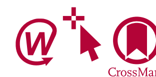
Identification of genetic risk loci and causal insights associated with Parkinson's disease in African and African admixed populations: a genome-wide association study



Mie Rizig†, Sara Bandres-Ciga*†, Mary B Makarious*†, Oluwadamilola Omolara Ojo, Peter Wild Crea, Oladunni Victoria Abiodun, Kristin S Levine, Sani Atta Abubakar, Charles Obiora Achoru, Dan Vitale, Olaleye Akinmola Adeniji, Osigwe Paul Agabi, Mathew J Koretsky, Uchechi Agulanna, Deborah A Hall, Rufus Olusola Akinyemi, Tao Xie, Mohammed Wulgo Ali, Ejaz A Shamim, Ifeyinwa Ani-Osheku, Mahesh Padmanaban, Ohwotemu Michael Arigbodi, David G Standaert, Abiodun Hamzat Bello, Marissa N Dean, Cyril Oshomah Erameh, Inas Elsayed, Temitope Hannah Farombi, Olaitan Okunoye, Michael Bimbola Fawale, Kimberley J Billingsley, Frank Aiwansoba Imarhiagbe, Pilar Alvarez Jerez, Emmanuel Uzodinma Iwuozo, Breeana Baker, Morenikeji Adeyoyin Komolafe, Laksh Malik, Paul Osemeke Nwani, Kensuke Daida, Ernest Okwundu Nwazor, Abigail Miano-Burkhardt, Yakub Wilberforce Nyandaiti, Zih-Hua Fang, Yahaya Olugbo Obiabo, Jillian H Kluss, Olanike Adedoyin Odeniyi, Dena G Hernandez, Francis Ehidiamen Odiase, Nahid Tayebi, Francis Ibe Ojini, Ellen Sidranksy, Gerald Awele Onwuegbuzie, Andrea M D'Souza, Godwin Osawaru Osaigbovo, Bahafta Berhe, Nosakhare Osemwegie, Xylena Reed, Olajumoke Olufemi Oshinaike, Hampton L Leonard, Folajimi Morenikeji Otubogun, Chelsea X Alvarado, Shyngle Imiewan Oyakhire, Simon Izuchukwu Ozomma, Sarah Chabiri Samuel, Funmilola Tolulope Taiwo, Kolawole Wasiu Wahab, Yusuf Agboola Zubair, Hirotaka Iwaki, Jonggeol Jeffrey Kim, Huw R Morris, John Hardy, Mike A Nalls, Karl Heilbron, Lucy Norcliffe-Kaufmann, Nigeria Parkinson Disease Research Network§, International Parkinson's Disease Genomics Consortium Africa§, Black and African American Connections to Parkinson's Disease Study Group§, the 23andMe Research Team§, Cornelis Blauwendraat, Henry Houlden, Andrew Singleton*‡, Njideka Ulunma Okubadejo*‡, on behalf of the Global Parkinson's Genetics Program§*

Summary

Background An understanding of the genetic mechanisms underlying diseases in ancestrally diverse populations is an important step towards development of targeted treatments. Research in African and African admixed populations can enable mapping of complex traits, because of their genetic diversity, extensive population substructure, and distinct linkage disequilibrium patterns. We aimed to do a comprehensive genome-wide assessment in African and African admixed individuals to better understand the genetic architecture of Parkinson's disease in these underserved populations.

Methods We performed a genome-wide association study (GWAS) in people of African and African admixed ancestry with and without Parkinson's disease. Individuals were included from several cohorts that were available as a part of the Global Parkinson's Genetics Program, the International Parkinson's Disease Genomics Consortium Africa, and 23andMe. A diagnosis of Parkinson's disease was confirmed clinically by a movement disorder specialist for every individual in each cohort, except for 23andMe, in which it was self-reported based on clinical diagnosis. We characterised ancestry-specific risk, differential haplotype structure and admixture, coding and structural genetic variation, and enzymatic activity.

Findings We included 197 918 individuals (1488 cases and 196 430 controls) in our genome-wide analysis. We identified a novel common risk factor for Parkinson's disease (overall meta-analysis odds ratio for risk of Parkinson's disease 1·58 [95% CI 1·37–1·80], p=2·397×10^{-14}) and age at onset at the *GBA1* locus, rs3115534-G (age at onset β=–2·00 [SE=0·57], p=0·0005, for African ancestry; and β=–4·15 [0·58], p=0·015, for African admixed ancestry), which was rare in non-African or non-African admixed populations. Downstream short-read and long-read whole-genome sequencing analyses did not reveal any coding or structural variant underlying the GWAS signal. The identified signal seems to be associated with decreased glucocerebrosidase activity.

Interpretation Our study identified a novel genetic risk factor in *GBA1* in people of African ancestry, which has not been seen in European populations, and it could be a major mechanistic basis of Parkinson's disease in African populations. This population-specific variant exerts substantial risk on Parkinson's disease as compared with common variation identified through GWAS and it was found to be present in 39% of the cases assessed in this study. This finding highlights the importance of understanding ancestry-specific genetic risk in complex diseases, a particularly crucial point as the Parkinson's disease field moves towards targeted treatments in clinical trials. The distinctive genetics of African populations highlights the need for equitable inclusion of ancestrally diverse groups in future trials, which will be a valuable step towards gaining insights into novel genetic determinants underlying the causes of Parkinson's disease. This finding opens new avenues towards RNA-based and other therapeutic strategies aimed at reducing lifetime risk of Parkinson's disease.

Funding The Global Parkinson's Genetics Program, which is funded by the Aligning Science Across Parkinson's initiative, and the Michael J Fox Foundation for Parkinson's Research.

Lancet Neurol 2023; 22: 1015–25

Published **Online** August 23, 2023 https://doi.org/10.1016/S1474-4422(23)00283-1

See **Comment** page 975

*Contributed equally

†Joint first authors

‡Joint last authors

§Members are listed in appendix 1 (pp 22–57)

Department of Neuromuscular Diseases (M Rizig PhD, O Okunoye PhD, P Alvarez Jerez BSc, Prof H R Morris PhD FRCP, Prof H Houlden MD) **and Department of Neurodegenerative Disease** (Prof J Hardy PhD), **UCL Queen Square Institute of Neurology, London, UK; UCL Movement Disorders Centre, University College London, London, UK** (M B Makarious BSc)**; 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, USA** (S Bandres-Ciga PhD, P W Crea BSc, K S Levine MS, M J Koretsky BSc, K J Billingsley PhD, P Alvarez Jerez, B Baker BSc, L Malik MFS, K Daida MD, A Miano-Burkhardt BSc, X Reed PhD, H L Leonard MS, C X Alvarado MS, H Iwaki MD, J J Kim BA, C Blauwendraat PhD,

A Singleton PhD); Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA (M B Makarious, P W Crea, K J Billingsley, K Daida, A Miano-Burkhardt, J H Kluss PhD, D G Hernandez PhD, J J Kim, C Blauwendraat, A Singleton); College of Medicine, University of Lagos, Idi Araba, Lagos State, Nigeria (O O Ojo MD, O P Agabi MBBS, Prof F I Ojini MSc, Prof N U Okubadejo MD); General Hospital, Isolo, Lagos State, Nigeria (O V Abiodun FWACP); Data Tecnica International, Washington, DC, USA (K S Levine, D Vitale MS, H L Leonard, C X Alvarado, H Iwaki, M A Nalls PhD); Ahmadu Bello University, Zaria, Kaduna State, Nigeria (Prof S A Abubakar MBBS); Jos University Teaching Hospital, Jos, Plateau State, Nigeria (C O Achoru MBBS, G O Osaigbovo MBBS); Federal Medical Centre, Abeokuta, Ogun State, Nigeria (O A Adeniji MBBS); Lagos University Teaching Hospital, Idi Araba, Lagos State, Nigeria (U Agulanna MBBS, Prof N U Okubadejo); Department of Neurological Sciences, Rush University Medical Center, Chicago, IL, USA (D A Hall MD PhD); Neuroscience and Ageing Research Unit, Institute for Advanced Medical Research and Training, College of Medicine, University of Ibadan, Ibadan, Oyo State, Nigeria (Prof R O Akinyemi PhD); Department of Neurology, University of Chicago Medicine, Chicago, IL, USA (T Xie MD PhD, M Padmanaban MD); Federal Teaching Hospital Gombe, Gombe State, Nigeria (M W Ali MBBS); Human Motor Control Section, National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA (E A Shamim MD); Kaiser Permanente Mid-Atlantic States, Largo, MD, USA (E A Shamim); MidAtlantic Permanente Research Institute, Rockville, MD, USA (E A Shamim); Asokoro District Hospital, Asokoro, Abuja, Nigeria (I Ani-Osheku FMCP); Delta State University, Abraka, Delta State, Nigeria (O M Arigbodi MBBS); Department of Neurology, University of Alabama at

Copyright © 2023 Elsevier Ltd. All rights reserved.

Introduction

Genome-wide association studies (GWASs) have been instrumental for identifying common variants and unravelling the genetics and heritability of complex diseases such as Parkinson's disease in European populations. The largest published GWAS meta-analysis of risk of Parkinson's disease so far included individuals of European ancestry and identified 90 independent genome-wide significant risk signals that explain 16–36% of the heritable risk of Parkinson's disease.[1,2] However, very little is known about the genetics of Parkinson's disease in non-European populations. Considerable ethnic variability in the distribution of monogenic causes and genetic risk variants has been documented across populations. For example, the relatively common *LRRK2* Gly2019Ser mutation remains unreported in some sub-Saharan African populations, despite being most commonly associated with familial and sporadic Parkinson's disease in Zambia and northern Africa.[3–6]

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 blocks, and abundant alleles that are private to these populations.[7] In addition to promoting scientific equity to address health disparities, diverse representation provides a platform for replication studies to explore the strength and relevance of findings reported from other populations. Additionally, studying diverse populations has the potential to facilitate the identification of novel or unique loci and investigate genotype–phenotype correlations that can further expand our understanding of pathological and pathogenetic disease mechanisms in Parkinson's disease.

We aimed to provide the first GWAS-based insights into the genetics of Parkinson's disease in the African and African admixed populations. Our objective was to do a comprehensive genome-wide assessment of Parkinson's disease risk and age at onset, characterising ancestry-specific risk, haplotype structure, and genetic admixture.

Methods

Study design and participants

Three sources of data were included in this GWAS. First, we used individual-level data from the International Parkinson's Disease Genomics Consortium Africa

Research in context

Evidence before this study

We searched original research articles on PubMed written in English between Jan 1, 2013, and Jan 1, 2023, with the keywords "Parkinson's disease", "genome-wide association study", and "diversity". Our current understanding of Parkinson's disease is disproportionately based on studies of populations of European ancestry, leading to a substantial gap in our knowledge about the genetics, clinical characteristics, and pathophysiology in under-represented populations. This dearth is particularly notable in individuals of African and African admixed ancestries. Over the past two decades, we have witnessed a revolution in the research area of complex genetic diseases. In the field of Parkinson's disease, large-scale genome-wide association studies (GWASs) in the European, Asian, and Latin American populations have identified multiple risk loci associated with the disease. These include 78 loci and 90 independent signals associated with risk of Parkinson's disease 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 Parkinson's disease genetics.

Added value of this study

To address the lack of diversity in our research field, we aimed to provide the first genome-wide assessment of Parkinson's disease genetics in African and African admixed populations. We identified a genetic risk factor linked to the risk of Parkinson's disease, dissected ancestry-specific differences in risk and age at onset, characterised known genetic risk factors, and highlighted the use of the African and African admixed risk haplotype substructure for future fine-mapping efforts.

Implications of all the available evidence

We nominate a novel signal affecting *GBA1*, which encodes glucocerebrosidase, as the major genetic risk factor for Parkinson's disease in African and African admixed populations. Our study could inform future *GBA1* clinical trials, improving patient stratification. In this regard, genetic testing can help to design trials that are likely to provide meaningful and actionable answers. We identified a novel disease mechanism via expression changes consistent with decreased glucocerebrosidase activity. This novel mechanism could hold promise for future efficient RNA-based therapeutic strategies, such as antisense oligonucleotides or short interfering RNAs, aimed at preventing and decreasing disease risk. This work represents a valuable resource in an underserved population, supporting pioneering research within the Global Parkinson's Genetics Program and beyond. Deciphering causal and genetic risk factors in people of African ancestries will help to determine whether interventions, potential targets for disease-modifying treatments, and prevention strategies that are being studied in the European populations are relevant to African and African admixed populations.

(IPDGC Africa). For our study, data were obtained between January, 2008, and January, 2022. Second, we used individual-level data from the Global Parkinson's Genetics Program (GP2). This cohort includes participant data from 147 different cohorts in 59 sites as of Aug 15, 2023. For our study, data were obtained from release 5. Third, we obtained GWAS summary statistics from 23andMe (figure 1; appendix 1 p 64). All three sources provided data for both cases and controls.

All participants in IPDGC Africa and GP2 underwent a neurological examination by the study's respective neurologists to document clinical and neurological status. A diagnosis of Parkinson's disease was based on fulfilment of the Parkinson's UK Brain Bank criteria (excluding the requirement for not more than one affected relative).[8] People without a diagnosis of Parkinson's disease were designated as controls and were assessed to detect overall signs of neurological conditions. Controls presenting any clinical signs of neurodegenerative diseases were excluded from the series. For the 23andMe dataset, the diagnosis of Parkinson's disease was self-reported based on a previous clinical diagnosis (appendix 1 p 12). Summary statistics for individuals with or without Parkinson's disease were provided through a collaborative agreement with 23andMe.

For our study, we studied individuals of both African and African admixed ancestry. For our study, we studied individuals of both African and African admixed ancestry as genetically designated by 1000 Genomes. We defined African admixed as individuals ancestrally similar to the following 1000 Genomes ancestry labels: African ancestry in Southwest United States of America (abbreviated as ASW in the 1000 Genomes project) and African Caribbean in Barbados (abbreviated as ACB in the 1000 Genomes project).

For the IPDGC Africa and the GP2 cohorts, the respective ethics committees for medical research approved involvement in genetic studies, and participants gave informed written consent. All cohorts recruited to the GP2

Birmingham, Birmingham, AL, USA (D G Standaert MD PhD, M N Dean MD); **University of Ilorin Teaching Hospital, Ilorin, Kwara State, Nigeria** (A H Bello FWACP, Prof K W Wahab MD); **Irrua Specialist Teaching Hospital, Irrua, Edo State, Nigeria** (C O Erameh MBBS); **Faculty of Pharmacy, University of Gezira, Wadmadani, Sudan** (I Elsayed PhD); **University College Hospital, Ibadan, Oyo State, Nigeria** (T H Farombi MBBS, F T Taiwo MBCHB); **Obafemi Awolowo University, Ile-Ife, Osun State, Nigeria** (M B Fawale MSc, Prof M A Komolafe MBBS); **University of Benin, Benin City, Edo State, Nigeria**

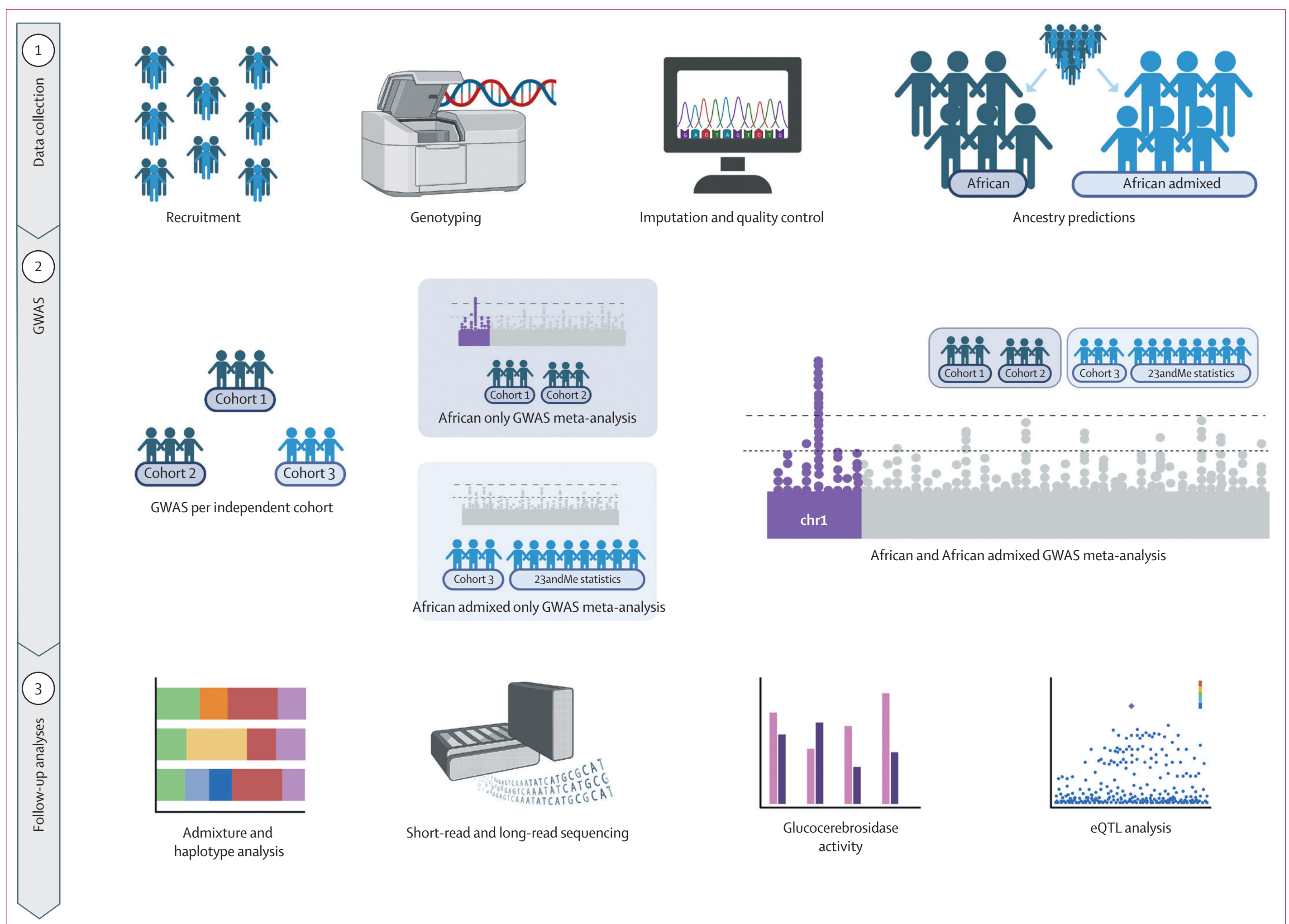


Figure 1: **Study design**
Figure created with BioRender.com. eQTL=expression quantitative trait locus. GWAS=genome-wide association study.

(Prof F A Imarhiagbe MBCHB, F E Odiase MBBS); **Benue State University, Makurdi, Benue State, Nigeria** (E U Iwuozo FMCP); **Nnamdi Azikiwe University Teaching Hospital, Nnewi, Anambra State, Nigeria** (P O Nwani MBBS); **Rivers State University Teaching Hospital, Port Harcourt, Rivers State, Nigeria** (E O Nwazor FMCP); **University of Maiduguri Teaching Hospital, Maiduguri, Borno State, Nigeria** (Prof Y W Nyandaiti MBBS, S C Samuel MBBS); **German Center for Neurodegenerative Diseases, Tuebingen, Germany** (Z-H Fang PhD); **Federal University of Health Sciences, Otukpo, Benue State, Nigeria** (Prof Y O Obiabo MBCHB); **General Hospital, Lagos Island, Lagos State, Nigeria** (O A Odeniyi MBBS); **Medical Genetics Branch, National Human Genome Research Institute, National Institutes of Health, Bethesda, MD, USA** (N Tayebi PhD, E Sidransky MD, A M D'Souza BSc, B Berhe BSc); **University of Abuja, Abuja, Federal Capital Territory, Nigeria** (G A Onwuegbuzie MBBS); **University of Port Harcourt, Port Harcourt, Rivers State, Nigeria** (N Osemwegie MBBS); **Lagos State University College of Medicine, Ikeja, Lagos State, Nigeria** (Prof O O Oshinaike FWACP); **Federal Medical Center, Ebute Metta, Lagos State, Nigeria** (F M Otubogun MBCHB); **National Hospital, Abuja, Federal Capital Territory, Nigeria** (S I Oyakhire MBBS, Y A Zubair MSc); **University of Calabar Teaching Hospital, Calabar, Cross River State, Nigeria** (S I Ozomma FMCP); **University of Ilorin, Ilorin, Kwara State, Nigeria** (Prof K W Wahab); **23andMe, Sunnyvale, CA, USA** (K Heilbron PhD, L Norcliffe-Kaufmann PhD)

Correspondence to: Prof Njideka U Okubadejo, College of Medicine, University of Lagos and Lagos University Teaching Hospital, Idi Araba, Lagos State, Nigeria **nokubadejo@unilag.edu.ng**

or

initiative undergo a thorough review of the consent forms in the Operations and Compliance working group, ensuring that each contributing study abides by the ethical 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. Participants in 23andMe provided informed consent and volunteered to participate in the research online, under a protocol approved by the external Association for Accreditation of Human Research Protection Programs-accredited institutional review board, Ethical & Independent Review Services. As of 2022, Ethical & Independent Review Services is part of the Salus institutional review board (appendix 1 p 10).

Databases

In this study, we focused primarily on individuals from African and African admixed ancestries. To compare population frequencies in other populations, we queried the gnomAD database and dbSNP on May 4, 2023, to access variant frequencies from curated datasets, such as the Allele Frequency Aggregator initiative and the 1000 Genomes project. Additionally, we queried the OpenTargets database on May 4, 2023, to confirm whether *GBA1* is functionally implicated by this variant and to assess the variant's association with gene expression.

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

The IPDGC Africa and GP2 blood or saliva samples were genotyped using two different genotyping platforms—ie, NeuroBooster (version 1.0; Illumina, San Diego, CA, USA) and NeuroChip (version 1.0; Illumina). The NeuroBooster array contains a backbone of 1 914 935 variants that densely cover ancestry informative markers, markers for determination of identity by descent, and X-chromosome single-nucleotide polymorphisms (SNPs) for sex determination. The NeuroBooster array contains 96 517 customised variants. The NeuroChip array contains a backbone of 306 670 variants and customised content comprising 179 467 variants.[9] Samples collected as part of the GP2 initiative were genotyped on the NeuroBooster array. Samples collected as part of the IPDGC Africa initiative were genotyped using both the NeuroBooster array and the NeuroChip array. Samples collected as part of 23andMe were genotyped on one of five genotyping platforms (appendix 1 p 12).

Raw genotype data were passed through a custom ancestry prediction and pruning machine learning method as a part of the GenoTools pipeline, as described elsewhere.[10] All samples, other than those from 23andMe (appendix 1 p 14), underwent similar standardised quality control (appendix 1 p 10).

Estimation of risk, age at onset, and admixture

Principal component analysis is a dimensionality reduction method that can be used to identify differences in ancestry among populations and samples, regardless of the historical patterns underlying the structure. To estimate the effect of genetic variation on risk associated with Parkinson's disease, imputed dosages (ie, genotype probabilities for a variant to be A/A, A/B, or B/B, ranging from 0 to 2, which account for some uncertainty) were analysed using a logistic regression model that was adjusted for sex, age, and the first ten principal components as covariates. The principal components were fit on the set of overlapping SNPs between the datasets and the reference panels, before being transformed by uniform manifold approximation and projection (UMAP) to represent the population substructure (appendix 1 p 11). Age at onset was used for cases, and age at recruitment was used for controls. If age at onset was not available for cases, age at recruitment was used instead (<6% of individuals). For individuals who had no age information provided, average age was imputed (<5% of cases and <2% of controls). Summary statistics were generated using PLINK (versions 1.9 and 2.0)[11] and filtered for inclusion after meeting a minimum imputation quality of 0·30 and minor allele frequency (MAF) of more than 5%. To explore the effect of genetic variation on the age at onset of Parkinson's disease cases, a linear regression model, adjusted for the same covariates as for the analysis of disease risk, was performed. In this model, age at onset was defined as the self-reported date of first motor symptom. Additionally, we did linear regression analyses to explore how the potential GWAS signal would correlate with admixture levels. All analyses were performed on Terra. GWAS was conducted on African and African admixed ancestries independently using PLINK and a Bonferroni threshold of 5×10^{-8} before meta-analysis. We used fixed-effect meta-analyses as implemented in METAL (version 1.0)[12] to leverage summary statistics across all sources. Pairwise linkage disequilibrium values were calculated using 1000 Genomes African population data through LDlink. Co-localisation of association summary statistics were visualised using LocusCompareR (version 1.0).

Haplotype and fine-mapping analyses

Haplotype size was compared using individual-level data across African, African admixed, and European Parkinson's disease cases from GP2, data release 5. After standardising the three datasets with the same genotyped SNPs, passing identical quality-control steps, we determined the size of the haplotype blocks using default parameters in PLINK (version 1.9). This analysis estimates haplotype blocks using Haploview, which interprets the block definition. By default, only pairs of variants within 200 kb of each other were considered. Two variants are considered by this procedure to be in strong linkage disequilibrium if the lower bound of the 90% D′ CI was more than 0·70, and the upper bound of the CI was at least 0·98. Fine-mapping analyses were conducted using the R package coloc (version 5.2.2; appendix 1 p 18).

Short-read and long-read whole-genome sequencing

To further dissect the novel identified GWAS signal, we performed whole-genome sequencing (WGS). Short-read WGS DNA sequencing was performed by Psomagen (Rockville, MD, USA; appendix 1 p 15). Long-read WGS data were generated by Oxford Nanopore Technologies (Oxford, UK). High-molecular-weight DNA was extracted from either frozen blood samples or cell lines (appendix 1 p 16). To further understand the functional consequence underlying the GWAS signal, we leveraged existing whole-blood expression quantitative trait locus (eQTL) summary statistics from Kachuri and colleagues[13] based on RNA sequencing from 2733 samples of predominantly African American, Puerto Rican, and Mexican ancestries.

Glucocerebrosidase activity

To investigate whether the novel intronic *GBA1* variant had an effect on glucocerebrosidase activity, we obtained patient-derived lymphoblastoid cell lines from the Coriell Institute for Medical Research (Camden, NJ, USA) National Institute on Neurological Disease and Stroke (NINDS) repository. Lymphoblastoid cell lines were maintained (as directed) in suspension with RPMI 1640 medium (Invitrogen, Carlsbad, CA, USA) containing 2 mmol/L GlutaMAX (Thermo Fisher Scientific), and 15% fetal bovine serum (Thermo Fisher Scientific, Cincinnati, OH, USA) at 37°C in 5% carbon dioxide. Protein was extracted from lymphoblastoid cell lines using a citrate-phosphate buffer (0·2 mol/L disodium phosphate, 0·1 mol/L citrate, protease inhibitor, pH 5·8; Millipore Sigma Aldrich, St Louis, MO, USA) that was activated with 0·25% Triton X-100. Cells were subjected to a 4-methylumbelliferone (Millipore Sigma Aldrich) fluorometric glucocerebrosidase activity assay in technical quadruplicate, as previously reported in the literature,[14] with an adjusted incubation time of 2·5 h. A total of 5×10^6 cells were used per sample with protein concentrations normalised to 0·7 mg/mL via BCA Protein Assay (Thermo Fisher Scientific). Care was taken to avoid freeze-thaw cycles, which can decrease enzymatic activity, and all cells and lysates were stored in identical conditions. All samples were adjusted for background signal.

Role of the funding source

The funders of the study supported clinical data collection and genotyping data generation. The funders had no role in study design, data analysis, data interpretation, or writing of the report.

Results

From the three cohorts, 1488 cases and 196 430 controls with African and African admixed ancestry were included in the GWAS meta-analyses (appendix 1 p 64). In addition, 9230 cases and 4966 controls with European ancestry were also included in comparative analyses. The demographic and clinical characteristics of the cohorts under study are provided in table 1.

A total of 35 SNPs within the *GBA1* locus were significantly associated with Parkinson's disease risk, with consistent directionality of effect across all cohorts, the two most distant SNPs being 639 773 bp apart from each other (appendix 2 tab 2). In the overall meta-analysis, the odds ratio for risk of Parkinson's disease was 1·58 (95% CI 1·37–1·80, $p=2{\cdot}397 \times 10^{-14}$). Conditional analyses on the top two SNPs suggested that there was only one causal signal, which was driven by rs3115534 as the leading SNP (figure 2). Of note, rs3115534-G is much more common in individuals of African ancestry (appendix 1 p 52) or African admixed ancestry (appendix 1 p 1 53) relative to other populations (figure 3). The African and African admixed datasets in this study yielded similar allele frequencies: in the African dataset, the cohort MAF was 0·25, affected MAF was 0·33, and unaffected MAF was 0·19; in the African admixed dataset, respective frequencies were 0·14, 0·22, and 0·13. By comparison, the allele frequency was 0·16 according to gnomAD and 0·21 according to the African 1000 Genomes panel. Within our research

Dr Andrew Singleton, 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 20814, USA
singleta@nih.gov

For more on **IPDGC Africa** see https://www.ipdgc-africa.com/

For more on **GP2** see https://gp2.org/

For **23andMe** see https://www.23andme.com/en-gb

For the **1000 Genomes project** see https://www.internationalgenome.org/

For more on the **Salus institutional review board** see https://versiticlinicaltrials.org/salusirb

	African predicted ancestry		African admixed predicted ancestry*	
	Nigerian origin (IPDGC Africa cohort)†	African, broad unspecified origin (GP2 dataset)‡	African admixed origin (GP2 dataset)§	African admixed origin (23andMe dataset)
Total participants	589	1722	1334	194 273
Recruited from Nigerian sites	589 (100%)	1330 (77%)	50 (4%)	NA
Cases	304	711	185	288
Recruited from Nigerian sites	304 (100%)	672 (95%)	16 (9%)	NA
Female	80 (26%)	206 (29%)	80 (43%)	NA
Male	224 (74%)	505 (71%)	105 (57%)	NA
Controls	285	1011	1149	193 985
Recruited from Nigerian sites	285 (100%)	658 (65%)	34 (3%)	NA
Female	97 (34%)	448 (44%)	714 (62%)	NA
Male	188 (66%)	563 (56%)	435 (38%)	NA
Case age at onset, years	58·20 (9·67)	59·31 (11·37)	57·84 (14·69)	NA
Control age at examination, years	64·4 (7·56)	65·09 (9·55)	66·34 (8·71)	NA
Array	NeuroChip (version 1.0; Illumina, San Diego, CA, USA)	NeuroBooster (version 1.0; Illumina, San Diego, CA, USA)	NeuroBooster (version 1.0; Illumina, San Diego, CA, USA)	OmniExpress (Illumina, San Diego, CA, USA), Infinium GSA (Illumina), and HumanHap550+ (Illumina)

Data are n, n (%), or mean (SD). BLAAC PD=Black and African American Connections to Parkinson's Disease Study. GP2=Global Parkinson's Genetics Program. GSA=Global Screening Array. IPDGC Africa=International Parkinson's Disease Genomics Consortium Africa. MDGAP-KINGS=Movement Disorders Genotypes and Phenotypes—King's College London. NA=not available. PAGE=Parkinson's Genes and Environment Study. PPMI=Parkinson's Progression Markers Initiative. *African admixed defined as individuals ancestrally similar to the following 1000 Genomes project ancestry labels: African ancestry in Southwest United States of America (abbreviated as ASW in the 1000 Genomes project) and African Caribbean in Barbados (abbreviated as ACB in the 1000 Genomes project). †See appendix 1 (p 37) for a complete list of Nigerian hospitals and institutions contributing to this cohort. ‡GP2 cohorts with predicted African ancestry include Baylor College of Medicine (Houston, TX, USA), BioFIND, BLAAC PD, Coriell, MDGAP-KINGS (further details available online), PPMI , PAGE , University of Maryland (College Park, MD, USA), and IPDGC Africa. §GP2 cohorts with predicted African admixed ancestry include Baylor College of Medicine, BioFIND, BLAAC PD, Coriell, MDGAP-KINGS, PPMI, PAGE, University of Maryland, Systemic Synuclein Sampling Study,[15] and IPDGC Africa.

Table 1: **Characteristics of cohorts in the study**

For the **gnomAD database** see https://gnomad.broadinstitute.org/

For **dbSNP** see https://www.ncbi.nlm.nih.gov/snp

For the **Allele Frequency Aggregator** see https://www.ncbi.nlm.nih.gov/snp/docs/gsr/alfa/

For the **OpenTargets database** see https://genetics.opentargets.org/

For the **GenoTools pipeline** see https://github.com/GP2code/GenoTools

For more on **Terra** see https://terra.bio/

For **LDlink** see https://ldlink.nih.gov/?tab=home

For **LocusCompareR** see https://github.com/boxiangliu/locuscomparer

For more on **Haploview** see https://www.broadinstitute.org/haploview/haploview

For the **R package coloc** see https://CRAN.R-project.org/package=coloc

For the **repository** see https://www.coriell.org/1/NINDS

See **Online** for appendix 2

For **BioFIND** see https://biofind.loni.usc.edu/

For more on **BLAAC PD** see https://www.blaacpd.org/

For more on **PPMI** see https://www.ppmi-info.org/

For more on **PAGE** see https://www.pagestudy.org/

For more on **MDGAP-KINGS** see https://gp2.org/the-components-of-gp2s-third-data-release/

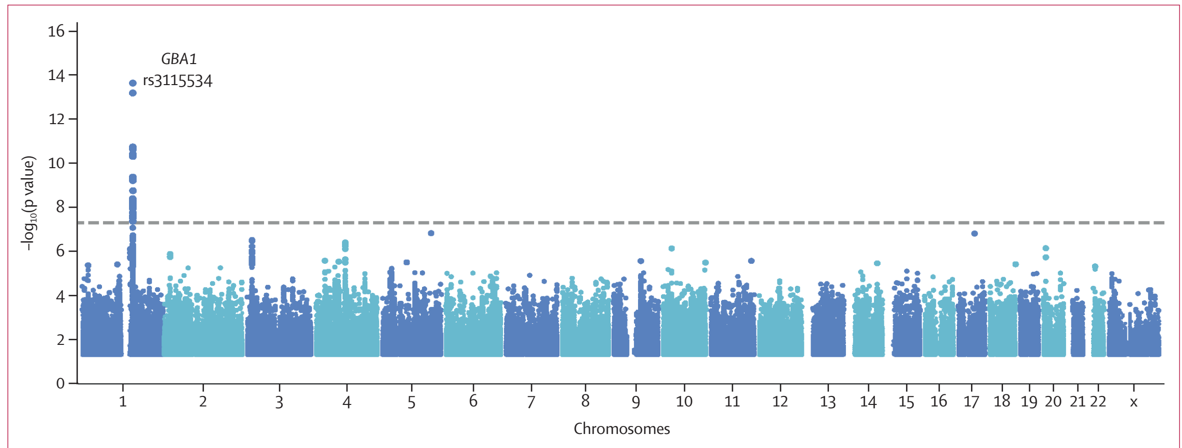


Figure 2: **African and African admixed GWAS meta-analysis assessing Parkinson's disease risk**
Manhattan plot showing the significance of the association as $-\log_{10}$(p value) against chromosomes at a genomic scale (Bonferroni correction highlighted in grey at 5×10^{-8}). GWAS=genome-wide association study.

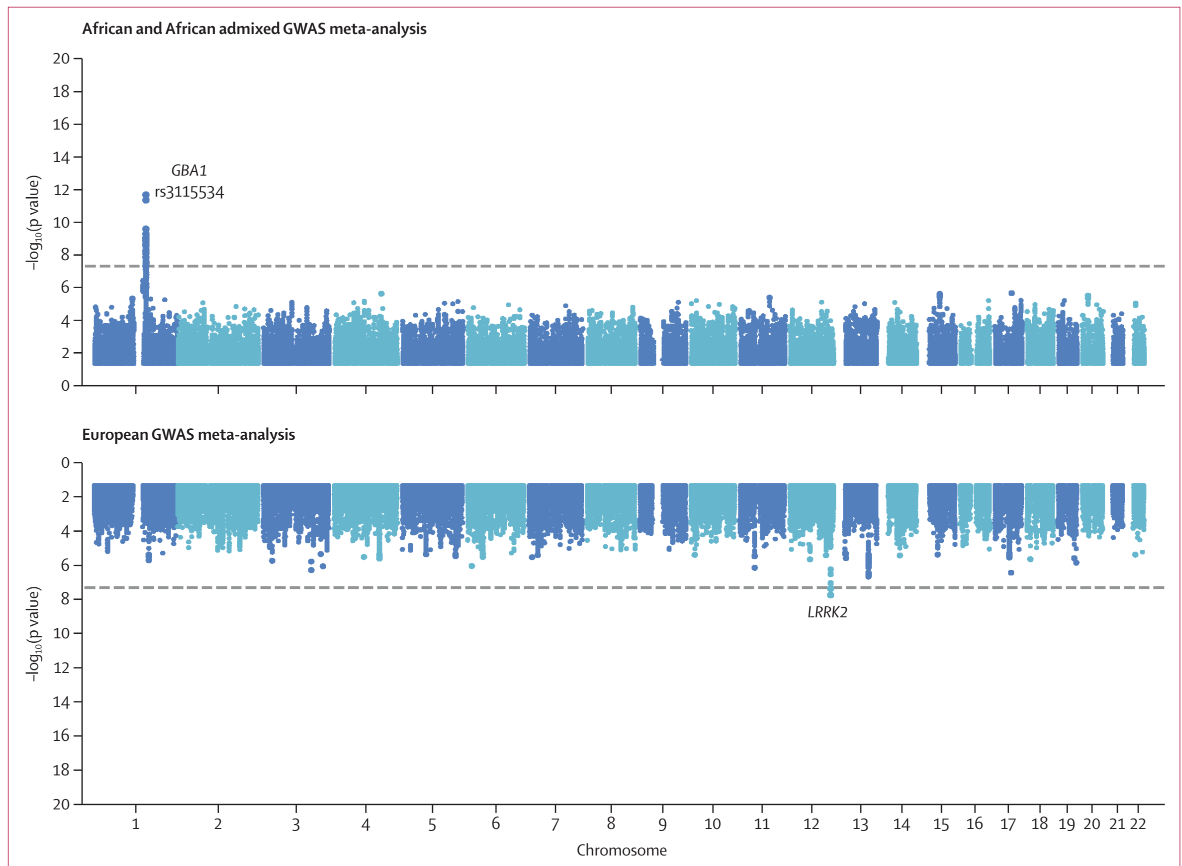


Figure 3: **Miami plot comparing European versus African and African admixed GWAS meta-analyses of Parkinson's disease risk**
Randomly sampled 1200 cases and 2445 controls were included in each of the GWAS meta-analyses. The grey horizontal lines indicate the significance threshold of 5×10^{-8}. GWAS=genome-wide association study.

cohorts, we found that rs3115534-G was more frequent in Nigerian populations (appendix 2 tab 3). Linear regression analyses showed that the *GBA1* rs3115534 variant was positively associated with a higher percentage of African ancestry (β=0·0385 [SE=0·0064], $p=2\cdot002\times10^{-9}$).

To test whether the effect of the risk allele was additive, and to investigate risk ratios, the frequency of homozygotes and heterozygotes was calculated in cases versus controls (appendix 1 p 17; appendix 2 tab 8). As a follow-up analysis, the association of this *GBA1* variant with age

	Base change	Functional consequence	Genetic variant	Cases with variant (n)	Controls with variant (n)	rs3115534-GG carriers (n)	rs3115534-GT carriers (n)	rs3115534-TT carriers (n)
chr1:155236249:A:C*	A→C	Non-synonymous SNV	Ile320Ser	1	0	0	1	0
rs149487315	C→T	Non-synonymous SNV	Met313Ile	1	0	0	0	1
rs143222798	C→T	Synonymous SNV	Gly277Gly	6	3	0	6	3
rs61748906	A→G	Non-synonymous SNV	Trp136Arg	1	0	1	0	0
rs368786234	G→T	Non-synonymous SNV	Ser77Arg	1	0	0	1	0
rs761621516	GTA→deleted	Non-frameshift deletion	Trp75del (222_224del)	1	0	0	1	0
rs150466109	T→C	Non-synonymous SNV	Lys13Arg	12	8	0	10	10

Analyses were done in 141 cases and 65 controls. All variants were on chromosome 1, were exonic, and were heterozygous. SNV=single-nucleotide variant. *Base pair positions reported on hg38 genome build.

Table 2: **Coding variants identified by short-read whole-genome sequencing in carriers and non-carriers of the novel *GBA1* rs3115534 variant**

at onset was assessed. Linear regression analyses in 711 African ancestry cases and 185 African admixed ancestry cases showed that *GBA1* rs3115534-G was also an age-at-onset disease modifier (β=–2·00 [SE=0·57], p=0·0005, for African ancestry; β=–4·15 [SE=0·58], p=0·015, for African admixed ancestry; mean β=–3·06 [SE=0·40], p=0·0077), resulting in onset of Parkinson's disease 3 years earlier per risk allele (appendix 1 p 55).

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 (appendix 1 p 57), suggesting a founder effect and therefore that this haplotype might have originated in these populations. Fine mapping of this locus showed the lead SNP had a posterior probability of being the causal variant of 71·4% (rs3115534; appendix 2 tab 4).

To identify a putative functional coding variant undetectable through genotyping or imputation that could account for the novel GWAS signal, short-read WGS analyses were done on 206 individuals (141 cases and 65 controls), of whom 39 individuals were *GBA1* rs3115534-GG carriers, 69 were rs3115534-GT carriers, and 98 were rs3115534-TT carriers. A 96·6% correlation was observed between short-read WGS and imputed genotyped data for rs3115534, validating the high quality of our imputed data. Long-read WGS data were generated for five *GBA1* rs3115534-GG carriers, two heterozygotes, and six *GBA1* rs3115534-TT carriers. No differences in coding variation were observed between carriers and non-carriers of the GWAS signal (table 2).

Further functional analyses showed a strong eQTL signal at rs3115534, located 8821 bp from the canonical transcription start site (β=0·238 [SE=0·022], p=9·99×10^{-25}; figure 4). The rs3115534-G risk allele was found to be associated with increased *GBA1* expression. Our data suggest a decreasing trend in glucocerebrosidase activity estimates when comparing rs3115534-GG homozygous risk allele carriers (mean 762·50 U [SD 273·50]) versus rs3115534-GT heterozygous carriers (2743·76 U [1960·83]; Welch two-sample *t* test for GG *vs* GT: *t*=–4·3138, df=21·583, p=0·00029) and rs3115534-TT homozygous non-risk allele carriers (1879·94 U [1010·84]) versus rs3115534-GG homozygous risk allele carrier (GG *vs* TT: *t*=–4·7564, df=18·363, p=0·00014; appendix 1 pp 59–60).

Discussion

Although several studies exploring Parkinson's disease genetics in African and African admixed populations have been published,[6,16–29] we gathered—to our knowledge—the largest collection of patients with Parkinson's disease and controls from African and African admixed ancestry populations to comprehensively assess the genetic architecture of Parkinson's disease on a genome-wide scale. We identified a novel African-specific GWAS signal in the *GBA1* locus, significantly associated with risk of Parkinson's disease and age at onset, to be the most important genetic risk factor for Parkinson's disease in these African and African admixed populations. Remarkably, almost a four times larger sample size in cases was required to nominate *GBA1* as one of the major risk factors for Parkinson's disease in the European ancestry population through GWAS,[30] showing the power and benefit of using diverse ancestry data.

GBA1 is a classic pleomorphic locus, showing coding, structural, and non-coding variants that exert different degrees of risk.[31] 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 account for this signal.[32–35]

Strikingly, by leveraging existing eQTL data predominantly of African American ancestry, we found the rs3115534-G risk allele to be associated with increased *GBA1* expression in whole blood, but paradoxically linked with a trend towards decreased glucocerebrosidase activity, which might be due to challenges with RNA sequencing in this locus. We questioned whether this observation could be accounted for 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 such as GTEx. Indeed, transcript diversity is a common and known biological phenomena[36] that could account for the fact that rs3115534-G might increase the expression of a non-functional transcript that in turn would decrease the

For **GTEx** see https://gtexportal.org/

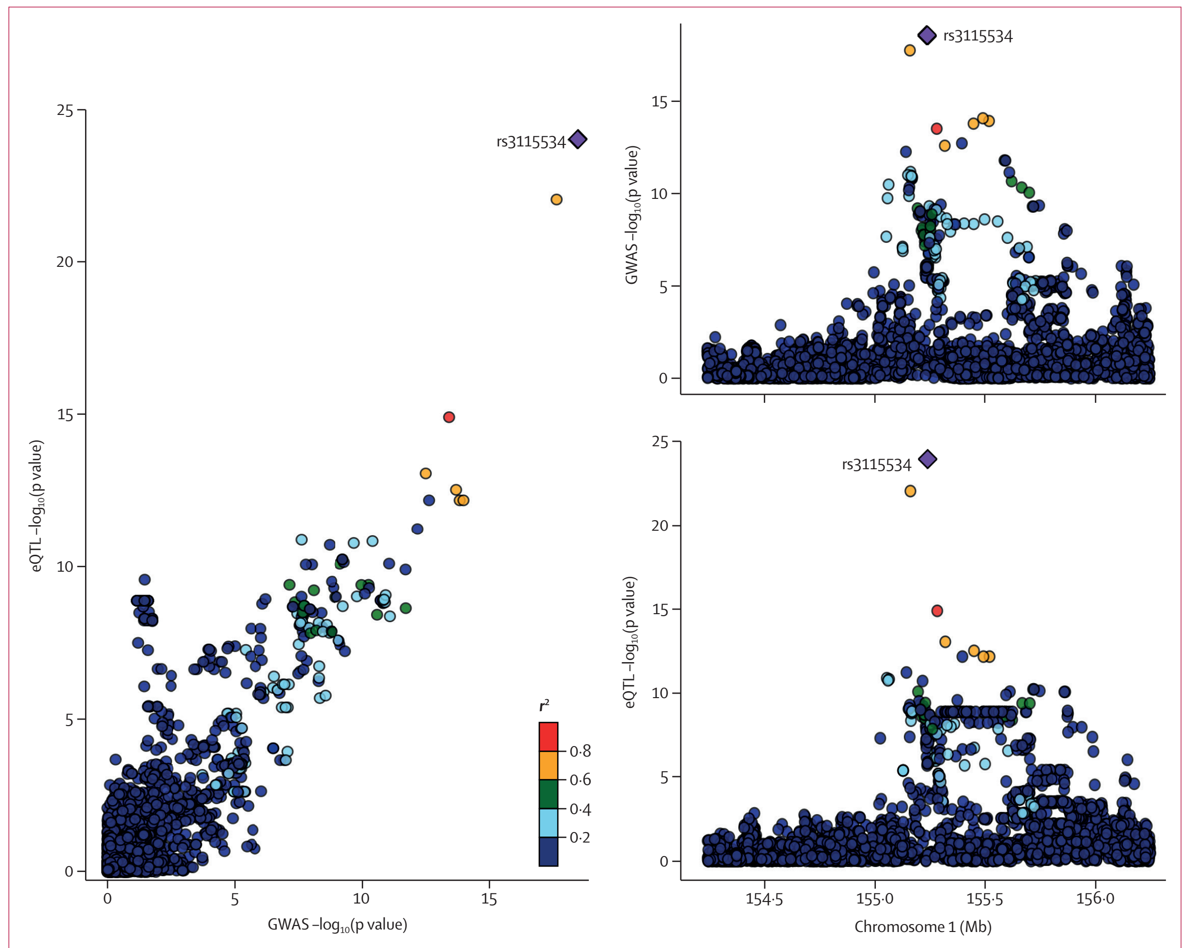


Figure 4: **LocusZoom plot displaying African and African admixed Parkinson's disease GWAS meta-analysis summary statistics versus African American eQTL summary statistics from blood**

eQTL=expression quantitative trait locus. GWAS=genome-wide association study.

levels of the transcript responsible for optimal production of the protein isoform with glucocerebrosidase activity. Future large-scale single-cell expression studies should investigate in which brain cell types these expression differences are most prominent. This potential novel mechanism opens new avenues towards efficient RNA-based therapeutic strategies, such as antisense oligonucleotides or short interfering RNAs, aimed at reducing lifetime risk.

Notably, given the high population frequency of the identified signal and the phenotypic characteristics of the homozygous Africans and African admixed carriers, our study does not support the notion that this variant causes Gaucher disease. Furthermore, the rs3115534 variant has been found to be very rare in non-African and non-African admixed populations. To further dissect the novel signal identified in the *GBA1* locus, effect estimates and directionality of effect were compared, leveraging summary statistics from the largest GWAS meta-analysis of Parkinson's disease in European,[1] Latin American,[37] and east and south Asian populations.[38] The rs3115534-G allele is very rare in European (allele frequency of 0·0015), east Asian (0·0005), south Asian (0·0017), and Ashkenazi Jewish populations (0·0009) according to gnomAD. These findings suggest an African founder effect and reinforce the notion that the genetic architecture of this locus and its influence on risk and onset is different across populations. The variant rs3115534 was also found to be associated with age at onset of Parkinson's disease in our study. The *GBA1* locus in African and African admixed populations differs substantially from European populations (figure 3; appendix 1 p 56), whose association with disease risk is driven by two independent signals, including rs35749011 (*GBA1*-Glu326Lys) and rs76763715 (*GBA1*-Asn370Ser). These variants are very rare in individuals of African and African admixed ancestry (appendix 1 p 62). Similarly, the *GBA1* locus considerably differs from that in the east Asian population, for which the rs3115534 variant was also not imputed in the largest east Asian GWAS meta-analysis (appendix 1 p 62). These differences are less noticeable when assessing the Amerindian and Latin

American and indigenous populations, which harbour higher levels of African admixture (appendix 1 p 62): rs3115534-G; odds ratio 1·13 [95% CI 0·41–1·86], p=0·72;[37] rs3115534-G; 1·56 [1·55–1·88], p=0·01, in the Amerindian and Latin American and indigenous 23andMe GWAS [unpublished]).

Our findings provide crucial insights into targeted construction of African ancestral haplotypes and potential novel pathogenic mechanisms underlying Parkinson's disease. The utility of genetically characterising populations of African and African admixed ancestry is unquestionable. This study demonstrates the importance of haplotype substructure discoveries for future fine-mapping efforts, showing how leveraging unique populations can benefit our understanding of complex diseases.

Overall, by addressing the genetic complexity underlying these under-represented populations, our study represents a valuable resource for identifying and tracking *GBA1* carriers that might prove to be relevant for enrolment in target-specific Parkinson's disease clinical trials. We envisage that these data generated under the GP2 initiative will be key to shed light on the molecular mechanisms involved in the disease process and might pave the way for future clinical trials and therapeutic interventions.

Although we have made progress in assessing genetic risk factors for Parkinson's disease in an under-served population, our study has several limitations. Unravelling additional susceptibility genetic risk and phenotypic relationships would have been possible if a larger cohort had been analysed. The largest Parkinson's disease GWAS and multi-ancestry GWAS meta-analyses so far identified a total of 104 independent significant risk variants.[1,38,39] Of the 104 variants, 91 variants passed quality control, imputation filters, and were present in the African and African admixed GWAS meta-analysis (table 2; appendix 1 p 63). Of the 91 variants, 16 were nominally significant ($p<0·05$; appendix 2 tab 5) in the African and African admixed meta-GWAS reported here. Considering our limited sample size, we lacked statistical power to detect common genetic variants of smaller effect sizes (appendix 1 p 61).

Another limitation is that an important proportion of the genetic risk contributing to the missing heritability of Parkinson's disease in the African and African admixed populations might result from rare alleles and structural variants that have not been assessed in this study. Owing to the scarcity of well powered and African or African admixed RNA sequencing datasets, the added complexity of multi-mapping reads between *GBA1* and *GBAP1*, and the shortage of lymphoblastoid cell lines to explore glucocerebrosidase activity in a large-scale manner, this potential novel functional mechanism merits further study. We are aware that, although our study represents the first GWAS on Parkinson's disease in the African and African admixed populations, two-thirds of the cases are of Nigerian descent, and therefore probably unrepresentative of the substantial genetic diversity across the continent. Parkinson's disease cases recruited from cohorts in Africa—ie, IPDGC Africa—are predominantly from west Africa, specifically Nigeria, and are therefore unlikely to be representative of the whole of Africa. However, most controls in this study were recruited from global efforts—ie, 23andMe—and have higher percentages of admixture. Some of the individuals predicted to be of African descent cannot with certainty be defined as from Nigeria, but nonetheless are unmistakably African, as indicated by a principal component analysis comparing our data with those from the 1000 Genomes reference panel (appendix 1 p 51).

In summary, we performed the largest genome-wide association study evaluating risk of Parkinson's disease in individuals of African ancestry. We identified an intronic *GBA1* variant, rs3115534, not previously associated with risk of Parkinson's disease. Short-read and long-read sequencing did not identify coding or structural variants, and fine-mapping prioritised this variant with a posterior probability of more than 70%. Our haplotype analyses indicate that rs3115534 is frequent in west African populations and suggests a possible founder effect, underscoring the significance once again of ancestral diversity in genetic studies. Because of the intronic nature of this variant, gene expression is potentially modified, paving the way for explorations in RNA-based or other therapeutic interventions that target the reduction of lifetime risk.

Contributors

MR, NUO, AS, CB, JH, HH, MAN, MBM, SB-C, PWC, EAS, and NT contributed to the study concept or design. All authors were involved in sample and data acquisition and access to raw data (not including 23andMe). MR, MBM, SB-C, KSL, DV, MJK, CB, KH, MAN, AS, OOOj, and NUO verified the underlying data. MBM, SB-C, DV, KSL, CXA, MAN, MJK, CB, KJB, PAJ, KD, JJK, HLL, HI, JHK, OOOj, IE, OOk, and KH did the analysis. All authors contributed to critical review and had final responsibility for the decision to submit for publication.

Declaration of interests

DV, HI, HLL, KSL, CXA, and MAN's participation in this project was part of a competitive contract awarded to Data Tecnica International by the US National Institutes of Health (NIH) to support open science research. MAN also currently serves on the scientific advisory board for Character Biosciences and Neuron 23. KH is employed by 23andMe and holds stock or stock options in 23andMe. AS, MBM, PAJ, CB, KD, MJK, JJK, and PWC are employed by the NIH. AS also declares funding for the present work from the Michael J Fox Foundation (MJFF) and Aligning Science Across Parkinson's (ASAP); royalties for a diagnostic for stroke (unrelated to the current work); honoraria for associate editorial work for the journals *Movement Disorders* and *npj Parkinson's Disease*; and travel support from the Chan Zuckerberg Initiative to attend annual investigators' meeting, MJFF to attend Parkinson's Progression Marker Initiative annual meeting, and Weill Cornell to give grand rounds. The spouse of AS is an employee of GeneDx. OOk, IE, HI, JHK, and DV declare funding from the NIH (1ZIA AG000534-04). HI also declares honoraria from GP2 for a steering committee meeting and from MJFF for a data community meeting. DGS declares support for the present work from ASAP (for the BLAAC-PD study); research support from the NIH (P50 108675), the MJFF, the Parkinson Foundation of the National Capital Area, the American Parkinson Disease Association, AbbVie, and Genentech; book royalties from McGraw Hill; consulting fees from AbbVie, Curium

Pharma, F Hoffman-La Roche, Appello Pharma, and Blue Rock Therapeutics; participation on a data safety monitoring board or advisory board for Sanofi-Aventis, Theravance, and Alnylam Pharmaceuticals; being Deputy Editor of the journal *Movement Disorders*; and being Chair of the Scientific Advisory Board of the American Parkinson Disease Association. DAH declares support for the present work from the NIH, the MJFF, the CHDI Foundation, Parkinson's Foundation, Lundbeck, Uniqure, and Neurocrine. TX declares research funding from the MJFF, the American Parkinson's Disease Association, and the NIH; and consulting fees from Parkinson's Foundation and CVS Caremark. HH and JH declare research funding by the Medical Research Council (UK), the Wellcome Trust, the MSA Trust, NIHR University College London Hospitals Biomedical Research Centre (NIHR-BRC), the MJFF, the Fidelity Trust, the Rosetrees Trust, the Guarantors of Brain, SOLVE-RD, and the Dolby Family Fund. HRM declares support from the MJFF related to this work; grants not related to this work from PSP Association, CBD Solutions, Drake Foundation, and Cure Parkinson's Trust; consulting fees from Roche, Amylyx, and Aprinoia; speaker's honoraria from Kyowa-Kirin, the *British Medical Journal*, and the Movement Disorders Society; travel support from the MJFF and the Movement Disorders Society; and being on the advisory board for Cure PSP Association, the Association of British Neurologists Movement Disorders special interest group, and the Association of British Neurologists Neurogenetics advisory group. HRM is a co-applicant on a patent application related to C9ORF72—Method for diagnosing a neurodegenerative disease (PCT/GB2012/052140). ES declares funding for the present work from the US National Human Genome Research Institute (NHGRI) Intramural Research Program; grants from ASAP and MJFF; and a cooperative research and development agreement with Roche. EAS declares funding for the present work from the MJFF; and travel support from the MJFF. MND declares funding for the present work from MJFF; grant support from ASAP; speaker's honoraria from the Parkinson's Foundation; and is a member of the Parkinson's Foundation Gulf Coast advisory board. Z-HF declares salary support from the MJFF. OOOj declares a study grant from the NIHR; honoraria for educational courses from the International Parkinson and Movement Disorder Society (IPMDS); travel support from P2 for an annual investigators' meeting and from the IPMDS for congress attendance; and is a member of the executive committee of the IPMDS. NUO declares a study grant from the NIHR; travel support and honoraria for educational courses from the IPMDS; and being Chair of the IPMDS Africa Section. UA declares a study grant from the NIHR. RA declares grant support from the NIH (U01HG010273, U19AG074865), the UK Royal Society and the African Academy of Sciences Future Leaders—African Independent Research (FLR/R1/191813, FCG/R1/201034), and the Global Brain Health Institute, Alzheimer's Association, and Alzheimer's Society UK (GBHI ALZ UK-21- 24204). TF declares speaker's honoraria from Roche; travel support from the Alzheimer's Association; and is a member of the Alzheimer's Disease International medical and scientific advisory panel. CB declares support from ASAP. KD declares a Japan Society for the Promotion of Science Research Fellowship. JJK declares participation in the graduate school programme for Queen Mary University London (London, UK). PAJ declares participation in the graduate school programme for University College London (London, UK). MBM declares participation in a summer internship at Genentech/Roche (unrelated to the current work). All other authors declare no competing interests.

For the **Biowulf Linux cluster** see http://hpc.nih.gov

For the **Accelerating Medicines Partnership in Parkinson's Disease website** see https://www.amp-pd.org/

For the **23andMe summary statistics** see https://research.23andme.com/dataset-access/

For **GenoTools** see https://github.com/GP2code/GenoTools

For **all scripts** see https://github.com/GP2code/GP2-AFR-AAC-metaGWAS

Data sharing

All GP2 data are hosted in collaboration with the Accelerating Medicines Partnership in Parkinson's Disease and are available via application on the website. The GWAS summary statistics from this study, excluding 23andMe, are available as of GP2's release 5. 23andMe summary statistics are available via application on the website. Genotyping imputation, quality control, ancestry prediction, and processing were performed using GenoTools (version 10), publicly available on GitHub. All scripts for analyses are publicly available on GitHub.

Acknowledgments

We thank all the participants who contributed to this study. This work was supported in part by the Intramural Research Program of the NIH, the National Institute on Aging (NIA), the NIH, the US Department of Health and Human Services (project number ZO1 AG000535 and ZIA AG000949), the National Institute of Neurological Disorders and Stroke (NINDS), and NHGRI. This work was also supported in part by the GP2. GP2 is funded by the ASAP initiative and is implemented by the MJFF for Parkinson's Research. Additional funding was provided by the MJFF for Parkinson's Research (grant MJFF-009421/17483). A complete list of GP2 members is available on the GP2 website. We thank the research participants and employees of 23andMe. Members of the 23andMe Research Team who contributed to this study are listed in appendix 1 (p 47–48). We also thank members of the BLAAC-PD team (appendix 1 p 45). We thank the Biowulf team at the NIH, because this study used the high-performance computational capabilities of the Biowulf Linux cluster.

References

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

2 Blauwendraat C, Nalls MA, Singleton AB. The genetic architecture of Parkinson's disease. *Lancet Neurol* 2020; **19:** 170–78.

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

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

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

6 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:** 567–71.

7 Choudhury A, Aron S, Botigué LR, et al. High-depth African genomes inform human migration and health. *Nature* 2020; **586:** 741–48.

8 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:** 181–84.

9 Blauwendraat C, Faghri F, Pihlstrom L, et al. NeuroChip, an updated version of the NeuroX genotyping platform to rapidly screen for variants associated with neurological diseases. *Neurobiol Aging* 2017; **57:** 247.e9–13.

10 Koretsky MJ, Alvarado C, Makarious MB, et al. Genetic risk factor clustering within and across neurodegenerative diseases. *Brain* 2023; published online May 16. https://doi.org/10.1093/brain/awad161.

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

12 Willer CJ, Li Y, Abecasis GR. METAL: fast and efficient meta-analysis of genomewide association scans. *Bioinformatics* 2010; **26:** 2190–91.

13 Kachuri L, Mak ACY, Hu D, et al. Gene expression in African Americans, Puerto Ricans and Mexican Americans reveals ancestry-specific patterns of genetic architecture. *Nat Genet* 2023; **55:** 952–63.

14 Peters SP, Lee RE, Glew RH. A microassay for Gaucher's disease. *Clin Chim Acta* 1975; **60:** 391–96.

15 Visanji NP, Mollenhauer B, Beach TG, et al. The Systemic Synuclein Sampling Study: toward a biomarker for Parkinson's disease. *Biomark Med* 2017; **11:** 359–68.

16 Ross OA, Wilhoite GJ, Bacon JA, et al. *LRRK2* variation and Parkinson's disease in African Americans. *Mov Disord* 2010; **25:** 1973–76.

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

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:** 296–99.

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

20 Milanowski LM, Oshinaike O, Walton RL, et al. Screening of *GBA* mutations in Nigerian patients with Parkinson's disease. *Mov Disord* 2021; **36:** 2971–73.

21 Nishioka K, Ross OA, Vilariño-Güell C, et al. Glucocerebrosidase mutations in diffuse Lewy body disease. *Parkinsonism Relat Disord* 2011; **17:** 55–57.

22 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:** 116–21.

23 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:** 861–66.

24 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:** 305–12.

25 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:** 591–94.

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

27 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.

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

29 Okunoye O, Ojo O, Abiodun O, et al. *MAPT* allele and haplotype frequencies in Nigerian Africans: population distribution and association with Parkinson's disease risk and age at onset. *medRxiv* 2023; published online March 24. https://doi.org/10.1101/2023.03.24.23287684 (preprint).

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

31 Singleton A, Hardy J. A generalizable hypothesis for the genetic architecture of disease: pleomorphic risk loci. *Hum Mol Genet* 2011; **20:** R158–62.

32 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:** 670.

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

34 Tayebi N, Park J, Madike V, Sidransky E. Gene rearranagement on 1q21 introducing a duplication of the glucocerebrosidase pseudogene and a metaxin fusion gene. *Hum Genet* 2000; **107:** 400–03.

35 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.e11–14.

36 Gustavsson EK, Sethi S, Gao Y, et al. The annotation and function of the Parkinson's and Gaucher disease-linked gene *GBA1* has been concealed by its protein-coding pseudogene *GBAP1*. *bioRxiv* 2023; published online March 21. https://doi.org/10.1101/2022.10.21.513169 (preprint).

37 Loesch DP, Horimoto ARVR, Heilbron K, et al. Characterizing the genetic architecture of Parkinson's disease in Latinos. *Ann Neurol* 2021; **90:** 353–65.

38 Foo JN, Chew EGY, Chung SJ, et al. Identification of risk loci for Parkinson disease in Asians and comparison of risk between Asians and Europeans: a genome-wide association study. *JAMA Neurol* 2020; **77:** 746–54.

39 Kim JJ, Vitale D, Véliz Otani D, et al. Multi-ancestry genome-wide meta-analysis in Parkinson's disease. *medRxiv* 2022; published online Aug 6. https://doi.org/10.1101/2022.08.04.22278432 (preprint).