



Published in final edited form as:

*Mov Disord.* 2024 October ; 39(10): 1843–1855. doi:10.1002/mds.29931.

## Parkinson's disease gene screening in familial cases from Central and South America

Oswaldo Lorenzo-Betancor, MD, PhD<sup>1,2,#</sup>, Seysha Mehta, BSc<sup>3,#</sup>, Janvi Ramchandra<sup>4,5,#</sup>, Sekinat Mumuney, BSc<sup>3</sup>, Artur F. S. Schuh, MD, PhD<sup>6,7</sup>, Mario Cornejo-Olivas, MD<sup>8,9</sup>, Elison H. Sarapura-Castro, MD<sup>8,9</sup>, Luis Torres, MD<sup>10</sup>, Miguel A. Inca-Martinez, BSc<sup>4</sup>, Pilar Mazzetti, MD<sup>9,11</sup>, Carlos Cosentino, MD<sup>10,11</sup>, Federico Micheli, MD, PhD<sup>12,13</sup>, Vitor Tumas, MD, PhD<sup>14</sup>, Elena Dieguez, MD<sup>15</sup>, Victor Raggio, MD<sup>16</sup>, Vanderci Borges, PhD<sup>17</sup>, Henrique B. Ferraz, MD, PhD<sup>17</sup>, Pedro Chana-Cuevas, MD<sup>18</sup>, Marlene Jimenez-Del-Rio, PhD<sup>19</sup>, Carlos Velez-Pardo, PhD<sup>19</sup>, Sonia Moreno, PhD<sup>19</sup>, Francisco Lopera, MD<sup>19</sup>, Jorge Luis Orozco

\*Correspondence to: Ignacio F. Mata, Associate Staff, Genomic Medicine Institute (GMI), Lerner Research Institute (LRI), R4-006, Assistant Professor, Molecular Medicine, Cleveland Clinic Lerner College of Medicine (CCLCM) of Case Western Reserve University (CWRU), LRI, R4-006 | Cleveland Clinic | 9500 Euclid Ave. | Cleveland, OH 44195, Fax: (216) 636-0009, Lab Website: <https://www.lerner.ccf.org/gmi/mata/>.

#These authors contributed equally to the paper.

### Author roles

1. Research project: A. Conception, B. Methodology, C. Software analyses, D. Validation.

2. Manuscript Preparation: A. Writing of the first draft, B. Review and Critique.

3. Resources: A. Sample provision, B. Funding acquisition, C. Project coordination.

O.L.B.: 1A, 1B, 1C, 2A.

S.M.: 1D, 2A.

J.R.: 1D, 2A.

S.M.: 1D, 2B.

A.F.S.S.: 2B, 3A.

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

E.H.S.C.: 2B, 3A.

L.T.: 2B, 3A.

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

P.M.: 2B, 3A.

C.C.: 2B, 3A.

F.M.: 2B, 3A.

V.T.: 2B, 3A.

E.D.: 2B, 3A.

V.R.: 2B, 3A.

V.B.: 2B, 3A.

H.B.F.: 2B, 3A.

P.C.C.: 2B, 3A.

M.J.D.R.: 2B, 3A.

C.V.P.: 2B, 3A.

S.M.: 2B, 3A.

F.L.: 2B, 3A.

J.L.O.V.: 2B, 3A.

B.M.O.: 2B, 3A.

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

J.L.O.V.: 2B, 3A.

B.M.O.: 2B, 3A.

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

A.M.E.: 2B, 3A.

D.Y.: 2B, 3C.

C.P.Z.: 1A, 2B, 3B, 3C.

I.F.M.: 1A, 2B, 3B, 3C.

Financial Disclosures for the previous 12 months

The authors declare that there are no additional disclosures to report.

**Velez, MD<sup>20,21</sup>, Beatriz Muñoz Ospina, PhD<sup>20,21</sup>, Carlos R. M. Rieder, MD<sup>22</sup>, Alex Medina Escobar, MD<sup>23,24</sup>, Dora Yearout, BSc<sup>1,2</sup>, Cyrus P. Zabetian, MD, MS<sup>1,2</sup>, Ignacio F. Mata, PhD<sup>1,2,3,4,\*</sup>, Latin American Research Consortium on the Genetics of PD (LARGE-PD)**

<sup>1</sup>Veterans Affairs Puget Sound Health Care System, Seattle, WA, United States

<sup>2</sup>Department of Neurology, University of Washington School of Medicine, Seattle, WA, United States

<sup>3</sup>Cleveland Clinic Lerner College of Medicine, Case Western Reserve University, Cleveland, OH, United States

<sup>4</sup>Genomic Medicine Institute, Cleveland Clinic Foundation Lerner Research Institute, Cleveland, OH, United States

<sup>5</sup>Department of Biochemistry, Case Western Reserve University, Cleveland, OH, United States

<sup>6</sup>Department of Pharmacology, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

<sup>7</sup>Department of Neurology, Clinics Hospital of Porto Alegre, Porto Alegre, Brazil

<sup>8</sup>Neurogenetics Working Group, Universidad Científica del Sur, Lima, Peru

<sup>9</sup>Neurogenetics Research Center, National Institute of Neurological Sciences, Lima, Peru

<sup>10</sup>Movement Disorders Unit, National Institute of Neurological Sciences, Lima, Peru

<sup>11</sup>School of Medicine, Universidad Nacional Mayor de San Marcos, Lima, Peru

<sup>12</sup>Parkinson's Disease and Movement Disorders Center, University of Buenos Aires, Buenos Aires, Argentina

<sup>13</sup>Centro de Parkinson y Movimientos Anormales, Fundación San Gabriel, Córdoba, Argentina

<sup>14</sup>Ribeirão Preto Medical School, University of São Paulo, Brazil

<sup>15</sup>Neurology Institute, Universidad de la Republica, Montevideo, Uruguay

<sup>16</sup>Department of Genetics, Facultad de Medicina, Universidad de la Republica, Montevideo, Uruguay

<sup>17</sup>Movement Disorders Unit, Department of Neurology and Neurosurgery, Universidade Federal de São Paulo, São Paulo, Brazil

<sup>18</sup>Centro de Trastornos del Movimiento (CETRAM), Facultad de Ciencias Médicas, Universidad de Santiago de Chile, Santiago de Chile, Chile

<sup>19</sup>Neuroscience Research Group, Medical Research Institute, Faculty of Medicine, Universidad de Antioquia, Medellín, Colombia

<sup>20</sup>Department of Neurology, Valle del Lili Foundation, Cali, Colombia

<sup>21</sup>Department of Human Sciences, Icesi University, Cali, Colombia

<sup>22</sup>Departamento de Neurologia, Universidade Federal de Ciências da Saúde de Porto Alegre, Porto Alegre, Brazil

<sup>23</sup>Department of Neurology, Universidad Nacional Autónoma de Honduras, Tegucigalpa, Honduras

<sup>24</sup>The Moncton City Hospital, Moncton, Canada

## Abstract

**Background.**—Parkinson's disease (PD) is the second most common neurodegenerative disease following Alzheimer's disease. Nearly 30 causative genes have been identified for PD and related disorders. However, most of these genes were identified in European-derived families, and little is known about their role in Latin American populations.

**Objectives.**—Our goal was to assess the spectrum and frequency of pathogenic variants in known PD genes in familial PD patients from Latin America.

**Methods.**—We selected 335 PD patients with a family history of PD from the Latin American Research Consortium on the Genetics of PD. We capture-sequenced the coding regions of 26 genes related to neurodegenerative parkinsonism. Of the 335 PD patients, 324 had sufficient sequencing coverage to be analyzed.

**Results.**—We identified pathogenic variants in 41 individuals (12.7%) in *FBXO7*, *GCHI*, *LRRK2*, *PARK7*, *PINK1*, *PLA2G6*, *PRKN*, *SNCA*, and *TARDBP*, *GBA1* risk variants in 25 individuals (7.7%), and variants of uncertain significance in another 24 individuals (7.4%) in *ATP13A2*, *ATP1A3*, *DNAJC13*, *DNAJC6*, *GBA1*, *LRKK2*, *PINK1*, *VPS13C*, and *VPS35*. Of the 70 unique variants identified, 19 were more frequent in Latin Americans than in any other population.

**Conclusions.**—This is the first screening of known PD genes in a large cohort of patients with familial PD from Latin America. There were substantial differences in the spectrum of variants observed in comparison to previous findings from PD families of European origin. Our data provide further evidence that differences exist between the genetic architecture of PD in Latinos and European-derived populations.

## Keywords

Parkinson's disease; Latino; Hispanic; genetics; pathogenic variant

## 1. Introduction

Parkinson's disease ([PD] OMIM #168600) is the second most common neurodegenerative disease after Alzheimer's disease.<sup>1</sup> The prevalence of PD is approximately 1–2% among individuals over the age of 60 years.<sup>2</sup> The incidence of PD varies by ancestry, which can be attributed to different environmental exposures and distinct genetic background for each population.<sup>3</sup> Pathogenic variants in nearly 30 genes causing autosomal dominant (AD), autosomal recessive (AR), or X-linked parkinsonism have been described to date.<sup>4</sup> In addition, homozygous mutations in *GBA1* cause Gaucher disease (GD), an autosomal recessive lysosomal storage disorder that is sometimes accompanied by parkinsonism<sup>5</sup> and heterozygous carriers of pathogenic mutations for GD have an increased risk of developing PD.<sup>6</sup> However, most of these genes have been identified by analyzing large families of

European origin.<sup>7-9</sup> There are very few familial studies in other populations, especially in Latin America. It has been estimated that about 15-25% of sporadic PD patients have an affected PD relative,<sup>10</sup> though there is little information on how much the proportion of familial PD varies across different ethnic and racial groups. Studies targeting established PD genes have reported that the prevalence of several causal or risk alleles can vary greatly across populations. For example, the carrier frequency of LRRK2 p.G2019S among patients with familial PD is 3–6% in populations of European origin<sup>11</sup> and about 14% in Ashkenazi Jews,<sup>12</sup> but this pathogenic variant is exceedingly rare in Asians. Conversely, p.G2385R and p.R1628P are the most common LRRK2 risk variants in East Asians but occur at lower frequencies in other populations.<sup>13</sup> LRRK2 p.G2019S frequency in familial PD patients from Latin America varies widely between countries and its frequency correlates directly with the amount of European ancestry observed.<sup>14</sup> Our group reported a substantially higher frequency of *GBA1* gene pathogenic variants in PD patients from Colombia (10%) versus Peru (4%). This difference was driven by the *GBA1* p.K198E, which is unique to the Colombian population.<sup>15</sup> Thus, the distribution of PD-related variants may differ even among subpopulations of Latinos.

Unfortunately, most genetics studies performed in Latin America have only genotyped a limited number of PD-related variants identified in European derived individuals rather than screening the entire coding region of genes, losing the opportunity of finding novel variants specific to Latinos such as *GBA1* p.K198E.<sup>16</sup> Thus, the aims of this study were to determine the frequency of pathogenic variants in known PD genes in Latin American patients with familial PD and discover population-specific variants in those genes.

## 2. Methods

### 2.1 Study design and participants inclusion criteria

This multicenter, retrospective study included participants who were recruited between 2009-2021 from 11 centers in seven countries from Latin America (Argentina, Brazil, Chile, Colombia, Honduras, Peru, and Uruguay) that are part of the Latin American Research Consortium on the Genetics of Parkinson's Disease (LARGE-PD).

Participants were included in the study if they 1) had been diagnosed with PD according to the UK Brain bank (UKBB) criteria, and 2) had a family history of PD (defined as having an affected first- or second-degree relative). All individuals underwent a comprehensive neurological examination performed by a movement disorder specialist. A blood or saliva sample was collected from each participant for DNA extraction. The study was approved by the Institutional Review Board at each enrollment site and the VA Puget Sound. All subjects provided written informed consent. Ultimately, samples from 335 participants met initial inclusion criteria. Eleven samples had less than 80% of the targeted region with more than 100X coverage and were removed from the analysis (Supplementary Figure 1). The average coverage for the remaining 324 samples was >1000X.

## 2.2 Targeted sequencing and copy number variation analysis

Genomic DNA was extracted from peripheral blood or saliva using standard methods at each location. The probands were screened using a targeted next generation sequencing (NGS) panel including 26 genes related to neurodegenerative parkinsonism: *ATP13A2*, *ATP1A3*, *DNAJC13*, *DNAJC6*, *EIF4G1*, *FBXO7*, *GBA1*, *GCH1*, *GIGYF2*, *HTRA2*, *LRRK2*, *PARK7*, *PINK1*, *PLA2G6*, *PRKN*, *RAB39B*, *SLC6A3*, *SNCA*, *SNCB*, *SYNJ1*, *TAF1*, *TARDBP*, *TH*, *TMEM230*, *VPS13C*, and *VPS35* (gene names, associated phenotypes, and gene coordinates based on genome build GRCh37 are provided in Supplementary Table 1). This panel includes the same genes as prior publications<sup>17</sup> and does not include some of the latest PD-related genes, such as *CHCHD2*, *PSAP*, or *PTPA*.

Targeted NGS was performed between 2017-2021 at the University of Washington Medicine Center for Precision Diagnostics with 1 µg of DNA on a HiSeq 2000 Sequencing System (Illumina, San Diego, CA) using standard 150 bp paired-end reads. Alignment and base calling was performed according to the GATK4 best practices pipeline.<sup>18</sup> Variant annotation was performed using ANNOVAR software.<sup>19</sup> Variants were numbered according to standard nomenclature (<http://www.hgvs.org/mutnomen/>)<sup>20</sup> and **excluded**, for all genes except *GBA1*, if they matched any of the following filtering criteria: 1) low quality variants: genotype quality <20 or read depth <10; 2) single nucleotide polymorphisms with a minor allele frequency (MAF) >1% in the Exome Sequencing Project 6500 (<http://evs.gs.washington.edu/EVS/>), 1000 Genomes database<sup>21</sup> or in any of the populations from the Genome Aggregation Database v.4.0.0 (gnomAD v.4.0.0);<sup>22</sup> 3) variants reported as “Benign” or “Likely benign” based on a consensus that included the ClinVar variant classification (<https://www.ncbi.nlm.nih.gov/clinvar/>),<sup>23</sup> the Human Genomics variant search engine Varsome (<https://varsome.com/>),<sup>24</sup> and the Franklin by genoox tool (<https://franklin.genoox.com/clinical-db/home>), 4) variants with a Combined Annotation Dependent Depletion (CADD) score <15;<sup>25</sup> 5) variants considered polymorphisms in the literature; and 6) splicing variants that did not damage the splicing region according to Human Splicing Finder algorithm.<sup>26</sup> Only variants fulfilling the following criteria were kept in the analysis: missense variants, coding deletions or insertions, frameshift variants, or intronic and synonymous variants near splice sites (≤6 nucleotides; see Fluxogram in Supplementary Figure 2).

Candidate variants that have never been reported before in gnomAD v4.0.0, nor in any prior publications, were validated, when additional DNA was available, by Sanger sequencing performed either by Eurofins Genomics with the Applied Biosystems Big-Dye Terminator v3.1 Cycle Sequencing Kit using the manufacturer’s recommended methodology. Primer pairs were designed using Primer3 (<http://bioinfo.ut.ee/primer3-0.4.0/>) and NCBI Primer Blast. Sequences were aligned and analyzed with Mutation Surveyor (v5.1.2, SoftGenetics, State College, PA).

*PRKN*, *PARK7*, *PINK1*, and *SNCA* were also screened for copy number variations (CNV) using the panelcn.MOPS R package.<sup>27</sup> The other recessive genes were not screened for CNVs because there is currently no evidence that deletions or duplications in any of those genes are associated with PD, and their validation would be difficult as they are not included

in the existing multiplex-ligation dependent probe amplification (MLPA) kits. The pipeline of this R package uses BAM files as inputs. A BED file is provided to specify the regions of interest (ROIs) that were used as count windows. For these four genes, each ROI consisted of a coding exon  $\pm 31$  nucleotides of flanking intronic sequences (Supplementary Table 2) leading to ROIs with a size between 69 and 449 bp. The package `panelcn.MOPS` models the depths of coverage across samples at each genomic position.<sup>27</sup> To select the reference control group, all samples (n=324) were initially considered individuals without CNVs. To limit the variance of the read counts (RCs), `panelcn.MOPS` uses the 25 samples from the reference controls with higher correlation of RCs compared to the RCs of each test sample. For every gene analysis, ROIs that are part of a candidate gene for the test sample are excluded from the correlation calculation. The `panelcn.MOPS` model sets a constant copy number of two for all samples and uses the linear relation between copy numbers and RCs to determine any trend that moves further away from the posterior probability than expected. We identified 109 robust reference samples that had no CNVs. When genome-wide array data was available, CNVs were validated utilizing a method described elsewhere,<sup>28</sup> showing good reliability of the `panelcn.MOPS` CNV detection algorithm (Table 1, and Supplementary Table 5).

Variants were considered pathogenic based on a consensus that included the ClinVar variant classification (<https://www.ncbi.nlm.nih.gov/clinvar/>),<sup>23</sup> the Human Genomics variant search engine Varsome (<https://varsome.com/>),<sup>24</sup> and the Franklin by genoox tool (<https://franklin.genoox.com/clinical-db/home>), which uses the criteria of the American College of Medical Genetics (ACMG) for variant classification. Variants in all genes except *GBAI* that were classified as “pathogenic” or “likely pathogenic” by one or more of these tools were kept in the main study, while the rest were considered Variants of Uncertain Significance (VUS) and are listed separately in Supplementary Table 3. All duplications and deletions in *PARK7*, *PINK1*, *PRKN*, and *SNCA* were considered pathogenic. *GBAI* variants were separated into three categories: (1) variants classified as “pathogenic” or “likely pathogenic” for GD by one or more of the previous tools were considered risk factors for PD (Table 2A); (2) variants that have been linked to PD risk previously but are not pathogenic for GD are listed separately (Table 2B); (3) the rest of the variants were considered VUS (Supplementary Table 4). All *GBAI* variants were assessed for PD severity using the *GBAI*-PD Browser (<https://pdgenetics.shinyapps.io/gba1browser/>)<sup>29</sup>

### 3. Results

#### 3.1 Cohort description

The final LARGE-PD familial cohort included 176 (54.3%) males and 148 (45.7%) females. The average age at onset (AAO) of the probands was 50.3 years (SD = 15.5; range = 8-85) and the average disease duration was 9.1 years (SD = 7.7; range = 0-64). Most of the patients were from Brazil (n = 107, n females = 48) and Peru (n = 107, n females = 50), followed by Colombia (n = 56, n females = 25), Argentina (n = 30, n females = 13), Uruguay (n = 12, n females = 7), Honduras (n = 9, n females = 4), and Chile (n = 3, n females = 1; Figure 1A). The AAO and disease duration were similar across countries

(Supplementary Table 4), except for the Peruvian cohort that had an earlier AAO ( $40.0 \pm 15.6$ ) and a slight longer disease duration ( $10.4 \pm 9.5$ ).

### 3.2 Diagnostic Yield

We identified pathogenic variants in “causal” genes, *GBA1* risk variants, and variants of uncertain significance (VUS) in 41 (12.7%), 25 (7.7%), and 24 (7.4%) individuals, respectively (Figure 1B). The prevalence of pathogenic variant carriers was highest for *PRKN* (n=22), *LRRK2* (n=9), and *SNCA* (n=3) (Figure 1B). We identified 25 carriers of PD risk variants in *GBA1*. The carrier frequency for VUS was highest for *LRRK2* (n=9), and *DNAJC13* (n=6). All these counts include carriers of single variants in autosomal dominant genes, or *GBA1* risk variants, and homozygous or compound heterozygous variants in autosomal recessive genes (Tables 1 and 2). Carriers of single variants in recessive genes are listed in Supplementary Table 5.

In total, we identified 70 unique variants (pathogenic, risk variants, or VUS) in the *ATP13A2*, *ATP1A3*, *DNAJC13*, *DNAJC6*, *FBXO7*, *GBA1*, *GCH1*, *LRRK2*, *PARK7*, *PINK1*, *PLA2G6*, *PRKN*, *SNCA*, *TARDBP*, *VPS13C*, or *VPS35* genes (Figure 1C; Tables 1 and 2, and Supplementary Tables 3 and 4). Of the 70 unique variants, 56 were single nucleotide variants [SNVs] and 14 were copy number variants. Of the 56 SNVs (22 pathogenic variants; 11 *GBA1* risk factors, and 23 VUS; Figure 1C), six have never been described in any public databases. Nineteen of the 56 variants were more frequent in the Latino population than any other population according to gnomAD v4.0 (Tables 1 and 2, and Supplementary Tables 3 and 4). We were able to validate five of the six novel SNVs (*PARK7* p.G75C, *PRKN* p.G430\_C436del, *PRKN* p.K27X, *PRKN* c.7+1G>T, and *PLA2G6* p.Y643X). We did not have enough DNA remaining to attempt validation of one SNV (*GBA1* p.V457D).

Two of the 14 CNVs observed were of particular interest. We identified two individuals with a partial *SNCA* duplication and a deletion: *SNCA* exon 1 heterozygous duplication in a male with early onset (AAO = 42) levodopa-responsive, tremor dominant PD from Brazil, and a *SNCA* exon 2 heterozygous deletion in an Argentinian male (AAO = 63; Figure 3). Unfortunately, lack of additional DNA precluded our ability to validate either of these variants using a different approach.

The other 12 CNVs have been reported multiple times to date. We were able to validate six of them using available genome-wide genotyping data (Table 1 and Supplementary Table 5). The other 6 did not have additional genotyping data for validation.

We observed different variant distributions by country. While most cases in Peru were caused by *PRKN* (n=14), the number of cases in Brazil caused by *PRKN* (n = 5), and *LRRK2* (n = 4) were very similar (Figure 2A). The prevalence of *GBA1* risk variants also differed by country. While there were 12 *GBA1* carriers in Brazil and seven in Colombia, there were only four carriers in Peru (Figure 2A). VUS distribution also varied between countries (Figure 2B).

Additionally, there were 19 individuals who had a single pathogenic variant in one of the PD recessive genes that could not fully explain the disease because they were missing a second variant in the same gene (*PRKN* = 15; *PLA2G6* = 2; *PARK7* = 1; *TH* = 1; Supplementary Table 5A). Another 38 individuals carried a single VUS in a recessive PD gene (Supplementary Table 5B). CNV analyses of *PARK7*, *PINK1*, and *PRKN* genes ruled out a deletion or a duplication as the second pathogenic variant in 24 individuals carrying a single variant in those genes. However, a partial or complete deletion or duplication could not be ruled out in any of the individuals carrying a single variant in any of the other recessive genes (n = 33) because CNV analyses for those genes were not performed.

Furthermore, seven subjects had potential compound heterozygous variants in a known recessive gene, but one of the two variants did not pass filtering criteria and was not included in the analysis (Supplementary Table 7).

#### 4. Discussion

To our knowledge, this is the first study in which a large Latino familial PD cohort was screened for variants in known PD genes. It provides a broad estimate of the frequency of pathogenic variants and VUSs in these genes in familial cases in several Latino populations.

We identified pathogenic variants (either single variants in autosomal dominant genes or homozygous or compound heterozygous variants in autosomal recessive genes) in 41 participants (12.7%) and *GBA1* risk variants in 25 individuals (7.7%). In comparison, the initial report from the Rostock International Parkinson's Disease (ROPAD) study which included 1,360 PD patients (96.4% Europeans, 1.7% Hispanics, 0.9% Asians, 0.5% Black Africans, and 0.5% other ancestry) reported a genetic diagnosis in about 5.2% of their participants, and *GBA1* risk variants, including p.E326K, and p.T369M, in 8.5% of their cohort.<sup>30</sup> Interestingly, the ROPAD study screened for variants in 47 different PD-related genes, nineteen of which overlap with our study. This study did not provide a list of all identified variants nor which ones were more common in each ethnic group, so a direct comparison between studies is not possible. The most plausible explanation for their lower diagnostic yield, even if they screened a larger number of genes (47 vs 26), is that only 27.4% of their participants had a family history of disease, whereas all our participants had at least one first- or second-degree affected relative. Furthermore, some clinically significant variants might have been missed since only 71% (807/1,136) of their cohort underwent panel sequencing. In a similar study that assessed genes associated with different movement disorders (e.g., tremor, dystonia, myoclonus, chorea, parkinsonism, tics, and paroxysmal movement disorders) in a cohort of 378 individuals of mostly European descent, pathogenic variants were identified in 25.4% of individuals with parkinsonism (n = 46/181). The later study reported a family history of disease in 37% of all participants, but specific details for the parkinsonian subgroup were not available for comparison.<sup>31</sup> Since all participants in our cohort reported at least one affected relative, we expected to obtain a much higher diagnostic yield.

However, it is important to highlight that in our study we only screened genes that have been previously identified in European derived populations. Nine of the pathogenic variants

that were identified in these known PD-related genes are more frequent in Latinos (Admixed Americans in gnomAD) than in any other population (Tables 1 and 2), confirming the hypothesis that there are Latino specific pathogenic variants. Furthermore, there could also be Latino specific PD genes yet to be discovered that were not sequenced.

In the present cohort, two of the 57 (3.5%) Colombian PD probands were carriers of the GBA1 p.K198E pathogenic variant. The differences in previous studies regarding the association of this variant with risk for PD may be due to sample heterogeneity. While the cohort in the original study showing an association had been recruited in Medellín (Antioquia),<sup>15</sup> the cohort from the replication study was more heterogeneous.<sup>32</sup> Importantly, all four variant carriers from the replication study were of Amerindian ancestry.<sup>32</sup> Antioquia residents, and especially the “paisa” population, are a genetically isolated community,<sup>33</sup> suggesting a probable founder effect which would explain the high p.K198E frequency in this population. In our study, one p.K198E variant carrier was from Cali, Colombia and the second one was from Antioquia (Medellín, Colombia), further reinforcing the hypothesis that this variant is specific to the Colombian population. The Cali individual had a first degree relative who also carried p.K198E variant.

Interestingly, some *PRKN* variants were observed only in certain countries. In this regard, *PRKN* p.N52fs was only present in Brazilian patients (two homozygote and two compound heterozygote carriers; Table 1), and the *PRKN* c.619-1G>A splicing variant was only present in Peruvian individuals (2 homozygote and 3 compound heterozygote carriers; Table 1). These findings suggest that the *PRKN* variant spectrum in Latin American countries differs across regions and that there are pathogenic variants that are population specific. In fact, the MAF for the *PRKN* p.N52fs in the gnomAD v4 database is almost five times higher in Admixed Americans compared to other populations and *PRKN* c.619-1G>A was only observed in an Admixed American individual.

Recently, complex structural variants in the *PRKN* region, such as inversions within the gene or involving breakpoints outside of *PRKN*, have been described as a cause of levodopa-responsive dystonia<sup>34</sup> and young onset dystonia-parkinsonism<sup>35</sup> using whole genome sequencing and long-read sequencing, respectively. The capture sequencing technique that we used does not allow for the identification of these types of structural variants. It is possible that some of the single variant carriers in the *PRKN* gene (n = 17) in our cohort may be explained by these types of structural variants.

The interpretation of *DNAJC13* gene variants requires some caution. Six *DNAJC13* VUS that have never been reported to date were identified in this gene in our dataset (Supplementary Table 3). Considering that the original manuscript describing the first “pathogenic” *DNAJC13* variant (p.N855S) included six asymptomatic relative carriers and two phenocopies in the index family,<sup>36</sup> it is uncertain whether missense variants in *DNAJC13* gene cause PD. An association study performed in 2,408 unrelated PD patients and 3,444 controls from European and Ashkenazi-Jewish ancestry, that included *DNAJC13*, *UCHL1*, *HTRA2*, *GIGYF2*, and *EIF4G1* genes was unable to reveal a significant association between any of these genes and PD after correcting for multiple comparisons.<sup>37</sup> Therefore, it

is possible that the VUS identified in our study reflect only the normal *DNAJC13* variation in Latino populations.

In total, we identified nine *GBA1* pathogenic or likely pathogenic variants for GD and two risk factors for PD in 25 individuals (Table 2). Five of the variants have been reported to cause GD and are also risk factors for PD (p.W184R, p.K198E, p.G202R, p.N370S, and p.L444P),<sup>15, 38</sup> four are likely pathogenic variants for GD whose risk for PD is either uncertain or has not been assessed, and two do not cause GD, but are risk factors for PD (p.E326K, and p.T369M; Table 2B). Additionally, we identified three *GBA1* VUS (Supplementary Table 4). From the likely pathogenic variants, two have never been reported in GD patients before (p.L160fs, and p.V457D). Of these, one was identified in a PD patient from Colombia (p.L160fs), where prevalence studies for GD pathogenic variants have been carried out. Conversely, the *GBA1* p.V457D variant was present in a PD individual from Chile. To the best of our knowledge, no screening for GD variants has been performed in this country, and this variant has not been described elsewhere. Therefore, this variant may be specific to this population. This further emphasizes that the spectrum of variants in known PD genes in the Latino population includes several variants that have not been studied before and require further characterization.

The current study has certain limitations. First, the sample size for some of the countries is low, and it may be hard to have a good estimation of variant frequencies, especially for rare genes, in those countries. Therefore, the generalization of the results of this study to certain countries is limited. Second, since our gene panel was assembled based on PD genes identified before 2016, we might have missed pathogenic variants in more recently discovered PD-related genes. Third, we identified a large number of VUS because there is less genomic data variation available for Latino populations than for other ethnicities. This makes the pathogenic interpretation for some variants harder to establish. Fourth, we did not perform CNV analyses for all recessive genes included in the panel to identify a second hit in single variant carriers. Finally, we did not have DNA for affected relatives of most probands, so we were not able to perform co-segregation analyses, which would have been helpful to elucidate their pathogenicity.

The results of the current work emphasize the importance of performing more studies in underrepresented populations, such as those in Latin American countries. The identification of additional individuals carrying the new variants reported in our study may help establish pathogenicity. In addition, more resources are needed for functional validation especially for VUS.

A global movement with the support of the Aligning Science Across Parkinson's Global Parkinson's Genetics Program (ASAP-GP2) and Michael J Fox Foundation is currently ongoing and will further advance our understanding of the genetics of PD in all populations.

In summary, this work shows that the genetic risk factors for familial PD in Latinos only partially overlap with European populations, and pathogenic variants are sometimes unique to certain Latino populations. The extent to which these specific variants originated in

the ancestral Amerindian population or occurred recently and are therefore geographically limited requires analyses of larger cohorts including individuals from additional countries.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## Acknowledgments

We sincerely thank all the individuals who participated in this study. There are no financial conflicts of interest to disclose.

## Funding Sources and Conflict of Interest

Recruitment of LARGE-PD participants was originally funded by an International Research Program Grant (2010-2012) to C.P.Z. and I.F.M. and by a Stanley Fahn Junior Faculty Award (2016-2019) to I.F.M., both from the Parkinson's Disease Foundation. This work was funded by a research grant from the American Parkinson's Disease Association to I.F.M. and with resources and the use of facilities at the Veterans Affairs Puget Sound Health Care System (O.L.B., D.Y., C.P.Z., and I.F.M.).

I.F.M. also receives funding from the National Institutes of Health (NIH) grant R01 1R01NS112499-01A1, the Michael J. Fox Foundation, and the Aligning Science Across Parkinson's initiative. M.C.-O. and E.H.S.-C. were supported by the D43TW009345 Global Health Research Fellowship.

## Appendix A

Members of the Latin American Research Consortium on the Genetics of PD (LARGE-PD)

**Argentina:** Emilia Gatto, Federico Micheli, Clarisa Marchetti, Marcelo Kauffman, Claudia Perandones, Sergio Rodriguez, Alejandro Pellene, Marcela Montiel, Alejandro San Juan, Natalia Gonzalez, Gustavo Da Prat, Martin Radrizzani, Emmanuel Franchello, Cesar Avila, Griselda Alvarado, Lucia Wang, Lucila Falcone, Marcello Merello, Maria Cecilia Peralta, Valentina Muller, Matias Lopez, Federico Capparelli, Carolina Villa, Marcela Tela, Dario Adamec

**Bolivia:** Erick Gonzalez

**Brazil:** Vitor Tumas, Delson José da Silva, Francisco Eduardo Costa Cardoso, Helio Afonso Ghizoni Teive, Artur Francisco Schumacher-Schuh, Carlos Roberto de Mello Rieder, Marcus Vinicius Della Coletta, Bruno Lopes dos Santos Lobato, Egberto Reis Barbosa, Pedro Renato de Paula Brandão, Clécio de Oliveira Godeiro Júnior, Vanderci Borges, Pedro Braga Neto, Ana Lucía Zuma de Rosso, Grace Helena Letro, Maria Gabriela dos Santos Ghilardi, Sarah Camargos

**Chile:** Pedro Chana, Patricio Olguin, Paula Saffie, Alicia Colombo, Maria Eugenia Pinto, Floria Pancetti, Elias Fernandez,

**Colombia:** Carlos Velez-Pardo, Gonzalo Arboleda Bustos, Ruth Eliana Pineda Mateus, Sonia Catalina Cerquera Cleves, Jorge Luis Orozco Velez, Beatriz Muñoz, Marlene Jimenez del Rio

**Costa Rica:** Jaime Fornaguera, Alvaro Hernandez, Gabriel Torrealba

**Dominican Republic:** Rossy Cruz Vicioso

**Ecuador:** Edison Vasquez

**El Salvador:** Susana Peña, Tatiana Ascencio, Roberto Ayala

**Grenada:** Andrew Sobering

**Honduras:** Reyna Duron Martinez, Alex Medina

**Mexico:** Daniel Martinez Ramirez, Mayela Rodriguez, Renteria Miguel, Sarael Alcauter, Paula Reyes, Ana Jimena Hernandez-Medrano, Alejandra Ruiz Contreras, Alejandra Medina Rivera, Alejandra Lázaro Figueroa

**Peru:** Mario Cornejo-Olivas, Angel Medina, Julia Esther Rios Pinto, Ivan Fernando Cornejo Herrea, Edward Ochoa Valle, Nicanor Mori, Koni Mejia, Carlos Cosentino, Luis Torres, Cintia Armas, Maryenela Illanes, Freddy Requejo

**Puerto Rico:** Angel Viñuela, Esther Colon

**Uruguay:** Elena Dieguez, Victor Raggio, Andres Lescano

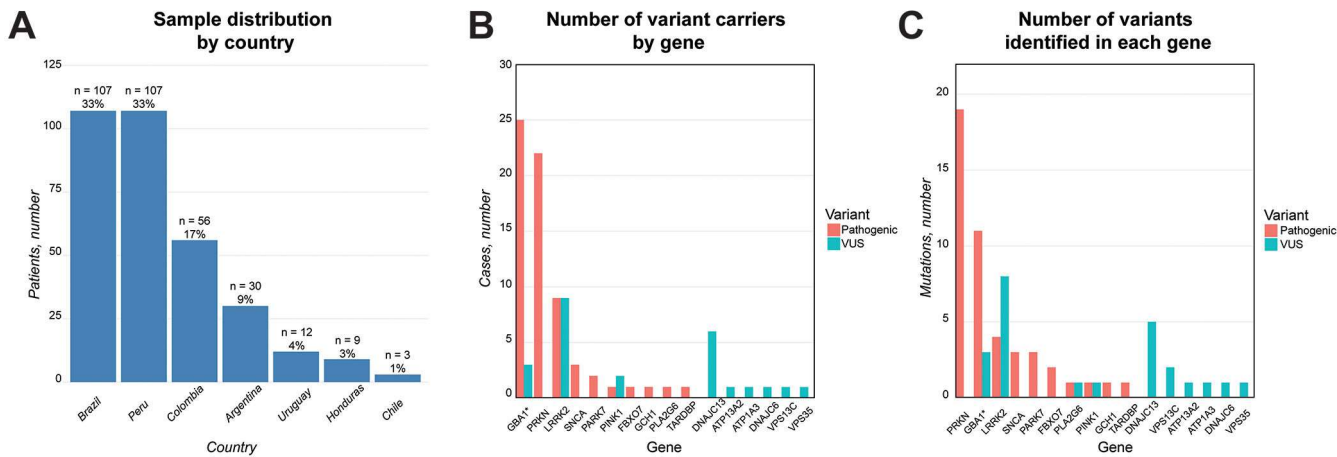
**USA:** Ignacio Fernandez Mata, Mauricio Chaparro Solano, Nicolas Gutierrez, Maria Rivera, Emily Waldo, Miguel Inca-Martinez, Karen Nuytemans

## REFERENCES

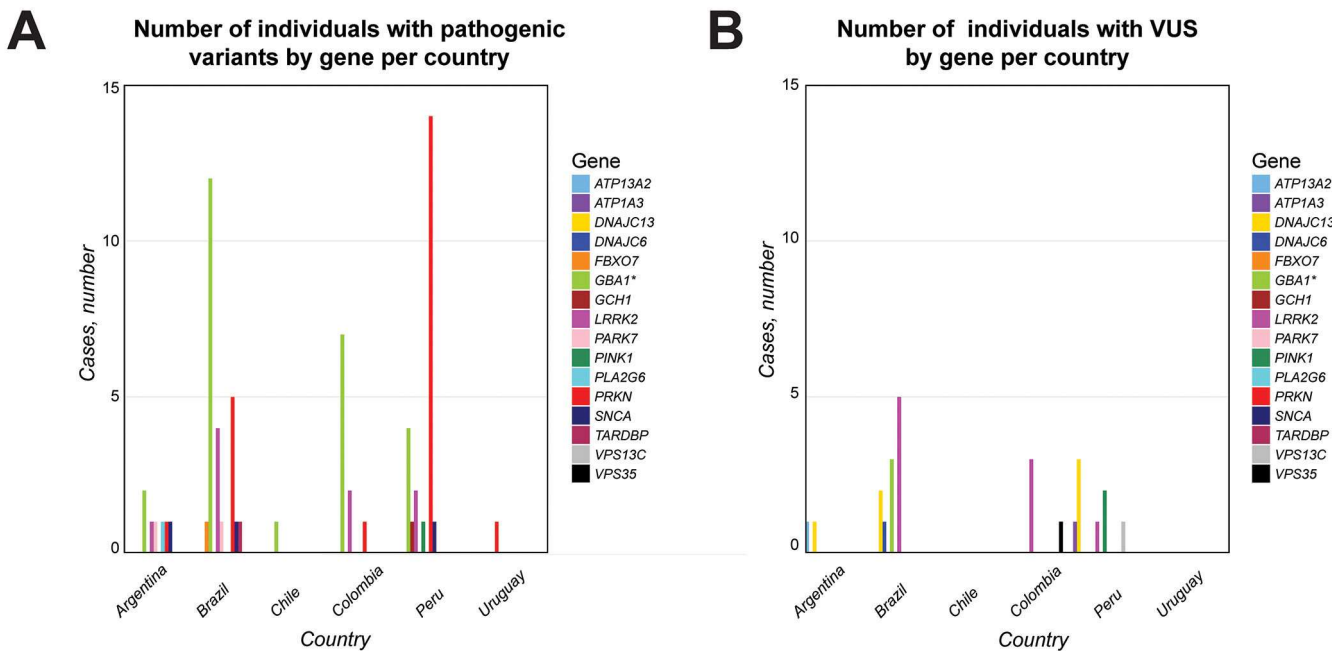
1. Poewe W, Seppi K, Tanner CM, et al. Parkinson disease. Nat Rev Dis Primers 2017;3:17013. [PubMed: 28332488]
2. de Rijk MC, Launer LJ, Berger K, et al. Prevalence of Parkinson's disease in Europe: A collaborative study of population-based cohorts. Neurologic Diseases in the Elderly Research Group. Neurology 2000;54(11 Suppl 5):S21–23.
3. Van Den Eeden SK, Tanner CM, Bernstein AL, et al. Incidence of Parkinson's disease: variation by age, gender, and race/ethnicity. Am J Epidemiol 2003;157(11):1015–1022. [PubMed: 12777365]
4. Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. Lancet Neurol 2020;19(2):170–178. [PubMed: 31521533]
5. Cormand B, Montfort M, Chabas A, Vilageliu L, Grinberg D. Genetic fine localization of the beta-glucocerebrosidase (GBA) and prosaposin (PSAP) genes: implications for Gaucher disease. Hum Genet 1997;100(1):75–79. [PubMed: 9225972]
6. Lwin A, Orvisky E, Goker-Alpan O, LaMarca ME, Sidransky E. Glucocerebrosidase mutations in subjects with parkinsonism. Mol Genet Metab 2004;81(1):70–73. [PubMed: 14728994]
7. Domingo A, Klein C. Genetics of Parkinson disease. Handb Clin Neurol 2018;147:211–227. [PubMed: 29325612]
8. Deng H, Wang P, Jankovic J. The genetics of Parkinson disease. Ageing research reviews 2018;42:72–85. [PubMed: 29288112]
9. Bekris LM, Mata IF, Zabetian CP. The genetics of Parkinson disease. J Geriatr Psychiatry Neurol 2010;23(4):228–242. [PubMed: 20938043]
10. Parkinson's Disease: Hope Through Research. In: (NINDS) NINDaS, editor. 2020.
11. Guo L, Wang W, Chen SG. Leucine-rich repeat kinase 2: relevance to Parkinson's disease. Int J Biochem Cell Biol 2006;38(9):1469–1475. [PubMed: 16600664]

12. Luzon-Toro B, Rubio de la Torre E, Delgado A, Perez-Tur J, Hilfiker S. Mechanistic insight into the dominant mode of the Parkinson's disease-associated G2019S LRRK2 mutation. *Hum Mol Genet* 2007;16(17):2031–2039. [PubMed: 17584768]
13. Fu X, Zheng Y, Hong H, et al. LRRK2 G2385R and LRRK2 R1628P increase risk of Parkinson's disease in a Han Chinese population from Southern Mainland China. *Parkinsonism Relat Disord* 2013;19(3):397–398. [PubMed: 22981185]
14. Cornejo-Olivas M, Torres L, Velit-Salazar MR, et al. Variable frequency of LRRK2 variants in the Latin American research consortium on the genetics of Parkinson's disease (LARGE-PD), a case of ancestry. *NPJ Parkinson's disease* 2017;3:19.
15. Velez-Pardo C, Lorenzo-Betancor O, Jimenez-Del-Rio M, et al. The distribution and risk effect of GBA variants in a large cohort of PD patients from Colombia and Peru. *Parkinsonism Relat Disord* 2019;63:204–208. [PubMed: 30765263]
16. Santos-Lobato BL, Schumacher-Schuh AF, Mata IF. Lack of full sequencing GBA1 studies for patients with Parkinson's disease in Latin America. *NPJ Parkinson's disease* 2022;8(1):101.
17. Lorenzo-Betancor O, Lin YH, Samii A, et al. Novel compound heterozygous FBXO7 mutations in a family with early onset Parkinson's disease. *Parkinsonism Relat Disord* 2020;80:142–147. [PubMed: 33002721]
18. Poplin R, Ruano-Rubio V, DePristo MA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv* 2018:201178.
19. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res* 2010;38(16):e164. [PubMed: 20601685]
20. den Dunnen JT, Antonarakis SE. Nomenclature for the description of human sequence variations. *Hum Genet* 2001;109(1):121–124. [PubMed: 11479744]
21. Auton A, Brooks LD, Durbin RM, et al. A global reference for human genetic variation. *Nature* 2015;526(7571):68–74. [PubMed: 26432245]
22. Karczewski KJ, Francioli LC, Tiao G, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *bioRxiv* 2020:531210.
23. Landrum MJ, Chitipiralla S, Brown GR, et al. ClinVar: improvements to accessing data. *Nucleic Acids Res* 2020;48(D1):D835–D844. [PubMed: 31777943]
24. Kopanos C, Tsiolkas V, Kouris A, et al. VarSome: the human genomic variant search engine. *Bioinformatics* 2019;35(11):1978–1980. [PubMed: 30376034]
25. 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. [PubMed: 30371827]
26. Desmet FO, Hamroun D, Lalande M, Collod-Beroud G, Claustres M, Beroud C. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Res* 2009;37(9):e67. [PubMed: 19339519]
27. Povysil G, Tzika A, Vogt J, et al. panelcn.MOPS: Copy-number detection in targeted NGS panel data for clinical diagnostics. *Hum Mutat* 2017;38(7):889–897. [PubMed: 28449315]
28. 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. [PubMed: 33150996]
29. Parlar SC, Grenn FP, Kim JJ, Baluwendrat C, Gan-Or Z. Classification of GBA1 Variants in Parkinson's Disease: The GBA1-PD Browser. *Mov Disord* 2023;38(3):489–495. [PubMed: 36598340]
30. Skrahina V, Gaber H, Vollstedt EJ, et al. The Rostock International Parkinson's Disease (ROPAD) Study: Protocol and Initial Findings. *Mov Disord* 2021;36(4):1005–1010. [PubMed: 33314351]
31. Montaut S, Tranchant C, Drouot N, et al. Assessment of a Targeted Gene Panel for Identification of Genes Associated With Movement Disorders. *JAMA neurology* 2018;75(10):1234–1245. [PubMed: 29913018]
32. Tipton PW, Soto-Beasley AI, Walton RL, et al. Prevalence of GBA p.K198E mutation in Colombian and Hispanic populations. *Parkinsonism Relat Disord* 2020;73:16–18. [PubMed: 32197197]

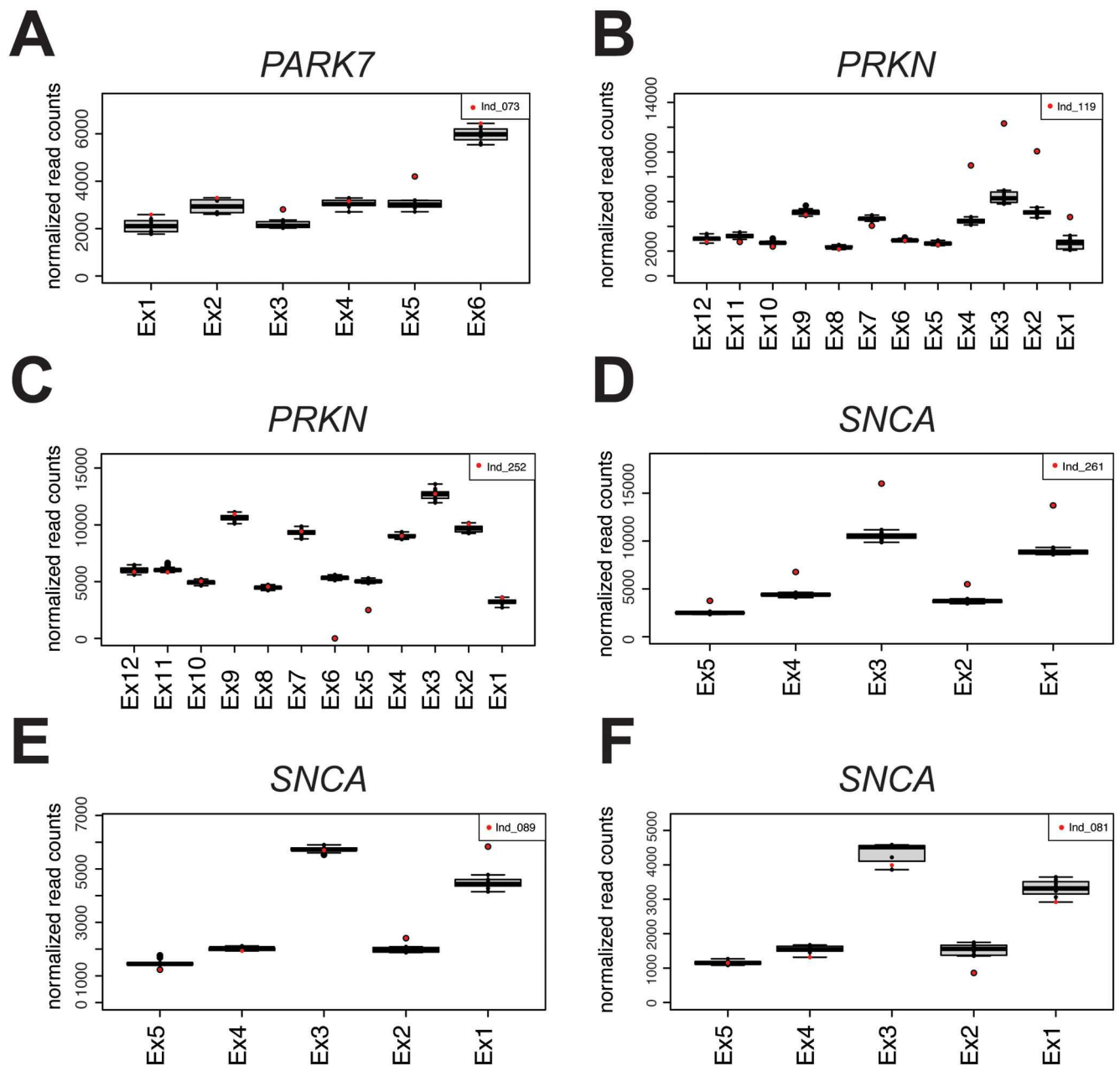
33. De Castro M, Restrepo CM. Genetics and genomic medicine in Colombia. *Molecular genetics & genomic medicine* 2015;3(2):84–91. [PubMed: 25802878]
34. Mor-Shaked H, Paz-Ebstein E, Basal A, et al. Levodopa-responsive dystonia caused by biallelic PRKN exon inversion invisible to exome sequencing. *Brain Commun* 2021;3(3):fcab197. [PubMed: 34514401]
35. Daida K, Funayama M, Billingsley KJ, et al. Long-Read Sequencing Resolves a Complex Structural Variant in PRKN Parkinson's Disease. *Mov Disord* 2023;38(12):2249–2257. [PubMed: 37926948]
36. Vilarino-Guell C, Rajput A, Milnerwood AJ, et al. DNAJC13 mutations in Parkinson disease. *Hum Mol Genet* 2014;23(7):1794–1801. [PubMed: 24218364]
37. Saini P, Rudakou U, Yu E, et al. Association study of DNAJC13, UCHL1, HTRA2, GIGYF2, and EIF4G1 with Parkinson's disease. *Neurobiol Aging* 2021;100:119 e117–119 e113.
38. Hoglinger G, Schulte C, Jost WH, et al. GBA-associated PD: chances and obstacles for targeted treatment strategies. *J Neural Transm (Vienna)* 2022.



**Figure 1.**  
A. Sample distribution across countries. B. Pathogenic variants and variants of uncertain significance carriers' distribution by gene. C. Pathogenic variants and variants of uncertain significance distribution by gene. VUS = Variant of uncertain significance. \*Pathogenic *GBA1* counts refer to individuals carrying variants that cause GD and/or are considered risk factors for PD.



**Figure 2. Distribution of pathogenic variants and variants of uncertain significance by countries.**  
A. Pathogenic variants. \*Pathogenic *GBA1* counts refer to individuals carrying variants that cause GD and/or are considered risk factors for PD. B. Variants of uncertain significance.



**Figure 3. Copy number variants examples identified using the panelcn.MOPS package.** The normalized RCs of each test sample and all controls are displayed as boxplots for each exon. Exons are displayed in chromosomal order for each gene. The RCs of each control sample are symbolized by black dots, whereas the RCs of the test sample are highlighted by red dots. A deviation of the test sample over the boxplot and whiskers represents a duplication, while its deviation below the boxplot and whiskers, represents a deletion. Depending on the degree of deviation the duplication or deletion can be either heterozygous or homozygous. A. *PARK7* Ex3 and Ex5 compound heterozygous duplication; B. *PRKN* Ex1-Ex4 homozygous duplication; C. *PRKN* Ex5 heterozygous deletion and

Ex6 homozygous deletion; D. Entire *SNCA* heterozygous duplication; E. *SNCA* Ex1 heterozygous duplication; F. *SNCA* Ex2 heterozygous deletion.

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Pathogenic variants in known PD genes in the LARGE-PD familial cohort.

| Sample ID | Sex | AAO, y | Country   | Gene          | Variant; status           | CADD score | Pathogenicity * consensus | MAF Admixed American gnomAD <sup>§</sup> | MAF NFE gnomAD <sup>§</sup> | Max MAF Population gnomAD <sup>§</sup> |
|-----------|-----|--------|-----------|---------------|---------------------------|------------|---------------------------|------------------------------------------|-----------------------------|----------------------------------------|
| Ind_073   | F   | 38     | Argentina | <i>PARK7</i>  | Ex3 duplication; het      | NA         | Pathogenic                | NA                                       | NA                          | NA                                     |
| Ind_109   | M   | 44     | Brazil    | <i>PARK7</i>  | Ex5 duplication; het      | NA         | Pathogenic                | NA                                       | NA                          | NA                                     |
| Ind_029   | F   | 69     | Brazil    | <i>TARDBP</i> | p.G75C; hom               | 29.1       | Likely Pathogenic         | Absent                                   | Absent                      | NA                                     |
| Ind_005   | F   | 24     | Peru      | <i>PINK1</i>  | p.N267S; het              | 15.75      | Likely Pathogenic         | 0.00008332                               | 0.00004831                  | Middle Eastern                         |
| Ind_081   | M   | 63     | Argentina | <i>SNCA</i>   | Ex2 deletion; het         | NA         | Pathogenic                | NA                                       | 0.00001435                  | Remaining                              |
| Ind_089   | M   | NA     | Brazil    | <i>SNCA</i>   | Ex1 duplication; het      | NA         | Pathogenic                | NA                                       | NA                          | NA                                     |
| Ind_261   | M   | 45     | Peru      | <i>SNCA</i>   | Entire duplication; het   | NA         | Pathogenic                | NA                                       | NA                          | NA                                     |
| Ind_302   | M   | 39     | Peru      | <i>PRKN</i>   | p.W445X; het              | 44         | Pathogenic                | 0                                        | 0.000003597                 | NFE                                    |
| Ind_270   | F   | 32     | Peru      | <i>PRKN</i>   | Ex2 deletion; het         | NA         | Pathogenic                | NA                                       | NA                          | NA                                     |
| Ind_108   | M   | 40     | Brazil    | <i>PRKN</i>   | p.G430_C436del; hom       | NA         | Likely Pathogenic         | Absent                                   | Absent                      | NA                                     |
|           |     |        |           | <i>PRKN</i>   | <b>c.872-1G&gt;C; het</b> | 33         | <b>Pathogenic</b>         | <b>0.00004472</b>                        | <b>0</b>                    | <b>Admixed American</b>                |
|           |     |        |           | <i>PRKN</i>   | <b>p.N52fs; het</b>       | NA         | <b>Pathogenic</b>         | <b>0.001383</b>                          | <b>0.0002941</b>            | <b>Admixed American</b>                |
| Ind_253   | F   | 29     | Peru      | <i>PRKN</i>   | p.R275W; het              | 34         | Pathogenic                | 0.001849                                 | 0.003732                    | NFE                                    |
|           |     |        |           | <i>PRKN</i>   | <b>c.619-1G&gt;A; het</b> | 35         | <b>Likely Pathogenic</b>  | <b>0.00002239</b>                        | <b>0</b>                    | <b>Admixed American</b>                |
| Ind_214   | F   | 26     | Colombia  | <i>PRKN</i>   | <b>p.C212Y; het</b>       | 33         | <b>Likely Pathogenic</b>  | <b>0.00006665</b>                        | <b>0.00000255</b>           | <b>Admixed American</b>                |
|           |     |        |           | <i>PRKN</i>   | Ex3 duplication; het      | NA         | Pathogenic                | NA                                       | NA                          | NA                                     |
| Ind_134   | M   | 14     | Peru      | <i>PRKN</i>   | <b>c.619-1G&gt;A; hom</b> | 35         | <b>Likely Pathogenic</b>  | <b>0.00002239</b>                        | <b>0</b>                    | <b>Admixed American</b>                |
| Ind_268   | M   | 20     | Peru      | <i>PRKN</i>   | <b>c.619-1G&gt;A; het</b> | 35         | <b>Likely Pathogenic</b>  | <b>0.00002239</b>                        | <b>0</b>                    | <b>Admixed American</b>                |
|           |     |        |           | <i>PRKN</i>   | Ex3 duplication; het      | NA         | Pathogenic                | NA                                       | NA                          | NA                                     |
| Ind_296   | M   | 13     | Peru      | <i>PRKN</i>   | <b>c.619-1G&gt;A; hom</b> | 35         | <b>Likely Pathogenic</b>  | <b>0.00002239</b>                        | <b>0</b>                    | <b>Admixed American</b>                |

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

| Sample ID | Sex | AAO, y | Country   | Gene         | Variant; status                         | CADD score | Pathogenicity consensus * | MAF Admixed American gnomAD\$ | MAF NFE gnomAD\$ | Max MAF Population gnomAD\$ |
|-----------|-----|--------|-----------|--------------|-----------------------------------------|------------|---------------------------|-------------------------------|------------------|-----------------------------|
| Ind_313   | F   | 48     | Peru      | <i>PRKN</i>  | c.619-1G>A; het                         | 35         | Likely Pathogenic         | 0.00002239                    | 0                | Admixed American            |
|           |     |        |           | <i>PRKN</i>  | p.K27X; het                             | 38         | Likely Pathogenic         | Absent                        | Absent           | NA                          |
| Ind_107   | M   | 12     | Uruguay   | <i>PRKN</i>  | p.W74fs; het                            | NA         | Pathogenic                | 0.00006675                    | 0.00003051       | Admixed American            |
|           |     |        |           | <i>PRKN</i>  | Ex3-Ex6 deletion; het <del>de</del>     | NA         | Pathogenic                | NA                            | NA               | NA                          |
| Ind_034   | M   | 55     | Brazil    | <i>PRKN</i>  | p.N52fs; het                            | NA         | Pathogenic                | 0.001383                      | 0.0002941        | Admixed American            |
|           |     |        |           | <i>PRKN</i>  | c.7+1G>T; het                           | 36         | Pathogenic                | Absent                        | Absent           | NA                          |
| Ind_126   | F   | 27     | Brazil    | <i>PRKN</i>  | p.N52fs; hom                            | NA         | Pathogenic                | 0.001383                      | 0.0002941        | Admixed American            |
| Ind_127   | M   | 33     | Brazil    | <i>PRKN</i>  | p.N52fs; hom                            | NA         | Pathogenic                | 0.001383                      | 0.0002941        | Admixed American            |
| Ind_269   | F   | 45     | Peru      | <i>PRKN</i>  | p.K27X; het                             | 38         | Likely Pathogenic         | Absent                        | Absent           | NA                          |
|           |     |        |           | <i>PRKN</i>  | Ex3-Ex6 deletion; het                   | NA         | Pathogenic                | NA                            | NA               | NA                          |
| Ind_182   | M   | NA     | Argentina | <i>PRKN</i>  | Ex8-Ex9 deletion; homo                  | NA         | Pathogenic                | NA                            | NA               | NA                          |
| Ind_276   | M   | 30     | Peru      | <i>PRKN</i>  | Ex8-Ex9 deletion; homo                  | NA         | Pathogenic                | NA                            | NA               | NA                          |
| Ind_314   | F   | 28     | Peru      | <i>PRKN</i>  | Ex8-Ex9 deletion; homo                  | NA         | Pathogenic                | NA                            | NA               | NA                          |
| Ind_311   | F   | 35     | Peru      | <i>PRKN</i>  | Ex6 duplication; homo                   | NA         | Pathogenic                | NA                            | NA               | NA                          |
| Ind_252   | F   | 8      | Peru      | <i>PRKN</i>  | Ex6 deletion; homo                      | NA         | Pathogenic                | NA                            | NA               | NA                          |
|           |     |        |           | <i>PRKN</i>  | Ex5 deletion; het                       | NA         | Pathogenic                | NA                            | NA               | NA                          |
| Ind_260   | M   | 22     | Peru      | <i>PRKN</i>  | Ex5 deletion; homo                      | NA         | Pathogenic                | NA                            | NA               | NA                          |
|           |     |        |           | <i>PRKN</i>  | Ex4+Ex6 deletion; het                   | NA         | Pathogenic                | NA                            | NA               | NA                          |
| Ind_250   | M   | 18     | Peru      | <i>PRKN</i>  | Ex3 duplication; het                    | NA         | Pathogenic                | NA                            | NA               | NA                          |
|           |     |        |           | <i>PRKN</i>  | Ex6 duplication; het                    | NA         | Pathogenic                | NA                            | NA               | NA                          |
| Ind_119   | F   | 31     | Brazil    | <i>PRKN</i>  | Ex1-Ex4 duplication; homo <del>de</del> | NA         | Pathogenic                | NA                            | NA               | NA                          |
| Ind_233   | M   | 29     | Peru      | <i>LRRK2</i> | p.R1441G; het                           | 25         | Pathogenic                | 0.00004473                    | 0.000002698      | Admixed American            |
| Ind_318   | F   | 48     | Brazil    | <i>LRRK2</i> | p.R1441C; het                           | 26.7       | Pathogenic                | 0                             | 0.0000195        | Middle Eastern              |
| Ind_168   | M   | 43     | Colombia  | <i>LRRK2</i> | p.E1948Q; het                           | 28.6       | Likely Pathogenic         | 0.00006861                    | 0                | Admixed American            |

| Sample ID | Sex | AAO, y | Country   | Gene          | Variant; status | CADD score | Pathogenicity consensus * | MAF Admixed American gnomAD <sup>§</sup> | MAF NFE gnomAD <sup>§</sup> | Max MAF Population gnomAD <sup>§</sup> |
|-----------|-----|--------|-----------|---------------|-----------------|------------|---------------------------|------------------------------------------|-----------------------------|----------------------------------------|
| Ind_051   | F   | 48     | Brazil    | <i>LRRK2</i>  | p.G2019S; het   | 31         | Pathogenic                | 0.0005168                                | 0.0002721                   | Ashkenazi                              |
| Ind_129   | M   | 41     | Brazil    | <i>LRRK2</i>  | p.G2019S; het   | 35         | Pathogenic                | 0.0005168                                | 0.0002721                   | Ashkenazi                              |
| Ind_180   | F   | 43     | Argentina | <i>LRRK2</i>  | p.G2019S; het   | 35         | Pathogenic                | 0.0005168                                | 0.0002721                   | Ashkenazi                              |
| Ind_191   | F   | NA     | Colombia  | <i>LRRK2</i>  | p.G2019S; het   | 35         | Pathogenic                | 0.0005168                                | 0.0002721                   | Ashkenazi                              |
| Ind_264   | F   | 53     | Peru      | <i>LRRK2</i>  | p.G2019S; het   | 35         | Pathogenic                | 0.0005168                                | 0.0002721                   | Ashkenazi                              |
| Ind_317   | M   | 55     | Brazil    | <i>LRRK2</i>  | p.G2019S; het   | 35         | Pathogenic                | 0.0005168                                | 0.0002721                   | Ashkenazi                              |
| Ind_245   | F   | 19     | Peru      | <i>GCHI</i>   | p.G203R; het    | 29.2       | Pathogenic                | 0                                        | 0.000001328                 | NFE                                    |
| Ind_016   | F   | 29     | Brazil    | <i>FBXO7</i>  | p.T22M; het     | 24.6       | Likely Pathogenic         | 0                                        | 0.000005566                 | Remaining                              |
|           |     |        |           | <i>FBXO7</i>  | p.R321X; het    | 40         | Pathogenic                | 0.00003336                               | 0.00003696                  | Remaining                              |
| Ind_170   | M   | 23     | Argentina | <i>PLA2G6</i> | p.S698F; het    | 24.4       | VUS                       | 0                                        | 0                           | Finnish                                |
|           |     |        |           | <i>PLA2G6</i> | p.Y643X; het    | 41         | Likely Pathogenic         | Absent                                   | Absent                      | NA                                     |

Variants are listed in chromosomal order.

\* Pathogenicity consensus based on a combination of the ClinVar variant classification,<sup>23</sup> the Human Genomics variant search engine Varsome,<sup>24</sup> and the Franklin by genoox database;

§ gnomAD was last accessed 11/13/2023; ‘Absent’ = variant has not been described in any individuals from the gnomAD cohort. Therefore, the ‘Max MAF in gnomAD’ = ‘NA’. ‘0’ = The variant has been described in the gnomAD cohort before, but not in Admixed Americans or NFEs. In this case, the MAF can be 0 and there is an ethnic group with a ‘Max MAF’.

& CNV validated using genome-wide array data.

AAO = age at onset; gnomAD = Genome Aggregation Database v4.0; MAF = minor allele frequency; Max = maximum; NA = not available; NFE = non-Finnish European. Variants that are more frequent in Latinos compared to other populations are highlighted in bold. Note: starting with gnomAD v4.0.0 the Latino population has changed its name to Admixed American population to avoid linking race and ethnicity with genetic ancestry inference.

**Table 2.**

Summary of *GBA1* variant carriers in the LARGE-PD familial cohort

| <b>2A. Pathogenic variants</b>     |                        |             |                          |                                          |                             |                                        |                                 |
|------------------------------------|------------------------|-------------|--------------------------|------------------------------------------|-----------------------------|----------------------------------------|---------------------------------|
| Legacy Name                        | Number of het carriers | CADD score  | Pathogenicity consensus* | MAF Admixed American gnomAD <sup>§</sup> | MAF NFE gnomAD <sup>§</sup> | Max MAF Population gnomAD <sup>§</sup> | PD risk assessment <sup>‡</sup> |
| p.V457D                            | 1                      | 25.9        | Likely pathogenic        | Absent                                   | Absent                      | NA                                     | VUS                             |
| p.L444P                            | 2                      | 23.6        | Pathogenic               | 8.34E-05                                 | 7.55E-05                    | Middle Eastern                         | Severe                          |
| p.N370S <sup>#</sup>               | 3                      | 22.7        | Pathogenic               | 0.00095                                  | 0.001728                    | Ashkenazi                              | Mild                            |
| p.M361I                            | 1                      | 19.31       | Likely pathogenic        | 0.0002                                   | 2.54E-06                    | African                                | VUS                             |
| p.R262H                            | 3                      | 15.74       | Likely pathogenic        | 3.33E-05                                 | 0.000226                    | NFE                                    | VUS                             |
| p.G202R <sup>†</sup>               | 1                      | 23.5        | Pathogenic               | 1.67E-05                                 | 3.05E-05                    | East Asian                             | Severe                          |
| <b>p.K198E</b>                     | <b>2</b>               | <b>24.1</b> | <b>Pathogenic</b>        | <b>0.000119</b>                          | <b>4.78E-06</b>             | <b>Admixed American</b>                | <b>Severe</b>                   |
| <b>p.W184R</b>                     | <b>1</b>               | <b>22</b>   | <b>Pathogenic</b>        | <b>3.33E-05</b>                          | <b>1.19E-05</b>             | <b>Admixed American</b>                | <b>Severe</b>                   |
| p.L160fs                           | 1                      | NA          | Likely pathogenic        | 0                                        | 5.48E-06                    | NFE                                    | NA                              |
| <b>2B. Non-Pathogenic variants</b> |                        |             |                          |                                          |                             |                                        |                                 |
| p.T369M                            | 4                      | 22.2        | Likely Benign            | 0.003149                                 | 0.008175                    | Amish                                  | Risk factor for PD              |
| p.E326K                            | 7                      | 17.33       | VUS                      | 0.002286                                 | 0.04325                     | Finnish                                | Risk factor for PD              |

Variants are listed in chromosomal order.

\* Pathogenicity consensus refers to Gaucher's disease and it is based on a combination of the ClinVar variant classification,<sup>23</sup> the Human Genomics variant search engine Varsome,<sup>24</sup> and the Franklin by genoox database.

<sup>‡</sup>PD risk assessment according to the *GBA1*-PD Browser (<https://pdgenetics.shinyapps.io/gba1browser/>);<sup>29</sup>

<sup>§</sup> gnomAD was last accessed 11/13/2023;

<sup>#</sup> two Peruvian individuals carried the *GBA1* N370S variant and two different compound heterozygous *PRKN* variants;

<sup>†</sup> this Brazilian individual also carried E326K;

gnomAD = Genome Aggregation Database v4.0; MAF = minor allele frequency; NFE = non-Finnish European. Variants that are more frequent in Latinos compared to other populations are highlighted in bold. Note: starting with gnomAD v4.0.0 the Latino population has changed its name to Admixed American population to avoid linking race and ethnicity with genetic ancestry inference.