2020;91:795–808.
3. Polymeropoulos MH, Lavedan C, Leroy E, et al. Mutation in the α -synuclein gene identified in families with Parkinson's disease. *Science* 1997;276(5321):2045–2047.
4. Arienti F, Casazza G, Franco G, et al. Family history in Parkinson's disease: a national cross-sectional study. *Mov Disord Clin Pract* 2024;11(11):1434–1440.

5. Aasly JO. Long-term outcomes of genetic Parkinson's disease. *J Mov Disord* 2020;13(2):81–96.
6. Kim JJ, Vitale D, Otani DV, et al. Multi-ancestry genome-wide association meta-analysis of Parkinson's disease. *Nat Genet* 2024;56(1):27–36.
7. Trevisan L, Gaudio A, Monfrini E, Avanzino L, Di Fonzo A, Mandich P. Genetics in Parkinson's disease, state-of-the-art and future perspectives. *Br Med Bull* 2024;149(1):60–71.
8. Menon PJ, Sambin S, Criniere-Boizet B, et al. Genotype–phenotype correlation in PRKN-associated Parkinson's disease. *NPJ Parkinsons Dis* 2024;10(1):72.
9. Denison SR, Callahan G, Becker NA, Phillips LA, Smith DI. Characterization of FRA6E and its potential role in autosomal recessive juvenile parkinsonism and ovarian cancer. *Genes Chromosomes Cancer* 2003;38(1):40–52.
10. Kitada T, Asakawa S, Hattori N, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature* 1998;392(6676):605–608.
11. Borsche M, König IR, Delcambre S, et al. Mitochondrial damage-associated inflammation highlights biomarkers in PRKN/PINK1 parkinsonism. *Brain* 2020;143(10):3041–3051.
12. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov Disord* 2015;30(12):1591–1601.
13. 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.
14. Layer RM, Chiang C, Quinlan AR, Hall IM. LUMPY: a probabilistic framework for structural variant discovery. *Genome Biol* 2014;15(6):1–19.
15. Abyzov A, Urban AE, Snyder M, Gerstein M. CNVnator: an approach to discover, genotype, and characterize typical and atypical CNVs from family and population genome sequencing. *Genome Res* 2011;21(6):974–984.
16. Chen K, Wallis JW, McLellan MD, et al. BreakDancer: an algorithm for high-resolution mapping of genomic structural variation. *Nat Methods* 2009;6(9):677–681.
17. Jeffares DC, Jolly C, Hoti M, et al. Transient structural variations have strong effects on quantitative traits and reproductive isolation in fission yeast. *Nat Commun* 2017;8:1–11.
18. Nicholas TJ, Cormier MJ, Quinlan AR. Annotation of structural variants with reported allele frequencies and related metrics from multiple datasets using SVAFotate. *BMC Bioinformatics* 2022;23(1):1–17.
19. Geoffroy V, Herenger Y, Kress A, et al. AnnotSV: an integrated tool for structural variations annotation. *Bioinformatics* 2018;34(20):3572–3574.
20. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* 2012;13(1):134.
21. Pillay NS, Ross OA, Christoffels A, Bardien S. Current status of next-generation sequencing approaches for candidate gene discovery in familial Parkinson's disease. *Front Genet* 2022;13:781816.
22. Di Fonzo A, Percetti M, Monfrini E, et al. Harmonizing genetic testing for Parkinson's disease: results of the PARKNET multicentric study. *Mov Disord* 2023;38(12):2241–2248.
23. Lincoln SE, Hambuch T, Zook JM, et al. One in seven pathogenic variants can be challenging to detect by NGS: an analysis of 450,000 patients with implications for clinical sensitivity and genetic test implementation. *Genet Med* 2021;23(9):1673–1680.
24. Liu Z, Roberts R, Mercer TR, Xu J, Sedlazeck FJ, Tong W. Towards accurate and reliable resolution of structural variants for clinical diagnosis. *Genome Biol* 2022;23(1):1–25.
25. Wasner K, Grünewald A, Klein C. Parkin-linked Parkinson's disease: from clinical insights to pathogenic mechanisms and novel therapeutic approaches. *Neurosci Res* 2020;159:34–39.
26. Ruiz-Lopez M, Freitas ME, Oliveira LM, et al. Diagnostic delay in Parkinson's disease caused by PRKN mutations. *Parkinsonism Relat Disord* 2019;63:217–220.
27. Towns C, Fang ZH, Tan MMX, et al. Parkinson's families project: a UK-wide study of early onset and familial Parkinson's disease. *NPJ Parkinsons Dis* 2024;10(1):1–13.
28. Cogan G, Daida K, Billingsley KJ, et al. Long-read sequencing unravels the complexity of structural variants in PRKN in two individuals with early-onset Parkinson's disease. *Mov Disord* 2024;39(9):1647–1648.
29. Daida K, Cogan G, Tesson C, et al. Screening of hidden pathogenic structural variants in PRKN. *Mov Disord* 2025;23:3–5.
30. Daida K, Yoshino H, Malik L, et al. The utility of long-read sequencing in diagnosing early onset Parkinson's disease. *Ann Neurol* 2025;97(4):753–765.
31. Zhao X, Collins RL, Lee WP, et al. Expectations and blind spots for structural variation detection from long-read assemblies and short-read genome sequencing technologies. *Am J Hum Genet* 2021;108(5):919–928.
32. Guan P, Sung WK. Structural variation detection using next-generation sequencing data: a comparative technical review. *Methods* 2016;102:36–49.
33. Proukakis C, Shoaee M, Morris J, et al. Analysis of Parkinson's disease brain-derived DNA for alpha-synuclein coding somatic mutations. *Mov Disord* 2014;29(8):1060–1064.
34. Proukakis C. Somatic mutations in neurodegeneration: an update. *Neurobiol Dis* 2020;144:105021.
35. Lobon I, Solís-Moruno M, Juan D, et al. Somatic mutations detected in Parkinson disease could affect genes with a role in synaptic and neuronal processes. *Front Aging* 2022;3:851039.
36. Iourov IY, Vorsanova SG, Kurinnaia OS, Kutsev SI, Yurov YB. Somatic mosaicism in the diseased brain. *Mol Cytogenet* 2022;15(1):1–11.

Supporting Data

Additional Supporting Information may be found in the online version of this article at the publisher's web-site.