Genome-wide Association Identifies Novel Etiological Insights Associated with Parkinson's Disease in African and African Admixed Populations

Mie Rizig PhD*[1,2], Sara Bandres-Ciga PhD*[3], Mary B Makarious BSc*[2,4], Oluwadamilola Ojo MD[5], Peter Wild Crea, BSc[3,4], Oladunni Abiodun, FWACP[6], Kristin S Levine, MS[3,7], Sani Abubakar, MBBS[8], Charles Achoru, MBBS[9], Dan Vitale, MS[7], Olaleye Adeniji, MBBS[10], Osigwe Agabi, MBBS[5], Mathew J Koretsky, BSc[3], Uchechi Agulanna, MBBS[11], Deborah A. Hall, MD, PhD[12], Rufus Akinyemi, PhD[13], Tao Xie, MD, PhD[14], Mohammed Ali, MBBS[15], Ejaz A. Shamim, MD[16,17,18], Ifeyinwa Ani-Osheku, FMCP[19], Mahesh Padmanaban, MD[14], Owotemu Arigbodi, MBBS[20], David G Standaert, MD, PhD[21], Abiodun Bello, FWACP[22], Marissa Dean, MD[21], Cyril Erameh, MBBS[23], Inas Elsayed, PhD[24], Temitope Farombi, MBBS[25], Olaitan Okunoye, PhD[1], Michael Fawale, MSc[26], Kimberley J Billingsley, PhD[3,4], Frank Imarhiagbe, MBCHB[27], Pilar Alvarez Jerez, BSc[1,3], Emmanuel Iwuozo, FMCP[28], Breeana Baker, BSc[3], Morenikeji Komolafe, MBBS[26], Laksh Malik, MFS[3], Paul Nwani, MBBS[29], Kensuke Daida, MD[3,4], Ernest Nwazor, FMCP[30], Abigail Miano-Burkhardt, BSc[3,4], Yakub Nyandaiti, MBBS[31], Zih-Hua Fang, PhD[32], Yahaya Obiabo, MBCHB[33], Jillian H. Kluss, PhD[4], Olanike Odeniyi, MBBS[34], Dena Hernandez, PhD[4], Francis Odiase, MBBS[27], Nahid Tayebi, PhD[35], Francis Ojini, MSc[5], Ellen Sidranksy, MD[35], Gerald Onwuegbuzie, MBBS[36], Andrea M. D'Souza, BSc[35], Godwin Osaigbovo, MBBS[9], Bahafta Berhe[35], Nosakhare Osemwegie, MBBS[37], Xylena Reed, PhD[3], Olajumoke Oshinaike, FWACP[38], Hampton Leonard, MS[3,7], Folajimi Otubogun, MBCHB[39], Chelsea X Alvarado, MS[3,7], Shyngle Oyakhire, MBBS[40], Simon Ozomma, FMCP[41], Sarah Samuel, MBBS[31], Funmilola Taiwo, MBCHB[25], Kolawole Wahab, MD[22,42], Yusuf Zubair, MSc[40], Hirotaka Iwaki, MD[3,7], Jonggeol Jeffrey Kim, BA[3,4], Huw R Morris, PHD FRCP,[1,2], John Hardy, PhD[1], Mike Nalls, PhD[7], Karl Heilbron, PhD[43], Lucy Norcliffe-Kaufmann, PhD[43], Nigeria Parkinson Disease Research Network, International Parkinson's Disease Genomics Consortium - Africa (IPDGC Africa), Black and African American Connections to Parkinson's Disease (BLAAC PD) Study Group, the 23andMe Research Team, Cornelis Blauwendraat, PhD[3,4], Henry Houlden, MD[1], Andrew Singleton, PhD[#,CA,3,4], Njideka Okubadejo, MD[#,CA,5] on behalf of the Global Parkinson's Genetics Program.

*joint first
[#]joint last
[CA]Corresponding

40
41
42 **Authorship Affiliations:**
43

44 1. Department of Neuromuscular Diseases, UCL Queen Square Institute of Neurology, London WC1N
45 3BG, UK
46 2. UCL Movement Disorders Centre, University College London, London, WC1N 3BG, UK
47 3. Center for Alzheimer's and Related Dementias (CARD), National Institute on Aging and National
48 Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA,
49 20814
50 4. Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda,
51 MD, USA
52 5. College of Medicine, University of Lagos, Idi Araba, Lagos State, Nigeria
53 6. General Hospital, Isolo, Lagos State, Nigeria
54 7. Data Tecnica International, Washington, DC, USA
55 8. Ahmadu Bello University, Zaria, Kaduna State, Nigeria
56 9. Jos University Teaching Hospital, Jos, Plateau State, Nigeria
57 10. Federal Medical Centre, Abeokuta, Ogun State, Nigeria
58 11. Lagos University Teaching Hospital, Idi Araba, Lagos State, Nigeria
59 12. Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA
60 13. Neuroscience and Ageing Research Unit, Institute for Advanced Medical Research and Training,
61 College of Medicine, University of Ibadan, Ibadan, Oyo State, Nigeria
62 14. Department of Neurology, University of Chicago Medicine, Chicago, Illinois, USA
63 15. Federal Teaching Hospital Gombe, Gombe State, Nigeria
64 16. Human Motor Control Section, National Institute of Neurological Disorders and Stroke, National
65 Institutes of Health, Bethesda, Maryland, USA
66 17. Kaiser Permanente Mid-Atlantic States, Largo, Maryland, USA
67 18. MidAtlantic Permanente Research Institute, Rockville, Maryland, USA
68 19. Asokoro District Hospital, Asokoro, Abuja, Nigeria
69 20. Delta State University, Abraka, Delta State, Nigeria
70 21. Department of Neurology, University of Alabama at Birmingham, Birmingham, AL, USA
71 22. University of Ilorin Teaching Hospital, Ilorin, Kwara State, Nigeria
72 23. Irrua Specialist Teaching Hospital, Irrua, Edo State, Nigeria

73 24. Faculty of Pharmacy, University of Gezira, Wadmadani, 20, Sudan
74 25. University College Hospital, Ibadan, Oyo State, Nigeria
75 26. Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria
76 27. University of Benin, Benin City, Edo State, Nigeria
77 28. Benue State University, Makurdi, Benue State, Nigeria
78 29. Nnamdi Azikiwe University Teaching Hospital, Nnewi, Anambra State, Nigeria
79 30. Rivers State University Teaching Hospital, Port Harcourt, Rivers State, Nigeria
80 31. University of Maiduguri Teaching Hospital, Maiduguri, Borno State, Nigeria
81 32. German Center for Neurodegenerative Diseases (DZNE), Tuebingen, Germany
82 33. Federal University of Health Sciences, Otukpo, Benue State, Nigeria
83 34. General Hospital, Lagos Island, Lagos State, Nigeria
84 35. Medical Genetics Branch, National Human Genome Research Institute, National Institutes of
85 Health, Bethesda, Maryland, USA
86 36. University of Abuja, Abuja, Federal Capital Territory, Nigeria
87 37. University of Port Harcourt, Port Harcourt, Rivers State, Nigeria
88 38. Lagos State University College of Medicine, Ikeja, Lagos State, Nigeria
89 39. Federal Medical Center, Ebute Metta, Lagos State, Nigeria
90 40. National Hospital, Abuja, Federal Capital Territory, Nigeria
91 41. University of Calabar Teaching Hospital, Calabar, Cross River State, Nigeria
92 42. University of Ilorin, Ilorin, Kwara State, Nigeria
93 43. 23andMe, Inc., Sunnyvale, CA, USA
94
95 **Correspondence to:**
96 Njideka Okubadejo, MD | nokubadejo@unilag.edu.ng
97 Professor & Consultant Neurologist
98 College of Medicine, University of Lagos & Lagos University Teaching Hospital, Idi Araba, Lagos State,
99 Nigeria
100
101 Andrew Singleton, PhD | singleta@nih.gov
102 Director, Center for Alzheimer's and Related Dementias
103 National Institute on Aging and National Institute of Neurological Disorders and Stroke, National
104 Institutes of Health, Bethesda, MD, USA
105
106 **Keywords:** genetics, Parkinson's disease, genome-wide association study, African, African Admixed,
107 *GBA1,* expression quantitative trait locus, therapeutic interventions.

108 **Funding:**
109 Data used in the preparation of this article were obtained from Global Parkinson's Genetics Program
110 (GP2). GP2 is funded by the Aligning Science Across Parkinson's (ASAP) initiative and implemented by
111 The Michael J. Fox Foundation for Parkinson's Research (https://gp2.org). For a complete list of GP2
112 members see https://gp2.org. Additional funding was provided by The Michael J. Fox Foundation for
113 Parkinson's Research through grant MJFF-009421/17483.
114
115
116
117
118
119
120
121
122
123
124
125
126
127

Summary

Background

Understanding the genetic mechanisms underlying diseases in ancestrally diverse populations is a critical step towards the realization of the global application of precision medicine. The African and African admixed populations enable mapping of complex traits given their greater levels of genetic diversity, extensive population substructure, and distinct linkage disequilibrium patterns.

Methods

Here we perform a comprehensive genome-wide assessment of Parkinson's disease (PD) in 197,918 individuals (1,488 cases; 196,430 controls) of African and African admixed ancestry, characterizing ancestry-specific risk, differential haplotype structure and admixture, coding and structural genetic variation and polygenic risk profiling.

Findings

We identified a novel common risk factor for PD and age at onset at the *GBA1* locus (risk, rs3115534-G; OR=1.58, 95% CI = 1.37 - 1.80, P=2.397E-14; age at onset, BETA =-2.004, SE =0.57, P = 0.0005), that was found to be rare in non-African/African admixed populations. Downstream short- and long-read whole genome sequencing analyses did not reveal any coding or structural variant underlying the GWAS signal. However, we identified that this signal mediates PD risk via expression quantitative trait locus (eQTL) mechanisms. While previously identified *GBA1* associated disease risk variants are coding mutations, here we suggest a novel functional mechanism consistent with a trend in decreasing glucocerebrosidase activity levels. Given the high population frequency of the underlying signal and the phenotypic characteristics of the homozygous carriers, we hypothesize that this variant may not cause Gaucher disease. Additionally, the prevalence of Gaucher's disease in Africa is low.

Interpretation

The present study identifies a novel African-ancestry genetic risk factor in *GBA1* as a major mechanistic basis of PD in the African and African admixed populations. This striking result contrasts to previous work in Northern European populations, both in terms of mechanism and attributable risk. This finding highlights the importance of understanding ancestry-specific genetic risk in complex diseases, a particularly crucial point as the field moves toward precision medicine in PD clinical trials and while recognizing the need for equitable inclusion of ancestrally diverse groups in such trials. Given the distinctive genetics of these underrepresented populations, their inclusion represents a valuable step towards insights into novel genetic determinants underlying PD etiology. This opens new avenues towards RNA-based and other therapeutic strategies aimed at reducing lifetime risk.

Research in Context

Evidence Before this Study

Our current understanding of Parkinson's disease (PD) is disproportionately based on studying populations of European ancestry, leading to a significant gap in our knowledge about the genetics, clinical characteristics, and pathophysiology in underrepresented populations. This is particularly notable in individuals of African and African admixed ancestries. Over the last two decades, we have witnessed a revolution in the research area of complex genetic diseases. In the PD field, large-scale genome-wide association studies in the European, Asian, and Latin American populations have identified multiple risk loci associated with disease. These include 78 loci and 90 independent signals associated with PD risk in the European population, nine replicated loci and two novel ancestry-specific signals in the Asian population, and a total of 11 novel loci recently nominated through multi-ancestry GWAS efforts. Nevertheless, the African and African admixed populations remain completely unexplored in the context of PD genetics.

Added Value of this Study

To address the lack of diversity in our research field, this study aimed to conduct the first genome-wide assessment of PD genetics in the African and African admixed populations. Here, we identified a genetic risk factor linked to PD etiology, dissected African-specific differences in risk and age at onset, characterized known genetic risk factors, and highlighted the utility of the African and African admixed risk haplotype substructure for future fine-mapping efforts. We identified a novel disease mechanism via expression changes consistent with decreased *GBA1* activity levels. Future large scale single cell expression studies should investigate the neuronal populations in which expression differences are most prominent. This novel mechanism may hold promise for future efficient RNA-based therapeutic strategies such as antisense oligonucleotides or short interfering RNAs aimed at preventing and decreasing disease risk. We envisage that these data generated under the umbrella of the Global Parkinson's Genetics Program (GP2) will shed light on the molecular mechanisms involved in the disease process and might pave the way for future clinical trials and therapeutic interventions. This work represents a valuable resource in an underserved population, supporting pioneering research within GP2 and beyond. Deciphering causal and genetic risk factors in all these ancestries will help determine whether interventions, potential targets for disease modifying treatment, and prevention strategies that are being studied in the European populations are relevant to the African and African admixed populations.

Implications of all the Available Evidence

We nominate a novel signal impacting *GBA1* as the major genetic risk factor for PD in the African and African admixed populations. The present study could inform future *GBA1* clinical trials, improving patient stratification. In this regard, genetic testing can help to design trials likely to provide meaningful

and actionable answers. It is our hope that these findings may ultimately have clinical utility for this underrepresented population.

Introduction

Parkinson's disease (PD) is a complex, heterogeneous neurodegenerative disorder that manifests with progressive motor and non-motor features, including resting tremor, bradykinesia, mood disorders, olfactory dysfunction, and cognitive impairment. Globally, about 6.1 million people had PD in 2016[1], and as a result of an aging world population and increased longevity, this figure is expected to rise to 17.5 million by 2040[2] as a result of an aging world population and increased longevity.

Genome-wide association studies (GWAS) have been instrumental for identifying common variants associated with complex diseases like PD, unraveling the genetics and heritability of PD in European populations[3]. The largest published GWAS meta-analysis of PD risk to date was performed on individuals of European ancestry and identified 90 independent genome-wide significant risk signals that explain 16-22% of the heritable risk of PD[4,5]. However, very little is known about the genetics of PD in non-European populations. The largest PD GWAS meta-analysis in the East Asian population recently identified two ancestry-specific signals,[6] and the first PD GWAS in Latin Americans has suggested two potential novel loci that warrant further study[7]. The first multi-ancestry PD GWAS meta-analysis has nominated 11 novel loci, providing a foundation for future efforts aimed at fine-mapping novel genetic regions linked to PD[8]. GWAS are powerful tools in the creation of better prediction models and broadening our biological knowledge of specific diseases[9].

Nearly one-third of the genetic heritability of PD can be explained by polygenic risk scores (PRSs) according to the most recent genetic studies conducted in Europeans. However, the heritability explained by PRSs is totally unknown in under-researched and underserved populations, as is the total heritability[10,11]. There has been considerable ethnic variability in the distribution of monogenic causes and genetic risk variants documented across populations. For instance, the relatively common LRRK2 p.G2019S mutation remains unreported in some sub-Saharan African populations, despite being most commonly associated with familial and sporadic PD in Zambia and Northern Africa[12–16].

African and African admixed populations offer unique opportunities for studying the genetics of both monogenic and complex diseases because they contain the largest portion of the within-population genetic variability in the world, shorter linkage disequilibrium (LD) blocks, and abundant alleles that are private to these populations[17,18,19]. Africa is an ethnically diverse continent, with several ethno-linguistic groups across the geographical regions of Africa. The West African population (of which Nigeria is the largest) belongs to the Niger-Congo phylum, whereas, for instance, the North African and Northeast African populations are predominantly Afroasiatic and Nilo-Saharan[20]. High quality data on the prevalence of PD in Africans remains sparse. The Global Burden of Disease (GBD) 2016 data catalog indicates that the age-standardized prevalence rates of PD are lowest in sub-Saharan Africa (30 to <60

247 per 100,000 population). In contrast, higher rates are reported for North Africa, where the prevalence is
248 more similar to that of Europe and the Middle East[1,21,22,23]. PD prevalence exhibits a male preponderance
249 and is about 1.4 times more frequent in males according to the 2016 GBD estimates[1]. Previous studies
250 from Africa are in concordance with these estimates but generally report a higher male to female ratio,
251 between 1.32 and 1.39 overall male:female prevalence ratios [24–26]. However, most of the data is derived
252 from hospital-based studies with the inherent bias of hospital attendance and health seeking for non-
253 obstetric indications being higher in men due to a combination of social and cultural factors. Delayed
254 diagnosis at a later stage of disease characterises the clinical scenario in sub-Saharan Africa in which
255 motor and non-motor manifestations seem to be similar as compared to other populations[27,28]. In
256 contrast, African Americans have been reported to be more likely to have higher rates of cognitive
257 impairment as compared to Whites as well as being less likely to have parkinsonism[29] and less
258 medication use[30].
259
260 In addition to promoting scientific equity to address health disparities, diverse representation provides a
261 platform for replication studies to explore the strength and relevance of findings reported from other
262 populations. Additionally it has the potential to facilitate the identification of novel or unique loci and
263 investigate genotype-phenotype correlations that can further expand our understanding of pathological
264 and pathogenetic disease mechanisms in PD[17,31].
265
266 This study provides the first GWAS-based insights into the genetics of PD in the African and African
267 admixed populations **(Figure 1)**. Here we performed a comprehensive genome-wide assessment of PD
268 risk and age at onset, characterizing population specific cumulative risk profiling, haplotype structure,
269 and genetic admixture. Leveraging this unique population genetic structure, our analyses identified a
270 novel association signal in *GBA1, t*he gene encoding the lysosomal enzyme glucocerebrosidase (GCase).
271 This led to the investigation of a novel disease mechanism of expression changes consistent with
272 decreased glucocerebrosidase activity levels relating to increased risk. Finally, we compare our findings
273 in the context of other global populations. We envisage that these data generated under the umbrella of
274 the Global Parkinson's Genetics Program (GP2) will shed light on the molecular mechanisms involved in
275 the disease process and might pave the way for future RNA-based therapeutic strategies aimed at
276 reducing lifetime risk.

277 Methods

278 Study participants

279 The demographic and clinical characteristics of the cohorts under study are provided in **Table 1.** Cohorts,
280 in the context of this study, are defined as groups of individuals with similarly predicted ancestry that
281 have been genotyped, imputed, and processed following the same quality control parameters. Three
282 sources of data were included in this study: Individual level data from the International Parkinson's
283 Disease Genomics Consortium - Africa *(IPDGCAN)* and the Global Parkinson's Disease Genetics Program
284 *(GP2),* and GWAS summary statistics from *23andMe, Inc*. The samples provided from efforts in Africa are

285 predominantly from West Africa, specifically Nigeria, and therefore unlikely to be representative of the
286 entirety of Africa. Some of the individuals predicted to be of African descent cannot with certainty be
287 defined as from Nigeria, but nonetheless unmistakably African (**Supplementary Figure 3**). Additionally,
288 we define African admixed as individuals ancestrally similar to the following 1000 Genomes ancestry
289 labels: African ancestry in Southwest United States of America (abbreviated as ASW in the 1000
290 Genomes project) and African Carribean in Barbados (abbreviated as ACB in the 1000 Genomes project).
291
292 For the *IPDGCAN* and the *GP2* cohorts, the diagnosis of PD was based on fulfillment of the United
293 Kingdom PD Society Brain Bank criteria (excluding the requirement for not more than one affected
294 relative)[32]. The respective ethical committees for medical research approved involvement in genetic
295 studies, and participants gave informed written consent. All participants underwent a neurological
296 examination conducted by a study neurologist to document clinical and neurological status. Controls
297 were generally assessed to detect overall signs of neurological condition and samples presenting any
298 clinical signs of neurodegenerative diseases were excluded from the control series.
299
300 Summary statistics for individuals with or without PD were provided through a collaborative agreement
301 with 23andMe, Inc. Participants provided informed consent and volunteered to participate in the
302 research online, under a protocol approved by the external AAHRPP-accredited IRB, Ethical &
303 Independent (E&I) Review Services. As of 2022, E&I Review Services is part of Salus IRB
304 (https://www.versiticlinicaltrials.org/salusirb). PD diagnosis was self-reported in this instance. PD
305 patients were recruited through a targeted email campaign in conjunction with the Michael J. Fox
306 Foundation, The Parkinson's Institute and Clinical Center, and many other PD patient groups and clinics.
307 Emails or hard copy mailings were sent to all individuals who had registered with these groups as PD
308 patients. PD cases were individuals who self-reported having been diagnosed with PD. Individuals
309 reporting a change in diagnosis or uncertainty about their diagnosis were removed from further
310 analyses. It has been previously shown that self- reported PD case status is approximately as reliable as
311 clinically-diagnosed PD with 50 out of 50 cases confirmed via telemedicine interview[33]. Individuals who
312 self-reported having ever been diagnosed with an atypical parkinsonism (e.g., dementia with Lewy
313 bodies, progressive supranuclear palsy, multiple system atrophy, corticobasal degeneration) or a non-
314 parkinsonian tremor disorder were removed. Additionally, a previous PD GWAS meta-analysis found a
315 strong genetic correlation between 23andMe GWAS data using self-reported cases and non-23andMe
316 GWAS data with PD cases ascertained by clinicians (genetic correlation from LDSC (rG) = 0.85, SE =
317 0.06)[4]. Age distributions of cohorts under study are illustrated in **Supplementary Figure 1.**

318 Genotype data generation, quality control, ancestry predictions, and imputation

319 *The IPDGCAN* and *GP2* samples were genotyped using two different genotyping platforms **(Table 1).** The
320 NeuroBooster array (v.1.0, Illumina, San Diego, CA) contains a backbone of 1,914,935 variants densely
321 covering ancestry informative markers, markers for determination of identity by descent, and X-
322 chromosome SNPs for sex determination. In addition, it contains 96,517 customized variants. Samples
323 collected as part of the *GP2* initiative were genotyped on this array. Samples collected as part of the
324 *IPDGCAN* initiative **(Table 1)** were genotyped using two different platforms; the Neurochip array,

containing a backbone of 306,670 variants and customized content comprising 179,467 variants[34], and the previously described NeuroBooster array.

Raw genotype data was passed through a custom ancestry prediction and pruning machine learning method as a part of the GenoTools pipeline (https://github.com/GP2code/GenoTools), as described elsewhere[35]. All samples underwent similar standardized quality control (QC) as follows: Samples were excluded from the analysis if: call rate was <95%, genetically determined sex did not match that from clinical data, or excess heterozygosity was detected (|F| statistics > 0.25).

Samples were subset by ancestry estimates (see **Supplementary Materials** for details). In brief, ancestry was defined using reference panels from the 1000 Genomes Project[36], Human Genome Diversity Project[37], and an Ashkenazi Jewish population dataset[38]. In total, 39,302 reference panel SNPs were genotyped on the NeuroBooster array and 24,404 reference panel SNPs were also genotyped on the NeuroChip array (see **Supplementary Materials** for details). Ancestry estimates were carried out using a uniform protocol across all samples.

Next, we removed those samples that were IBD for > 12.5% of the genome (approximately related at a first cousins level or closer). Once preliminary sample-level QC was completed, SNPs with Hardy-Weinberg Equilibrium (HWE) P value <1E-4 in control samples were removed. Next, variants were pruned for missingness by case-control status at P≤1E-4 to remove variants with non-random missingness. Finally, variants were pruned for non-random missingness by haplotype at P≤1E-4.

For the *GP2* data, variants were further filtered by minor allele frequency (MAF) < 0.005 and HWE P < 1E-5 prior to being submitted to the TOPMed Imputation server. The TOPMed reference panel version r^2 contains information from 97,256 reference samples and more than 300 million genetic variants across the 22 autosomes and the X chromosome. As of October 2022, the TOPMed panel consists of about 180,000 participants of which 29% are of African ancestry, 19% of Latin American ancestry, 8% of Asian ancestry, and 40% of European ancestry. More information about the TOPMed Study[39], Imputation Server[40], and Minimac Imputation[41] can be found at https://imputation.biodatacatalyst.nhlbi.nih.gov. The imputed files were then pruned applying a minor allele count (MAC) threshold of 10 and an imputation Rsq of 0.3. For additional information regarding *GP2* ancestry prediction as well as 23andMe data generation and processing, please see the **Supplementary Materials.**

Estimation of PD risk, age at onset and admixture

To estimate risk associated with PD, imputed dosages (meaning genotype probabilities for a variant to be A/A, A/B, or B/B from 0 to 2 that account for some uncertainty) were analyzed using a logistic regression model adjusted for sex, age, and the first ten PCs as covariates. The PCs were fit on the set of overlapping SNPs between the datasets and the reference panels before being transformed by UMAP to represent the population substructure (see **Supplementary Materials).** Age at onset (AAO) was used for cases and age at recruitment was used for controls. In instances where AAO was not available for cases, age at recruitment was used instead (less than 6% of individuals). For individuals who had no age

365 information provided, average age was imputed (less than 5% and 2% of cases and controls,
366 respectively). To explore the influence of genetic variation on the AAO of PD cases, a linear regression
367 model adjusted for the same covariates was performed. Here, AAO was defined as the self-reported
368 date of first motor symptom. Additionally, we conducted linear regression analyses to explore how
369 potential GWAS signals would correlate with admixture levels. All the analyses were performed on Terra
370 (https://terra.bio/). GWAS was conducted on African and African admixed ancestries independently and
371 then meta-analyzed. We utilized fixed-effects meta-analyses as implemented in METAL[43] to leverage
372 summary statistics across all sources. Pairwise LD values were calculated using 1000 genomes African
373 population data through LD link (https://ldlink.nci.nih.gov/?tab=home).

374 Haplotype and fine-mapping analyses

375

376 Haplotype size was compared using individual level data across African, African admixed, and European
377 PD cases. After standardizing the three datasets with the same genotyped SNPs passing identical QC
378 steps, we determined the size of the haplotype blocks using default parameters in PLINK 1.9. This
379 analysis estimates haplotype blocks by Haploview's interpretation of the block definition. By default,
380 only pairs of variants within 200 kilobases (kb) of each other were considered. Two variants are
381 considered by this procedure to be in strong LD if the lower bound of the 90% D-prime confidence
382 interval (CI) was >0.70, and the upper bound of the CI was at least 0.98.
383
384 In an attempt to prioritize putative causal variants within the identified *GBA1* risk haplotype, we
385 performed fine-mapping analyses across the LD block where the genome-wide signal was located by
386 using the "Approximate Bayes Factor fine mapping under a single causal variant assumption" method
387 provided by the R package coloc (https://CRAN.R-project.org/package=coloc). This analysis assesses the
388 posterior probability of each SNP being the causal variant within a locus. We derived posterior
389 probabilities (PP) for this region using the default prior probability of 1E-4 under the assumption of a
390 single causative variant per locus.

391 Short-read Whole Genome Sequencing

392

393 To further dissect the novel identified GWAS signal, we performed whole-genome sequencing (WGS)
394 analyses in 206 individuals (141 cases and 65 controls) of which 39 individuals were *GBA1* rs3115534-GG
395 carriers, 69 were rs3115534-GT and 98 were rs3115534-TT carriers. Short-read WGS DNA sequencing
396 was performed by Psomagen (detailed in Supplementary methods). We used the functional equivalence
397 pipeline[44] implemented at the Broad Institute to produce alignments and small variant calls against the
398 GRCh38DH reference genome. For sample-level WGS quality control, we followed the quality metrics
399 defined by the Accelerating Medicines Partnership Parkinson's Disease initiative (AMP-PD; https://amp-
400 pd.org)[45]. To produce a set of joint-genotyped variants for all the samples that passed quality control,
401 we ran the Broad Institute's joint discovery pipeline and retained only the high-quality variants flagged
402 as "PASS" after variant quality score recalibration, with a call rate > 0.95, genotype quality >20, read

depth >5, and heterozygous allele balance between 0.25 and 0.75 as described previously [46]. Additionally, we called *GBA1* variants using Gauchian v1.0.2[47] and genotyped known neurological repeat expansions using STRipy v2.2[48]. All the pipelines and scripts used are available via GitHub (https://github.com/GP2code). Data passing quality control metrics were annotated using ANNOVAR[49]. A comprehensive assessment of known and potential novel pathogenic variants driving the *GBA1* signal was performed. CRAM files were visualized using the Integrative Genomics Viewer (IGV) web browser[50].

The Gauchian algorithm[47] was then applied to nominate potential structural variants driving the *GBA1* signal. Briefly, this algorithm is a targeted variant caller for the *GBA1* gene based on WGS BAM files. Gauchian aims to solve the problems caused by the high sequence similarity with the pseudogene paralog *GBAP1*. This algorithm has been reported to be able to detect variants in the exons 9-11 homology region, such as large deletions or duplications between *GBA1* and *GBAP1*, and *GBAP1*-like variants in *GBA1*, including p.A495P, p.L483P, p.D448H, c.1263del, RecNciI, RecTL and c.1263del+RecTL.

Long-read Whole Genome Sequencing

Oxford Nanopore Technologies (ONT) long-read whole-genome sequencing data was generated for five *GBA1* rs3115534-GG carriers, two heterozygotes and six *GBA1* rs3115534-TT carriers. High molecular weight DNA was extracted from either frozen blood samples or cell-lines. For the blood samples DNA was extracted from 1ml per sample using the Kingfisher APEX instrument with the Nanobind CBB Big DNA kit (HBK-CBB-001). For the frozen cell-pellets DNA was extracted manually with the Nanobind CBB Big DNA kit (HBK-CBB-001) using the following protocol [51](https://dx.doi.org/10.17504/protocols.io.q26g74169gwz/v1).

The DNA then went through a size selection step using the Circulomics Short Read Eliminator Kit (SS-100-101-01) to remove fragments up to 25kb. Finally a library was prepared with the SQK-LSK 110 Ligation Sequencing Kit from ONT and each library was loaded onto a separate PromethION R9.4.1 flow cell following ONT standard operating procedures and ran for a total of 72 hours on a PromethION device.

Fast5 files containing raw signal data were obtained from sequencing performed using minKNOW v22.10.7 (ONT). All fast5 files were used to conduct super accuracy basecalling on each sample with Guppy v6.12. Fastq files that passed quality control filters in the super accuracy base calling step were then mapped to the GRChg38 reference genome using winnowmap v2.03[52]. Structural variants were called with Sniffles2[53] v2.0.3 using default parameters and the "–tandem-repeats'' option.

Glucocerebrosidase activity

Patient-derived lymphoblastoid cell lines (LCLs) were obtained from the Coriell repository (https://www.coriell.org/). LCLs were maintained as directed in suspension with RPMI 1640 (ThermoFisher Scientific, 11875093) containing 2mM Glutamax (ThermoFisher Scientific, 35050061), and 15% FBS (ThermoFisher Scientific, A3160501) at 37°C in 5% CO_2. Protein was extracted from LCLs using a citrate-phosphate buffer (0.2 M Na_2HPO_4, 0.1 M citrate, protease inhibitor, pH 5.8, Millipore Sigma, 11836170001) that was activated with 0.25% Triton X-100. Cells were subjected to a 4-methylumbelliferone (4-MU, Sigma Aldrich, M1381) fluorometric glucocerebrosidase (GCase) activity assay in quadruplicate as previously reported in the literature[54] with adjusted incubation time of 2.5 hours. A total of 5E6 cells were used per sample with protein concentrations normalized to 0.7 mg/ml via BCA Protein Assay (Thermo Fisher Scientific 23225).

Role of Funding Source

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. Additional funding was provided by The Michael J. Fox Foundation for Parkinson's Research through grant MJFF-009421/17483.

Results

GWAS reveals a novel genome-wide significant signal associated with PD risk and age at onset

We first performed a GWAS of PD risk in the African population, predominantly consisting of individuals of Nigerian descent which included a total of 997 PD cases and 1,294 controls. Of these individuals, 693 PD cases and 1,009 controls were genotyped on the NeuroBooster array, and 304 PD cases and 285 controls were screened on the NeuroChip array (λ=1.01; **Supplementary Figure 4**). A genome-wide significant SNP at the *GBA1* locus was associated with an increase in PD risk; rs3115534, a variant located in intron 8 of *GBA1* (34 nucleotides upstream of exon 9) was the top hit **(Supplementary Table 1, Supplementary Figure 4;** rs3115534; OR=1.58; 95% CI = 1.35 - 1.84, P=3.44E-09). Contrary to what we would expect when assessing common variation linked to PD risk (MAF > 5%), a high odds ratio was identified for this signal. Our study indicated that each additional risk allele, G, conferred a 1.58 increase in the odds of PD.

In parallel, we performed a GWAS in the African admixed population, leveraging the African-American and Afro-Caribbean datasets available as a part of the GP2 initiative combined with 23andMe African-American summary statistics. The PD African admixed GWAS included a total of 467 PD cases and 195,120 controls (λ=1.01; **Supplementary Figure 5**). No genome-wide significant hits were nominated.

479 Next, we performed a GWAS meta-analysis of all of the African and African admixed datasets **(Figure 2),**
480 totaling 1,488 cases and 196,430 controls. This revealed that a total of 35 SNPs near the *GBA1* gene
481 were significantly associated with PD risk with consistent directionality of effect, the two most distant
482 SNPs being 639,773 base pairs apart from each other. Conditional analyses on the top two SNPs
483 suggested that there is only one causal signal driven by rs3115534 as the leading SNP. Of note,
484 rs3115534-G is much more common in individuals of African or African admixed ancestry relative to
485 other populations; allele frequency = 0.16 according to gnomAD[55] and allele frequency = 0.21 according
486 to the African 1000 Genomes panel [36]. The African and African admixed datasets used in this study
487 yielded similar frequencies (African dataset; cohort MAF = 0.25, affected MAF = 0.33, unaffected MAF =
488 0.19), (African admixed datasets; cohort MAF = 0.14, affected MAF = 0.22, unaffected MAF = 0.13).
489 Within our research cohorts, we found that rs3115534-G was more frequent in Nigerian populations
490 **(Supplementary Table 3).** We performed power calculations to forecast the sample size needed to
491 achieve genome-wide significance for rs3115534 considering an effect estimate of OR = 1.58 and
492 assuming a disease prevalence of 0.6% (as per the Global Burden of Disease Study) for the following
493 minor allele frequencies: African cohort MAF = 0.25; African admixed MAF = 0.14 and average African
494 and African admixed MAF = 0.195; (Supplementary Figure 13). Our calculations showed a power <50 %
495 for the African admixed cohort confirming that we were underpowered to detect such a signal when
496 conducting GWA studies specifically on this dataset. Estimates showed power close to 70% to nominate
497 rs3115534 as a GWAS hit in the African cohort alone. By meta-analysing all cohorts we were able to
498 reach adequate power.
499
500 Linear regression analyses showed that the *GBA1* rs3115534 variant was positively associated with the
501 genome-wide percentage of African ancestry (BETA = -0.001, SE= 0.0005, P= 0.011).
502
503 We tested whether the effect of the risk allele was additive by calculating the frequency of homozygotes
504 for the risk allele and heterozygotes in cases versus controls. Notably, our analyses conducted on
505 individual level data from *IPDGCAN* and *GP2* showed that rs3115534-GG was 3.39 times more frequent
506 in African cases (130/1015) than controls (49/1296) and 3.80 times more frequent in African admixed
507 cases (11/185) than controls (18/1149), while rs3115534-GT was 1.17 times more frequent in African
508 cases (398/1015) than controls (435/1296) and 1.38 times more frequent in African admixed cases
509 (61/185) than controls (274/1149). Zygosity analysis of 23andMe data showed that rs3115534-GG was
510 1.92 times more frequent in African admixed cases (10/288) than controls (3,537/193,985) while
511 rs3115534-GT was 1.27 times more frequent in African admixed cases (85/288) than controls
512 (44,967/193,985). We also analyzed rs3115534 under a dominant model (African ancestry - dominant
513 model: OR = 1.74; 95% CI = 1.40 - 2.15; P = 3.467E-07; African admixed ancestry - dominant model: OR =
514 1.96; 95% CI = 1.40 - 2.75; P =7.65E-5). Despite the large differences observed in frequencies, effect
515 estimates from the additive model are extremely similar to the dominant model with largely overlapping
516 confidence intervals. This suggests that this variant is additive, and not increasing the risk for PD
517 following a dominant inheritance pattern (African ancestry - additive model: OR =1.75; 95% CI = 1.47 -
518 2.07, P = 1.40E-10; African admixed ancestry - additive model: OR =1.95; 95% CI =1.47 - 2.60; P =4.12E-
519 6).
520

521 As a follow-up analysis, we assessed whether this *GBA1* variant is associated with AAO. Linear regression
522 analyses in 711 African ancestry cases and 185 African admixed ancestry cases showed that *GBA1*
523 rs3115534-G is also an AAO disease modifier (African ancestry: BETA =-2.004, SE = 0.57, P = 0.0005;
524 African-admixed: BETA = -4.15, SE =0.58, P =0.015; Meta-analysis: BETA =-3.06, SE =0.40, P = 0.008)
525 resulting in onset of PD three years earlier per risk allele **(Supplementary Figure 7).** The African-admixed
526 estimates should be taken with caution due to small sample size and low number of GG carriers. No
527 differences in age at PD onset were found between *GBA1* rs3115534-GG and *GBA1* rs3115534-GT
528 carriers (T-test; P = 0.25).
529

530 Genome-wide comparison of the *GBA1* locus across populations suggests an African 531 founder effect

532
533 In an attempt to further dissect the novel signal identified in the *GBA1* locus, we next compared effect
534 estimates and directionality of effect leveraging summary statistics from the largest PD GWAS meta-
535 analysis of PD in Europeans[4], Latin American[7], and East Asian populations[6]. The rs3115534-G allele is
536 extremely rare in Europeans (allele frequency = 0.0015), East Asians (allele frequency = 0.0005), South
537 Asians (allele frequency =0.0017), and Ashkenazi Jewish populations (allele frequency = 0.0009)
538 according to gnomAD.
539
540 When looking at *GP2* European data, the rs3115534 variant was found to be poorly imputed in 13,186
541 samples (R^2 = 0.16, MAF = 0.009). In fact, the *GBA1* locus in African and African admixed populations
542 differs substantially from Europeans **(Figure 3; Supplementary Figure 8),** whose association with disease
543 risk is driven by two independent signals, including rs35749011 (*GBA1*-E326K) and rs76763715 (*GBA1*-
544 N370S). These variants are very rare in individuals of African and African admixed ancestry **(Figure 4B).**
545 Similarly, the *GBA1* locus considerably differs from the East Asian population, for which the rs3115534
546 variant was also not imputed in the largest East Asian GWAS meta-analysis[6] **(Figure 4C).** These
547 differences are less noticeable when assessing the Amerindian and indigenous populations, which
548 harbor higher levels of African admixture **(Figure 4D)** (Loesch et al. GWAS[7]; rs3115534-G; OR = 1.13, 95%
549 CI =0.41-1.86, P= 0.72; Amerindian and indigenous 23andMe GWAS; rs3115534-G; OR = 1.56, 95% CI =
550 1.55-1.88, P= 0.01).
551
552 Furthermore, we assessed the rs3115534-G variant on individual level data from the *GP2* initiative. The
553 variant was not imputed in individuals of European, Ashkenazi Jewish, South Asian, East Asian and
554 Central Asian ancestries, likely due to its low frequency. On the other hand, the rs3115534-G variant was
555 imputed in 230 cases and 182 controls of Amerindian and indigenous ancestry (MAF = 0.027; P = 0.43).
556 Notably, linear regression analyses versus genomic admixture revealed that rs3115534-G was positively
557 correlated with percentage of African ancestry (BETA = 0.064, SE = 0.024, P = 0.01), confirming an
558 African founder effect. At consensus genotyped variants, haplotype size at the *GBA1* risk locus spanning
559 the rs3115534 variant substantially differed across populations when comparing African, African
560 admixed and European PD cases from the *GP2* initiative (European haplotype length = 79.19, European
561 N SNPs = 90; African haplotype length = 19.30, African N SNPs = 29; African admixed haplotype length =

15.15, African admixed N SNPs = 22). Interestingly, the larger sub-African population haplotypes spanning the rs3115534 variant were found in the Esan and the Yoruba in Ibadan (Nigerian) populations according to 1000 Genomes **(Supplementary Figure 9),** suggesting that this haplotype might have originated in these populations, given that founder effects result in decreased genetic diversity and therefore larger haplotype block sizes. Fine-mapping analyses showed the lead SNP had a PP of 71.4% (rs3115534; **Supplementary Table 4).**

Short- and long-read whole genome sequencing did not identify any coding or structural variant explaining the novel signal at *GBA1*

In an effort to identify a functional coding variant undetectable through genotyping or imputation that could explain the novel GWAS signal, we conducted WGS short-read analyses on a total of 206 individuals (141 cases and 65 controls) of which 39 individuals were *GBA1* rs3115534-GG carriers, 69 were rs3115534-GT and 98 were rs3115534-TT carriers. A 96.6 % correlation was observed between WGS-short read and imputed genotyped data for rs3115534, validating the high quality of our imputed data. No differences in coding variation were observed between carriers and non-carriers of the GWAS signal **(Table 2).** We next applied the Gauchian algorithm, a targeted variant caller for the *GBA1* gene based on WGS BAM files. Gauchian aims to solve the problems caused by the high sequence similarity with the pseudogene paralog *GBAP1 (see methods)*. The Gauchian algorithm did not identify any genetic rearrangement that could explain this signal. Then, Oxford Nanopore Technologies (ONT) WGS long-read sequencing data was generated for a total of five rs3115534-GG PD cases, two rs3115534-GT and six rs3115534-TT controls. Long-read data was compared to short-read WGS for a known structural variant carrier that was previously reported in African American populations in 2000 by Tayebi and colleagues **(Supplementary Figure 10)**[56]. No structural variants explaining this signal were identified. Splice prediction tools (www.phenosystems.com) predicted no significant impact on normal splicing.

Expression quantitative trait locus analysis provides novel mechanistic insights into risk at the *GBA1* locus

We leveraged existing whole blood expression quantitative trait locus (eQTL) summary statistics from Mak et al., 2021 based on RNA sequencing from 2,733 samples of predominantly African American and Indigenous American ancestries[57]. Of note, we identified a strong eQTL signal at rs3115534, located 8,821 bp from the canonical transcription start site (**Figure 5;** MAF = 0.15; P= 9.99E-25, BETA = 0.238, SE = 0.022). The rs3115534-G risk allele was found to be associated with increased *GBA1* expression levels. We questioned whether this observation could be explained by the existence of multi-mapping reads between *GBA1* and its pseudogene, *GBAP1*, which are often discarded in standard processing and do not contribute to gene-level quantification of expression in many publicly available datasets like GTEx (https://gtexportal.org/). Gustavsson et al., reported that only 42% of all reads mapping to *GBA1* did so uniquely, with the remaining reads mapping primarily to *GBAP1*[58]. This resulted in a significant misestimation of the relative expression of *GBA1* to *GBAP1.* The authors demonstrated the ability of

these transcripts to generate stable protein that lacked lysosomal GCase function, which would support our hypothesis. Indeed, transcript diversity is a common and known biological phenomena that could explain the fact that rs3115534-G may increase the expression of a non-functional transcript that in turn would decrease the levels of the transcript responsible for optimal production of the protein isoform with GCase activity. Our data suggests a decreasing trend in GCase activity estimates when comparing rs3115534-GG homozygous risk allele (762.50 ± 273.50 U) versus rs3115534-GT heterozygous carriers (2743.76 ± 1960.83 U); (Welch Two Sample t-test - GG versus GT; t = -4.3138, df = 21.583, p-value = 0.00029) and rs3115534-TT homozygous non-risk allele carriers (1879.94 ± 1010.84 U) versus rs3115534-GG homozygous risk allele carrier; (Welch Two Sample t-test - GG versus TT; t = -4.7564, df = 18.363, p-value = 0.00014). Furthermore, in PD cases alone, the trend in GCase activity between rs3115534-GG homozygous risk allele carriers (762.50 ± 273.50 U), rs3115534-GT heterozygous carriers (3749.47 ± 2620.82 U) and rs3115534-TT homozygous non-risk allele carriers (1976.20 ± 1415.99 U) remained consistent with rs3115534-GG homozygous risk allele displaying the lowest activity; (Welch Two Sample t-test: GG versus GT; t = -3.189, df = 7.3002, p-value = 0.01446; GG versus TT; t = -2.8158, df = 13.003, p-value = 0.01458; GT versus TT; t = 1.7509, df = 9.7545, p-value = 0.1113). All samples were screened for known *GBA1* pathogenic mutations that could bias these estimates. A total of two carriers (one heterozygous for *GBA1* p.I320S and one heterozygous for *GBA1* p.T75del) were removed from our analyses. We assume the limitation that LCLs were only available for one homozygous risk allele. Further research is needed to corroborate this hypothesis and understand the functional consequences of this variant in disease etiology **(Supplementary Figure 11).**

Discussion

Although there have been a number of published studies exploring PD genetics in the African and African admixed populations[61,62,63,64,65,66,67,68,69,70,16,71,72,73,74,75,60], in the present study, we have gathered the largest collection of PD patients and controls from African and African admixed ancestry populations to comprehensively assess the genetic architecture of PD on a genome-wide scale. Here, we identified a novel African-specific GWAS signal on the *GBA1* locus, significantly associated with PD risk and AAO, to be the most important risk factor for PD in this African and African admixed populations. In contrast, initial well powered GWAS in European populations nominated the *SNCA* and *MAPT* loci as the most significant contributors to PD genetic risk in Europeans. Remarkably, almost a four times larger sample size in cases was required to nominate *GBA1* as one of the major PD risk factors in the European ancestry population through GWAS[76], showing the power and benefit of using diverse ancestry data.

We suggest a novel disease mechanism via expression changes consistent with a trend towards decreased GCase activity levels. The *GBA1* c.1225-34C>A (rs3115534) GWAS hit alters a non-conserved intronic nucleotide (GERP++ score = -2.04). Despite the large effect size driven by this signal, our study did not identify an association with any previously reported or new *GBA1* coding or structural aberration that could explain this signal[47,77,56,78]. Splice prediction tools predicted no significant impact on normal splicing, while rs3115534 has been reported to be an expression quantitative trait locus (eQTL) in

642 several tissues[79,57]. Additionally, a large-scale pQTL study in African Americans with chronic kidney
643 disease suggests that at the protein level the risk allele for PD in our GWAS (G) is associated with a
644 reduction in the level of GCase protein in blood, as defined by the SOMAscan assay. This finding
645 supports the concept that the risk allele leads to a partial loss of both GCase protein and GCase enzyme
646 activity[80].
647
648 Strikingly, by leveraging existing eQTL data predominantly of African American ancestry, we found the
649 rs3115534-G risk allele to be associated with increased *GBA1* expression levels in whole blood, but
650 paradoxically linked with a trend towards decreased GCase activity, which may be due to challenges
651 with RNA-seq in this locus. This interesting finding, possibly explained by transcript diversity leading to
652 the expression of a protein with diminished lysosomal GCase activity, warrants further study. Previously,
653 *GBA1* variants associated with PD risk have all been coding mutations, but here we identify a novel
654 functional mechanism involved in disease etiology. Our findings are limited by the absence of brain QTL
655 data in non-European populations, underscoring the importance of increasing representation from
656 ancestrally diverse populations to enable new discoveries and ensure their equitable translation. Future
657 large scale single cell expression studies should investigate in which brain cell types these expression
658 differences are most prominent. This novel mechanism opens new avenues towards efficient RNA-based
659 therapeutic strategies, such as antisense oligonucleotides or short interfering RNAs aimed at reducing
660 lifetime risk.
661
662 Interestingly, given the high population frequency of the identified signal and the phenotypic
663 characteristics of the homozygous Africans and African admixed carriers, our study does not support the
664 notion that this variant causes Gaucher disease. Furthermore, the rs3115534 variant has been found to
665 be extremely rare in non-African/African admixed populations. These findings suggest an African
666 founder effect, and reinforce that the genetic architecture of this locus and its influence in risk and
667 onset is different across populations. Interestingly, rs3115534 was also found to be associated with PD
668 AAO in our study. The largest GWAS meta-analysis investigating the role of genetic determinants on PD
669 onset in European populations[81] nominated p.N409S as an AAO disease modifier. This variant, which is
670 one of the most common *GBA1* risk factors in European and Ashkenazi Jewish populations, is 100 times
671 less frequent in individuals of African and African admixed ancestry. In support of this notion, we did not
672 find any of the common *GBA1* pathogenic variants through WGS in this study.
673
674
675 Overall, addressing the genetic complexity underlying these underrepresented populations, our study
676 represents a valuable resource for identifying and tracking *GBA1* carriers that may prove relevant for
677 enrollment in target-specific PD clinical trials. *GBA1* genetic testing in the African and African admixed
678 populations can help to design an optimized trial with the highest likelihood of providing meaningful
679 results and actionable answers. We envisage that these data generated under the Global Parkinson's
680 Genetics Program initiative will be key to shed light on the molecular mechanisms involved in the
681 disease process and might pave the way for future clinical trials and therapeutic interventions.
682

683 This would be helpful to further improve our granularity in association testing and ability to fine-map
684 through integration of omics data while also evaluating population specific associations.
685

686 Limitations

687
688 Although we have made progress in assessing genetic risk factors for PD in an African-specific manner,
689 there are a number of limitations to our study. Unraveling additional susceptibility genetic risk and
690 phenotypic relationships would have been possible if a larger cohort had been analyzed. Considering our
691 limited sample size, we lacked statistical power to detect common genetic variants of smaller effect
692 sizes (**Supplemental Figure 13)**. Additionally, an important proportion of the genetic risk contributing to
693 the missing heritability of PD in the African and African admixed populations might result from rare
694 alleles and structural variants that have not been assessed in the present study. Downstream genetic
695 analyses such as gene-level burden analyses, heritability estimates, enrichment pathway analysis,
696 phenotypic and functional inferences should be conducted in subsequent well-powered studies as more
697 additional data becomes available. Additionally, due to lack of well-powered and African-specific RNA
698 sequencing datasets, the added complexity of multi-mapping reads between *GBA* and *GBAP1* and the
699 limited number of LCLs to explore GCase activity in a large scale manner, we assume the limitation that
700 this potential novel functional mechanism merits further study. We are aware that although this
701 represents the first PD GWAS in the African and African admixed populations, two-thirds of the sample
702 size is of Nigerian descent, therefore not representative of the substantial genetic diversity across the
703 continent. From the analytical perspective, another constraint is our inability to discriminate specific
704 sub-African population ancestry levels when it comes to principal component analyses (**Supplementary**
705 **Figure 3**). Future additional data will yield more and better refined clusters.
706
707 While this study marks major progress in assessing genetic risk factors for PD, there remains a great deal
708 to be done. Most study participants did not have in-depth phenotype information limiting our ability to
709 conduct further clinical and genetic characterization analyses. Future studies should explore the effect
710 of this variant on cognitive impairment in PD.
711

712 Data Sharing

713 All GP2 data is hosted in collaboration with the Accelerating Medicines Partnership in Parkinson's
714 disease, and is available via application on the website (https://amp-pd.org/register-for-amp-pd). The
715 GWAS summary statistics from this study, excluding 23andMe, are available as of GP2's release 5.
716 23andMe summary statistics are available via application on the website
717 (https://research.23andme.com/dataset-access/). Genotyping imputation, quality control, ancestry
718 prediction, and processing was performed using GenoTools v1.0, publicly available on GitHub
719 (https://github.com/GP2code/GenoTools). All scripts for analyses are publicly available on GitHub
720 [https://github.com/GP2code/GP2-AFR-AAC-metaGWAS; 10.5281/zenodo.7888141].
721

Ethics Statement

All cohorts recruited to the GP2 initiative undergo a thorough review of the consent forms in the Operations and Compliance working group, ensuring that each contributing study abided by the ethics guidelines set out by their institutional review boards, and all participants gave informed consent for inclusion in both their initial cohorts and subsequent studies within local law constraints. All GP2 data is hosted in collaboration with the Accelerating Medicines Partnership in Parkinson's disease, and is available via application on the website (https://amp-pd.org/register-for-amp-pd).

Summary statistics for individuals with or without PD were provided through a collaborative agreement with 23andMe, Inc. Participants provided informed consent and volunteered to participate in the research online, under a protocol approved by the external AAHRPP-accredited IRB, Ethical & Independent (E&I) Review Services. As of 2022, E&I Review Services is part of Salus IRB (https://www.versiticlinicaltrials.org/salusirb). 23andMe summary statistics are available via application on the website (https://research.23andme.com/dataset-access/).

Declaration of Interests

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) and the National Human Genome Research Institute (NHGRI).

This work was supported in part by the Global Parkinson's Genetics Program (GP2). GP2 is funded by the Aligning Science Against Parkinson's (ASAP) initiative and implemented by The Michael J. Fox Foundation for Parkinson's Research (https://gp2.org). Additional funding was provided by The Michael J. Fox Foundation for Parkinson's Research through grant MJFF-009421/17483. For a complete list of GP2 members see https://gp2.org

D.V., H.I., H.L.L., K.L. and M.A.N.'s participation in this project was part of a competitive contract awarded to Data Tecnica International LLC by the National Institutes of Health to support open science research.
K.H. and members of the 23andMe Research Team are employed by and hold stock or stock options in 23andMe, Inc. M.A.N. also currently serves on the scientific advisory board for Character Biosciences Inc and Neuron 23 Inc.

DGS is a member of the faculty of the University of Alabama at Birmingham and is supported by endowment and University funds, is an investigator in studies funded by Abbvie, Inc., the American Parkinson Disease Association, the Michael J. Fox Foundation for Parkinson Research, The National Parkinson Foundation, Alabama Department of Commerce, Alabama Innovation Fund, Genentech, the Department of Defense, and NIH grants P50NS108675 and R25NS079188 and has a clinical practice and

is compensated for these activities through the University of Alabama Health Services Foundation. He serves as Deputy Editor for the journal Movement Disorders and is compensated for this role by the International Parkinson and Movement Disorders Society. In addition, since January 1, 2022 he has served as a consultant for or received honoraria from Abbvie Inc., Curium Pharma, Appello, Theravance, Sanofi-Aventis, Alnylam Pharmaceutics, Coave Therapeutics, BlueRock Therapeutics and F. Hoffman-La Roche.

We thank the research participants and employees of 23andMe. The following members of the 23andMe Research Team contributed to this study:

Stella Aslibekyan, Adam Auton, Elizabeth Babalola, Robert K. Bell, Jessica Bielenberg, Katarzyna Bryc, Emily Bullis, Paul Cannon, Daniella Coker, Gabriel Cuellar Partida, Devika Dhamija, Sayantan Das, Sarah L. Elson, Nicholas Eriksson, Teresa Filshtein, Alison Fitch, Kipper Fletez-Brant, Pierre Fontanillas, Will Freyman, Julie M. Granka, Alejandro Hernandez, Barry Hicks, David A. Hinds, Ethan M. Jewett, Yunxuan Jiang, Katelyn Kukar, Alan Kwong, Keng-Han Lin, Bianca A. Llamas, Maya Lowe, Jey C. McCreight, Matthew H. McIntyre, Steven J. Micheletti, Meghan E. Moreno, Priyanka Nandakumar, Dominique T. Nguyen, Elizabeth S. Noblin, Jared O'Connell, Aaron A. Petrakovitz, G. David Poznik, Alexandra Reynoso, Madeleine Schloetter, Morgan Schumacher, Anjali J. Shastri, Janie F. Shelton, Jingchunzi Shi, Suyash Shringarpure, Qiaojuan Jane Su, Susana A. Tat, Christophe Toukam Tchakouté, Vinh Tran, Joyce Y. Tung, Xin Wang, Wei Wang, Catherine H. Weldon, Peter Wilton, Corinna D. Wong

We thank Cynthia J. Casaceli, Debbie Baker and Christi Alessi-Fox from the University of Rochester Clinical Trials Coordination Center for its contribution to the coordination of the BLAAC PD Study. We thank Lisa Shulman for her contribution to the design of the protocol for BLAAC PD clinical assessments.

We thank the Biowulf team, as this study used the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health (http://hpc.nih.gov).

Figure 1 was generated on www.biorender.com.

Author contributions

M.R, S.B.C, M.B.M, N.O, and A.S contributed equally to this work.
Concept or design: M.R, N.O, A.S, C.B, J.H, H.H, M.N, M.B.M, S.B.C, P.W.C, E.S, N.T
Sample and data acquisition: All
Analysis: M.B.M, S.B.C, D.V, K.S.L, C.X.A, M.N, M.J.K, C.B, K.B, P.A, K.D, J.J.K, H.L, H.I, J.K, O.O, I.E, O.O, K.H
Critical review: All

799 # References

800 1 GBD 2016 Parkinson's Disease Collaborators. Global, regional, and national burden of Parkinson's
801 disease, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet*
802 *Neurol* 2018; **17**: 939–53.

803 2 Dorsey ER, Sherer T, Okun MS, Bloem BR. The Emerging Evidence of the Parkinson Pandemic. *J*
804 *Parkinsons Dis* 2018; **8**: S3–8.

805 3 Chang M, He L, Cai L. An Overview of Genome-Wide Association Studies. *Methods Mol Biol* 2018;
806 **1754**: 97–108.

807 4 Nalls MA, Blauwendraat C, Vallerga CL, *et al.* Identification of novel risk loci, causal insights, and
808 heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet*
809 *Neurol* 2019; **18**: 1091–102.

810 5 Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. *Lancet*
811 *Neurol* 2020; **19**: 170–8.

812 6 Foo JN, Chew EGY, Chung SJ, *et al.* Identification of Risk Loci for Parkinson Disease in Asians and
813 Comparison of Risk Between Asians and Europeans: A Genome-Wide Association Study. *JAMA*
814 *Neurol* 2020; **77**: 746–54.

815 7 Loesch DP, Horimoto ARVR, Heilbron K, *et al.* Characterizing the Genetic Architecture of Parkinson's
816 Disease in Latinos. *Ann Neurol* 2021; **90**: 353–65.

817 8 Kim JJ, Vitale D, Véliz Otani D, *et al.* Multi-ancestry genome-wide meta-analysis in Parkinson's
818 disease. bioRxiv. 2022; published online Aug 6. DOI:10.1101/2022.08.04.22278432.

819 9 Bandres-Ciga S, Diez-Fairen M, Kim JJ, Singleton AB. Genetics of Parkinson's disease: An
820 introspection of its journey towards precision medicine. *Neurobiol Dis* 2020; **137**: 104782.

821 10 Elsayed I, Martinez-Carrasco A, Cornejo-Olivas M, Bandres-Ciga S. Mapping the Diverse and
822 Inclusive Future of Parkinson's Disease Genetics and Its Widespread Impact. *Genes* 2021; **12**.
823 DOI:10.3390/genes12111681.

824 11 Schumacher-Schuh AF, Bieger A, Okunoye O, *et al.* Underrepresented Populations in Parkinson's
825 Genetics Research: Current Landscape and Future Directions. *Mov Disord* 2022; **37**: 1593–604.

826 12 Okubadejo N, Britton A, Crews C, *et al.* Analysis of Nigerians with apparently sporadic Parkinson
827 disease for mutations in LRRK2, PRKN and ATXN3. *PLoS One* 2008; **3**: e3421.

828 13 Cilia R, Sironi F, Akpalu A, *et al.* Screening LRRK2 gene mutations in patients with Parkinson's
829 disease in Ghana. *J Neurol* 2012; **259**: 569–70.

830 14 Okubadejo NU, Rizig M, Ojo OO, *et al.* Leucine rich repeat kinase 2 (LRRK2) GLY2019SER mutation is
831 absent in a second cohort of Nigerian Africans with Parkinson disease. *PLoS One* 2018; **13**:
832 e0207984.

833 15 du Toit N, van Coller R, Anderson DG, Carr J, Bardien S. Frequency of the LRRK2 G2019S mutation in

834 South African patients with Parkinson's disease. *Neurogenetics* 2019; **20**: 215–8.

835 16 Yonova-Doing E, Atadzhanov M, Quadri M, *et al.* Analysis of LRRK2, SNCA, Parkin, PINK1, and DJ-1 in
836 Zambian patients with Parkinson's disease. *Parkinsonism Relat Disord* 2012; **18**: 567–71.

837 17 Bentley AR, Callier SL, Rotimi CN. Evaluating the promise of inclusion of African ancestry
838 populations in genomics. *NPJ Genom Med* 2020; **5**: 5.

839 18 Choudhury A, Aron S, Botigué LR, *et al.* High-depth African genomes inform human migration and
840 health. *Nature* 2020; **586**: 741–8.

841 19 Campbell MC, Tishkoff SA. African genetic diversity: implications for human demographic history,
842 modern human origins, and complex disease mapping. *Annu Rev Genomics Hum Genet* 2008; **9**:
843 403–33.

844 20 Schlebusch CM, Jakobsson M. Tales of Human Migration, Admixture, and Selection in Africa. *Annu*
845 *Rev Genomics Hum Genet* 2018; **19**: 405–28.

846 21 Safiri S, Noori M, Nejadghaderi SA, *et al.* The burden of Parkinson's disease in the Middle East and
847 North Africa region, 1990-2019: results from the global burden of disease study 2019. *BMC Public*
848 *Health* 2023; **23**: 107.

849 22 Okubadejo NU, Bower JH, Rocca WA, Maraganore DM. Parkinson's disease in Africa: A systematic
850 review of epidemiologic and genetic studies. *Mov Disord* 2006; **21**: 2150–6.

851 23 Blanckenberg J, Bardien S, Glanzmann B, Okubadejo NU, Carr JA. The prevalence and genetics of
852 Parkinson's disease in sub-Saharan Africans. *J Neurol Sci* 2013; **335**: 22–5.

853 24 Khedr EM, Fawi G, Abbas MAA, *et al.* Prevalence of Parkinsonism and Parkinson's disease in Qena
854 governorate/Egypt: a cross-sectional community-based survey. *Neurol Res* 2015; **37**: 607–18.

855 25 El-Tallawy HN, Farghaly WM, Shehata GA, *et al.* Prevalence of Parkinson's disease and other types
856 of Parkinsonism in Al Kharga district, Egypt. *Neuropsychiatr Dis Treat* 2013; **9**: 1821–6.

857 26 Zirra A, Rao SC, Bestwick J, *et al.* Gender Differences in the Prevalence of Parkinson's Disease. *Mov*
858 *Disord Clin Pract* 2023; **10**: 86–93.

859 27 Ojo OO, Abubakar SA, Iwuozo EU, *et al.* The Nigeria Parkinson Disease Registry: Process, Profile,
860 and Prospects of a Collaborative Project. *Mov Disord* 2020; **35**. DOI:10.1002/mds.28123.

861 28 Ojo OO, Wahab KW, Bello AH, *et al.* A Cross-Sectional Comprehensive Assessment of the Profile and
862 Burden of Non-motor Symptoms in Relation to Motor Phenotype in the Nigeria Parkinson Disease
863 Registry Cohort. *Mov Disord Clin Pract* 2021; **8**: 1206–15.

864 29 Bailey M, Shulman LM, Ryan D, *et al.* Frequency of Parkinsonism and Parkinson Disease in African
865 Americans in the Chicago Community. *J Gerontol A Biol Sci Med Sci* 2021; **76**: 1340–5.

866 30 Xie T, Liao C, Lee D, *et al.* Disparities in diagnosis, treatment and survival between Black and White
867 Parkinson patients. *Parkinsonism Relat Disord* 2021; **87**: 7–12.

868 31 Peterson RE, Kuchenbaecker K, Walters RK, *et al.* Genome-wide Association Studies in Ancestrally

869 Diverse Populations: Opportunities, Methods, Pitfalls, and Recommendations. *Cell* 2019; **179**: 589–
870 603.

871 32 Hughes AJ, Daniel SE, Kilford L, Lees AJ. Accuracy of clinical diagnosis of idiopathic Parkinson's
872 disease: a clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* 1992; **55**: 181–4.

873 33 Heilbron K, Noyce AJ, Fontanillas P, *et al.* The Parkinson's phenome-traits associated with
874 Parkinson's disease in a broadly phenotyped cohort. *NPJ Parkinsons Dis* 2019; **5**: 4.

875 34 Blauwendraat C, Faghri F, Pihlstrom L, *et al.* NeuroChip, an updated version of the NeuroX
876 genotyping platform to rapidly screen for variants associated with neurological diseases. *Neurobiol*
877 *Aging* 2017; **57**: 247.e9–247.e13.

878 35 Koretsky MJ, Alvarado C, Makarious MB, *et al.* Genetic risk factor clustering within and across
879 neurodegenerative diseases. bioRxiv. 2022; published online Dec 3.
880 DOI:10.1101/2022.12.01.22282945.

881 36 1000 Genomes Project Consortium, Auton A, Brooks LD, *et al.* A global reference for human genetic
882 variation. *Nature* 2015; **526**: 68–74.

883 37 Siva N. 1000 Genomes project. *Nat Biotechnol* 2008; **26**: 256.

884 38 Bray SM, Mulle JG, Dodd AF, Pulver AE, Wooding S, Warren ST. Signatures of founder effects,
885 admixture, and selection in the Ashkenazi Jewish population. *Proc Natl Acad Sci U S A* 2010; **107**:
886 16222–7.

887 39 Taliun D, Harris DN, Kessler MD, *et al.* Sequencing of 53,831 diverse genomes from the NHLBI
888 TOPMed Program. *Nature* 2021; **590**: 290–9.

889 40 Das S, Forer L, Schönherr S, *et al.* Next-generation genotype imputation service and methods. *Nat*
890 *Genet* 2016; **48**: 1284–7.

891 41 Fuchsberger C, Abecasis GR, Hinds DA. minimac2: faster genotype imputation. *Bioinformatics* 2015;
892 **31**: 782–4.

893 42 Purcell S, Neale B, Todd-Brown K, *et al.* PLINK: a tool set for whole-genome association and
894 population-based linkage analyses. *Am J Hum Genet* 2007; **81**: 559–75.

895 43 Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association
896 scans. *Bioinformatics* 2010; **26**: 2190–1.

897 44 Regier AA, Farjoun Y, Larson DE, *et al.* Functional equivalence of genome sequencing analysis
898 pipelines enables harmonized variant calling across human genetics projects. *Nat Commun* 2018; **9**:
899 4038.

900 45 Iwaki H, Leonard HL, Makarious MB, *et al.* Accelerating Medicines Partnership: Parkinson's Disease.
901 Genetic Resource. *Mov Disord* 2021; **36**: 1795–804.

902 46 Lange LM, Avenali M, Ellis M, *et al.* Global Parkinson's Genetics Program (GP2) Monogenic Network
903 Protocol: Elucidating causative gene variants in hereditary Parkinson's disease. bioRxiv. 2022;

904 published online Dec 5. DOI:10.1101/2022.12.01.22282794.

905 47 Toffoli M, Chen X, Sedlazeck FJ, *et al.* Comprehensive short and long read sequencing analysis for
906 the Gaucher and Parkinson's disease-associated GBA gene. *Commun Biol* 2022; **5**: 670.

907 48 Halman A, Dolzhenko E, Oshlack A. STRipy: A graphical application for enhanced genotyping of
908 pathogenic short tandem repeats in sequencing data. *Hum Mutat* 2022; **43**: 859–68.

909 49 Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-
910 throughput sequencing data. *Nucleic Acids Res* 2010; **38**: e164.

911 50 Robinson JT, Thorvaldsdóttir H, Winckler W, *et al.* Integrative genomics viewer. *Nat Biotechnol*
912 2011; **29**: 24–6.

913 51 Kolmogorov M, Billingsley KJ, Mastoras M, *et al.* Scalable Nanopore sequencing of human genomes
914 provides a comprehensive view of haplotype-resolved variation and methylation. *bioRxiv* 2023;
915 published online Jan 15. DOI:10.1101/2023.01.12.523790.

916 52 Jain C, Rhie A, Hansen NF, Koren S, Phillippy AM. Long-read mapping to repetitive reference
917 sequences using Winnowmap2. *Nat Methods* 2022; **19**: 705–10.

918 53 Smolka M, Paulin LF, Grochowski CM, *et al.* Comprehensive Structural Variant Detection: From
919 Mosaic to Population-Level. bioRxiv. 2022; : 2022.04.04.487055.

920 54 Peters SP, Lee RE, Glew RH. A microassay for Gaucher's disease. *Clin Chim Acta* 1975; **60**: 391–6.

921 55 Karczewski KJ, Francioli LC, Tiao G, *et al.* Author Correction: The mutational constraint spectrum
922 quantified from variation in 141,456 humans. *Nature* 2021; **590**: E53.

923 56 Tayebi N, Park J, Madike V, Sidransky E. Gene rearranagement on 1q21 introducing a duplication of
924 the glucocerebrosidase pseudogene and a metaxin fusion gene. *Hum Genet* 2000; **107**: 400–3.

925 57 Mak ACY, Kachuri L, Hu D, *et al.* Gene expression in African Americans and Latinos reveals ancestry-
926 specific patterns of genetic architecture. bioRxiv. 2021; : 2021.08.19.456901.

927 58 Gustavsson EK, Sethi S, Gao Y, *et al.* Pseudogenes limit the identification of novel common
928 transcripts generated by their parent genes. bioRxiv. 2022; : 2022.10.21.513169.

929 59 Donnelly MP, Paschou P, Grigorenko E, *et al.* The distribution and most recent common ancestor of
930 the 17q21 inversion in humans. *Am J Hum Genet* 2010; **86**: 161–71.

931 60 Okunoye O, Ojo O, Abiodun O, *et al.* allele and haplotype frequencies in Nigerian Africans:
932 population distribution and association with Parkinson's disease risk and age at onset. *medRxiv*
933 2023; published online March 24. DOI:10.1101/2023.03.24.23287684.

934 61 Ross OA, Wilhoite GJ, Bacon JA, *et al.* LRRK2 variation and Parkinson's disease in African Americans.
935 *Mov Disord* 2010; **25**: 1973–6.

936 62 McGuire V, Van Den Eeden SK, Tanner CM, *et al.* Association of DRD2 and DRD3 Polymorphisms
937 with Parkinson's Disease in a Multiethnic Consortium. *J Neurol Sci* 2011; **307**: 22.

938 63 Clark LN, Levy G, Tang M-X, *et al.* The Saitohin 'Q7R' polymorphism and tau haplotype in multi-
939 ethnic Alzheimer disease and Parkinson's disease cohorts. Neuroscience Letters. 2003; **347**: 17–20.

940 64 Gwinn-Hardy K, Singleton A, O'Suilleabhain P, *et al.* Spinocerebellar ataxia type 3 phenotypically
941 resembling parkinson disease in a black family. *Arch Neurol* 2001; **58**.
942 DOI:10.1001/archneur.58.2.296.

943 65 Okubadejo NU, Okunoye O, Ojo OO, *et al.* APOE E4 is associated with impaired self-declared
944 cognition but not disease risk or age of onset in Nigerians with Parkinson's disease. *npj Parkinson's*
945 *Disease* 2022; **8**: 1–6.

946 66 Lesage S, Condroyer C, Hecham N, *et al.* Mutations in the glucocerebrosidase gene confer a risk for
947 Parkinson disease in North Africa. *Neurology* 2011; **76**: 301–3.

948 67 Milanowski LM, Oshinaike O, Walton RL, *et al.* Screening of GBA Mutations in Nigerian Patients with
949 Parkinson's Disease. *Mov Disord* 2021; **36**: 2971–3.

950 68 Nishioka K, Ross OA, Vilariño-Güell C, *et al.* Glucocerebrosidase mutations in diffuse Lewy body
951 disease. *Parkinsonism Relat Disord* 2011; **17**: 55–7.

952 69 Bardien S, Keyser R, Yako Y, Lombard D, Carr J. Molecular analysis of the parkin gene in South
953 African patients diagnosed with Parkinson's disease. *Parkinsonism Relat Disord* 2009; **15**: 116–21.

954 70 Hashad DI, Abou-Zeid AA, Achmawy GA, Allah HMOS, Saad MA. G2019S mutation of the leucine-
955 rich repeat kinase 2 gene in a cohort of Egyptian patients with Parkinson's disease. *Genet Test Mol*
956 *Biomarkers* 2011; **15**: 861–6.

957 71 Keyser RJ, Lombard D, Veikondis R, Carr J, Bardien S. Analysis of exon dosage using MLPA in South
958 African Parkinson's disease patients. *Neurogenetics* 2010; **11**: 305–12.

959 72 Hulihan MM, Ishihara-Paul L, Kachergus J, *et al.* LRRK2 Gly2019Ser penetrance in Arab-Berber
960 patients from Tunisia: a case-control genetic study. *Lancet Neurol* 2008; **7**: 591–4.

961 73 Ishihara-Paul L, Hulihan MM, Kachergus J, *et al.* PINK1 mutations and parkinsonism. *Neurology*
962 2008; **71**: 896–902.

963 74 Bouhouche A, Tesson C, Regragui W, *et al.* Mutation Analysis of Consanguineous Moroccan Patients
964 with Parkinson's Disease Combining Microarray and Gene Panel. *Front Neurol* 2017; **8**: 567.

965 75 Trinh J, Gustavsson EK, Vilariño-Güell C, *et al.* DNM3 and genetic modifiers of age of onset in LRRK2
966 Gly2019Ser parkinsonism: a genome-wide linkage and association study. *Lancet Neurol* 2016; **15**:
967 1248–56.

968 76 Simón-Sánchez J, Schulte C, Bras JM, *et al.* Genome-wide association study reveals genetic risk
969 underlying Parkinson's disease. *Nat Genet* 2009; **41**: 1308–12.

970 77 Park JK, Koprivica V, Andrews DQ, *et al.* Glucocerebrosidase mutations among African-American
971 patients with type 1 Gaucher disease. *Am J Med Genet* 2001; **99**: 147–51.

972 78 Mahungu AC, Anderson DG, Rossouw AC, *et al.* Screening of the glucocerebrosidase (GBA) gene in

973 South Africans of African ancestry with Parkinson's disease. *Neurobiol Aging* 2020; **88**: 156.e11–
974 156.e14.

975 79 Raj T, Rothamel K, Mostafavi S, *et al.* Polarization of the effects of autoimmune and
976 neurodegenerative risk alleles in leukocytes. *Science* 2014; **344**. DOI:10.1126/science.1249547.

977 80 Identification of 969 protein quantitative trait loci in an African American population with kidney
978 disease attributed to hypertension. *Kidney Int* 2022; **102**: 1167–77.

979 81 Blauwendraat C, Heilbron K, Vallerga CL, *et al.* Parkinson's disease age at onset genome-wide
980 association study: Defining heritability, genetic loci, and α-synuclein mechanisms. *Mov Disord* 2019;
981 **34**: 866–75.

982 82 Sirugo G, Williams SM, Tishkoff SA. The Missing Diversity in Human Genetic Studies. *Cell* 2019; **177**:
983 26–31.

984 83 Lake J, Warly Solsberg C, Kim JJ, *et al.* Multi-ancestry meta-analysis and fine-mapping in Alzheimer's
985 disease. *Mol Psychiatry* 2023; published online May 18. DOI:10.1038/s41380-023-02089-w.