

# Large-scale genetic characterization of Parkinson's disease in the African and African admixed populations

Fulya Akçimen,<sup>1</sup> Kimberly Paquette,<sup>2</sup> Peter Wild Crea,<sup>1,2</sup> Kathryn Step,<sup>3,4</sup> Emily Waldo,<sup>5</sup> Mathew J. Koretsky,<sup>2,6</sup> Paula Saffie-Awad,<sup>7,8</sup> Charles Achoru,<sup>9</sup> Funmilola Taiwo,<sup>10</sup> Simon Ozomma,<sup>11</sup> Gerald Onwuegbuzie,<sup>12</sup> Marzieh Khani,<sup>2</sup> Spencer Grant,<sup>1,2</sup> Lukman Owolabi,<sup>13</sup> Chiamaka Okereke,<sup>14</sup> Olajumoke Oshinaike,<sup>15</sup> Emmanuel Iwuzo,<sup>16</sup> Suleyman Can Akerman,<sup>17,18</sup> Paul Suhwan Lee,<sup>1</sup> Shyngle Oyakhire,<sup>19</sup> Nosakhare Osemwegie,<sup>20</sup> Kensuke Daida,<sup>1,2,21</sup> Sani Abubakar,<sup>22</sup> Adedunni Olusanya,<sup>23,24</sup> Mariam Isayan,<sup>25</sup> Christiane Alvarez,<sup>2</sup> Rami Traurig,<sup>2</sup> Adebimpe Ogunmodede,<sup>26</sup> Sarah Samuel,<sup>27</sup> Mary B. Makarious,<sup>2,6</sup> Fadimatu Sa'ad,<sup>28</sup> Rashidat Olanigan,<sup>29</sup> Kristin Levine,<sup>2,6</sup> Ewere Marie Ogbimi,<sup>30</sup> Dan Vitale,<sup>2,6</sup> Francis Odiase,<sup>31</sup> Francis Ojini,<sup>22,32</sup> Olanike Odeniyi,<sup>33</sup> Zih-Hua Fang,<sup>34</sup> Nkechi Obianozie,<sup>35</sup> Deborah A. Hall,<sup>36</sup> Ernest Nwazor,<sup>37</sup> Tao Xie,<sup>38</sup> Francesca Nwaokorie,<sup>23</sup> Mahesh Padmanaban,<sup>38</sup> Paul Nwani,<sup>39</sup> Ejaz A. Shamim,<sup>40,41,42</sup> Alero Nnama,<sup>20</sup> David Standaert,<sup>43</sup> Morenikeji Komolafe,<sup>44</sup> Marissa Dean,<sup>43</sup> Godwin Osaigbovo,<sup>9</sup> Elizabeth Disbrow,<sup>45</sup> Ismaila Ishola,<sup>31</sup> Ashley Rawls,<sup>46</sup> Frank Imarhiagbe,<sup>30</sup> Shivika Chandra,<sup>47</sup> Cyril Eramah,<sup>48</sup> Vanessa Hinson,<sup>49</sup> Naomi Louie,<sup>50</sup> Ahmed Idowu,<sup>51</sup> J Solle,<sup>50</sup> Scott A. Norris,<sup>52</sup> Abdullahi Ibrahim,<sup>53</sup> Camilla Kilbane,<sup>54</sup> Gauthaman Sukumar,<sup>55,56</sup> Lisa M. Shulman,<sup>57</sup> Daniel Ezuduemoih,<sup>32</sup> Julia Staisch,<sup>58</sup> Sarah Breaux,<sup>58</sup> Clifton Dalgard,<sup>59,60</sup> Erin R. Foster,<sup>52</sup> Abiodun Bello,<sup>61</sup> Andrew Ameri,<sup>49</sup> Raquel Real,<sup>62,63</sup> Erica Ikweni,<sup>32</sup> Huw R Morris,<sup>62,64,65</sup> Roosevelt Anyanwu,<sup>23</sup> Erin Furr Stimming,<sup>47</sup> Kimberley Billingsley,<sup>2</sup> Wemimo Alaofin,<sup>66</sup> Pilar Alvarez Jerez,<sup>1,62</sup> Osigwe Agabi,<sup>23,32</sup> Dena G. Hernandez,<sup>1</sup> Rufus Akinyemi,<sup>67</sup> Sampath Arepalli,<sup>1</sup> Laksh Malik,<sup>2</sup> Raymond Owolabi,<sup>26</sup> Yakub Nyandaiti,<sup>27</sup> Hampton L. Leonard,<sup>2,6</sup> Kolawole Wahab,<sup>66</sup> Oladunni Abiodun,<sup>68</sup> Carlos F. Hernandez,<sup>69</sup> Fatima Abdulai,<sup>35</sup> Hirotaka Iwaki,<sup>2,6</sup> Soraya Bardien,<sup>3</sup> Christine Klein,<sup>70</sup> John Hardy,<sup>71</sup> Henry Houlden,<sup>72</sup> Kamalini Ghosh Galvelis,<sup>73</sup> Mike A. Nalls,<sup>2,6</sup> Nabila Dahodwala,<sup>74</sup> Whitley Aamodt,<sup>74</sup> Emily Hill,<sup>75</sup> Alberto Espay,<sup>75</sup> Stewart Factor,<sup>76</sup> Chantale Branson,<sup>77</sup> Cornelis Blauwendraat,<sup>1,2</sup> Andrew B. Singleton,<sup>1,2</sup> Oluwadamilola Ojo,<sup>23,32</sup> Lana M. Chahine,<sup>78</sup> the Black and African American Connections to Parkinson's Disease Study (BLAAC PD), the Nigeria Parkinson's Disease

Research Network (NPDRN), the Racial Disparities in Parkinson Disease (RaD-PD), the Global  
Parkinson's Genetics Program (GP2), Njideka Okubadejo<sup>23,32,†</sup> and Sara Bandres-Ciga<sup>2,†</sup>

<sup>†</sup>These authors contributed equally to this work.

## Abstract

Elucidating the genetic contributions to Parkinson's disease (PD) etiology across diverse  
ancestries is a critical priority for the development of targeted therapies in a global context. We  
conducted the largest sequencing characterization of potentially disease-causing, protein-altering  
and splicing mutations in 710 cases and 11,827 controls from genetically predicted African or  
African admixed ancestries. We explored copy number variants (CNVs) and runs of  
homozygosity (ROHs) in prioritized early onset and familial cases. Our study identified rare  
*GBA1* coding variants to be the most frequent mutations among PD patients, with a frequency of  
4% in our case cohort.

Out of the 18 *GBA1* variants identified, ten were previously classified as pathogenic or likely  
pathogenic, four were novel, and four were reported as of uncertain clinical significance. The  
most common known disease-associated *GBA1* variants in the Ashkenazi Jewish and European  
populations, p.Asn409Ser, p.Leu483Pro, p.Thr408Met, and p.Glu365Lys, were not identified  
among the screened PD cases of African and African admixed ancestry. Similarly, the European  
and Asian *LRRK2* disease-causing mutational spectrum, including *LRRK2* p.Gly2019Ser and  
p.Gly2385Arg genetic risk factors, did not appear to play a major role in PD etiology among  
West African-ancestry populations. However, we found three heterozygous novel missense  
*LRRK2* variants of uncertain significance overrepresented in cases, two of which — p.Glu268Ala  
and p.Arg1538Cys — had a higher prevalence in the African ancestry population reference  
datasets. Structural variant analyses revealed the presence of *PRKN* CNVs with a frequency of  
0.7% in African and African admixed cases, with 66% of CNVs detected being compound  
heterozygous or homozygous in early-onset cases, providing further insights into the genetic  
underpinnings in early-onset juvenile PD in these populations. Short tandem repeat analysis also  
identified *ATXN3* repeat expansions within the pathogenic range ( $CAG_n > 45$ ) in three PD  
patients of African ancestry. Novel genetic variation overrepresented in cases versus controls

among screened genes warrants further replication and functional prioritization to unravel their pathogenic potential.

Here, we created the most comprehensive genetic catalog of both known and novel coding and splicing variants potentially linked to PD etiology in an underserved population and further conducted global and local ancestry analyses to further explore population-specific effects. Our study has the potential to guide the development of targeted therapies in the emerging era of precision medicine. By expanding genetics research to involve underrepresented populations, we hope that future PD treatments are not only effective but also inclusive, addressing the needs of diverse ancestral groups.

#### **Author affiliations:**

1 Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD 20892, USA

2 Center for Alzheimer's and Related Dementias, National Institute on Aging and National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD 20892, USA

3 Department of Biomedical Sciences, Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa

4 South African Medical Research Council Centre for Tuberculosis Research, Stellenbosch University, Cape Town, South Africa

5 Genomic Medicine, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH 44106, United States

6 DataTecnica LLC, Washington, DC, 20037, USA

7 Programa de Pós-Graduação em Ciências Médicas, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil

8 Clínica Santa María, Santiago, Chile

9 Jos University Teaching Hospital, Jos, Plateau State, Nigeria

- 1 10 University College Hospital, Ibadan, Oyo State, Nigeria
- 2 11 University of Calabar Teaching Hospital, Calabar, Cross River State, Nigeria
- 3 12 University of Abuja, Gwagwalada, Federal Capital Territory, Nigeria
- 4 13 Bayero University, Kano, Kano State, Nigeria
- 5 14 University of Nigeria Teaching Hospital, Ituku-Ozalla, Enugu State, Nigeria
- 6 15 Lagos State University College Of Medicine, Ikeja, Lagos State, Nigeria
- 7 16 Benue State University, Makurdi, Benue State, Nigeria
- 8 17 Brain Science Institute, Johns Hopkins University School of Medicine, Baltimore, MD 21205,
- 9 USA
- 10 18 Department of Neurology, Johns Hopkins University School of Medicine, Baltimore, MD
- 11 21205, USA
- 12 19 National Hospital Abuja, Federal Capital Territory, Nigeria
- 13 20 University of Port Harcourt Teaching Hospital, Rivers State, Nigeria
- 14 21 Department of Neurology, Faculty of Medicine, Juntendo University, Tokyo, Japan
- 15 22 Ahmadu Bello University, Zaria, Kaduna State, Nigeria
- 16 23 College of Medicine University of Lagos, Idi-araba, Lagos State, Nigeria
- 17 24 R-Jolad Hospital, Gbagada, Lagos, Nigeria
- 18 25 Department of Neurology and Neurosurgery, National Institute of Health, Yerevan, Armenia
- 19 26 Federal Medical Center, Owo, Ondo State, Nigeria
- 20 27 University of Maiduguri Teaching Hospital, Maiduguri, Borno State, Nigeria
- 21 28 Federal Teaching Hospital, Gombe, Gombe State, Nigeria
- 22 29 Lagos State University Teaching Hospital, Ikeja, Lagos State, Nigeria
- 23 30 Delta State University Abraka, Delta State, Nigeria
- 24 31 University of Benin, Benin City, Edo State, Nigeria
- 25 32 Lagos University Teaching Hospital, Idi-araba, Lagos State, Nigeria

- 1 33 General Hospital, Lagos Island, Lagos State, Nigeria
- 2 34 German Center for Neurodegenerative Diseases, DZNE, Tübingen, Germany
- 3 35 University of Abuja Teaching Hospital, Gwagwalada, Federal Capital Territory, Nigeria
- 4 36 Department of Neurological Sciences, Rush University Medical Center, Chicago, IL 60612,  
5 USA
- 6 37 Rivers State University Teaching Hospital, Port Harcourt, Rivers State, Nigeria
- 7 38 University of Chicago Medicine, Department of Neurology, Chicago, IL 60637, USA
- 8 39 Nnamdi Azikiwe University Teaching Hospital, Nnewi, Anambra State, Nigeria
- 9 40 National Institute of Neurological Disorders and Stroke, Bethesda, MD 20892, USA
- 10 41 Mid-Atlantic Permanente Medical Group, Department of Neurology, Largo, MD 20774, USA
- 11 42 Kaiser Permanente, MidAtlantic Permanente Research Institute, Washington, DC 20002,  
12 USA
- 13 43 University of Alabama at Birmingham, Department of Neurology, Birmingham, AL 35294,  
14 USA
- 15 44 Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria
- 16 45 Department of Neurology, LSU Health Shreveport, LSU Health Shreveport Center for Brain  
17 Health, Shreveport, LA 71103, USA
- 18 46 University of Florida Norman Fixel Institute for Neurological Diseases, Neurology  
19 Movement Disorders, Gainesville, FL 32608, USA
- 20 47 The University of Texas Health Science Center at Houston, Houston, TX 20036, USA
- 21 48 Irrua Specialist Teaching Hospital, Irrua, Edo State, Nigeria
- 22 49 Medical University of South Carolina, Charleston, SC 29425, USA
- 23 50 Michael J Fox Foundation for Parkinson's Research, Department of Clinical Research, New  
24 York, NY 10163, USA
- 25 51 Obafemi Awolowo University Teaching Hospitals Complex, Ile-Ife, Osun State, Nigeria
- 26 52 Washington University in St Louis, St Louis, MO 63130, USA

- 1 53 Federal University of Health Sciences Teaching Hospital, Azare, Bauchi State, Nigeria
- 2 54 University Hospital in Cleveland Medical Center/Case Western Reserve University (UH), OH
- 3 44106, USA
- 4 55 Department of Anatomy, Physiology and Genetics, School of Medicine, Uniformed Services,
- 5 Bethesda, MD 20814, USA
- 6 56 University of the Health Sciences, 4301 Jones Bridge Rd, Bethesda, MD, 20814, USA
- 7 57 University of Maryland, Baltimore, MD 21201, USA
- 8 58 Ochsner Clinic Foundation, New Orleans, LA 70124, USA
- 9 59 Henry M Jackson Foundation for the Advancement of Military Medicine, Uniformed Services
- 10 University of the Health Sciences, Bethesda, MD 20814, USA
- 11 60 The American Genome Center, Collaborative Health Initiative Research Program, Uniformed
- 12 Services University of the Health Sciences, Bethesda, MD 20814, USA
- 13 61 University of Ilorin Teaching Hospital, Ilorin, Kwara State, Nigeria
- 14 62 Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of
- 15 Neurology, London WC1N 3BG, UK
- 16 63 UCL Movement Disorders Centre, University College London, London WC1N 3BG, UK
- 17 64 National Hospital for Neurology and Neurosurgery, London, WC1N 3BG, UK
- 18 65 Department of Neurology, Royal Free Hospital, London, NW3 2QG, UK
- 19 66 University of Ilorin, Ilorin, Kwara State, Nigeria
- 20 67 Neuroscience and Ageing Research Unit, Institute for Advanced Medical Research and
- 21 Training, College of Medicine, University of Ibadan, Ibadan, Oyo State, Nigeria
- 22 68 General Hospital, Isolo, Lagos State, Nigeria
- 23 69 Universidad del Desarrollo, Centro de Genética y Genómica, Facultad de Medicina Clínica
- 24 Alemana, Santiago 7610658, Chile
- 25 70 Institute of Neurogenetics and Department of Neurology, University of Lübeck and
- 26 University Hospital Schleswig-Holstein, Lübeck, Germany

71 Reta Lila Weston Institute, University College London Institute of Neurology, Queen Square,  
London, WC1N 1PJ, UK

72 Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, London,  
WC1N 3BG, UK

73 Parkinson's Foundation, New York, NY 10018, USA

74 University of Pennsylvania, Philadelphia, PA 19104, USA

75 University of Cincinnati, Cincinnati, OH 45221, USA

76 Emory University, Atlanta, GA 30322, USA

77 Morehouse College, Atlanta, GA 30314, USA

78 University of Pittsburgh, 3471 Fifth Avenue, Pittsburgh, PA 15213, USA

Correspondence to: Njideka Okubadejo

College of Medicine, University of Lagos

Idi Araba, Lagos State, 12003, Nigeria

E-mail: nokubadejo@unilag.edu.ng

Correspondence may also be addressed to: Sara Bandres-Ciga

Center for Alzheimer's Disease and Related Dementias (CARD), National Institute on Aging  
(NIA), National Institutes of Health (NIH)

9000 Rockville Pike, NIH Building T44

Bethesda, MD 20892, USA

E-mail: sara.bandresciga@nih.gov

**Running title:** Parkinson's disease genetics in African ancestry

**Keywords:** genetics; Parkinson's disease; Black and African American population; African ancestry; disease-causing mutations

## Introduction

Parkinson's disease (PD) is a multifaceted neurodegenerative disorder influenced by genetics, environment, and other factors. The influence of genetics has been primarily studied in European populations, limiting our understanding of the genetic landscape of PD in the African and African admixed populations<sup>1</sup>. Bridging this gap is essential for advancing equitable precision medicine interventions and developing universally effective strategies for the prevention and treatment of PD.

Genetic risk factors and their impact on PD in individuals of African and African admixed genetic ancestry remain largely unknown. Around 77% of African genetic studies in PD have been conducted in North Africa (mainly Tunisia) and South Africa<sup>2</sup>. In contrast, there have been limited PD genetic studies in populations originating from Central, Eastern, and the French speaking West coast of Africa. Similarly, the genetics of African admixed individuals have been poorly studied in the context of PD genetics, with few published studies<sup>3-5</sup>. As a result, our understanding of the role of genetics in disease etiology among African admixed individuals lags behind that of European ancestry individuals. Further research with more diverse representation and standardization is crucial for drawing accurate conclusions on a global scale.

The first Genome-Wide Association Study (GWAS) exploring PD genetic risk in African and African admixed populations nominated a novel non-coding PD-associated variant (rs3115534) in *GBA1*<sup>6</sup>. This variant is present in 33% and 22% of the African and African admixed PD cohorts respectively<sup>6</sup> while it is almost completely absent in predominantly European ancestry cohorts. The rs3115534 intronic variant acts by interfering with the splicing of functional *GBA1* transcripts, resulting in reduced glucocerebrosidase activity. This represents a novel mechanism of *GBA1*-derived PD risk and an attractive candidate for precision-based therapeutics in a remarkably underserved population<sup>7</sup>, reinforcing the idea that distinct genetic architectures

1 contribute to disease susceptibility and in turn could inform optimized treatment options in the  
 2 future. Other *GBA1* coding variants, like p.Asn409Ser and p.Leu483Pro, are associated with PD  
 3 risk and commonly observed in European and Ashkenazi Jewish ancestry populations. These  
 4 same variants are notably rare among individuals of African and African admixed ancestries<sup>8,9</sup>.

5  
 6 To gain a deeper understanding of the genetic architecture of PD in underserved African  
 7 populations, we leveraged data from the Black African and African American Connections to  
 8 Parkinson's Disease Study (BLAAC PD)<sup>10</sup> as a part of the Global Parkinson's Genetics Program  
 9 (GP2)<sup>11</sup>, the Nigerian Parkinson's Disease Research Network (NPDRN)<sup>12</sup>, the PDGENeration  
 10 (PDGENE) initiative<sup>13</sup>, All of Us<sup>14</sup>, and the UK Biobank (UKB)<sup>15</sup>. Utilizing data from 710  
 11 cases and 11,827 controls, we conducted the largest sequencing characterization of potentially  
 12 disease-causing, protein-altering, and splicing mutations across 51 genes in African and African  
 13 admixed ancestry populations. We focused on known PD-associated genes linked to autosomal  
 14 dominant inheritance (*SNCA*, *LRRK2* and *VPS35*), autosomal recessive inheritance (*PRKN*,  
 15 *PINK1*, *DJ-1*, and *VPS13C*), and increased risk (*LRRK2*, *GBA1*), as well as other PD-associated  
 16 genes such as *RAB32*, *PSMF1*, and *ITSN1* from literature. Furthermore, we assessed putative  
 17 pleiotropic effects by examining genes that have been shown to manifest as parkinsonism<sup>16</sup> (i.e.  
 18 *ATP13A2*, *FBXO7*, *PLA2G6*, *SYNJ1*, *GCH1*, *TRPM7*, *GCH1*, *KCDT7*, *XPRI*, and *TOR1A*) or  
 19 other neurodegenerative conditions (i.e. *APP*, *MAPT*, and *DCTN1*). We leveraged short-read  
 20 sequencing datasets (whole-genome, whole-exome, and targeted-exome sequencing) to  
 21 comprehensively unravel the genetic architecture of rare variants contributing to PD in these  
 22 populations and further explored potential population-specific effects through global and local  
 23 ancestry analyses. In addition, we examined short tandem repeats (STRs) previously reported to  
 24 be associated with PD risk, including *ATXN2*, *ATXN3*, and *TBP*<sup>17,18,19</sup>. We further investigated  
 25 potential recessive variants by exploring runs of homozygosity (ROHs) enriched in the patient  
 26 population. We assessed potential copy number variation (CNVs) by investigating B allele  
 27 frequency (BAF) and Log2 Ratio (L2R) from genotyping data along with Multiplex ligand-  
 28 dependent probe amplification (MLPA), focusing on cases with positive family history and early  
 29 age at onset.

# Materials and methods

## Participants and study design

We obtained genetic data from All of Us, GP2-BLAAC PD, the NPDRN, PDGENE, and UKB (Supplementary Table 1). In total, the data comprised 710 cases and 11,827 controls of African or African admixed ancestry or Black or African American self-reported race. Please see Supplementary Table 1 for demographic characteristics of cohorts under study and see Figure 1 for a summary of our workflow, explained in further detail below.

BLAAC PD is a multicenter study in North America recruiting individuals who self-identify as Black and/or African American. In the GP2-BLAAC PD dataset, cases were defined as individuals aged 18 or older in the United States with a clinician-confirmed PD diagnosis (n=186), whereas controls were neurologically healthy individuals with no personal or family history of any neurodegenerative disease (n=236). Institutional review board approval was implemented at each study site, with written informed consent from each participant<sup>20</sup>. Assessments of cognitive status and clinical features were conducted and reported by respective clinicians. Cognitive status was determined by the cognition component of the Clinical Impression of Severity Index for PD (CISI-PD) with scores ranging from 0 to 6: 0 (normal), 1 (slowness and/or minimal cognitive problems), 2 (mild cognitive problems; no limitations), 3 (mild to moderate cognitive problems; does not need help for basic activities of daily living), 4 (moderate cognitive problems; help is required for some basic activities of daily living), 5 (severe cognitive problems; help is required for most or all basic activities of daily living), and 6 (severely disabled; helpless; complete assistance needed)<sup>21</sup>. Clinical symptoms such as bradykinesia, rest tremor, rigidity, and use/response to levodopa was assessed through the International Parkinson and Movement Disorder Society (MDS) Clinical Diagnostic Criteria for idiopathic PD (MDS-PD criteria)<sup>22</sup>.

For the NPDRN cohort (cases =108; controls=60), PD diagnosis was based on the United Kingdom PD Society Brain Bank criteria (excluding the requirement of not more than one

affected relative)<sup>12,23</sup>. Study assessments include the Movement Disorders Society Unified Parkinson's Disease Rating Scale (MDS-UPDRS), a one question, self-reported olfactory assessment, and a neurological examination, performed by a study neurologist. Controls were excluded if any neurological signs were present<sup>12</sup>. All participants provided written informed consent.

In addition, we mined existing data from PDGENE, a multi-center, observational study (NCT04057794, NCT04994015) offering CLIA-certified genetic testing and genetic counseling to PD patients in North America<sup>13</sup>. We included targeted exome sequencing data from 138 PD cases with genetically predicted African or African admixed ancestry and 145 self-reported Black or African American PD cases from the GP2 release 8. The study was approved by IRBs, and the Scientific Review and Executive Committees of the Parkinson Study Group. All participants signed informed consent forms.

In the All of Us dataset, we created a case-control cohort comprising 93 PD cases and 10,360 controls of African or African admixed ancestry. Controls were individuals aged 60 years or older with no family history of neurological disease and no present neurological condition in their electronic health records. All participants provided written informed consent. The All of Us IRB reviews the protocol, informed consent, and other materials for participants.

PD cases of genetically predicted African ancestry in the UKB (n=40) were defined by the UKB field ID 42032, using diagnoses according to the UKB's algorithmically defined outcomes v2.0 (<https://biobank.ndph.ox.ac.uk/ukb/refer.cgi?id=460>). The control cohort (n=1,171) includes individuals aged 60 years or older without personal or family history of any neurological disorders.

Sequencing, genotyping procedures, and quality assessment of 710 cases and 11,827 controls are detailed in Supplementary Methods.

## Genetic characterization of potentially rare disease-causing variants

We performed functional annotation of identified variants using ANNOVAR<sup>24</sup>, incorporating relevant datasets and prediction tools, and gnomAD v4.1. To distinguish *GBA1* variation from *GBAP1* pseudogene mutations, we implemented the Gauchian targeted caller<sup>25</sup> on whole genome sequencing (WGS) data to detect known *GBAP1* variation within the exons 9–11 homology region. The clinical significance of candidate variants was cross-checked on the ClinVar. Genes implicated in parkinsonism and related neurodegenerative diseases were annotated using the OMIM genemap2 file (downloaded from omim.org in September 2024). We focused on missense, loss-of-function variants including stopgain, stop loss, frameshift, start loss, splice site changes, and short insertions and deletions, with a minor allele frequency below 0.01 and a minimum Combined Annotation Dependent Depletion (CADD) PHRED-scaled score of 12.37. This threshold is predicted to be pathogenic by Amendola et al.<sup>26</sup> and falls within the top 5–6% most deleterious variants in the genome, according to Kircher et al.<sup>27</sup>. We assigned a confidence level for PD genes using the criteria described by Blauwendraat, et al.<sup>28</sup>.

## Results

### Baseline characteristics

An overview of our study is illustrated in Figure 1. Demographic and clinical characteristics of five datasets including 710 cases and 11,827 neurological healthy controls are summarized in Supplementary Table 1. All participants were of African and African admixed ancestry based on genetic prediction, except for the PDGENE dataset, where 138 participants genotyped on the NeuroBooster array were confirmed as African or African admixed, and 145 participants were included based on self-reported Black or African American race (Supplementary Figure 1).

## Genetic findings

We identified 443 rare variants, including 105 in eight PD genes and 338 in genes associated with parkinsonism or other neurodegenerative diseases (Supplementary Table 2). A total of 19 variants were previously reported as pathogenic or likely pathogenic in ClinVar, while 182 variants were novel. A total of 242 variants were reported as benign, likely benign or of uncertain clinical significance. The ClinVar dataset is largely based on mutation reports from European populations and so is likely to have incomplete information on pathogenicity in African individuals. Variants were prioritized based on clinical relevance, zygosity, and the presence of an early disease onset in cases involving recessive PD genes. This process identified 84 variants across 14% of the cases (n=97) that had no prior classification as benign (Supplementary Table 3).

The average age at onset (AAO) for cases in our study ranged from  $56.34 \pm 11.90$  years (NPDRN) to  $66.79 \pm 9.55$  years (UKB). Among carriers of prioritized variants, the AAO was notably younger, at  $52.41 \pm 13.65$  years. Furthermore, carriers with a family history of PD had an earlier AAO of  $44.43 \pm 13.77$  years, compared to  $53.64 \pm 13.73$  years for those without a family history. The distribution of variants in PD genes across the five datasets, excluding those previously reported as benign or likely benign, is presented in Figure 2.

Ten variants in *GBA1*, one in *PINK1*, two in *PRKN*, one in *FBXO7*, and one in *PLA2G6* were previously described as pathogenic or likely pathogenic (Figure 3, Supplementary Table 3). A total of 41 novel variants were identified in PD genes, including four in *GBA1*, seven in *PINK1*, two in *DNAJC6*, four in *DJ-1*, one in *SNCA*, five in *PRKN*, three in *LRRK2*, twelve in *VPS13C*, two in *VPS35*, and one in *SYNJ1* (Figure 3, Table 2, Supplementary Table 3). Notably, the *GBA1* p.His413Arg, *SNCA* p.Met116Ile, and *PRKN* p.Gly139Valfs\*38 have previously been reported only in individuals of African ancestry in the gnomAD and All of Us datasets (Table 1).

We obtained the pathogenicity scores of the candidate variants from CADD, AlphaMissense, and PrimateAI. A score  $>12.36$  for CADD,  $>0.564$  for AlphaMissense, and  $>0.803$  for PrimateAI are

recommended as pathogenic cutoffs. None of the previously reported pathogenic variants in known genes fulfilled the pathogenicity criteria for AlphaMissense and PrimateAI. Therefore, although we included scores from these tools, we did not apply any filtering based on the recommended thresholds. All identified variants in recessive PD genes were present in single heterozygous states (Figure 4, Figure 5, Supplementary Table 3). A female PD case from the PDGENE study with an AAO under 30 was identified as a carrier of two heterozygous variants in *PRKN* (p.Gly139Valfs\*38 and p.Pro113Thrfs\*51) that would suggest a potential compound heterozygous state (Table 2, Supplementary Table 4). We were unable to confirm compound heterozygosity due to the absence of phasing data. A distribution of AAO among prioritized variant carriers in PD genes (*PRKN*, *DJ-1*, *PINK1*, *VPS13C*, *DNAJC6*, *GBA1*, *SNCA*, *LRRK2*, *TRPM7*, and *VPS35*), is shown in Figure 4.

### **Rare *GBA1* variant landscape differs African and African admixed ancestry patients**

We identified a total of 18 heterozygous *GBA1* variants among 26 cases (4%), including ten previously reported as pathogenic or likely pathogenic, four novel variants, and four variants of uncertain clinical significance (Figure 2, Figure 3, Supplementary Figure 2, Table 1, Table 2, Supplementary Table 4). To identify potential known variants in the pseudogene paralog *GBAP1* exons 9-11 homology region, we screened 15 WGS samples. The Gaussian algorithm predicted four carrier mutations, including variants p.Asp448His, c.1263del and c.1263del+RecTL, i.e. deletion combined with pseudogene variant in exon 9 and three in exon 10. The algorithm predicted no structural variants: a total of four copies per *GBA1* and *GBAP1* variant in each sample were identified. Overall, four variants, including p.Gly416Ser, p.Trp223Arg, p.Arg170Leu, and p.Thr75del in *GBA1* were targeted outside the homology region.

The previously reported pathogenic p.Arg170Leu (rs80356763) variant was identified in a family with two early-onset female PD siblings of African genetic ancestry, with ages at onset in their early 30s and 40s. The p.Thr75del (rs761621516) variant was identified in two cases from PDGENE and one case from the GP2-BLAAC PD dataset with no familial history. The carriers

from the PDGENE were two Black or African/African American male patients (one with confirmed African genetic ancestry), who had onset in their 50s and early 70s. The carrier from the GP2-BLAAC PD dataset was a female patient with a disease onset in her late 50s. Both of these variants appear to be specific to African and African admixed ancestries, with no records in other ancestries according to gnomAD v4.1.0 and All of Us release 7 (Table 1, Supplementary Table 3).

Four additional known pathogenic *GBA1* variants, including p.Arg535His (1:155235002:C:T, rs75822236), p.Asp448His (1:155235727:C:G, rs1064651), p.Arg502Cys (1:155235196:G:A, rs80356771), and p.Asn227Ser (1:155238215:T:C, rs364897), were identified in patients from the GP2- BLAAC PD release 8 data (Table 1). The p.Arg535His variant was identified in a female individual of African ancestry with disease onset in her 70s, whereas the p.Asp448His and p.Arg502Cys variants were identified in female individuals of African admixed ancestry with onset in their 50s. The p.Asn227Ser variant was identified in a male patient of African ancestry with onset in his early 50s.

Two previously reported pathogenic variants, p.Ala495Pro (1:155235217:C:G, rs368060) and p.Gly416Ser (1:155235823:C:T, rs121908311), were found in two female patients in the All of Us dataset (Table 1). Both patients were of African ancestry and diagnosed in their 50s. Additionally, the known missense p.Trp223Arg (1:155238228:A:G, rs61748906) variant was identified in a male patient of African genetic ancestry from the NPDRN cohort. He presented with initial symptoms in his early 60s and reported no family history. The stopgain p.Arg202Ter (1:155238291:G:A, rs1009850780) variant was found in a female patient from the PDGENE dataset, with onset in her late 50s (Table 1).

Four novel variants in *GBA1* were identified (Table 2). Of note, none of these were found to be present in the *GBA1*-PD browser (<https://pdgenetics.shinyapps.io/gba1browser/>, accessed November 1, 2024). The missense *GBA1* p.His413Arg variant (1:155235831:T:C, rs911331923, CADD=23.9) was identified in a male case from the GP2-BLAAC PD cohort, with disease onset

in his 40s. This variant is specific to African/African American ancestry, reported only in individuals of these ancestries in the gnomAD v4.1.0 (allele count=3) and All of Us release 7 (allele count=7) browsers, respectively. The histidine to proline change at this position was previously reported as likely pathogenic, whereas the p.His413Arg variant identified in the current study has not been reported before. The second novel missense *GBA1* p.Ile407Ser variant (1:155236249:A:C, rs1057519358, CADD=26.5) was identified in a male patient of African admixed genetic ancestry from the GP2-BLAAC PD cohort with disease onset in his early 50s. While *GBA1* p.Ile407Thr was previously reported with an uncertain clinical significance, p.Ile407Ser has not been reported before. The third novel variant identified in this study was *GBA1* p.Gly364Arg (1:155236379:C:G, CADD=20.2) in a male patient with slowness and/or minimal cognitive problems from the GP2-BLAAC PD dataset, who had his initial symptoms in his late 50s. While a cysteine to thymine substitution at this position was previously reported, this specific variant is novel and absent in gnomAD v4.1.0. Lastly, a frameshift variant, *GBA1* p.Ala300Profs\*4 (1:155237441:GC:G, CADD=28.7), was identified in an early-onset female case of African genetic ancestry from the GP2-BLAAC PD dataset, who presented with initial PD symptoms in her early 40s and had a second-degree relative affected by the disease.

The *GBA1* p.Met400Ile variant (1:155236269:C:T, rs149487315), previously classified as having uncertain clinical significance, was the most prevalent variant among the screened PD genes, with a carrier frequency of 0.84% across all cases and an allele frequency of 0.0028 in the gnomAD African populations. This variant is more prevalent in individuals of African ancestry, with an allele count rate of 91% in the All of Us dataset (release 7) (Supplementary Table 4).

Although the focus of this study was not to screen for common or non-coding variants, we screened the recently identified genetic risk factor linked to PD risk in the African and African American populations (rs3115534-G). Our study identified 3,324 *GBA1*-rs3115534-G heterozygous (3,207 in African and 117 in African admixed) and 320 homozygous carriers (311 in African and nine in African admixed), with a risk allele frequency of 0.23 in cases and 0.17 in controls in African ancestry and a frequency of 0.15 in cases and 0.14 in controls in African

admixed individuals. The frequency of the G allele in each dataset tested is presented in Supplementary Table 5.

## European and Asian *LRRK2* variants are not major factors in West African or African admixed patients

The current study did not identify any pathogenic or likely pathogenic *LRRK2* variants previously reported in Europeans (p.Asn1437His, p.Arg1441Ser, p.Arg1441Gly, p.Arg1441Cys, p.Arg1441His, p.Val1447Met, p.Tyr1699Cys, p.Phe1700Leu, p.Ile2020Leu, and p.Ile2020Thr) or Asian ancestries (p.Gly2385Arg). Three novel heterozygous missense *LRRK2* variants of uncertain significance were found in the All of Us and PDGENE datasets (Table 2). A missense p.Glu268Ala substitution (12:40243646:A:C, rs373254349, CADD=24.6) was identified in a female patient with disease onset in her early 60s. A p.Ile1438Lys variant (12:40309228:A:T, CADD=24.2) was found in a PDGENE participant over 60 years old (age at onset was unavailable) with a family history of PD. The third missense variant, *LRRK2* p.Arg1538Cys (12:40314047:C:T, rs150620977, CADD=29.2), was detected in a female patient from All of Us, with disease onset in her early 40s and a disease duration of 15 years. The *LRRK2* p.Glu268Ala and p.Arg1538Cys variants were more prevalent in individuals of African ancestry in the All of Us dataset (release 7), with allele count rates of 93.75% African (6.25% European) and 87.5% African (8.3% American, 1.40% European, 2.78% other) respectively, across the entire dataset. The *LRRK2* p.Ile1438Lys variant was absent in both gnomAD v4.1.0 and All of Us.

## Novel variation in Parkinson's disease genes warrants further replication and functional validation

A p.Met116Ile missense variant (4:89729236:C:A, rs1378041201, CADD=18.9) in *SNCA* was identified in a male patient of African ancestry from the PDGENE dataset, with disease onset in his late 50s and no family history of PD (Supplementary Table 3). No clinical information about dementia was available. While a C to T transition at the same position was reported in a single individual of East Asian genetic ancestry in gnomAD, while the variant was absent in gnomAD v4.1.0.

The *VPS35* missense variant of uncertain significance p.Met607Val (16:46662991:T, rs1555523076, CADD=21.5) was identified in a male African ancestry PD patient in the NPDRN dataset, while the p.Asp205His variant (16:46679050:C, CADD=29.4) was identified in a male patient of African admixed ancestry in the GP2-BLAAC PD (Supplementary Table 4). The individual carrying the p.Met607Val variant had disease onset in his early 40s and a positive family history, with a second-degree relative affected by the disease. At the time of enrollment, he had Hoehn and Yahr stage 2 and had demonstrated a sustained response to levodopa without dyskinesia. In terms of non-motor features, he showed no signs of hyposmia or cognitive impairment (MDS-UPDRS Part 1.1 cognition score of 0). Notably, the p.Met607Val variant is 13 amino acids from the known pathogenic p.Asp620Asn variant and lies in the C terminal domain. The patient carrying the p.Asp205His variant of uncertain significance presented with initial symptoms in his late 40s and reported no family history of PD. Using AlphaFold and PyMOL, we visualized the predicted structural consequences of the novel amino acid changes. This also includes estimates of protein stability, with more negative  $\Delta\Delta G$  values generally indicating greater instability (Supplementary Figure 4).

## Integration of local ancestry analysis with variants of interest reveal population specific variants

Local ancestry inference revealed that the majority of prioritized variants were located on haplotypes of African ancestry (Table 1, Table 2, Supplementary Figure 6, Supplementary Figures 7a-7b). Specifically, several *GBA1* variants including p.Asp448His, p.Thr75del, p.Ala300Profs4, p.Arg170Leu, p.Arg535His, p.Asn227Ser, p.Gly364Arg, p.Trp223Arg, and p.His413Arg were all found on African-ancestry haplotypes. Additional variants of African ancestry origin included *VPS35* p.Asp205His and p.Met607Val, as well as *SNCA* p.Met116Ile. One variant (*PRKN* p.Asn52Metfs29) was found on a South Asian ancestry haplotype, and *GBA1* p.Arg502Cys was assigned to a European ancestry background. For *GBA1* p.Ile407Ser, the local ancestry could not be confidently resolved to either African or European origin. A stopgain variant, *FBXO7* p.Arg321Ter, and the missense variant *PLA2G6* p.Thr319Met were

located on haplotypes of Middle Eastern and African ancestry, respectively. Ancestry inference for the *PRKN* p.Pro113Thrfs\*51 deletion was not possible.

## ***PRKN* copy number variants linked to Parkinson's disease in African and African admixed populations**

We screened 96 individuals for CNVs using MLPA. We prioritized samples by individuals positive for both family history and early onset PD (n=4), only family history (n=15), only early onset PD (n=36), and randomly selected sporadic PD cases (n=41). We identified one early onset PD case from the GP2-BLAAC PD dataset having a *PRKN* exon 3 heterozygous deletion and a *PRKN* exon 4 homozygous deletion (Figure 5). This PD case had an AAO of 27 years and no family history of PD. Mimicking the *PRKN*-PD clinical phenotype previously reported, this individual was levodopa responsive, and was positive for asymmetric onset, bradykinesia, rest tremor, rigidity, and gait difficulties. In our additional screening of 1,167 PD cases of African and African admixed ancestry using genotyping data, we identified *PRKN* CNVs in eight samples (African-admixed =5, African =3), two of which were homozygous (African-admixed=1, African=1). Out of these, five were early-onset cases (Supplementary Table 6, Supplementary Figure 3). We did not identify any protein-altering or splicing *PRKN* variant among these samples. We did not identify any CNVs in *SNCA*.

## **Previously reported pathogenic variants in parkinsonism and neurodegenerative disease-related genes**

We identified heterozygous variants in genes associated with recessive PD and parkinsonism, including one in *DNAJC6*, six in *VPS13C*, two in *SYNJ1*, three in *PLA2G6* and one in *FBXO7* in early-onset cases. Among these, the homozygous *FBXO7* p.Arg321Ter variant (22:32491175:C:T, rs369105683) was reported as pathogenic in ClinVar. However, this variant was found to be heterozygous in the current study. We also reported rare heterozygous variants in neurodegenerative disease related genes, including *TRPM7*, *MAPT*, and *APP* that could be associated with parkinsonism phenotypes; however, none have been previously reported as pathogenic. Notably, we identified a homozygous p.Ala128Gly variant (17:45982962:C:G,

rs899291077) in *MAPT* in a male case from the GP2-BLAAC PD dataset. Variants in *MAPT* were previously reported in cases with progressive supranuclear palsy, an atypical parkinsonian and dementia syndrome<sup>29</sup>. The carrier of this variant in our study presented with resting tremor, bradykinesia, and rigidity in his 70s, with no evidence of dementia (Supplementary Table 3). None of the cases with *APP* variants exhibited dementia or cognitive impairment. We identified rare variants in parkinsonism and dystonia genes, including novel *VPS13D* homozygous variants in four cases, and heterozygous variants in *PD8EB*, *KCNN2*, *CSF1R*, *THAP1*, *TOR1A*, *ANO3*, *GCH1*, *VPS16*, *KCDT17*, *KMT2B*, and *ATP1A3* genes. These variants were novel or previously reported with unknown clinical significance (Supplementary Table 7).

## Runs of homozygosity highlight regions overlapping Parkinson's disease and atypical parkinsonism

In the African and African-admixed groups, 164 and 43 ROH pools, respectively, were identified, which overlap known PD and atypical parkinsonism gene regions and GWAS loci, respectively (Supplementary Tables 8 and 9). None of the ROH regions enriched in cases for either group passed the Bonferroni correction. However, in the African group, one ROH pool (S2421) that overlaps the *BAG3* and *INPP5F* genes was present in four cases and completely absent in controls. Moreover, in the African-admixed group, two ROH pools (S622 and S636) were found in cases and absent in controls. These pools overlap the *MEX3C* (n=2 cases), *SMAD4* (n=2 cases), and *PLA2G6* (n=2 cases) genes (Supplementary Tables 8 and 9).

## Spinocerebellar ataxia type 3 associated repeat expansions in African and African-admixed cases

*In silico* analysis using ExpansionHunter identified *ATXN3* repeat expansions within the pathogenic range ( $CAG_n > 45$ ) in three PD patients (1.02%) of African ancestry (two cases from the GP2-BLAAC PD dataset and one case from the NPDRN cohort) and in one control (0.34%) from the NPDRN cohort. RP-PCR validation confirmed the expansions in two cases and one control, with repeat lengths of ( $CAG_n = 72$ ,  $CAG_n = 58$ , and  $CAG_n = 52$ , respectively). The carrier of the  $CAG_n = 72$  expansion is a female patient of African ancestry with an AAO of 41 years. The

carrier of the  $CAG_n=58$  expansion is also a female patient with an AAO of 35 years. The control carrier, with  $CAG_n=52$  is a 40-year-old male. Due to the lack of a biospecimen, we did not perform RP-PCR for the fourth potential carrier, an African ancestry male patient from the NPDRN cohort with an AAO of 35 years, identified through *in silico* analysis. However, visualization of wild-type and expanded reads from short-read sequencing data is presented in Supplementary Figure 5. None of the *ATXN3* expansion carriers were found to carry any additional rare variants in known PD genes.

STR lengths in the *ATXN2* and *TBP* were within the normal range. The lengths of STRs in *ATXN2*, *ATXN3*, and *TBP* for all individuals from the GP2 BLAAC PD and NPDRN datasets are presented in Supplementary Table 10.

## Discussion

We undertook the largest sequencing characterization of PD, parkinsonism, and neurodegenerative disease-related genes in the African and African admixed populations. We created the most comprehensive catalog of both known and novel coding and splicing variants potentially linked to PD in these populations. Notably, our local ancestry analysis revealed that the majority of prioritized variants were enriched on African-derived haplotypes, since African populations harbor the greatest genetic diversity, and regions of the genome with African-derived haplotypes are expected to show SNP enrichment compared to non-African segments. Local-ancestry analyses that lump all African haplotypes together can miss substructure-specific differences (e.g., a variant common in West African populations but rare in East or Southern Africans), which constrains our ability to resolve sub-continental structure. While our study was not powered to systematically compare African and African admixed individuals, these remain important considerations because haplotype background can influence variant interpretation, pathogenicity, allele frequency distributions, and effect size estimates. For African and African admixed populations, this framework is especially critical given the high genetic diversity and complex admixture histories. Admixture results for the GP2-BLAAC PD, NPDRN, and

PDGENE samples are presented in Supplementary Figure 6 and Supplementary Table 11 and Supplementary Table 12.

Biallelic *GBA1* variants are associated with Gaucher's disease, while monoallelic *GBA1* variants are the most prevalent genetic risk factor for PD, with prevalence varying among diverse ethnic groups<sup>30–32,33–35</sup>. We identified that heterozygous rare and potentially disease-causing *GBA1* variants represent the most commonly identified mutations in PD among patients of African and African admixed ancestry, with a frequency of 4%, aligning with the frequency of 4.6% reported in North African populations<sup>36</sup>. Our study identified 18 heterozygous *GBA1* variants, five of which appear more frequent in African ancestry than other ancestry populations according to gnomAD v4.1 and All of Us (release 7), including p.His413Arg, p.Met400Ile, p.Asn227Ser, p.Ser77Arg, and p.Thr75del. Of those, p.Asn227Ser was classified as severe, p.Thr75del as mild, p.Met400Ile as unknown, while p.His413Arg and p.Ser77Arg variants have not been previously reported<sup>37</sup>. No carriers were identified for the most common disease-causing *GBA1* variants in the Ashkenazi Jewish and European populations (p.Asn409Ser, p.Leu483Pro, p.Thr408Met and p.Glu365Lys).

The recent PD GWAS in the African and African admixed ancestry populations revealed a non-coding variant within *GBA1* (rs3115534-G) conferring a major population attributable risk in patients<sup>6</sup>. Follow-up functional characterization highlighted a novel mechanism of *GBA1*-related PD risk, offering a promising target for precision medicine in a population that has been historically underserved in genetic research and therapeutic development<sup>7</sup>. Though this study was primarily focused on assessing the impact of coding variants, we specifically screened for this variant in our cohort. Our analyses showed a higher estimate of the risk G allele in predicted African ancestry individuals, with an overall frequency of 0.23 in cases and 0.17 in controls and a frequency of 0.15 in cases and 0.14 in controls among individuals of African admixed ancestry.

Our results showed that the previously reported European and Asian *LRRK2* mutation spectrum does not play a significant role in PD in West African and African American populations. For

instance, the *LRRK2* p.Gly2019Ser variant was absent in tested cases. It is known that *LRRK2* p.Gly2019Ser shows varying frequencies across North African genetic ancestry groups, ranging from 10-42%, with the highest rates reported in Tunisia<sup>38-44</sup>. Conversely, this variant is rare among Sub-Saharan Africans, with only a few cases reported in South Africa<sup>45</sup>. A study screening for p.Gly2019Ser in 109 Nigerian patients did not identify any variant carriers<sup>46</sup>. Additionally, *LRRK2* p.Gly2019Ser carriers were not detected among 16 South African patients<sup>47</sup> and 38 Zambian individuals with PD<sup>48</sup>. Our results align with a recent study examining *LRRK2* mutations in 22 African American PD patients<sup>3</sup>. Although the authors identified both known and novel variants, no previously reported pathogenic mutations were detected, in contrast to what has been observed in North African populations. Interestingly, we identified a heterozygous *LRRK2* p.Gly8Arg variant, a novel coding mutation not previously reported in the literature, considered to be of uncertain significance, present only in populations of African and Admixed American ancestry according to gnomAD v4.1 and All of Us browsers (release 7). Similarly, the *LRRK2* p.Glu1797Asp heterozygous variant was identified in our study among two male PD cases of African ancestry; this variant has been shown to be present in African and Admixed American ancestry populations and absent in Europeans (73% African and 23% Admixed American ancestry in All of Us release 7). Overall, these findings highlight ancestral differences in *LRRK2* variant distribution, underscoring the need for further research to explore these disparities. The discovery of novel variants opens new avenues for functional investigations.

Among the novel variants identified, *GBA1* p.Ile407Ser and *LRRK2* p.Ile1438Lys were predicted to cause greater protein instability (Supplementary Figure 4). The *GBA1* p.Ile407Ser showed the strongest destabilization ( $\Delta\Delta G = -2.98$ ), which may suggest a potential dual mechanism. First, reduced glucocerebrosidase activity suggest a potential loss-of-function mechanism, consistent with pathogenic variants in the same region such as p.Asn409Ser and p.Leu444Pro, which are known to reduce  $\beta$ -glucocerebrosidase activity and increase levels of glucosylceramide and  $\alpha$ -synuclein<sup>49</sup>. However, it is worth noting that several studies support the general hypothesis (not variant-specific), in which impaired trafficking may lead to accumulation of aberrant glucocerebrosidase species and proteostatic stress, implying a toxic gain-of-function component<sup>50</sup>. The *LRRK2* p.Ile1438Lys variant ( $\Delta\Delta G = -1.72$ ) lies in the ROC domain near the

1 p.Arg1441His mutational hotspot, where destabilization is thought to enhance kinase activity of  
2 *LRRK2*, supporting a gain-of-function mechanism<sup>51</sup>.

3  
4 Our study aligns with previous research showing that *SNCA* coding and structural variation is  
5 rare in African ancestry populations<sup>52</sup>. Our analysis did not detect any coding variants or  
6 structural rearrangements in *SNCA* across the cohorts studied. This finding is consistent with  
7 several other studies on PD in African and African admixed populations. For example, a study of  
8 38PD cases from Zambia (plus one with atypical parkinsonism) found no pathogenic *SNCA*  
9 mutations<sup>48</sup>. Similarly, a screening of 15 early-onset PD cases from Southwestern Nigeria<sup>46</sup>,  
10 and a study of 202 South African PD cases (8% Black) also found no exonic *SNCA* variants  
11 linked to disease<sup>53</sup>.

12  
13 Regarding recessive genes, nearly 20 PD cases associated with *PRKN* and *PINK1* mutations have  
14 been documented in Sub-Saharan Africa across multiple studies<sup>48,52,54–58</sup>. In contrast, fewer than  
15 10 PD cases with these mutations have been reported in North Africa<sup>44,59–64</sup>. Though compound  
16 heterozygous *PRKN* mutations cause recessive disease, heterozygous protein altering variants do  
17 not appear to confer PD risk, with multiple null associations reported<sup>65–67</sup>. In our study, we  
18 identified two heterozygous *PRKN* variants in an early-onset PD case, one of which had been  
19 previously reported as pathogenic, suggesting a potential compound heterozygous state that was  
20 not confirmed due to the lack of phasing data. Together with the CNVs identified in eight  
21 cases—three from African and five from African admixed ancestry—a total of 14 early-onset  
22 PD cases were found to carry potential disease-causing *PRKN* mutations (Table 1-2,  
23 Supplementary Table 4).

24  
25 *PRKN* carriers constituted 1.3% (9 out of 710) of all cases. *PRKN* variants constituted 13.6% of  
26 all observed variants in PD genes, with 13 variants in the GP2-BLAAC PD cohort and 10 in the  
27 NPDRN. Of those, 9 variants were more frequent in cases than controls and 7 were more  
28 frequent in African ancestry individuals, relative to non-Finnish European ancestry, according to  
29 gnomAD. We identified two heterozygous *PINK1* variants, previously reported as pathogenic, in

two late-onset PD cases, which were likely not the cause of underlying PD in these cases. While we did not report any pathogenic variants in *DJ-1* and *VPS13C*, we found four novel *DJ-1* and 13 novel *VPS13C* variants, all in heterozygous state, that warrant further study.

We identified SCA3-associated *ATXN3* repeat expansions in three PD patients and one African ancestry control. Although the control individual is currently unaffected, the presence of a pathogenic-range expansion raises the possibility of future disease manifestation, highlighting the importance of continued follow-up. Notably, a similar analysis of a large European-ancestry PD cohort from the Accelerating Medicines Partnership Program for Parkinson's Disease did not identify any *ATXN3* expansion carriers (personal communication, unpublished data). Previously, Gwinn-Hardy et al. reported PD-like parkinsonism in African American individuals carrying pathogenic *ATXN3* repeat expansions. In their study, three of four affected family members exhibited classic parkinsonian features and were responsive to levodopa<sup>18</sup>. Our findings further support the possibility that *ATXN3* expansions may contribute to early-onset or atypical parkinsonism in individuals of African and African-admixed ancestry and highlight the importance of incorporating STR analysis into PD genetic studies in diverse populations. Screening for *ATXN3* repeat expansions may be especially informative in parkinsonian cases lacking known pathogenic variants.

Several limitations should be acknowledged. First, our study includes a highly heterogeneous population. The current ancestry reference panels cannot accurately distinguish African subpopulations. African admixed study participants commonly exhibit diverse heritage encompassing African, European, and, in certain instances, Native American genetic components. The African component of African American ancestry is itself varied, reflecting the multitude of ethnic groups involved in the trans-Atlantic slave trade. Most of the PD studies encompassing individuals of African ancestry primarily stem from West African populations, which capture only part of Africa's genetic diversity and do not represent the continent in its entirety<sup>68</sup>. Second, most candidate variants identified in our study are singletons, observed only in a single individual or within a single dataset. Therefore, caution is warranted in interpreting their causality. In addition, variants across the *GBA1* and *GBAP1* homology region require

1 further analysis for validation using other tools, since Gaochuan cannot detect novel variants in  
2 this region. Third, structural variations were not comprehensively captured due to limitations in  
3 both detection strategy and sequencing technology. Our CNV analysis was restricted to *SNCA*  
4 and *PRKN*, two genes where CNVs were previously reported in PD. However, most samples did  
5 not undergo MPLA, reducing sensitivity and confidence of CNV detection and precluding  
6 formal association testing. Long-read technologies are more effective in detecting large  
7 insertions or deletions, which may not be captured by short-read sequencing or MLPA.  
8 Incorporating long-read sequencing in future studies could enhance our understanding of the  
9 genetic underpinnings of PD in these populations. Furthermore, since the data were being hosted  
10 on different platforms, joint calling was not possible. We therefore did not merge the sequencing  
11 datasets. Given these limitations in sample composition and cohort harmonization, we currently  
12 lack sufficient statistical power to conduct robust single-variant association or gene-based burden  
13 testing. Clinical information was also lacking; only a subset of samples had data from the  
14 Montreal Cognitive Assessment (MoCA) and few had data for the Identification and Intervention  
15 for Dementia in Elderly Africans (IDEA) cognitive screen and Rapid Eye Movement sleep  
16 behavior disorder (RBD). One question from the CISI-PD questionnaire was used to assess  
17 cognitive status. Future data collection efforts as part of the GP2-BLAAC PD study will include  
18 the MoCA and other clinical assessments.

19  
20 Our study, the largest of its kind to explore rare coding and splicing variations in African and  
21 African-admixed populations, aims to fill critical gaps in PD genetics by providing a  
22 comprehensive genetic characterization of these ancestries. While recognition of the importance  
23 of studying PD in diverse populations is growing, Black and African admixed communities have  
24 long been underrepresented in research, creating major gaps in understanding how PD manifests  
25 and progresses. By focusing on these underserved groups, our work seeks to expand knowledge  
26 of PD etiology, ultimately improving diagnostic accuracy and therapeutic development. As we  
27 advance toward precision medicine, therapies should be developed with a global perspective.  
28 Operating under the umbrella of the GP2, it is essential that data guiding individualized  
29 treatments reflects population diversity, ensuring equitable benefits from medical advancements  
30 in PD care.

## Data availability

All GP2 data for these analyses is available through collaboration with Accelerating Medicines Partnership in Parkinson's disease and is available through application on the AMP-PD website (<https://amp-pd.org/register-for-amp-pd>). Other publicly available data consortiums include All of Us Research Program (<https://www.researchallofus.org/register/>) and UK Biobank (<https://www.ukbiobank.ac.uk/enable-your-research/register>), in which data can be accessed after applying. We have received an exception to the Data and Statistics Dissemination Policy from the All of Us Resource Access Board.

Analyses such as genotyping quality control, ancestry prediction, and processing were performed using GenoTools (<https://github.com/dvitale199/GenoTools>). Local ancestry analysis was performed using pipelines developed and maintained by Dr. Thiago Peixoto Leal ([peixott@ccf.org](mailto:peixott@ccf.org)) and are available at <https://github.com/MataLabCCF.A> repository of all other code used for processing and analyzing is publicly available (<https://github.com/GP2code/PD-GeneticCharacterization-inAFRAndAAC/>; DOI 10.5281/zenodo.14579574).

## Acknowledgements

We thank Paige Brown Jarreau for her support with manuscript preparation.

Data used in the preparation of this article were obtained from Global Parkinson's Genetics Program (GP2). GP2 is funded by the Aligning Science Across Parkinson's (ASAP) initiative and implemented by The Michael J. Fox Foundation for Parkinson's Research (<https://gp2.org>). For a complete list of GP2 members see <https://gp2.org>. The All of Us Research Program is supported by the National Institutes of Health, Office of the Director: Regional Medical Centers: 1 OT2 OD026549; 1 OT2 OD026554; 1 OT2 OD026557; 1 OT2 OD026556; 1 OT2 OD026550; 1 OT2 OD 026552; 1 OT2 OD026553; 1 OT2 OD026548; 1 OT2 OD026551; 1 OT2 OD026555; IAA #: AOD 16037; Federally Qualified Health Centers: HHSN 263201600085U; Data and Research Center: 5 U2C OD023196; Biobank: 1 U24 OD023121; The Participant Center: U24 OD023176; Participant Technology Systems Center: 1 U24 OD023163;

Communications and Engagement: 3 OT2 OD023205; 3 OT2 OD023206; and Community Partners: 1 OT2 OD025277; 3 OT2 OD025315; 1 OT2 OD025337; 1 OT2 OD025276. In addition, the All of Us Research Program would not be possible without the partnership of its participants.

This research has been conducted using the UK Biobank Resource under application number 33601.

## Funding

This work utilized the computational resources of the NIH HPC Biowulf cluster (<https://hpc.nih.gov>). This research was supported in part by the Intramural Research Program of the NIH, National Institute on Aging (NIA), National Institutes of Health, Department of Health and Human Services; project number ZO1 AG000535 and ZIA AG000949, as well as the National Institute of Neurological Disorders and Stroke (NINDS, program # ZIANS003154) and the National Human Genome Research Institute (NHGRI). This work utilized the computational resources of the NIH STRIDES Initiative (<https://cloud.nih.gov>) through the Other Transaction agreement - Azure: OT2OD032100, Google Cloud Platform: OT2OD027060, Amazon Web Services: OT2OD027852. Data used in the preparation of this article were obtained from the Global Parkinson's Genetics Program (GP2). GP2 is funded by the Aligning Science Across Parkinson's (ASAP) initiative and implemented by The Michael J. Fox Foundation for Parkinson's Research (<https://gp2.org>). For a complete list of GP2 members see <https://gp2.org>. KS. is supported by The Michael J. Fox Foundation and Aligning Sciences Across Parkinson's Disease Global Parkinson Genetic Program. EW. is supported by the Michael J. Fox Foundation (MJFF-026283) and Alzheimer's Disease Sequencing Project (ADSP) (5U01AG076482-03). Additional funding was provided by The Michael J. Fox Foundation for Parkinson's Research through grant MJFF-009421/17483.

## Competing interests

KL, DV, MBM, HLL, HI, MJK and MAN declare that they are consultants employed by DataTecnica LLC, whose participation in this is part of a consulting agreement between the US National Institutes of Health and said company. MAN also owns stock from Neuron23 Inc and Character Biosciences.

## Supplementary material

Supplementary material is available at *Brain* online.

## References

1. Khani M, Cerquera-Cleves C, Kekenadze M, Wild Crea P, Singleton AB, Bandres-Ciga S. Towards a global view of Parkinson's disease genetics. *Ann Neurol*. 2024;95(5):831-842.
2. Okunoye O, Zewde YZ, Azar J, et al. The State of Play of Parkinson's Disease in Africa: A Systematic Review and Point of View. *medRxiv*. Published online July 8, 2023:2023.07.07.23292392. doi:10.1101/2023.07.07.23292392
3. Ross OA, Wilhoite GJ, Bacon JA, et al. LRRK2 variation and Parkinson's disease in African Americans. *Mov Disord*. 2010;25(12):1973-1976.
4. McGuire V, Van Den Eeden SK, Tanner CM, et al. Association of DRD2 and DRD3 polymorphisms with Parkinson's disease in a multiethnic consortium. *J Neurol Sci*. 2011;307(1-2):22-29.
5. Clark LN, Levy G, Tang MX, et al. The Saitohin "Q7R" polymorphism and tau haplotype in multi-ethnic Alzheimer disease and Parkinson's disease cohorts. *Neurosci Lett*. 2003;347(1):17-20.
6. Rizig M, Bandres-Ciga S, Makarious MB, et al. Identification of genetic risk loci and causal insights associated with Parkinson's disease in African and African admixed populations: a genome-wide association study. *Lancet Neurol*. 2023;22(11):1015-1025.
7. Álvarez Jerez P, Wild Crea P, Ramos DM, et al. African ancestry neurodegeneration risk variant disrupts an intronic branchpoint in GBA1. *Nature Structural & Molecular Biology*. Published online December 12, 2024:1-9.
8. Mahungu AC, Anderson DG, Rossouw AC, et al. Screening of the glucocerebrosidase

- (GBA) gene in South Africans of African ancestry with Parkinson's disease. *Neurobiol Aging*. 2020;88:156.e11-e156.e14.
9. Milanowski LM, Oshinaike O, Walton RL, et al. Screening of GBA Mutations in Nigerian Patients with Parkinson's Disease. *Mov Disord*. 2021;36(12):2971-2973.
  10. Bandres-Ciga S, Black and African American Connections to Parkinson's Disease (BLAAC PD) Study Group. Black and African American Connections to Parkinson's Disease Study: Addressing Missing Diversity in Parkinson's Disease Genetics. *Mov Disord*. 2022;37(7):1559-1561.
  11. Global Parkinson's Genetics Program. GP2: The global Parkinson's genetics program. *Mov Disord*. 2021;36(4):842-851.
  12. Ojo OO, Abubakar SA, Iwuzo EU, et al. The Nigeria Parkinson Disease Registry: Process, Profile, and Prospects of a Collaborative Project. *Mov Disord*. 2020;35(8):1315-1322.
  13. Cook L, Verbrugge J, Schwantes-An TH, et al. Parkinson's disease variant detection and disclosure: PD GENERation, a North American study. *Brain*. 2024;147(8):2668-2679.
  14. All of Us Research Program Investigators, Denny JC, Rutter JL, et al. The "All of Us" Research Program. *N Engl J Med*. 2019;381(7):668-676.
  15. Bycroft C, Freeman C, Petkova D, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature*. 2018;562(7726):203-209.
  16. Westenberger A, Skrahina V, Usnich T, et al. Relevance of genetic testing in the gene-targeted trial era: the Rostock Parkinson's disease study. *Brain*. 2024;147(8):2652-2667.
  17. Kim JM, Hong S, Kim GP, et al. Importance of low-range CAG expansion and CAA interruption in SCA2 Parkinsonism. *Arch Neurol*. 2007;64(10):1510-1518.
  18. Gwinn-Hardy K, Singleton A, O'Suilleabhain P, et al. Spinocerebellar ataxia type 3 phenotypically resembling parkinson disease in a black family. *Arch Neurol*. 2001;58(2):296-299.
  19. Kim JY, Kim SY, Kim JM, et al. Spinocerebellar ataxia type 17 mutation as a causative and susceptibility gene in parkinsonism. *Neurology*. 2009;72(16):1385-1389.
  20. Chahine LM, Louie N, Solle J, et al. The Black and African American Connections to Parkinson's Disease (BLAAC PD) study protocol. *BMC Neurol*. 2024;24(1):403.
  21. Martínez-Martín P, Forjaz MJ, Cubo E, Frades B, de Pedro Cuesta J, ELEG Project Members. Global versus factor-related impression of severity in Parkinson's disease: a new clinimetric index (CISI-PD). *Mov Disord*. 2006;21(2):208-214.
  22. Postuma RB, Berg D, Stern M, et al. MDS clinical diagnostic criteria for Parkinson's disease: MDS-PD Clinical Diagnostic Criteria. *Mov Disord*. 2015;30(12):1591-1601.
  23. Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic

- Parkinson's disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry*. 1992;55(3):181-184.
24. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164.
25. Toffoli M, Chen X, Sedlazeck FJ, et al. Comprehensive short and long read sequencing analysis for the Gaucher and Parkinson's disease-associated GBA gene. *Commun Biol*. 2022;5(1):670.
26. Amendola LM, Dorschner MO, Robertson PD, et al. Actionable exomic incidental findings in 6503 participants: challenges of variant classification. *Genome Res*. 2015;25(3):305-315.
27. Kircher M, Witten DM, Jain P, O'Roak BJ, Cooper GM, Shendure J. A general framework for estimating the relative pathogenicity of human genetic variants. *Nat Genet*. 2014;46(3):310-315.
28. Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. *Lancet Neurol*. 2020;19(2):170-178.
29. Hutton M, Lendon CL, Rizzu P, et al. Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. *Nature*. 1998;393(6686):702-705.
30. Migdalska-Richards A, Schapira AHV. The relationship between glucocerebrosidase mutations and Parkinson disease. *J Neurochem*. 2016;139 Suppl 1(Suppl Suppl 1):77-90.
31. De Marco EV, Annesi G, Tarantino P, et al. Glucocerebrosidase gene mutations are associated with Parkinson's disease in southern Italy. *Mov Disord*. 2008;23(3):460-463.
32. Kalinderi K, Bostantjopoulou S, Paisan-Ruiz C, Katsarou Z, Hardy J, Fidani L. Complete screening for glucocerebrosidase mutations in Parkinson disease patients from Greece. *Neurosci Lett*. 2009;452(2):87-89.
33. Tan EK, Tong J, Fook-Chong S, et al. Glucocerebrosidase mutations and risk of Parkinson disease in Chinese patients. *Arch Neurol*. 2007;64(7):1056-1058.
34. Mitsui J, Mizuta I, Toyoda A, et al. Mutations for Gaucher disease confer high susceptibility to Parkinson disease. *Arch Neurol*. 2009;66(5):571-576.
35. 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.
36. Lesage S, Condroyer C, Hecham N, et al. Mutations in the glucocerebrosidase gene confer a risk for Parkinson disease in North Africa. *Neurology*. 2011;76(3):301-303.
37. Parlar SC, Grenn FP, Kim JJ, Baluwendraat C, Gan-Or Z. Classification of GBA1 Variants in Parkinson's Disease: The GBA1-PD Browser. *Mov Disord*. 2023;38(3):489-495.
38. Belarbi S, Hecham N, Lesage S, et al. LRRK2 G2019S mutation in Parkinson's disease: A

- neropsychological and neuropsychiatric study in a large Algerian cohort. *Parkinsonism Relat Disord.* 2010;16:676-679.
39. Troiano AR, Elbaz A, Lohmann E, et al. Low disease risk in relatives of north african lrrk2 Parkinson disease patients. *Neurology.* 2010;75(12):1118-1119.
40. Hashad DI, Abou-Zeid AA, Achmawy GA, Allah HMOS, Saad MA. G2019S mutation of the leucine-rich repeat kinase 2 gene in a cohort of Egyptian patients with Parkinson's disease. *Genet Test Mol Biomarkers.* 2011;15(12):861-866.
41. Warren L, Gibson R, Ishihara L, et al. A founding LRRK2 haplotype shared by Tunisian, US, European and Middle Eastern families with Parkinson's disease. *Parkinsonism Relat Disord.* 2008;14:77-80.
42. Ishihara L, Gibson RA, Warren L, et al. Screening for Lrrk2 G2019S and clinical comparison of Tunisian and North American Caucasian Parkinson's disease families. *Mov Disord.* 2007;22(1):55-61.
43. Hulihan MM, Ishihara-Paul L, Kachergus J, et al. LRRK2 Gly2019Ser penetrance in Arab-Berber patients from Tunisia: a case-control genetic study. *Lancet Neurol.* 2008;7(7):591-594.
44. Nishioka K, Kefi M, Jasinska-Myga B, et al. A comparative study of LRRK2, PINK1 and genetically undefined familial Parkinson's disease. *J Neurol Neurosurg Psychiatry.* 2010;81(4):391-395.
45. du Toit N, van Coller R, Anderson DG, Carr J, Bardien S. Frequency of the LRRK2 G2019S mutation in South African patients with Parkinson's disease. *Neurogenetics.* 2019;20(4):215-218.
46. Milanowski LM, Oshinaike O, Broadway BJ, et al. Early-Onset Parkinson Disease Screening in Patients From Nigeria. *Front Neurol.* 2020;11:594927.
47. Mahne AC, Carr JA, Bardien S, Schutte CM. Clinical findings and genetic screening for copy number variation mutations in a cohort of South African patients with Parkinson's disease. *S Afr Med J.* 2016;106(6). doi:10.7196/SAMJ.2016.v106i6.10340
48. Yonova-Doing E, Atadzhanov M, Quadri M, et al. Analysis of LRRK2, SNCA, Parkin, PINK1, and DJ-1 in Zambian patients with Parkinson's disease. *Parkinsonism Relat Disord.* 2012;18(5):567-571.
49. Cullen V, Sardi SP, Ng J, et al. Acid  $\beta$ -glucosidase mutants linked to Gaucher disease, Parkinson disease, and Lewy body dementia alter  $\alpha$ -synuclein processing. *Ann Neurol.* 2011;69(6):940-953.
50. Mazzulli JR, Xu YH, Sun Y, et al. Gaucher disease glucocerebrosidase and  $\alpha$ -synuclein form a bidirectional pathogenic loop in synucleinopathies. *Cell.* 2011;146(1):37-52.
51. Liao J, Wu CX, Burlak C, et al. Parkinson disease-associated mutation R1441H in LRRK2

prolongs the “active state” of its GTPase domain. *Proc Natl Acad Sci U S A*. 2014;111(11):4055-4060.

52. Keyser RJ, Lombard D, Veikondis R, Carr J, Bardien S. Analysis of exon dosage using MLPA in South African Parkinson’s disease patients. *Neurogenetics*. 2010;11(3):305-312.

53. Keyser RJ, Oppon E, Carr JA, Bardien S. Identification of Parkinson’s disease candidate genes using CAESAR and screening of MAPT and SNCAIP in South African Parkinson’s disease patients. *J Neural Transm (Vienna)*. 2011;118(6):889-897.

54. Bardien S, Keyser R, Yako Y, Lombard D, Carr J. Molecular analysis of the parkin gene in South African patients diagnosed with Parkinson’s disease. *Parkinsonism Relat Disord*. 2009;15(2):116-121.

55. Haylett WL, Keyser RJ, du Plessis MC, et al. Mutations in the parkin gene are a minor cause of Parkinson’s disease in the South African population. *Parkinsonism Relat Disord*. 2012;18(1):89-92.

56. Keyser RJ, Lesage S, Brice A, Carr J, Bardien S. Assessing the prevalence of PINK1 genetic variants in South African patients diagnosed with early- and late-onset Parkinson’s disease. *Biochem Biophys Res Commun*. 2010;398(1):125-129.

57. Dekker MCJ, Suleiman JM, Bhwana D, et al. PRKN-related familial Parkinson’s disease: First molecular confirmation from East Africa. *Parkinsonism Relat Disord*. 2020;73:14-15.

58. Carr J, Guella I, Szu-Tu C, et al. Double homozygous mutations (R275W and M432V) in the ParkinGene associated with late-onset Parkinson’s disease. *Mov Disord*. 2016;31(3):423-425.

59. Bouhouche A, Tesson C, Regragui W, et al. Mutation Analysis of Consanguineous Moroccan Patients with Parkinson’s Disease Combining Microarray and Gene Panel. *Front Neurol*. 2017;8:567.

60. Gouider-Khouja N, Larnaout A, Amouri R, et al. Autosomal recessive parkinsonism linked to parkin gene in a Tunisian family. Clinical, genetic and pathological study. *Parkinsonism Relat Disord*. 2003;9(5):247-251.

61. Ibáñez P, Lesage S, Lohmann E, et al. Mutational analysis of the PINK1 gene in early-onset parkinsonism in Europe and North Africa. *Brain*. 2006;129(Pt 3):686-694.

62. Ishihara-Paul L, Hulihan MM, Kachergus J, et al. PINK1 mutations and parkinsonism. *Neurology*. 2008;71(12):896-902.

63. Lücking CB, Abbas N, Dürr A, et al. Homozygous deletions in parkin gene in European and North African families with autosomal recessive juvenile parkinsonism. The European Consortium on Genetic Susceptibility in Parkinson’s Disease and the French Parkinson’s Disease Genetics Study Group. *Lancet*. 1998;352(9137):1355-1356.

64. Norman BP, Lubbe SJ, Tan M, Warren N, Morris HR. Early Onset Parkinson’s Disease in a

family of Moroccan origin caused by a p.A217D mutation in PINK1: a case report. *BMC Neurol.* 2017;17(1):153.

65. Zhu W, Huang X, Yoon E, et al. Heterozygous PRKN mutations are common but do not increase the risk of Parkinson's disease. *Brain.* 2022;145(6):2077-2091.

66. Lesage S, Lohmann E, Tison F, et al. Rare heterozygous parkin variants in French early-onset Parkinson disease patients and controls. *J Med Genet.* 2008;45(1):43-46.

67. Yu E, Rudakou U, Krohn L, et al. Analysis of Heterozygous PRKN Variants and Copy-Number Variations in Parkinson's Disease. *Mov Disord.* 2021;36(1):178-187.

68. Tishkoff SA, Reed FA, Friedlaender FR, et al. The genetic structure and history of Africans and African Americans. *Science.* 2009;324(5930):1035-1044.

## Figure legends

**Figure 1 Workflow.** Our workflow begins with creating cohorts within the datasets. We utilized short-read sequencing data to characterize Parkinson's disease, parkinsonism and related neurodegenerative disease genes. We prioritized missense, frameshift, start loss, stop loss, stop gain, and splicing variants with a minor allele frequency < 0.01 and CADD score > 12.37.

## Figure 2 Distribution of variants in Parkinson's disease and parkinsonism-related genes.

Parkinson's disease and parkinsonism related genes included for analysis. †het indicates autosomal recessive inheritance. "Very high" indicates genes with very high confidence being related to disease, "High" indicates high confidence, and "Low" indicates low confidence based on replication in the literature (Blauwendraat et al., 2020). Five cohorts were included: Black African and African American Connections to Parkinson's Disease Study (BLAAC PD), as a part of the Global Parkinson's Genetics Program (GP2), Nigerian Parkinson's Disease Research (NPDR) Network, the PDGENERation initiative, the All of Us Program and the UK Biobank (UKB). The width of the connecting line between the cohort and each gene is correlated to the number of variants found.

**Figure 3 Prioritized rare protein-altering or splicing variants in known autosomal dominant, recessive and genetic risk factors for Parkinson's disease.** Autosomal recessive

genes for PD include: *PRKN*, *PINK1*, *DJ-1*, and *VPS13C*. Genetic risk factors for PD include: *GBA1* and *LRRK2*. Autosomal dominant genes for PD include: *SNCA* and *VPS35*. For the latter group, these genes were not individually plotted since only one variant in each autosomal dominant gene was detected. Rare variants were included if minor allele frequency (MAF) was lower than 0.01 and CADD score > 12.37. Novel indicates variants not previously reported in the literature. Likely benign and benign variants, as well as likely pathogenic and pathogenic variants according to pathogenicity predictors, are presented together for each gene. VUS=variant of uncertain significance. (Novel variants in other PD genes include two in *DNAJC6*, one in *SNCA*, two in *VPS35*, and one in *SYNJ1*).

**Figure 4 Distribution of age at onset among Parkinson's disease prioritized variant carriers.**

**Figure 5 Biallelic *PRKN* deletion identified in an early-onset Parkinson's disease case. A)** Multiplex Ligation-dependent Probe Amplification (MLPA) was performed in early-onset Parkinson's disease cases and/or individuals with positive family history. One early-onset Parkinson's disease with an age at onset of 27 years was detected to have a decrease in the height of the electropherogram peak at 202 and 477 nucleotides (red arrows), which correspond to *PRKN*-4 and *PRKN*-3 probes, respectively. B) The probe ratios indicate a *PRKN* exon 3 heterozygous deletion (0.5 ratio) and a *PRKN* exon 4 homozygous deletion (0 ratio).

1 **Table I Known pathogenic or likely pathogenic variants identified in Parkinson's disease genes**

| Variant                                                | Protein change   | Gene   | dbSNP        | gnomAD AF (AFR)      | gnomAD AF (ALL)      | Genotype (cases) | Dataset                 | Ancestry     | Sex                                      | Age range                               | Family history |
|--------------------------------------------------------|------------------|--------|--------------|----------------------|----------------------|------------------|-------------------------|--------------|------------------------------------------|-----------------------------------------|----------------|
| 1:155235002:C:T                                        | p.Arg535His      | GBA1   | rs75822236   | $5.0 \times 10^{-5}$ | $1.1 \times 10^{-4}$ | 1 het            | BLAAC PD                | AFR          | female                                   | 70s                                     | no             |
| 1:155235196:G:A                                        | p.Arg502Cys      | GBA1   | rs80356771   | $8.1 \times 10^{-5}$ | $2.3 \times 10^{-4}$ | 1 het            | BLAAC PD                | AAC          | female                                   | late 50s                                | no             |
| 1:155235217:C:G                                        | p.Ala495Pro      | GBA1   | rs368060     | $9.0 \times 10^{-4}$ | $6.0 \times 10^{-4}$ | 1 het            | All of Us               | AFR          | female                                   | late 50s                                | NA             |
| 1:155235727:C:G                                        | p.Asp448His      | GBA1   | rs1064651    | $2.5 \times 10^{-4}$ | $1.1 \times 10^{-4}$ | 1 het            | BLAAC PD                | AAC          | female                                   | early 50s                               | NA             |
| 1:155235823:C:T                                        | p.Gly416Ser      | GBA1   | rs121908311  | 0                    | $2.2 \times 10^{-5}$ | 1 het            | All of Us               | AFR          | female                                   | early 50s                               | NA             |
| 1:155238215:T:C                                        | p.Asn227Ser      | GBA1   | rs364897     | $2.0 \times 10^{-4}$ | $6.2 \times 10^{-5}$ | 1 het            | BLAAC PD                | AFR          | male                                     | early 50s                               | NA             |
| 1:155238228:A:G                                        | p.Trp223Arg      | GBA1   | rs61748906   | $1.3 \times 10^{-5}$ | $1.2 \times 10^{-5}$ | 1 het            | NPDRN                   | AFR          | male                                     | early 60s                               | no             |
| 1:155238291:G:A                                        | p.Arg202Ter      | GBA1   | rs1009850780 | $1.4 \times 10^{-5}$ | $1.1 \times 10^{-5}$ | 1 het            | PDGE NE                 | AFR          | female                                   | late 50s                                | no             |
| 1:155238596:C:A                                        | p.Arg170Leu      | GBA1   | rs80356763   | 0                    | $6.2 \times 10^{-7}$ | 2 hets           | BLAAC PD                | AFR, AFR     | female                                   | two siblings, early 30s and early 40s   | yes, siblings  |
| 1:155239968:GGTA:G                                     | p.Thr75del       | GBA1   | rs761621516  | $5.5 \times 10^{-4}$ | $3.2 \times 10^{-5}$ | 3 hets           | BLAAC P (1) PDGE NE (2) | AFR; AFR; NA | female in BLAAC PD, two males in PDGE NE | 50s in BLAAC PD; 50s and 70s in PDGE NE | NA             |
| 1:20645615:G:A                                         | p.Ala339Thr      | PINK1  | rs55831733   | $3.7 \times 10^{-4}$ | $1.3 \times 10^{-3}$ | 1 het            | PDGE NE                 | NA           | male                                     | 50s                                     | no             |
| 6:162262560:TCAGTGTGCAGATGACAGCCAGCCCCACAGAGTCTCCTGG:T | p.Pro113Thrfs*51 | PRKN   | rs771529549  | $4.0 \times 10^{-5}$ | $2.6 \times 10^{-4}$ | 1 het            | PDGE NE                 | AFR          | female                                   | 20s                                     | no             |
| 6:162443325:AT:A                                       | p.Asn52Metfs*29  | PRKN   | rs754809877  | $6.7 \times 10^{-5}$ | $2.8 \times 10^{-4}$ | 2 hets           | PDGE NE; All of Us      | NA; AFR      | female; male                             | 20s; 60s                                | no             |
| 22:32491175:C:T                                        | p.Arg321Ter      | FBXO7  | rs369105683  | $1.3 \times 10^{-5}$ | $3.3 \times 10^{-5}$ | 1 het            | PDGE NE                 | AAC          | female                                   | late 40s                                | no             |
| 22:38132952:G:A                                        | p.Thr319Met      | PLA2G6 | rs149653983  | $6.2 \times 10^{-4}$ | $2.9 \times 10^{-4}$ | 1 het            | PDGE NE                 | AFR          | female                                   | early 50s                               | no             |

2 dbSNP: The Single Nucleotide Polymorphism database; AF: allele frequency in gnomAD v4.1; AFR: African/African American; NPDRN: Nigerian  
 3 Parkinson's Disease Research Network; het: heterozygous; AAC: African admixed ancestry; NA: not present; Variant position: referred to hg38.

4 We have received an exception to the Data and Statistics Dissemination Policy from the All of Us Resource Access Board.

1 **Table 2 Novel variants identified in Parkinson's disease genes**

| Variant           | Protein change   | Gene  | dbSNP        | gnomAD AF (AFR)      | gnomAD AF (ALL)      | Genotype (cases) | Dataset   | Ancestry | Inferred ancestry window for the variant | Sex    | Age range | Family history           |
|-------------------|------------------|-------|--------------|----------------------|----------------------|------------------|-----------|----------|------------------------------------------|--------|-----------|--------------------------|
| 1:155235831: T:C  | p.His413Arg      | GBA1  | rs911331923  | $4.0 \times 10^{-5}$ | $1.9 \times 10^{-6}$ | I het            | BLAAC PD  | CAH      | AFR                                      | male   | 40-49     | NA                       |
| 1:155236249: A:C  | p.Ile407Ser      | GBA1  | rs1057519358 | NA                   | NA                   | I het            | BLAAC PD  | AAC      | Unconfirmed (either AFR or EUR)          | male   | 50-59     | no                       |
| 1:155236379: C:G  | p.Gly364Arg      | GBA1  | NA           | NA                   | NA                   | I het            | BLAAC PD  | AFR      | AFR                                      | male   | 50-59     | no                       |
| 1:155237441: GC:G | p.Ala300Profs*4  | GBA1  | NA           | NA                   | NA                   | I het            | BLAAC PD  | AFR      | AFR                                      | female | 40-49     | yes, 2nd degree relative |
| 4:89729236C:A     | p.Met116Ile      | SNC A | rs1378041201 | NA                   | NA                   | I het            | PDGENE    | AFR      | AFR                                      | male   | 50-59     | no                       |
| 6:162201248: AC:A | p.Gly139Valfs*38 | PRKN  | rs751790069  | $4.0 \times 10^{-5}$ | $1.9 \times 10^{-6}$ | I het            | PDGENE    | AFR      | NA                                       | female | 20-29     | no                       |
| 12:40243646: A:C  | p.Glu268Ala      | LRRK2 | rs373254349  | $1.6 \times 10^{-4}$ | $8.1 \times 10^{-6}$ | I het            | PDGENE    | NA       | NA                                       | female | 60-69     | no                       |
| 12:40309228: A:T  | p.Ile1438Lys     | LRRK2 | NA           | NA                   | NA                   | I het            | PDGENE    | NA       | NA                                       | female | 60-69     | no                       |
| 12:40314047: C:T  | p.Arg1538Cys     | LRRK2 | rs150620977  | $5.6 \times 10^{-4}$ | $3.1 \times 10^{-5}$ | I het            | All of Us | AFR      | NA                                       | female | 40-49     | no                       |
| 16:46662991: T:C  | p.Met607Val      | VPS35 | rs1555523076 | 0                    | $6.2 \times 10^{-7}$ | I het            | NPDRN     | AFR      | AFR                                      | male   | 40-49     | yes, 2nd degree relative |
| 16:46679050: C:G  | p.Asp205His      | VPS35 | NA           | NA                   | NA                   | I het            | BLAAC PD  | AAC      | AFR                                      | male   | 40-49     | no                       |

2 dbSNP: The Single Nucleotide Polymorphism database; AF: allele frequency in gnomAD v4.1; AFR: African/African American; NPDRN: Nigerian  
 3 Parkinson's Disease Research Network; het: heterozygous; AAC: African admixed ancestry; CAH: Complex admixed ancestry; NA: not present;  
 4 Variant position: referred to hg38. We have received an exception to the Data and Statistics Dissemination Policy from the All of Us Resource  
 5 Access Board.  
 6  
 7

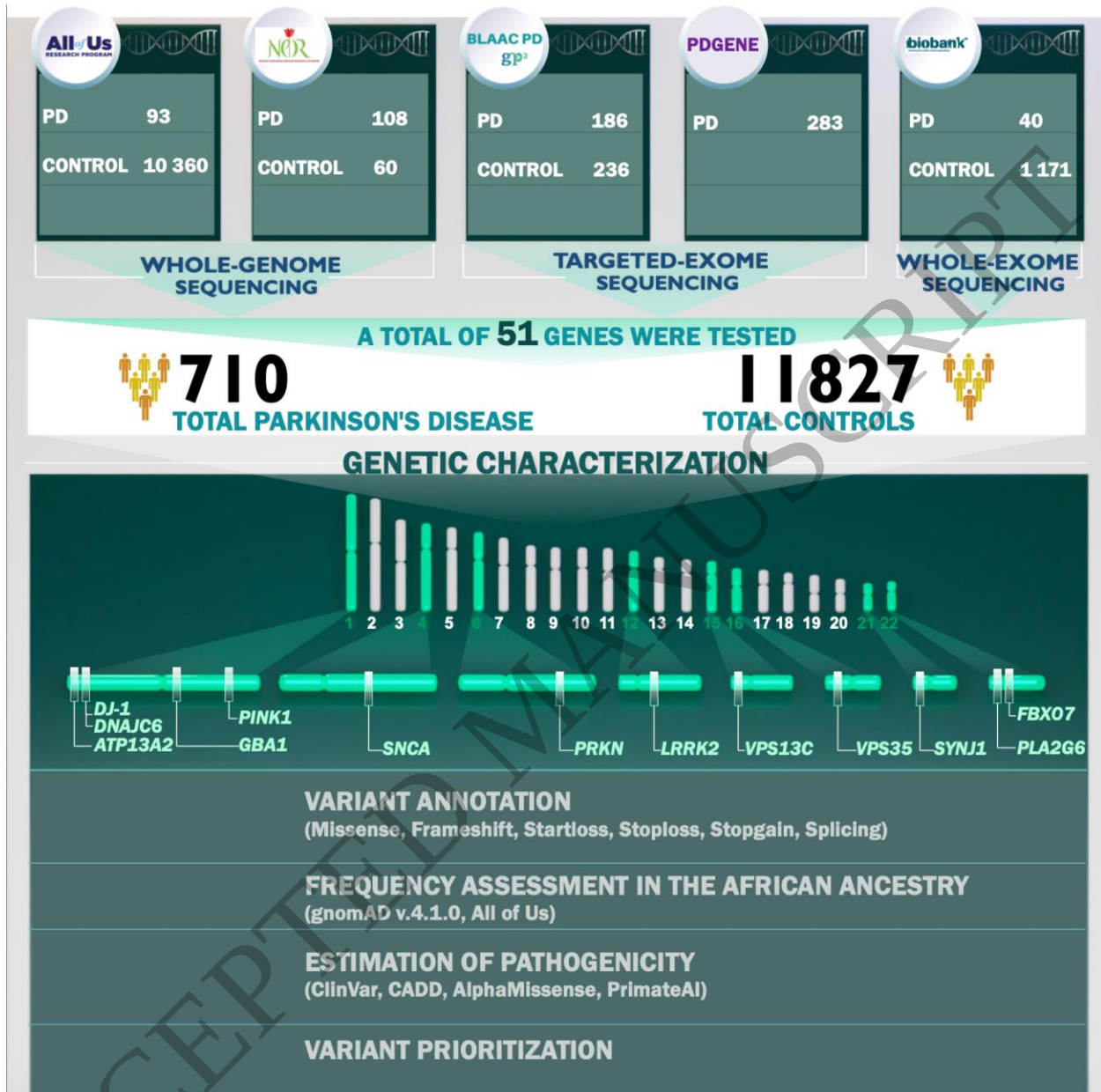


Figure 1  
165x168 mm (DPI)

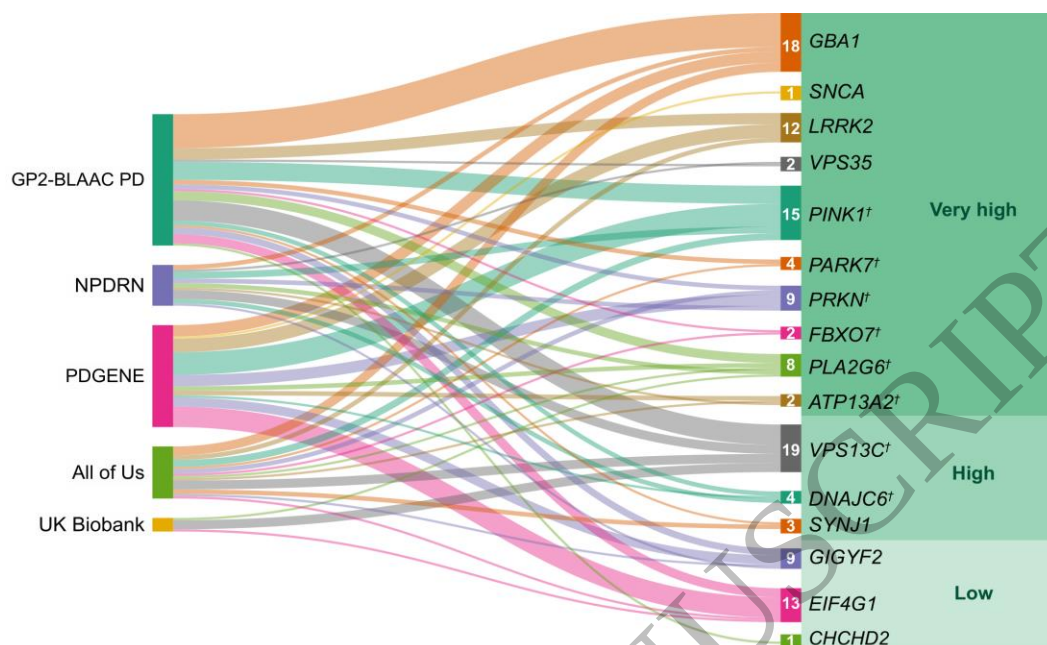


Figure 2  
71x44 mm (DPI)

1  
2  
3  
4

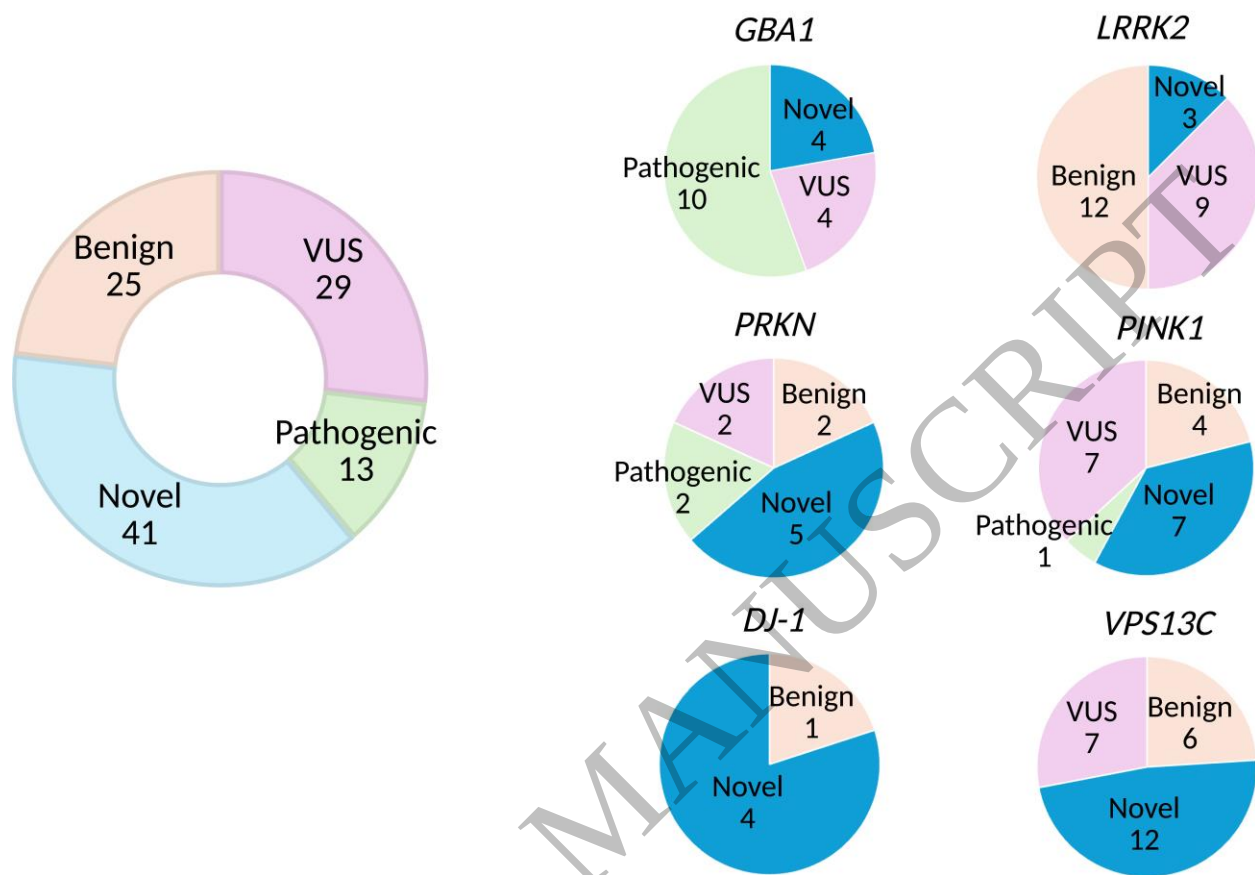


Figure 3  
165x118 mm (DPI)

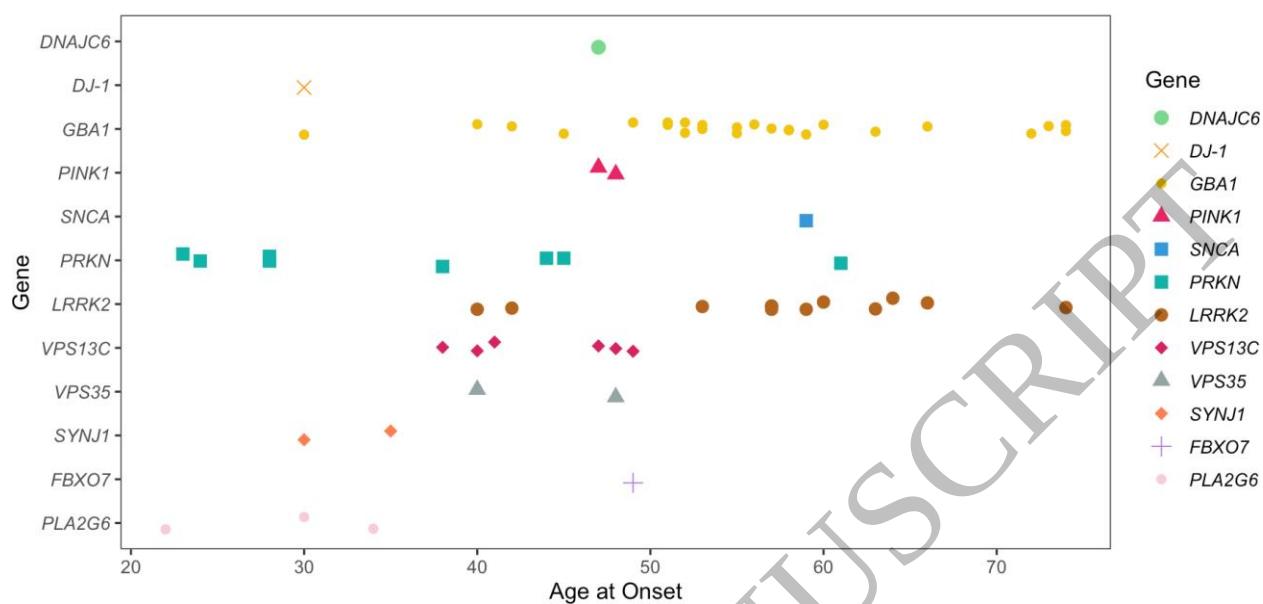


Figure 4  
165x83 mm (DPI)

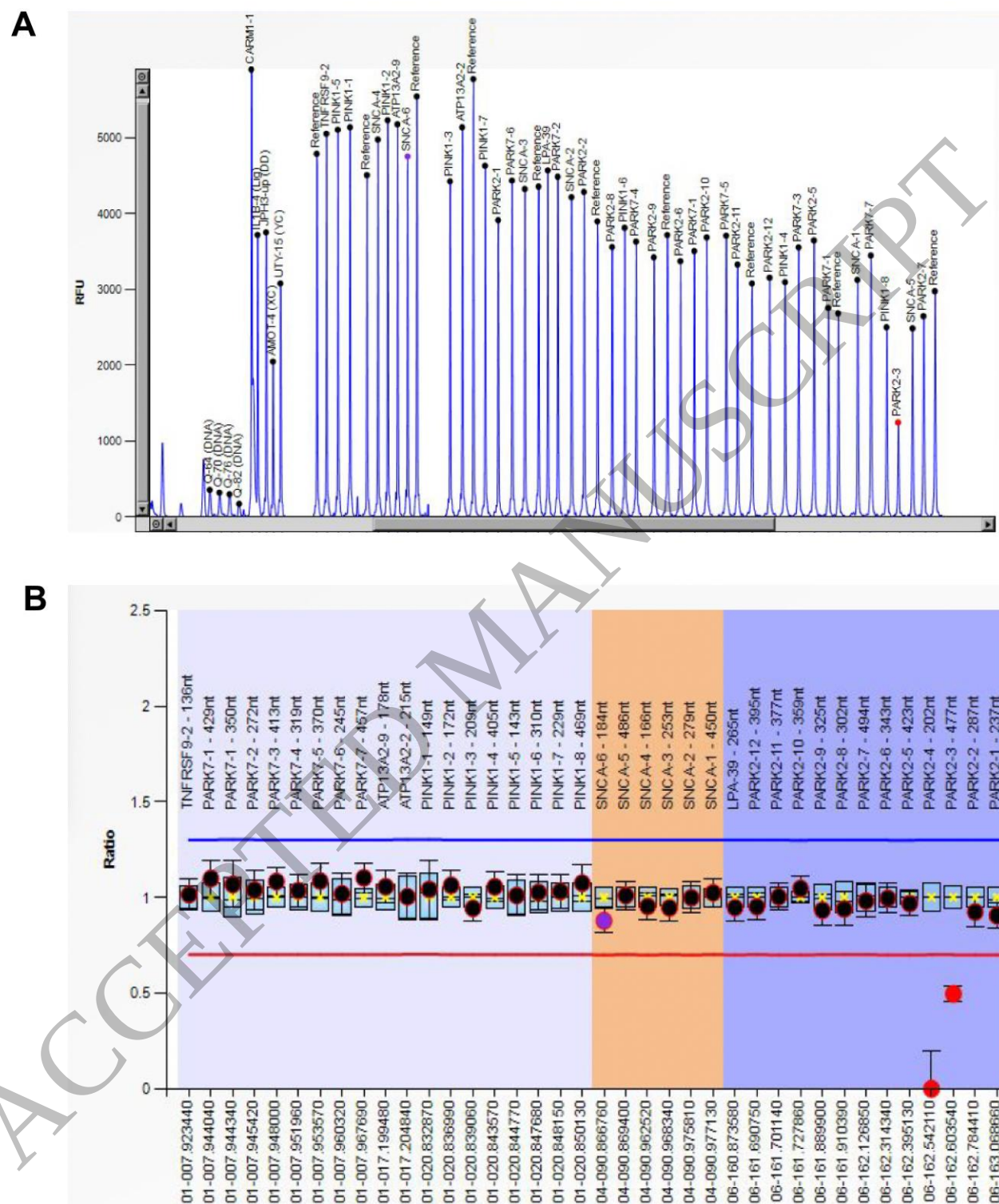


Figure 5  
165x201 mm (DPI)