

Genetic Screening of Colombian Patients With Early-Onset Parkinson Disease

Tobias M. Franz,^{1,2,3,*} Marios Gavrielatos,^{1,*} Ysabella Wijaya,¹ Alexandra I. Beasley,¹ Sarah Kim-Hellmuth,^{3,4} Audrey J. Strongosky,⁵ Claudia L. Moreno-López,^{6,7} Owen A. Ross,^{1,8} Zbigniew K. Wszolek,⁵ and Catalina Cerquera-Cleves⁹

Correspondence
Dr. Ross
ross.owen@mayo.edu

Neurol Genet 2026;12:e200427. doi:10.1212/NXG.0000000000200427

Abstract

Background and Objectives

Early-onset Parkinson disease (EOPD), defined as symptom onset before 50 years of age, accounts for approximately 10% of patients and is suggested to have a greater genetic component than typical late-onset forms of the disease. Recessive variants in *PRKN*, *PINK1*, and *DJ-1* are the most common genetic cause of EOPD, however, most studies are in patients of white ancestry. This study aims to analyze genetic variants in *PRKN*, *PINK1*, and *DJ-1* in Colombian patients to help address the gap in EOPD genetic research of South American populations.

Methods

We analyzed 43 unrelated patients with EOPD using Sanger sequencing for the *PRKN*, *PINK1*, and *DJ-1* genes and employed multiplex ligation-dependent probe amplification to detect copy number variants. Additionally, long-read whole-genome sequencing was conducted on 3 unresolved patients with age at onset before 30 years of age (long-read sequencing [LRS] patient A–C).

Results

We identified known pathogenic single-nucleotide variants and copy number variants in the *PRKN* gene accounting for 2 patients' disease (4.6% of patients). We observed 2 pathogenic variants in *PRKN* (c.155delA; p.N52Mfs*29 and c.1083+1G>A) in patient 1, who reported an age at onset of 16 years. We further detected a homozygous duplication of *PRKN* exons 5–6 in an additional patient, age at onset of 18 years.

Discussion

Our study helps characterize genetic contributors to EOPD in Colombian patients, demonstrating genetic forms (*PRKN*, *PINK1*, and *DJ-1*) are rare. Our results highlight a need to include diverse populations in research to improve genetic understanding of disease.

Introduction

Neurologic conditions are the leading source of disability worldwide, and the prevalence of Parkinson disease (PD) is increasing more rapidly than others.¹ PD can be subclassified as late-onset PD (LOPD) and early-onset PD (EOPD; ~10% of all patients with PD) forms. Patients with EOPD (symptom onset before 50 years) more frequently report a positive family history (LOPD: 7%, EOPD: 20%) and thus are more likely to have a genetic form of disease.² Studying the genetic architecture of EOPD and subsequently identifying biomarkers can provide

*These authors are co-first authors.

¹Department of Neuroscience, Mayo Clinic, Jacksonville, FL; ²Department of Dermatology and Allergy Biederstein, School of Medicine and Health, Technical University of Munich, Germany; ³Department of Pediatrics, Dr. von Hauner Children's Hospital, University Hospital LMU Munich, Germany; ⁴Computational Health Center, Institute of Translational Genomics, Helmholtz Munich, Neuherberg, Germany; ⁵Department of Neurology, Mayo Clinic, Jacksonville, FL; ⁶Movement Disorders Clinic, Hospital Universitario San Ignacio, Bogotá, Colombia; ⁷Fundación Cardioinfantil, Bogotá, Colombia; ⁸Department of Clinical Genomics, Mayo Clinic, Jacksonville, FL; and ⁹Department of Neurosciences, Neurology Unit, Hospital Universitario San Ignacio, Bogotá, Colombia.

The Article Processing Charge was funded by the authors.

This is an open access article distributed under the terms of the Creative Commons Attribution-Non Commercial-No Derivatives License 4.0 (CCBY-NC-ND), where it is permissible to download and share the work provided it is properly cited. The work cannot be changed in any way or used commercially without permission from the journal.

Copyright © 2026 The Author(s). Published by Wolters Kluwer Health, LLC. on behalf of the American Academy of Neurology.

e200427(1)

MORE ONLINE

Supplementary Material

Glossary

AMP = Association for Molecular Pathology; **CADD** = Combined Annotation Dependent Depletion; **CNV** = copy-number variant; **EOPD** = early-onset Parkinson's disease; **LOPD** = late-onset Parkinson's disease; **LRS** = long-read sequencing; **PD** = Parkinson's disease; **SNV** = single-nucleotide variant; **SV** = structural variant; **TR** = tandem repeat; **VUS** = variants of uncertain significance.

important insights into its early pathophysiology and is crucial as it can enable early diagnosis and therapeutic intervention during the preclinical stage.³

Genes involved in EOPD include dominant (*SNCA* triplication) and recessive (*PRKN*, *PINK1*, and *DJ-1*) mechanisms of inheritance, highlighting pathways such as lysosomal and mitochondrial dysfunction.^{4,5} Variants of these genes can range from simple point variants, single-exon dosage variants, copy number variants, and genomic multiplications.⁶ Most genetic studies in EOPD have focused on patients of European ancestry, while the genetic composition of EOPD in populations that remain underrepresented in clinical/genetic research, including those from South America, remains poorly characterized. Some studies have reported that *PRKN* variants are almost 2.7-fold more frequent in Hispanic EOPD populations compared with non-Hispanic White population.^{7,8} In a study encompassing 71 Ecuadorian and 26 Colombian EOPD cases, MLPA detected a duplication of *PRKN* exons 5 and 6, and a deletion in exon 7 in the series.⁹

Understudied populations should be investigated because gene variants may have a population-specific prevalence. Investigation of these populations will shed light on the pathophysiology of PD and contribute to a better understanding of the disease's clinical heterogeneity. To address the lack of genetic data on PD genetics in Latin America, we sought to further characterize *PRKN*, *PINK1*, and *DJ-1* variants in 43 unrelated patients with EOPD from Colombia. In addition, we conducted long-read sequencing whole-genome on 3 patients with age at onset younger than 30 years and unresolved from the initial screen, aiming to enhance our understanding of the genetic origins of the disease in the Colombian population.

Methods

Standard Protocol Approvals, Registrations, and Patient Consents

All participants provided written informed consent, with approval from the local IRBe and Mayo Clinic IRBe (1087-98).

Study Design

Whole blood was obtained from 200 consecutive patients with PD recruited from a movement disorder clinic in Bogotá, Colombia. All patients resided in the urban and suburban areas of Bogotá, primarily of mestizo descent. Of these 200 patients, 43 unrelated patients fulfilled the criteria for EOPD

(<50 years of age at symptom onset) and were included in this study. The remaining patients all presented with LOPD and were not included in the analysis.

Genetic Analysis

DNA extraction was performed using the QIAamp DNA Blood Midi Kit (Qiagen, Hilden, Germany) following standard protocols. Bidirectional Sanger Sequencing was performed on all coding exons plus 15 base pairs (bp) flanking each exon of *PRKN* (NM_004562.3), *PINK1* (NM_032409.3), and *DJ-1* (NM_007262.5) using custom primers. Polymerase chain reaction (PCR) products were purified using Agencourt Ampure XP and CleanSeq purification systems performed on the Beckman Coulter Biomek Fxp Automated workstation (Beckman Coulter Life Sciences, Indianapolis, IN). Sequence analysis was performed using Seqscape Software v3.0 (Thermo Fisher Scientific, Waltham, MA). Using the aggregated ClinVar pathogenicity scores¹⁰ and ACMG/AMP criteria, we categorized genetic variants within our cohort as pathogenic, likely pathogenic, or variants of uncertain significance (VUS). Furthermore, the Combined Annotation Dependent Depletion (CADD) scores¹¹ were used to evaluate the deleteriousness of single-nucleotide variants (SNV) in the human genome. MLPA version P051 Parkinson mix 1 was used to screen for copy number variants (including *SNCA* genomic multiplication) and analyzed using Coffalyser.Net™ (MCR Holland, Netherlands). Microsatellite markers were chosen to span the *PRKN* x5-6 locus between D6S386 and D6S255. One primer of each pair was labeled with a fluorescent tag and PCR reactions were performed under standard conditions. Products were run on an ABI 3730XL genetic analyzer, and the results were analyzed using GeneMapper 5.0 (Thermo Fisher Scientific). Marker sizes were normalized to those reported in the CEPH database.

Long-Read Sequencing Analysis

Data Generation

High molecular weight genomic DNA was extracted from whole blood using the Nanobind Cells, Culture and Blood Big DNA kit (Pacific Biosciences, California). DNA quantification was conducted with the Invitrogen Qubit fluorometry assay and NanoDrop 8000 spectrophotometer. The DNA was sheared using a MegaRuptor 3 system, and SMRTbell libraries were prepared by ligating hairpin adapters to both ends of the DNA fragments, creating a circular template for the polymerase to repeatedly incorporate nucleotides across the entire long fragment.

For sequencing, we used one SMRT cell per sample. Each SMRTbell library was sequenced on a PacBio Revio system using Sequel II 2.0 chemistry (v11.0). Unaligned BAM (uBAM) files are the native format that stores read information. These uBAMs were processed following the long-read analysis pipeline described in eMethods, resulting in database-annotated VCF files containing SNV, structural variants (SV), copy number variants (CNV), and tandem repeats (TR). Unlike conventional sequencing methods, long reads offer superior phasing capabilities by capturing multiple variant sites on single DNA molecules, directly revealing whether duplications and triplications occur on the same or different chromosomes. This technology enables complete haplotype reconstruction, definitively showing which alleles are inherited together on each parental chromosome. Furthermore, using long reads precisely identifies breakpoint junctions at single-nucleotide resolution. Finally, long reads can uncover complex SVs potentially missed by standard genetic testing.

CNV Validation and Haplotype Genotyping

We used a TaqMan Copy Number Assays to validate *PRKN* copy number variation using the QuantStudio 7 Flex Real-Time PCR system and QuantStudio Real-Time PCR Software (Applied Biosystems, Carlsbad, CA) according to manufacturer's protocol. CNV calls were accepted when call rates exceeded 98%. In the samples used for long-read sequencing, we performed an analysis of homozygosity in the intronic regions flanking the CNV to distinguish between a homozygous duplication and a heterozygous triplication. Specifically, we analyzed SNV calls within the region extending 300 bp before the start of the closest exon to 300 bp before the copy number event in both flanking regions of the CNV. In the case of a CNV-positive sample, the homozygosity analysis was also performed for the whole chromosome in which the variant was found.¹²

Data Availability

The scripts used to run the different tools are available in GitHub.¹³ The data that support the findings of this study are available from the corresponding author on reasonable request.

Results

The study cohort, consisted of 23 male and 20 female patients with EOPD (where sex was self-reported as sex assigned at birth), a mean age at onset of 41 years (range 16–49) and an average age of 47.2 ± 6.3 years at data collection. We performed targeted gene analysis (sequencing and exon dosage) and subsequently performed long-read whole genome sequencing on 3 unresolved patients (LRS patient A-C) with an age at onset younger than 30 years.

In our cohort of 43 patients, 8 variants were identified in *PRKN*, 7 in *PINK1*, and one in *DJ-1*, as detailed in Table 1. No *SNCA* genomic multiplications were detected. Notably,

we identified compound heterozygous pathogenic *PRKN* variants in patient 1 (age at onset 16 years), a splice site SNV (c.1083+1G>A; exon9+1) likely disrupting normal splicing, and a heterozygous single-base deletion in exon 2 (c.155delA; p.N52Mfs*29) causing a frameshift and premature truncation of the protein. Using MLPA, we detected *PRKN* CNVs in 2 patients, both carriers of either a possible homozygous duplication or heterozygous triplication of exons 5 and 6. These carriers had an age at onset of 18 years (LRS patient A) and 35 years (patient 2), respectively. Haplotype analysis (eTable 1) revealed that LRS patient A shares a haplotype with a previously reported Colombian patient.⁹ All reported cases present motor symptoms related to EOPD; however, LRS patient A in this study has also been diagnosed with bipolar disorder and reported nonmotor symptoms like anxiety and depression. Patient 2 appears to carry a distinct haplotype, based on the unique allele pattern at marker D6S305 (204/204; eTable 1). The *PRKN* CNV was verified using a TaqMan Copy Number Assay with 4 replicates run for each sample (eTable 2). All SNV listed in Table 1 were assessed for pathogenicity by applying American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP) criteria.

Three patients with the earliest age at onset (18, 23, and 27 years) who were negative/unresolved in our initial screen were selected for long-read sequencing analysis. These particular patients were prioritized to identify potential complex genetic factors that might explain such unusually early disease manifestation because most PD cases occur in later decades of life. This type of sequencing was necessary to precisely characterize the complex SVs observed in these patients, particularly to determine the homozygous/heterozygous orientation of the *PRKN* duplication/triplication (LRS patient A) that cannot be fully resolved using conventional sequencing methods.

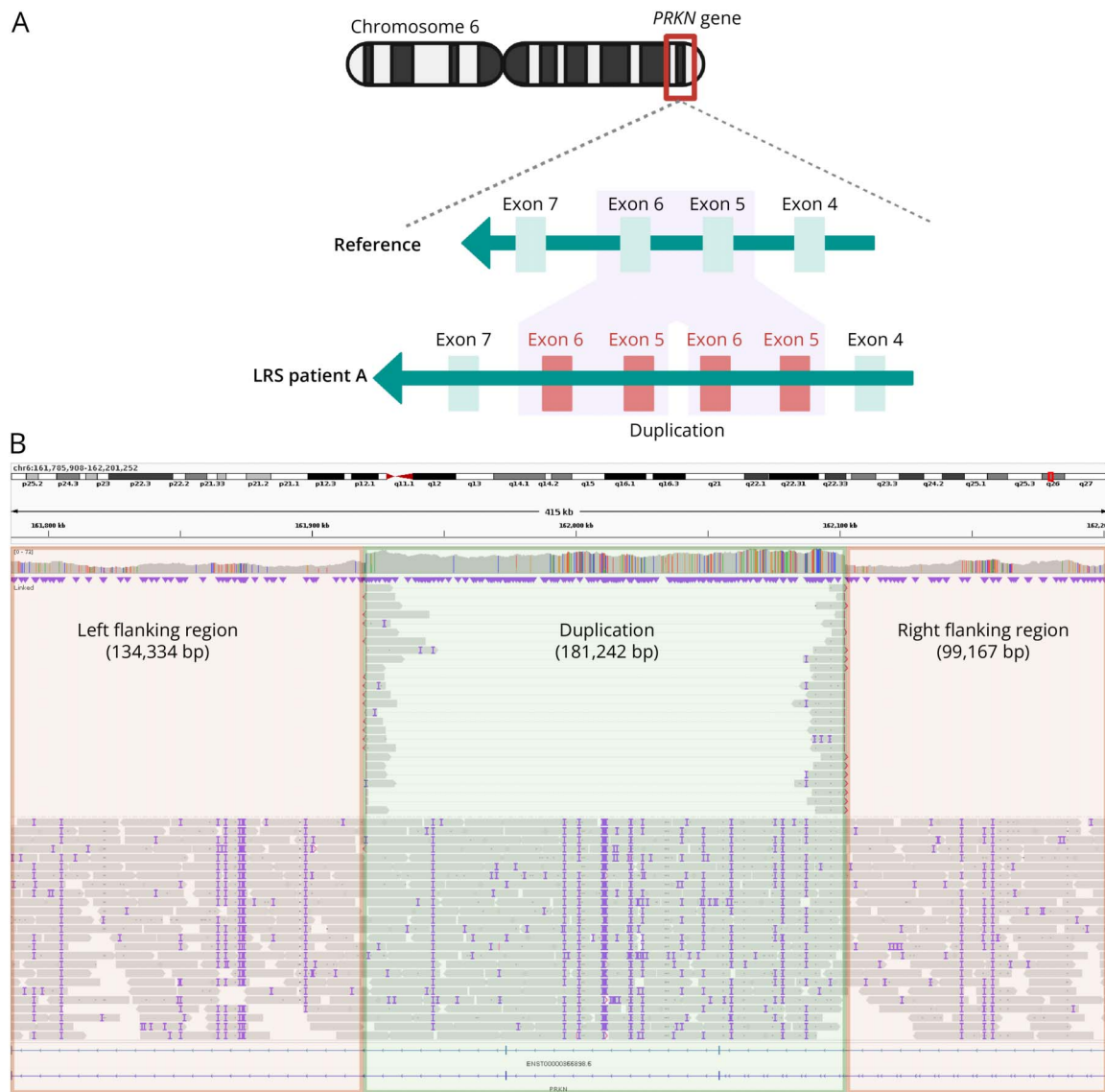
Technically, the analysis yielded robust data across various metrics. The HiFi read count varied from 3,989,793 to 5,439,449 per sample, with a mean of 4,574,084 reads and a median of 4,433,547. The total HiFi yield ranged from approximately 66.76 billion to 95.07 billion base pairs, with an average of 82.34 billion base pairs and a median of 83.77 billion base pairs. The HiFi read length N50, reflecting read length distribution, varied between 16,064 and 18,182 base pairs, with a mean N50 of 17,277 base pairs and a median of 17,432 base pairs. The median HiFi read quality was consistent across samples (Q31). Coverage across samples and chromosomes was generally uniform, with a mean coverage of 25.40x across the entire genome. Notable variations in coverage were observed, with higher mean coverage on chromosomes such as chromosome 4 (29.11x) and lower coverage on chr22 (17.55x). Autosomal chromosomes showed a mean coverage of 25.87x and a median of 27.25x, highlighting a relatively even distribution of sequencing depth across the autosomal genome. See eTables 3 and 4 for more details.

Table 1 *PRKN*, *PINK1*, and *DJ-1* Single-Nucleotide Variants Identified in Exons ±15 bp, Within the Colombian EOPD Cohort

Exon	dbSNP rsID	A.A pos	cDNA change	Major Hom	Het	Minor Hom	Total N = 43 patients	MAF (%) N = 86 alleles	MAF global (%)	MAF Admixed Americans (%)	CADD score	ACMG classification for SNVs
PRKN SNV (NM_004562.3)												
2	rs77795533	p.P37P	c.111G>A	42	1	0	43	1.2	0.6	1.0	4.5	Benign (BP7, BS2, BP6, BA1)
2	rs754809877	p.N52Mfs*29	c.155delA	42	1	0	43	1.2	0.03	0.1	29.2	Pathogenic (PVS1, PM3, PM2, PP5)
4	rs1801474	p.S167N	c.500G>A	33	8	2	43	13.9	3.8	12.7	14.4	Benign (BS2, BP6, BA1)
5	rs9456735	p.M192L	c.574A>C	42	1	0	43	1.2	0.3	0.4	20.7	Benign (BS2, BP6, BA1)
9	rs111546299	NA	c.1083+1G>A	42	1	0	43	1.2	0.0	0.0	32.0	Likely pathogenic (PVS1, PM2)
10	rs1801582	p.V380L	c.1139G>C	33	10	0	43	11.6	16.8	13.1	4.5	Benign (BS2, BP6, BA1)
11	rs1801334	p.D394N	c.1180G>A	41	2	0	43	2.4	3.4	1.3	22.9	Benign (BS2, BP6, BA1)
11	N/A	p.A398P	c.1192G>C	42	1	0	43	1.2	0.0	0.0	24.0	VUS (PM2)
PRKN CNV (NM_004562.3)												
5 and 6	Homozygous duplication or heterozygous triplication			41	2		43					
PINK1 SNV (NM_032409.3)												
1	rs45530340	p.L63L	c.189C>T	35	8	0	43	9.3	18.7	14.9	12.4	Benign (BP7, BS2, BP6, BA1)
1	rs148871409	p.Q115L	c.344A>T	42	1	0	43	1.2	4.6	2.4	22.0	Benign (BP4, BS2, BP6, BA1)
2	rs2298298	NA	c.388-7A>G	1	13	27	41	81.7	86.9	80.7	3.5	Benign (BP4, BS2, BP6, BA1)
5	rs3131713	NA	c.960-5G>A	29	11	3	43	19.8	86.9	80.8	0.3	Benign (BP4, BS2, BP6, BA1)
5	rs3738136	p.A340T	c.1018G>A	37	5	1	43	8.1	6.4	8.7	11.8	Benign (BP4, BS2, BP6, BA1)
7	rs115477764	p.E476K	c.1426G>A	42	1	0	43	1.2	0.2	0.3	4.9	Benign (BP4, BS2, BP6, BA1)
8	rs1043424	p.N521	c.1562A>C	26	17	0	43	19.8	28.5	23.1	17.3	Benign (BP4, BS2, BP6, BA1)
DJ-1 SNV (NM_007262.5)												
3	rs11548937	p.G78G	c.234C>T	42	1	0	43	1.2	0.63	0.64	5.0	Benign (BP4, BS2, BP6, BA1)

Abbreviations: A.A. pos = amino acid position; CADD score = Combined Annotation Dependent Depletion score; Het = number of carriers of heterozygous alleles; MAF = minor allele frequency in the Colombian patients with EOPD (n = 86 alleles); MAF admixed Americans = minor allele frequency annotated from the Latino/Hispanic population in gnomad; MAF global = minor allele frequency annotated from the global population in gnomad; Major Hom = number of carriers homozygous for major allele; Minor Hom = number of carriers homozygous for minor allele; NA = not applicable; SNV = single-nucleotide variant. A *PRKN* homozygous duplication or heterozygous triplication of exons 5 and 6 was identified in 2 patients.

Figure PRKN Exon 5 and 6 Duplication in LRS Patient A



(A) Visual representation of the duplication (Created in BioRender. Tamvaka, N. (2026) BioRender.com/t574o3i). (B) The IGV visualization displays the duplicated region (green) alongside the read alignments and flanking regions (red) used for genotyping analysis. The supplementary alignments at both edges of the duplication indicate that the 2 exons are duplicated together and not individually. IGV = integrative genomics viewer.

Our long-read sequencing analysis confirmed the pathogenic *PRKN* CNV called through MLPA in LRS patient A (Figure, A). The alignment data demonstrate that both exons are copied as a continuous segment (Figure, B). The analyses indicate a homozygous duplication or heterozygous triplication of exons 5 and 6, with a value consistent with other positive patients. Figure, B visualizes read alignments in the duplicated and flanking regions via integrative genomics viewer, where SNV genotyping was analyzed. The homozygosity analysis, summarized in Table 2, revealed ~95% SNV homozygosity in the flanking intronic regions of the CNV in LRS patient A, while the rest of chromosome 6 showed a significantly lower rate (57.82%). That, along with the lower percentage of homozygosity in the same region on the other participants of our long-read cohort, as

seen in the same table, indicates a homozygous tandem duplication.

Table 3 details the homozygous and compound heterozygous variants identified in LRS patients A-C. For all participants, we examined genes that have been related to early-onset forms of parkinsonism (eTable 5) to identify potential complex pathogenic variants.³ Apart from LRS patient A, in which we detected the *PRKN* exon 5–6 duplication, we did not observe pathogenic SNV or SV through our candidate gene approach. Accordingly, we carried out an unbiased analysis to detect potentially pathogenic variants. We filtered for rare variants using a minor allele cutoff (gnomAD total) <1%, (gnomAD Latino/Admixed American) <1%, CADD-score >20 and applying ACMG/AMP criteria, retaining only homozygous or

Table 2 Genotyping Analysis for the *PRKN* CNV

Sample name	Chromosomal region	Homozygosity (%)	Heterozygosity (%)
Control	chr6:161785908-161920242 and chr6:162102085-162201252	51.37	48.22
LRS patient A	chr6:161785908-161920242 and chr6:162102085-162201252	97.51	02.49
	Entire Chromosome 6	57.82	42.18
LRS patient B	chr6:161785908-161920242 and chr6:162102085-162201252	40.26	59.74
LRS patient C	chr6:161785908-161920242 and chr6:162102085-162201252	36.21	63.79

Abbreviation: LRS = long-read sequencing.
The CNV-positive patient (LRS patient A) exhibits high levels of homozygosity in the region compared with the full length of chromosome 6, as well as to other patients in the same, flanking, intronic regions of the CNV. The results designate a homozygous duplication.

compound heterozygous as candidates. For LRS patient B and LRS patient C, 3 and 4 genes, respectively, were identified as carrying biologically relevant candidate variants (Table 3).

Discussion

In this study, we examined the 3 key genes associated with EOPD (*PRKN*, *PINK1*, and *DJ-1*), in a Colombian cohort of 43 patients. We identified a compound heterozygous *PRKN* variant in patient 1 (C>T, Exon9+1A>; p.N52STOP79) with an age at onset of 16 years in our initial screen. In addition, we discovered a homozygous duplication of exons 5 and 6. We performed long-read whole-genome analysis (LRS patient A) and were able to validate and visualize the homozygous duplication, showcasing the capabilities of long-read sequencing in variant detection and characterization (Figure). Noting, a homozygous duplication of exons 5 and 6 CNV was previously reported in a cohort of 26 patients from Colombian and haplotype analysis confirmed a shared ancestry for LRS patient A.⁹

Our findings underscore the critical role of *PRKN* protein (encoded by *PRKN*) structure in its function, where protein-truncating, structural, and frameshift variants, as well as cumulative exon dysfunction, correlate with an early age at onset for PD in Colombia. The implications of homozygous and compound heterozygous variants, including CNVs in *PRKN*, are well-documented, particularly in EOPD.^{14,15} There is a growing body of evidence suggesting that heterozygosity in *PRKN* and *PINK1* also contributes to PD risk, often associated with a reduced age at onset.¹⁶ A recent investigation into *PRKN* variants indicated that individuals possessing a single CNV in the *PRKN* gene tend to experience an earlier average age at onset compared with those with only an SNV.¹⁷ Understanding the functional/biological perturbations when for example only one copy of *PRKN* and *PINK1* is lost, will be important to resolve as targeted clinical intervention trials are designed.

Sanger sequencing remains highly relevant in clinical settings because of its high accuracy and low cost, thus a logical practical choice for genetic diagnostics in particular cases. Novel long-read sequencing approaches may be particularly

valuable in studying populations underrepresented in clinical research, where genomic architecture data are often limited. In the context of EOPD, where genetic factors play a significant role, long-read sequencing has the potential to be a powerful tool for uncovering the complex genetic architecture that may contribute to disease risk.

Through our exploratory long-read whole-genome sequencing in LRS patients A-C notable variants were observed including a rare homozygous splice site insertion in *ADCY2* (c.2094+5_2094+6insCCCCAGAGCATGGCCGTGTTCAACATGGTAAG), which was present in LRS patient A and LRS patient B. *ADCY2* protein is involved in key signaling pathways that may increase PD risk¹⁸ and is associated with an increased susceptibility of patients with PD to levodopa-induced dyskinesia.¹⁹ Furthermore, in LRS patient A we found a homozygous frameshift truncation in exon 3 of *ADH1B*, potentially posing a higher PD risk.²⁰ The most notable variant in the data for LRS patient C is a missense SNV in *HCFC1* (c.3874C>A; p.P1292T), a gene previously associated with PD risk.²¹ *HCFC1* is also linked to developmental delays and neurologic disorders, with partial loss of function potentially disrupting brain development and contributing to X-linked disability.^{22,23}

This study design has limitations that may affect the interpretation and generalizability of the findings. The presence of a control group within the Colombian population would help strengthen evidence of potential pathogenicity and the lack of a double-blind design may introduce biases because of the reliance on physician objectivity during diagnosis. This study primarily focused on identifying monogenic disease-causing variants in recessive genes (*PRKN*, *PINK1*, and *DJ-1*), and a comprehensive genome large-scale sequencing approach would support both screening and discovery. As the sample sizes of Latin American studies in PD grow, future research will increasingly be able to examine the polygenic nature and burden of EOPD, including critical gene-environment (GxE) and gene-gene (GxG) interactions. These studies are important for a full understanding of the complex genetics underlying disease onset and

Table 3 List of Filtered Variants From Whole-Genome Long-Read Variant Discovery Analysis

Gene Name	Position	dbSNP rsID	Transcript: cDNA change	A.A. pos	Sequence ontology	Zygosity	MAF global (%)	MAF Admixed Americans (%)	Path.	PhenoSV	CADD-score	ACMG classification for SNVs
LRS patient A												
FZD5	chr2: 207767594-207767594	NA	ENST00000295417.4:c.1146C>A	p.D382E	Missense variant	Hom					22.4	VUS (PM2)
ADH1B	chr4: 99318064-99318064	rs768801751	ENST00000305046.13:c.241del	p.V81X	Stop gained	Hom	0	0.01			31	VUS (PM2)
ADCY2	chr5: 7757569-7757569	rs1560953526	ENST00000338316.9:c.2094 + 5_2094+6insCCCCAGAGCATGGCCGTGTTCAACATGGTAAG		Splice site variant	Hom	0	0			32	VUS (PM2)
PRKN	chr6: 161920455-162101786	NA	NA		Duplication	Hom			0.899	0.0001		
MALRD1	chr10: 19066730-19066730	NA	ENST00000454679.7:c.211G>T	p.E71X	Stop gained	Het					38	VUS (PM2)
MALRD1	chr10: 19352037-19352037	rs138930604	ENST00000454679.7:c.4181G>A	p.G1394D	Missense variant	Het	0.02	0.02			24.3	Likely Benign (PM2, BP4, BS2)
DMBT1	chr10: 122604478-122616984	NA	NA		Deletion	Het			0.7934	0.2766		
DMBT1	chr10: 122610109-122610109	NA	ENST00000338354.10:c.4312G>A	p.G1438S	Missense variant	Het					21.8	VUS (PM2)
DMBT1	chr10: 122614168-122614168	NA	ENST00000338354.10:c.4699G>A	p.G1567S	Missense variant	Het					23.3	VUS (PM2)
JAKMIP3	chr10: 132160000-132160001	NA	ENST00000684848.1:c.2221-3203_2221-3202del		Splice site variant	Hom						VUS (PM2)
OR4C6	chr11: 55597046-55666075	NA	NA		Inversion	Hom			0.6434	0.1279		
PWWP4	chrX: 153273425-153273425	rs868933258	NM_001395996.1:c.1808G>A	p.W603X	Stop gained	Hom					22.4	VUS (PM2)
WASH6P	chrX: 156022702-156022702	rs200340194	ENST00000359512.8:c.588C>G	p.Y196X	Stop gained	Hom	0.14	0.04			24.9	VUS (BP7)
LRS patient B												
ADCY2	chr5: 7757569-7757569	rs1560953526	ENST00000338316.9:c.2094 + 5_2094+6insCCCCAGAGCATGGCCGTGTTCAACATGGTAAG		Splice site variant	Hom					32	VUS (PM2)
PTPRD	chr9: 8127592-8803199	NA	NA		Duplication	Het			0.9563	0.0054		
DMBT1	chr10:122601724-122613901	NA	NA		Deletion	Hom			0.8719	0.304		

Continued

Table 3 List of Filtered Variants From Whole-Genome Long-Read Variant Discovery Analysis (continued)

Gene Name	Position	dbSNP rsID	Transcript: cDNA change	A.A. pos	Sequence ontology	Zygosity	MAF global (%)	MAF Admixed Americans (%)	Path.	PhenoSV	CADD-score	ACMG classification for SNVs
LRS patient C												
USP17L15	chr4: 9235977-9235977	rs181172003	ENST00000456464.2:c.1593G>A	p.W531X	Stop gained	Hom	0.08	0.27			26.7	VUS (PM2, BS2)
CNTNAP3B	chr9: 41898859-41898859	rs1554733872	ENST00000377561.7:c.3583G>C	p.G1195R	Missense variant	Het					22.1	VUS (PM2)
CNTNAP3B	chr9: 41960000-41960000	NA	ENST00000377561.7:c.1876+3del		Splice site variant	Het						VUS (PM2)
HCFC1	chrX: 153954525-153954525	NA	ENST00000310441.12:c.3874C>A	p.P1292T	Missense variant	Hom					23.9	VUS (PM2, PP2, BP6)
PWWP4	chrX: 153273425-153273425	rs868933258	NM_001395996.1:c.1808G>A	p.W603X	Stop gained	Hom	0	0			22.4	VUS (PM2)

Abbreviations: A.A. pos = amino acid position; CADD score = Combined Annotation Dependent Depletion score; Het = heterozygous allele count; Hom = homozygous allele count; LR = long-read; MAF admixed Americans = minor allele frequency annotated from the Latino/Hispanic population in gnomad; MAF global = minor Allele Frequency annotated from the global population in gnomad; NA = not applicable/available.; Path. = pathogenicity score given by PhenoSV; PhenoSV = pathogenicity score adjusted for the neurodegeneration ontology; SNV = single-nucleotide variant; SNV = single-nucleotide variant; VUS = variants of uncertain significance. Variants were retained if MAF (gnomAD total) < 1%, MAF (gnomAD Latino/Admixed American) < 1%, and CADD-score >20. Only homozygous or compound heterozygous variants were preserved. Phasing analysis was performed to determine the cis or trans configuration of compound heterozygous variants where possible. For *CNTNAP3B*, the phasing could not be resolved.

progression, and current efforts by the Global Parkinson's Genetics Program (GP2) are essential.

Future studies on other racial and ethnic populations will be imperative to assess the presence and relevance of the identified variants across populations. Finally, studies using long-read sequencing should be expanded to include more participants, allowing for a more comprehensive analysis of complex genetic variants and structural changes in EOPD within Latin American and across populations.

Acknowledgment

The authors would like to thank all those who have contributed to their research, particularly the patients and families who donated blood and DNA samples for this work. The authors would like to thank Ms. L. Camacho from Mayo Clinic International Business Office in Bogotá, Colombia. Mayo Clinic is an American Parkinson Disease Association (APDA) Information and Referral Center and APDA Center for Advanced Research.

Author Contributions

T.M. Franz: drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data. M. Gavrielatos: drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data. Y. Wijaya: analysis or interpretation of data. A.I. Beasley: drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data. S. Kim-Hellmuth: drafting/revision of the manuscript for content, including medical writing for content. A.J. Strongosky: drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data. C.L. Moreno-López: drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design. O.A. Ross: drafting/revision of the manuscript for content, including medical writing for content; study concept or design. Z.K. Wszolek: drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design. C. Cerquera-Cleves: drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; study concept or design.

Study Funding

This work was supported by the Mayo Clinic, an American Parkinson Disease Association (APDA) Information and Referral Center and APDA Center for Advanced Research. Support was also received from the Michael J. Fox Foundation. O.A. Ross is the Wayne and Kathryn Preisel Professor of Neuroscience Research. C. Cerquera-Cleves is funded by a Canadian Institutes of Health Research Vanier Scholarship.

Disclosure

S. Kim-Hellmuth was supported by the Emmy Noether Programme KI 2091/2-1 (459153572), SFB/TRR237-B29 (369799452) and SFB/TRR359-B06 (491676693) of the

Deutsche Forschungsgemeinschaft (DFG), by the Federal Ministry of Education and Research (BMBF) as part of the German Center for Child and Adolescent Health (DZKJ) under the funding code 01 GL2406A and a European Research Council (ERC) Starting Grant (101076303). O.A. Ross was supported in part by the Mayo Clinic LBD Center WithOut Walls (CWOW; U54-NS110435) and an American Parkinson Disease Association Center for Advanced Research; is supported by NIH (R01 AG087165; R01 AG089380; R01 AG056366, U19 AG071754), DOD (W81XWH-17-1-0249; HT9425-25-1-0288), American Brain Foundation, Alzheimer's Association, The Michael J. Fox Foundation, The Little Family Foundation, Ted Turner and family, and the Mayo Clinic Robert and Arlene Kogod Center on Aging and is an editorial board member of *American Journal of Neurodegenerative Disease* and *Frontiers of Neurology: Neurogenetics and Molecular Neurodegeneration*. Z.K. Wszolek was partially supported by the NIH/NIA and NIH/NINDS (1U19AG063911, FAIN: U19AG063911), Mayo Clinic Center for Regenerative Medicine, the gifts from the Donald G. and Jodi P. Heeringa Family, the Haworth Family Professorship in Neurodegenerative Diseases fund, The Albertson Parkinson's Research Foundation, PPND Family Foundation, and Margaret N. and John Wilchek Family; serves as PI or Co-PI on Biohaven Pharmaceuticals, Inc. (BHV4157-206), Vigil Neuroscience, Inc. (VGL101-01.002, VGL101-01.201, PET tracer development protocol, Csf1r biomarker and repository project, and ultra-high field MRI in the diagnosis and management of CSF1R-related adult-onset leukoencephalopathy with axonal spheroids and pigmented glia), and ONO-2808-03 projects/grants; serves as Co-PI of the Mayo Clinic APDA Center for Advanced Research and as an external advisory board member for the Vigil Neuroscience Inc.; and serves as a consultant for Eli Lilly & Company and NovoGlia Inc. The other authors report no relevant disclosures. Go to Neurology.org/NG for full disclosures.

Publication History

Received by *Neurology® Genetics* April 30, 2025. Accepted in final form June 18, 2026. Submitted and externally peer-reviewed. The handling editor was Editor Peter B. Kang, MD, FAAN.

References

1. Dulski J, Ross OA, Wszolek ZK. Genetics of Parkinson's disease: state-of-the-art and role in clinical settings. *Neurologia i Neurochirurgia Polska*. 2024;58(1):38-46. doi:10.5603/pjnns.97806
2. Mehanna R, Smilowska K, Fleisher J, et al. Age cutoff for early-onset Parkinson's disease: recommendations from the international parkinson and movement disorder Society task force on early onset Parkinson's disease. *Mov Disord Clin Pract*. 2022;9(7):869-878. doi:10.1002/mdc3.13523
3. Kolichski A, Turcano P, Tamvaka N, et al. Early-onset Parkinson's disease: creating the right environment for a genetic disorder. *J Parkinsons Dis*. 2022;12(8):2353-2367. doi:10.3233/jpd-223380
4. Hua P, Zhao Y, Zeng Q, et al. Genetic analysis of patients with early-onset Parkinson's disease in Eastern China. *Front Aging Neurosci*. 2022;14:849462. doi:10.3389/fnagi.2022.849462
5. Teixeira M, Sheta R, Idi W, Oueslati A. Alpha-synuclein and the endolysosomal system in Parkinson's disease: guilty by association. *Biomolecules*. 2021;11(9):1333. doi:10.3390/biom11091333
6. Milanowski LM, Lindemann JA, Hoffman-Zacharska D, et al. Frequency of mutations in PRKN, PINK1, and DJ1 in patients with early-onset Parkinson Disease from neighboring countries in Central Europe. *Parkinsonism Relat Disord*. 2021;86:48-51. doi:10.1016/j.parkreldis.2021.03.026

7. Marder KS, Tang MX, Mejia-Santana H, et al. Predictors of parkin mutations in early-onset Parkinson disease: the consortium on risk for early-onset Parkinson disease study. *Arch Neurol*. 2010;67(6):731-738. doi:10.1001/archneurol.2010.95
8. Sarihan EI, Perez-Palma E, Niestroj LM, et al. Genome-Wide analysis of copy number variation in Latin American Parkinson's Disease patients. *Mov Disord*. 2021;36(2):434-441. doi:10.1002/mds.28353
9. Tipton PW, Jaramillo-Koupermann G, Soto-Beasley AI, et al. Genetic characterization of Parkinson's disease patients in Ecuador and Colombia. *Parkinsonism Relat Disord*. 2020;75:27-29. doi:10.1016/j.parkreldis.2020.05.005
10. Landrum MJ, Lee JM, Riley GR, et al. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Res*. 2014;42(Database issue):D980-D985. doi:10.1093/nar/gkt1113
11. Rentzsch P, Witten D, Cooper GM, Shendure J, Kircher M. CADD: predicting the deleteriousness of variants throughout the human genome. *Nucleic Acids Res*. 2019;47(D1):D886-D894. doi:10.1093/nar/gky1016
12. Merico D, Pasternak Y, Zarrei M, et al. Homozygous duplication identified by whole genome sequencing causes LRBA deficiency. *NPJ Genom Med*. 2021;6(1):96. doi:10.1038/s41525-021-00263-z
13. Ross OA. *Colombian_EOPD_scripts Repository*. Accessed September 20, 2024. github.com/ORossLab/Colombian_EOPD_scripts
14. Kitada T, Asakawa S, Hattori N, et al. Mutations in the parkin gene cause autosomal recessive juvenile parkinsonism. *Nature*. 1998;392(6676):605-608. doi:10.1038/33416
15. Domingo A, Klein C. Genetics of Parkinson disease. *Handb Clin Neurol*. 2018;147:211-227. doi:10.1016/B978-0-444-63233-3.00014-2
16. Lucking CB, Durr A, Bonifati V, et al. Association between early-onset Parkinson's disease and mutations in the parkin gene. *N Engl J Med*. 2000;342(21):1560-1567. doi:10.1056/NEJM200005253422103
17. 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. doi:10.1093/hmg/ddaa273
18. Gray M, Nash KR, Yao Y. Adenylyl cyclase 2 expression and function in neurological diseases. *CNS Neurosci Ther*. 2024;30(7):e14880. doi:10.1111/cns.14880
19. Higgins AJ, Flynn AJ, Marconnet A, et al. Author correction: cycloalkane-modified amphiphilic polymers provide direct extraction of membrane proteins for CryoEM analysis. *Commun Biol*. 2022;5(1):1. doi:10.1038/s42003-021-02997-z
20. Garcia-Martin E, Diez-Fairen M, Pastor P, et al. Association between the missense alcohol dehydrogenase rs1229984T variant with the risk for Parkinson's disease in women. *J Neurol*. 2019;266(2):346-352. doi:10.1007/s00415-018-9136-9
21. Kelly K, Lewis PA, Plun-Favreau H, Manzoni C. Protein network analysis links the NSL complex to Parkinson's disease via mitochondrial and nuclear biology. *Mol Omics*. 2023;19(8):668-679. doi:10.1039/d2mo00325b
22. Jolly LA, Nguyen LS, Domingo D, et al. HCFC1 loss-of-function mutations disrupt neuronal and neural progenitor cells of the developing brain. *Hum Mol Genet*. 2015;24(12):3335-3347. doi:10.1093/hmg/ddv083
23. Castro VL, Quintana AM. The role of HCFC1 in syndromic and non-syndromic intellectual disability. *Med Res Arch*. 2020;8(6):10.18103/mra.v8i6.2122. doi:10.18103/mra.v8i6.2122