Exploring the role of underrepresented populations in polygenic risk scores for neurodegenerative disease risk prediction

Kathryn Step[1], Carene Anne Alene Ndong Sima[1], Ignacio Mata[2] and Soraya Bardien[1,3]*

[1]Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town, South Africa, [2]Genomic Medicine, Lerner Research Institute, Cleveland Clinic Foundation, Cleveland, OH, United States, [3]South African Medical Research Council/Stellenbosch University Genomics of Brain Disorders Research Unit, Stellenbosch University, Cape Town, South Africa

Introduction

Neurodegenerative diseases (NDs) are characterized by the progressive loss of neuronal function and structure (Lamptey et al., 2022). Additionally, neurodegeneration is associated with neural networks, synaptic dysfunction, and the accumulation of physiochemically altered proteins in the brain (Hoover et al., 2010; Milnerwood and Raymond, 2010). NDs affect ∼15% of the global population making them the leading cause of cognitive and physical disability worldwide (Feigin et al., 2020). Moreover, the prevalence of several of these diseases increases with age, therefore the global trend toward an increased life expectancy may also result in an increased burden of age-related NDs such as Alzheimer's disease (AD) and Parkinson's disease (PD; GBD 2016 Parkinson's Disease Collaborators, 2018). Of all the NDs, AD and related dementias and PD have been the most studied and genetically characterized. For this reason, these two disorders are highlighted in this article.

On an international scale, the majority of genetic research still primarily focuses on populations of European ancestry resulting in the relative absence of "non-European" populations (also known as underrepresented populations; URPs) from large-scale genomic study cohorts and subsequently, GWAS (Knerr et al., 2011). Many countries in the "Global South" are considered URPs, including but not limited to countries in Africa, Asia, Caribbean, Latin America, and Oceania (excluding Australia and New Zealand; Fatumo et al., 2022; Schumacher-Schuh et al., 2022; Bhattacharya et al., 2023). For the purposes of this article, URPs include countries in Asia such as Japan and Korea, since they have also been comparatively underrepresented in genetic studies.

There is an increased prevalence of PD and AD in URPs (GBD 2021 Nervous System Disorders Collaborators, 2024). Moreover, AD and related dementias are in the top five most prevalent nervous system disorders across URPs (GBD 2021 Nervous System Disorders Collaborators, 2024). It is predicted that worldwide there will be 65.7 million people in 2030 and 115.4 million people in 2050 living with dementia, with 63 and 71% living in URPs in 2030 and 2050, respectively (Ferri et al., 2005; Prince et al., 2013).

AD and PD have a complex etiology and while age is a common risk factor of certain disease development, many are considered multifactorial diseases with both environmental and genetic components. Exposure to environmental pollutants such as neurotoxic metals (e.g., lead, manganese, and mercury) and pesticides have been associated with disease pathogenesis by increasing beta-amyloid peptides, the phosphorylation of Tau protein, and the aggregation of α-synuclein (Chin-Chan et al., 2015). Although there have been great advances in elucidating the genetic architectures of AD and PD, the complete genetic etiology has yet to be defined further highlighting the complexity of these diseases. However, studies have shown that these diseases have a significant genetic component through the identification of rare variants with large effect in causal genes in familial cases, as well as common variants with smaller effect through association analysis studies in sporadic cases (Perrone et al., 2021).

Genetic risk variants, commonly known as susceptibility variants, that are statistically associated with the risk of developing a particular disease or trait, are identified using genome-wide association studies (GWAS; Uffelmann et al., 2021). In contrast to linkage analysis studies that utilize multiplex family datasets, GWAS use unrelated case-control datasets to perform an association analysis (Tan et al., 2014). These studies have revealed the polygenic architecture of several common complex disorders including NDs. For example, more than 90 and 75 genetic loci have been linked to PD (Nalls et al., 2019) and AD development (Bellenguez et al., 2022), respectively.

GWAS on NDs in URPs

Historically, GWAS have almost solely focused on participants of European ancestry (Lewis and Vassos, 2020), where ∼94.8% of previously conducted GWAS included European participants, while only 5.2% included all other population groups combined (Mills and Rahal, 2020). Since current risk prediction methods require an individual's genetic ancestry to be similar to the ancestries from a well-powered GWAS study, from which effect sizes can be estimated (Martin et al., 2019), this significantly limits the applicability of PRS done on Europeans, to URPs. A review of the GWAS Catalog (Sollis et al., 2023), revealed that as of September 11th, 2023, there were 355 ND GWAS studies reported with only 18% ($n = 63$) of these studies including individuals from URPs. This data is summarized in Figure 1, and a complete list of these studies is provided in Supplementary Table 1.
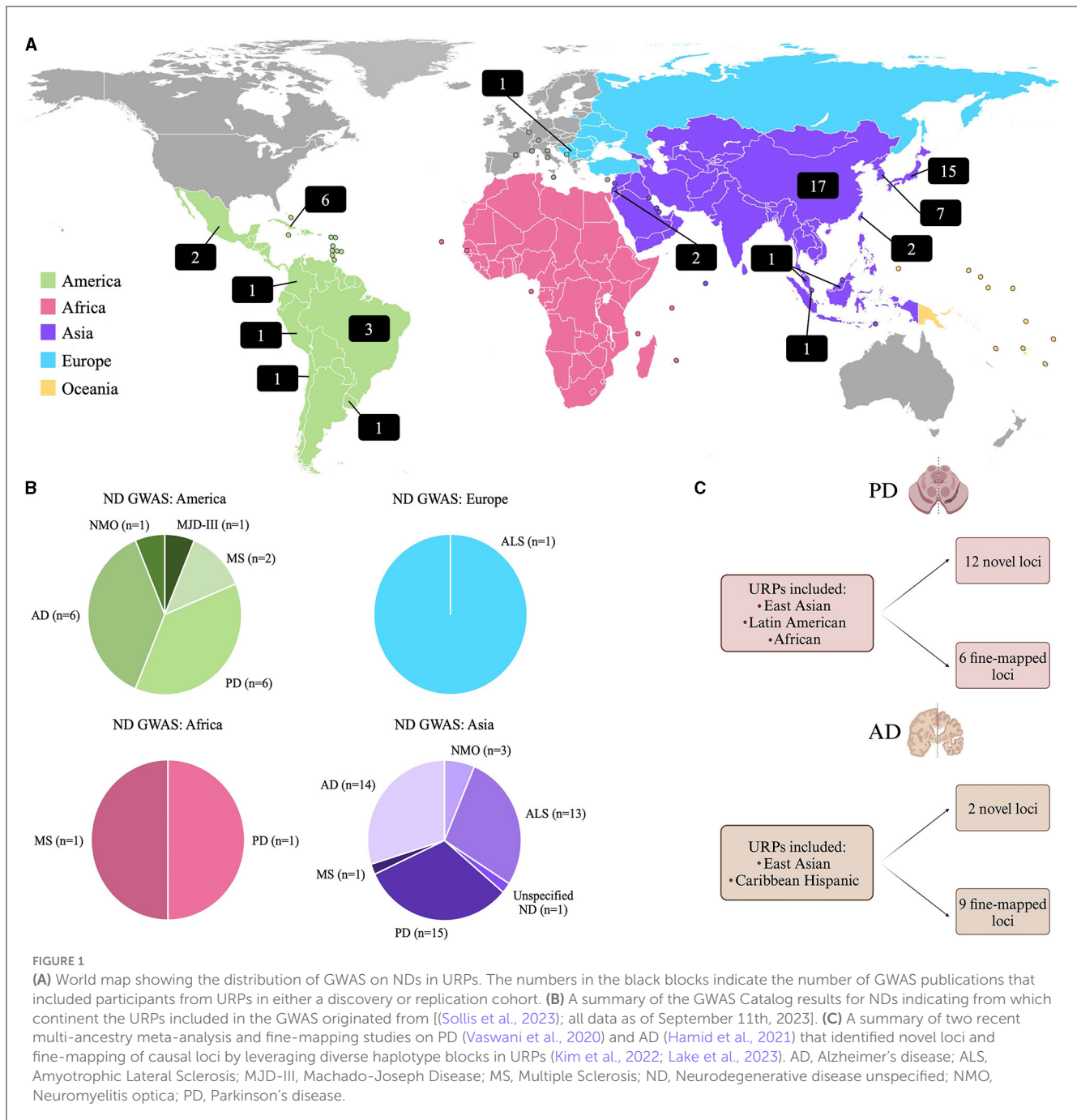
Moreover, there is a lack of diversity in ND research overall, including clinical trials, epidemiology studies, neuroimaging studies, and case studies (Vaswani et al., 2020; Franzen et al., 2022). This may be due to limited neurology services in URP regions, such as Africa, leading to people with NDs being undiagnosed or mis-diagnosed (Hamid et al., 2021). The disproportionate exclusion of ethnically diverse individuals in ND research and clinical trials can be attributed to the challenges with the diagnosis process, recruitment strategies, study enrolment, and participant retention practices (Gilmore-Bykovskyi et al., 2019).

Polygenic risk scores

Originally, risk prediction for complex diseases relied heavily on information on the family's medical history. More recently, polygenic risk scores (PRS) can be calculated and used to estimate an individual's lifetime risk of developing a particular disease. This is based on the combination of genetic susceptibility variants that they carry contributing either to an increased or decreased disease risk (Lewis and Vassos, 2020). The genetic susceptibility variants and the effect size estimates utilized in a PRS analysis are typically identified through GWAS. However, it should be noted that PRS only explains a small percentage of the risk, which needs to be complemented with a combination of factors including family history, environmental exposures, lifestyle factors, and clinical features, such as loss of smell and sleep disturbances as prodromal symptoms of disease (So et al., 2011; Seifan et al., 2019). It has been reported that PRS based on common genetic variants accounts for ∼7% of the variance in AD risk in European populations (Leonenko et al., 2021) and 2.2% variance for PD risk in the Latino population (Loesch et al., 2022). Ultimately, increasing representation from URPs may improve the predictive validity by capturing a broader range of genetic variation associated with the disease.

As an emerging precision medicine tool, accurate disease risk prediction could inform clinicians at different points of disease trajectory from providing information regarding the most efficient prevention strategies, to targeted screening interventions and disease diagnosis, as well as provide improved clinical management (Lewis and Vassos, 2020; Corpas et al., 2022). While there is evidence for the clinical utility and validity of PRS, there are still concerns relating to the transferability of these results across populations, the accuracy of disease prediction in heterogenous diseases, and whether PRS disclosure does indeed affect morbidity and mortality rates (Kumuthini et al., 2022). An individual's genetic ancestry has a significant impact on risk prediction. The distinct genetic architectures of different populations influence PRS performance due to differences in population-specific linkage disequilibrium patterns and allele frequencies (Sirugo et al., 2019). Additionally, the transferability of PRS across different populations can be questioned, as prevalence rates and genetic risk variants vary between populations (Kachuri et al., 2024). Therefore, a PRS developed using one ancestry may not be applicable to individuals of another ancestral group (Fatumo et al., 2023).

To address the issue of transferability, it is essential to develop methods and models that are inclusive of diverse populations and incorporate ancestry-specific genetic information (Cavazos and Witte, 2021; Fatumo et al., 2023). The incorporation of local ancestry in GWAS, using software such as TRACTOR (Atkinson et al., 2021), has shown promising results in identifying ancestry-specific risk loci which in turn can be used for PRS calculations (Swart et al., 2021). Recent efforts, such as PRS-CSx (Ruan et al., 2022) and BridgePRS (Hoggart et al., 2023), aim to improve PRS transferability by leveraging multi-ethnic datasets and accounting for population stratification and admixture. However, accurately predicting disease risk in admixed populations remains a challenge due to their complex genetic composition (Duan et al., 2018; Swart

FIGURE 1
(A) World map showing the distribution of GWAS on NDs in URPs. The numbers in the black blocks indicate the number of GWAS publications that included participants from URPs in either a discovery or replication cohort. **(B)** A summary of the GWAS Catalog results for NDs indicating from which continent the URPs included in the GWAS originated from [(Sollis et al., 2023); all data as of September 11th, 2023]. **(C)** A summary of two recent multi-ancestry meta-analysis and fine-mapping studies on PD (Vaswani et al., 2020) and AD (Hamid et al., 2021) that identified novel loci and fine-mapping of causal loci by leveraging diverse haplotype blocks in URPs (Kim et al., 2022; Lake et al., 2023). AD, Alzheimer's disease; ALS, Amyotrophic Lateral Sclerosis; MJD-III, Machado-Joseph Disease; MS, Multiple Sclerosis; ND, Neurodegenerative disease unspecified; NMO, Neuromyelitis optica; PD, Parkinson's disease.

et al., 2021). Therefore, the general PRS calculation method that utilizes only a single training (i.e., discovery) population would not be adequate to predict polygenic risk in admixed individuals since the choice of the discovery cohort to use is unknown. To tackle this issue, approaches like using a linear combination of PRS based on multiple training datasets show promise in improving prediction accuracy in admixed populations (Márquez-Luna and Loh, 2017). However, there are still limitations with the current methods, and it remains vital that the development of novel methods should go hand-in-hand with increasing the diversity of study cohorts.

Steady increase in genomic data from URPs

There has been a steady increase in the number of GWAS including URPs, specifically individuals of Asian ancestry (Mills and Rahal, 2020). This has been achieved through the implementation of large prospective population-based biobanks such as the China Kadoorie Biobank (Chen et al., 2011), BioBank Japan (Nagai et al., 2017) and major genome projects such as the Singapore 10K Genome project (Precise Health Research Singapore, 2023). In the future, the application of PRS in African

countries may be feasible through initiatives like the Three Million African Genome (3MAG) project (Wonkam, 2021).

Furthermore, as the field progresses, there is an increase in the establishment of disease-specific consortia. For example, the Multi-Partner Consortium to Expand Dementia Research in Latin America (ReDLat) is an initiative aiming to increase dementia research in Latin America and the Caribbean by combining clinical and genomic data (Ibanez et al., 2021). Likewise, the European Alzheimer & Dementia Biobank (EADB) consortium brings together various European GWAS consortia focused on investigating the genetic architecture of AD through a GWAS meta-analysis study (Bellenguez et al., 2022). Notably, collaborative efforts in the field of PD genetics have been particularly promising. Consortia such as the Latin American Research Consortium on the Genetics of Parkinson's Disease (LARGE-PD, Zabetian and Mata, 2017), the Genetic Epidemiology of Parkinson's Disease consortium (GEoPD; Farrer et al., 2021), the International Parkinson's Disease Genomics Consortium (IPDGC) Africa (Rizig et al., 2021), and the East Asian Parkinson's Disease Genomics Consortium (EAPDGC; Mok, 2021) were formed to investigate the genetic etiology of PD. Many of these consortia are in the process of integrating into a global effort to support the Global Parkinson's Genetics Program (GP2).

GP2 is an international collaboration investigating the genetic etiology of PD and making the knowledge globally relevant by increasing representation of URPs in PD research (Towns et al., 2023). GP2's goal is to collect standardized clinical data and genotype more than 200,000 participants with a minimum of 50,000 individuals from URPs for use in association studies. A recent GP2 GWAS, involving African and African admixed participants (not yet included in GWAS Catalog), revealed a novel ancestry-specific genetic risk factor for PD in *GBA1* (rs3115534; Rizig et al., 2023). This finding highlights the *value of leveraging genetic diversity* to detect novel disease associations. Collaborative large-scale international efforts, such as GP2, show promise in correcting the significant imbalance of European population biases in GWAS, thereby ultimately improving the accuracy of disease prediction in URPs.

Additional key factors to aid in the analysis of admixed populations and allow for the inclusion of URPs in GWAS include suitable genotyping arrays, reference panels for phasing and imputation, ancestry inference, and robust GWAS QC metrics tailored for URPs (Atkinson et al., 2021; Yang et al., 2023). Innovations in methodological development are ongoing to address the complexities of genetic studies involving URPs. Anticipated advancements in statistical methods, software tools, and genetic reference panels are poised to play pivotal roles as research endeavors increasingly embrace representative sample recruitment and the global population's growing diversity (Sariya et al., 2019; Tan and Atkinson, 2023).

Value of increasing URPs in genomic studies

It has been widely acknowledged that increasing diversity in genetic research will aid in understanding disease etiology, ultimately improving the effectiveness of genomic medicine

(Peterson et al., 2019). However, results from genetic research do not translate equally across populations, further highlighting the importance of including URPs in genetic research. For example, the apolipoprotein E (*APOE*) gene is a well-studied risk factor for AD and related dementias, however, its prevalence and influence on disease risk varies significantly among ancestral groups (Farrer et al., 1997; Tang et al., 2001). In East-Asian populations, *APOE* ε4 is the strongest risk effect, however, in African populations, it is the lowest risk effect, trending toward a protective effect (Hendrie et al., 2014; Rajabli et al., 2022). Moreover, many URPs are comprised of multiway admixed individuals with genomic architecture that consists of smaller haplotype blocks than those of the European genomes (Fatumo et al., 2023). Haplotype structure is useful in inferring information relating to gene flow and population structure (Shipilina et al., 2023).

Consequently, the inclusion of URPs can aid in fine-mapping known disease loci to improve post-GWAS functional studies by reducing the number of candidate variants prioritized. Additionally, the alleles observed in diverse admixed populations can facilitate the identification of novel population-specific disease risk variants (Rizig et al., 2023) and reveal disease risk variants for various populations, simultaneously (Swart et al., 2022). This utility was highlighted in recent multi-ancestry GWAS studies on PD and AD (Figure 1C). A PD meta-analysis identified 12 potentially novel loci and six putative causal variants (Kim et al., 2022). For AD, a similar approach was employed to leverage diverse haplotype structures to aid in variant identification and causal loci fine-mapping, where two novel disease loci were identified and nine loci were fine-mapped (Lake et al., 2023).

In addition to the unique ancestral makeup, there are a range of varying socio-economic factors affecting URPs, resulting in observed differences in disease prevalence and phenotypes. For example, in comparison to European populations, Latin American populations are 1.5 times more likely to develop AD and related dementias (Prince et al., 2013; Matthews et al., 2018). Factors affecting the increased prevalence may include lifestyle, work exposures, economic access to resources, access to healthcare, education on preventive and precautionary actions as well as living conditions (Epping-Jordan et al., 2005; Babulal et al., 2019). Consequently, the inclusion of URPs in genomic studies is vital to investigating the underlying genetic etiology of these diseases through the identification of pathogenic and susceptibility variants, in addition to identifying the social and environmental factors resulting in an increased or decreased disease risk.

It has been shown that there is limited clinical utility for PRS in non-European populations when calculating risk prediction using a European cohort as the discovery dataset (Saffie-Awad et al., 2023). This further highlights the importance of large-scale diverse GWAS and potentially improving PRS model prediction at a variant level by incorporating local ancestry proportions, and at a population level by incorporating global ancestry proportions (Márquez-Luna and Loh, 2017). Globally, to an extent, most individuals are multiway admixed, therefore the development and implementation of more inclusive GWAS and PRS models will allow a more accurate risk prediction by leveraging variant effect size distributions across populations (Song et al., 2020).

Need for integrated datasets for PRS

PRS models calculate an individual's risk of disease development using the cumulative estimated effect of all nominally associated genetic variants and not only genome-wide significant variants (Bellou et al., 2020). Ultimately, PRS uses the weighted sum of a large number of SNPs with small effect sizes. Risk allele patterns were previously examined using individual-level data where differences in allele frequency, directionality, and the magnitude of effect were observed across several ancestral populations (Saffie-Awad et al., 2023). This further highlights the allele frequency differences observed between populations as well as population-specific disease risk, emphasizing the need for diverse GWAS summary statistics to be employed in PRS models to improve risk prediction accuracy.

Due to the complexity of these diseases, there is a level of environmental risk factors that contribute independently to disease risk (Lewis and Vassos, 2020). Therefore, *integrated prediction models* using PRS with additional environmental exposures and lifestyle information enables researchers to more accurately quantify risk (Martin et al., 2019). However, cohort size, lack of population diversity in PRS research, and inconsistent clinical characteristics severely limit PRS estimates (Saffie-Awad et al., 2023). Hence, incorporation of environmental information (e.g., smoking status and pesticide exposure) as well as clinical information (e.g., age of onset and family history) can improve the risk prediction model's stability, accuracy, and sensitivity (Truong et al., 2023). This information needs to be consistently recorded and standardized between study cohorts and would need to be collected during study participant recruitment.

Standardization of the data

The medical and biological fields have advanced greatly over the past two decades resulting in a data-rich area of research (Holmes et al., 2010). Many large publicly available repositories are used in research by compiling several sources of data including demographic data, clinical data, environmental and lifestyle information, molecular genetic data, and genealogical information (Lowrance, 2001). However, many of these repositories often have extensive clinical information without genetic data or alternatively have sufficient molecular genetic data with a limited clinical characterization of phenotypes. While there are disease-specific databases like the Accelerating Medicines Partnership- Alzheimer's Disease (AMP-AD) and the Parkinson's Disease Biomarkers Program (PDBP; Hodes and Buckholtz, 2016; Ofori et al., 2016), the majority of individuals included in these databases are from European populations, further highlighting the lack of URP representation. Ultimately, there is an increased need for *standardized databases that are representative of the global population* with sufficient quality information for each data category that can be used in PRS models for increased sensitivity and specificity to ultimately improve risk prediction accuracy.

Recommendations

In summary, a list of the pertinent recommendations for future PRS studies is provided below:

- Inclusion of ethnically diverse individuals in research studies.
- Need for continuous advancements in statistical methods, software tools, and genetic reference panels to appropriately include URPs samples in PRS research.
- Leverage genetic data from biobanks and disease-specific consortia.
- Leverage genetic diversity to detect novel disease associations and fine-mapping of known disease loci.
- Use of integrated prediction models using PRS combined with environmental exposures, lifestyle factors, and traditional risk factors e.g., smoking status.
- Standardization of genomic databases that are representative of the global population.

Concluding remarks

The value of including URPs in genetic research is highlighted by building capacity for genomic research on a global scale and ensuring the equitable implementation of precision medicine tools. While it has been perceived that incorporating genetically diverse populations (mainly URPs) in genetic research is challenging, there are expanding analytical approaches and algorithms to accurately account for population substructure in GWAS (Peterson et al., 2019) aiding in the detection of true genetic associations that can be utilized in risk prediction. Ultimately, the inclusion of URPs has the potential to fast-track the field of risk prediction, to aid in novel susceptibility variant detection, disease prediction, and the development of precision medicine strategies that are both applicable and accessible to all populations.

Author contributions

KS: Conceptualization, Writing – original draft, Writing – review & editing. CN: Writing – original draft, Writing – review & editing. IM: Conceptualization, Writing – review & editing. SB: Conceptualization, Writing – review & editing.

Funding

Acknowledgments

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fnins.2024.1380860/full#supplementary-material

References

Atkinson, E. G., Maihofer, A. X., Kanai, M., Martin, A. R., Karczewski, K. J., Santoro, M. L., et al. (2021). Tractor uses local ancestry to enable the inclusion of admixed individuals in GWAS and to boost power. *Nat. Genet.* 53, 195–204. doi: 10.1038/s41588-020-00766-y

Babulal, G. M., Quiroz, Y. T., Albensi, B. C., Arenaza-Urquijo, E., Astell, A. J., Babiloni, C., et al. (2019). Perspectives on ethnic and racial disparities in Alzheimer's disease and related dementias: update and areas of immediate need. *Alzheimer's Dement.* 15, 292–312. doi: 10.1016/j.jalz.2018.09.009

Bellenguez, C., Küçükali, F., Jansen, I. E., Kleineidam, L., Moreno-Grau, S., Amin, N., et al. (2022). New insights into the genetic etiology of Alzheimer's disease and related dementias. *Nat. Genet.* 54, 412–436. doi: 10.1038/s41588-022-01024-z

Bellou, E., Stevenson-Hoare, J., and Escott-Price, V. (2020). Polygenic risk and pleiotropy in neurodegenerative diseases. *Neurobiol. Dis.* 142:104953. doi: 10.1016/j.nbd.2020.104953

Bhattacharya, R., Chen, N., Shim, I., Kuwahara, H., Gao, X., Alkuraya, F. S., et al. (2023). Massive underrepresentation of Arabs in genomic studies of common disease. *Genome Med.* 15:99. doi: 10.1186/s13073-023-01254-8

Cavazos, T. B., and Witte, J. S. (2021). Inclusion of variants discovered from diverse populations improves polygenic risk score transferability. *HGG Adv.* 2:100017. doi: 10.1016/j.xhgg.2020.100017

Chen, Z., Chen, J., Collins, R., Guo, Y., Peto, R., Wu, F., et al. (2011). China Kadoorie Biobank of 0.5 million people: survey methods, baseline characteristics and long-term follow-up. *Int. J. Epidemiol.* 40, 1652–1666. doi: 10.1093/ije/dyr120

Chin-Chan, M., Navarro-Yepes, J., and Quintanilla-Vega, B. (2015). Environmental pollutants as risk factors for neurodegenerative disorders: Alzheimer and Parkinson diseases. *Front. Cell Neurosci.* 9:124. doi: 10.3389/fncel.2015.00124

Corpas, M., Megy, K., Metastasio, A., and Lehmann, E. (2022). Implementation of individualised polygenic risk score analysis: a test case of a family of four. *BMC Med. Genom.* 15:207. doi: 10.1186/s12920-022-01331-8

Duan, Q., Xu, Z., Raffield, L. M., Chang, S., Wu, D., Lange, E. M., et al. (2018). A robust and powerful two-step testing procedure for local ancestry adjusted allelic association analysis in admixed populations. *Genet. Epidemiol.* 42, 288–302. doi: 10.1002/gepi.22104

Epping-Jordan, J. E., Galea, G., Tukuitonga, C., and Beaglehole, R. (2005). Preventing chronic diseases: taking stepwise action. *Lancet* 366, 1667–1671. doi: 10.1016/S0140-6736(05)67342-4

Farrer, L. A., Cupples, L. A., Haines, J. L., Hyman, B., Kukull, W. A., Mayeux, R., et al. (1997). Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. APOE and Alzheimer Disease Meta Analysis Consortium. *J. Am. Med. Assoc.* 278, 1349–1356. doi: 10.1001/jama.1997.03550160069041

Farrer, M. J., Bardien, S., Hattori, N., Lesage, S., Ross, O. A., Mellick, G. D., et al. (2021). Editorial: celebrating the diversity of genetic research to dissect the pathogenesis of Parkinson's disease. *Front. Neurol.* 12:648417. doi: 10.3389/fneur.2021.648417

Fatumo, S., Chikowore, T., Choudhury, A., Ayub, M., Martin, A. R., Kuchenbaecker, K. A., et al. (2022). roadmap to increase diversity in genomic studies. *Nat. Med.* 28, 243–250. doi: 10.1038/s41591-021-01672-4

Fatumo, S., Sathan, D., Samtal, C., Isewon, I., Tamuhla, T., Soremekun, C., et al. (2023). Polygenic risk scores for disease risk prediction in Africa: current challenges and future directions. *Genome Med.* 15:87. doi: 10.1186/s13073-023-01245-9

Feigin, V. L., Vos, T., Nichols, E., Owolabi, M. O., Carroll, W. M., Dichgans, M., et al. (2020). The global burden of neurological disorders: translating evidence into policy. *Lancet Neurol.* 19, 255–265. doi: 10.1016/S1474-4422(19)30411-9

Ferri, C. P., Prince, M., Brayne, C., Brodaty, H., Fratiglioni, L., Ganguli, M., et al. (2005). Global prevalence of dementia: a Delphi consensus study. *Lancet* 366, 2112–2117. doi: 10.1016/S0140-6736(05)67889-0

Franzen, S., Smith, J. E., van den Berg, E., Rivera Mindt, M., van Bruchem-Visser, R. L., Abner, E. L., et al. (2022). Diversity in Alzheimer's disease drug trials: the importance of eligibility criteria. *Alzheimer's Dement.* 18, 810–823. doi: 10.1002/alz.12433

GBD 2016 Parkinson's Disease Collaborators (2018). Global, regional, and national burden of Parkinson's disease, 1990-2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Neurol.* 17, 939–953. doi: 10.1016/S1474-4422(18)30295-3

GBD 2021 Nervous System Disorders Collaborators (2024). Global, regional, and national burden of disorders affecting the nervous system, 1990-2021: a systematic analysis for the Global Burden of Disease Study 2021. *Lancet Neurol.* 23, 344–381. doi: 10.1016/S1474-4422(24)00038-3

Gilmore-Bykovskyi, A. L., Jin, Y., Gleason, C., Flowers-Benton, S., Block, L. M., Dilworth-Anderson, P., et al. (2019). Recruitment and retention of underrepresented populations in Alzheimer's disease research: a systematic review. *Alzheimer's Dement.* 5, 751–770. doi: 10.1016/j.trci.2019.09.018

Hamid, E., Ayele, B. A., Massi, D. G., Ben Sassi, S., Tibar, H., Djonga, E. E., et al. (2021). Availability of therapies and services for Parkinson's disease in Africa: a continent-wide survey. *Mov. Disord.* 36, 2393–2407. doi: 10.1002/mds.28669

Hendrie, H. C., Murrell, J., Baiyewu, O., Lane, K. A., Purnell, C., Ogunniyi, A., et al. (2014). APOE ε4 and the risk for Alzheimer disease and cognitive decline in African Americans and Yoruba. *Int. Psychogeriatr.* 26, 977–985. doi: 10.1017/S1041610214000167

Hodes, R. J., and Buckholtz, N. (2016). Accelerating Medicines Partnership: Alzheimer's Disease (AMP-AD) knowledge portal aids Alzheimer's drug discovery through open data sharing. *Expert Opin. Ther. Targets* 20, 389–391. doi: 10.1517/14728222.2016.1135132

Hoggart, C., Choi, S. W., García-González, J., Souaiaia, T., Preuss, M., O'Reilly, P., et al. (2023). BridgePRS: a powerful trans-ancestry Polygenic Risk Score method. *bioRxiv*. doi: 10.1101/2023.02.17.528938

Holmes, C., McDonald, F., Jones, M., Ozdemir, V., and Graham, J. E. (2010). Standardization and omics science: technical and social dimensions are inseparable and demand symmetrical study. *Omics* 14, 327–332. doi: 10.1089/omi.2010.0022

Hoover, B. R., Reed, M. N., Su, J., Penrod, R. D., Kotilinek, L. A., Grant, M. K., et al. (2010). Tau mislocalization to dendritic spines mediates synaptic dysfunction independently of neurodegeneration. *Neuron* 68, 1067–1081. doi: 10.1016/j.neuron.2010.11.030

Ibanez, A., Yokoyama, J. S., Possin, K. L., Matallana, D., Lopera, F., Nitrini, R., et al. (2021). The multi-partner consortium to expand dementia research in Latin America (ReDLat): driving multicentric research and implementation science. *Front. Neurol.* 12:631722. doi: 10.3389/fneur.2021.631722

Kachuri, L., Chatterjee, N., Hirbo, J., Schaid, D. J., Martin, I., Kullo, I. J., et al. (2024). Principles and methods for transferring polygenic risk scores across global populations. *Nat. Rev. Genet.* 25, 8–25. doi: 10.1038/s41576-023-00637-2

Kim, J. J., Vitale, D., Veliz-Otani, D., Lian, M., and Heilbron, K. (2022). Multi-ancestry genome-wide meta-analysis in Parkinson's disease. *medRxiv*. doi: 10.1101/2022.08.04.22278432

Knerr, S., Wayman, D., and Bonham, V. L. (2011). Inclusion of racial and ethnic minorities in genetic research: advance the spirit by changing the rules? *J. Law Med. Ethics* 39, 502–512. doi: 10.1111/j.1748-720X.2011.00617.x

Kumuthini, J., Zick, B., Balasopoulou, A., Chalikiopoulou, C., Dandara, C., El-Kamah, G., et al. (2022). The clinical utility of polygenic risk scores in genomic medicine practices: a systematic review. *Hum. Genet.* 141, 1697–1704. doi: 10.1007/s00439-022-02452-x

Lake, J., Warly Solsberg, C., Kim, J. J., Acosta-Uribe, J., Makarious, M. B., Li, Z., et al. (2023). Multi-ancestry meta-analysis and fine-mapping in Alzheimer's disease. *Mol. Psychiatr.* 28, 3121–3132. doi: 10.1038/s41380-023-02089-w

Lamptey, R. N. L., Chaulagain, B., Trivedi, R., Gothwal, A., Layek, B., Singh, J., et al. (2022). A review of the common neurodegenerative disorders: current therapeutic approaches and the potential role of nanotherapeutics. *Int. J. Mol. Sci.* 23:31851. doi: 10.3390/ijms23031851

Leonenko, G., Baker, E., Stevenson-Hoare, J., Sierksma, A., Fiers, M., Williams, J., et al. (2021). Identifying individuals with high risk of Alzheimer's disease using polygenic risk scores. *Nat. Commun.* 12:4506. doi: 10.1038/s41467-021-24082-z

Lewis, C. M., and Vassos, E. (2020). Polygenic risk scores: from research tools to clinical instruments. *Genome Med.* 12:44. doi: 10.1186/s13073-020-00742-5

Loesch, D. P., Horimoto, A. R. V. R., Sarihan, E. I., Inca-Martinez, M., Mason, E., Cornejo-Olivas, M., et al. (2022). Polygenic risk prediction and SNCA haplotype analysis in a Latino Parkinson's disease cohort. *Parkinson. Relat. Disord.* 102, 7–15. doi: 10.1016/j.parkreldis.2022.06.010

Lowrance, W. W. (2001). The promise of human genetic databases. *Br. Med. J.* 322, 1009–1010. doi: 10.1136/bmj.322.7293.1009

Márquez-Luna, C., and Loh, P. R. (2017). Multiethnic polygenic risk scores improve risk prediction in diverse populations. *Genet. Epidemiol.* 41, 811–823. doi: 10.1002/gepi.22083

Martin, A. R., Kanai, M., Kamatani, Y., Okada, Y., Neale, B. M., Daly, M. J., et al. (2019). Clinical use of current polygenic risk scores may exacerbate health disparities. *Nat. Genet.* 51, 584–591. doi: 10.1038/s41588-019-0379-x

Matthews, K. A., Xu, W., Gaglioti, A. H., Holt, J. B., Croft, J. B., Mack, D., et al. (2018). Racial and ethnic estimates of Alzheimer's disease and related dementias in the United States (2015–2060) in adults aged ≥65 years. *Alzheimer's Dement.* 15, 17–24. doi: 10.1016/j.jalz.2018.06.3063

Mills, M. C., and Rahal, C. (2020). The GWAS Diversity Monitor tracks diversity by disease in real time. *Nat. Genet.* 52, 242–243. doi: 10.1038/s41588-020-0580-y

Milnerwood, A. J., and Raymond, L. A. (2010). Early synaptic pathophysiology in neurodegeneration: insights from Huntington's disease. *Trends Neurosci.* 33, 513–523. doi: 10.1016/j.tins.2010.08.002

Mok, K. Y. (2021). The East Asian Parkinson disease genomics consortium. *Lancet Neurol.* 20:982. doi: 10.1016/S1474-4422(21)00373-2

Nagai, A., Hirata, M., Kamatani, Y., Muto, K., Matsuda, K., Kiyohara, Y., et al. (2017). Overview of the BioBank Japan Project: study design and profile. *J. Epidemiol.* 27, 2–8. doi: 10.1016/j.je.2016.12.005

Nalls, M. A., Blauwendraat, C., Vallerga, C. L., Heilbron, K., Bandres-Ciga, S., Chang, D., et al. (2019). Identification of novel risk loci, causal insights, and heritable risk for Parkinson's disease: a meta-analysis of genome-wide association studies. *Lancet Neurol.* 18, 1091–1102. doi: 10.1016/S1474-4422(19)30320-5

Ofori, E., Du, G., Babcock, D., Huang, X., and Vaillancourt, D. E. (2016). Parkinson's disease biomarkers program brain imaging repository. *Neuroimage* 124, 1120–1124. doi: 10.1016/j.neuroimage.2015.05.005

Perrone, F., Cacace, R., van der Zee, J., and Van Broeckhoven, C. (2021). Emerging genetic complexity and rare genetic variants in neurodegenerative brain diseases. *Genome Med.* 13:59. doi: 10.1186/s13073-021-00878-y

Peterson, R. E., Kuchenbaecker, K., Walters, R. K., Chen, C. Y., Popejoy, A. B., Periyasamy, S., et al. (2019). Genome-wide association studies in ancestrally diverse populations: opportunities, methods, pitfalls, and recommendations. *Cell* 179, 589–603. doi: 10.1016/j.cell.2019.08.051

Precise Health Research Singapore (2023). *SG10K_Health Reference Genome Databse*. Available online at: https://www.npm.sg/collaborate/partners/sg10k/ (accessed November 3, 2023).

Prince, M., Bryce, R., Albanese, E., Wimo, A., Ribeiro, W., Ferri, C. P., et al. (2013). The global prevalence of dementia: a systematic review and metaanalysis. *Alzheimer's Dement.* 9, 63–75.e2. doi: 10.1016/j.jalz.2012.11.007

Rajabli, F., Beecham, G. W., Hendrie, H. C., Baiyewu, O., Ogunniyi, A., Gao, S., et al. (2022). A locus at 19q13.31 significantly reduces the ApoE ε4 risk for Alzheimer's Disease in African Ancestry. *PLoS Genet.* 18:e1009977. doi: 10.1371/journal.pgen.1009977

Rizig, M., Bandres-Ciga, S., Makarious, M. B., Ojo, O., Crea, P. W., Abiodun, O., et al. (2023). Genome-wide association identifies novel etiological insights associated with Parkinson's disease in African and African Admixed Populations. *medRxiv*. doi: 10.1101/2023.05.05.23289529

Rizig, M., Okubadejo, N., Salama, M., Thomas, O., Akpalu, A., Gouider, R., et al. (2021). The international Parkinson disease genomics consortium Africa. *Lancet Neurol.* 20:335. doi: 10.1016/S1474-4422(21)00100-9

Ruan, Y., Lin, Y. F., Feng, Y. C. A., Chen, C. Y., Lam, M., Guo, Z., et al. (2022). Improving polygenic prediction in ancestrally diverse populations. *Nat. Genet.* 54, 573–580. doi: 10.1038/s41588-022-01054-7

Saffie-Awad, P., Elsayed, I., Sanyaolu, A. O., Crea, P. W., Schumacher Schuh, A. F., Levine, K. S., et al. (2023). Evaluating the performance of polygenic risk profiling across diverse ancestry populations in Parkinson's disease. *medRxiv*. doi: 10.1101/2023.11.28.23299090v1

Sariya, S., Lee, J. H., Mayeux, R., Vardarajan, B. N., Reyes-Dumeyer, D., Manly, J. J., et al. (2019). Rare variants imputation in admixed populations: comparison across reference panels and bioinformatics tools. *Front. Genet.* 10:239. doi: 10.3389/fgene.2019.00239

Schumacher-Schuh, A. F., Bieger, A., Okunoye, O., Mok, K. Y., Lim, S. Y., Bardien, S., et al. (2022). Underrepresented populations in parkinson's genetics research: current landscape and future directions. *Mov. Disord.* 37, 1593–1604. doi: 10.1002/mds.29126

Seifan, A., Ganzer, C. A., Ryon, K., Lin, M., Mahmudur, R., Adolfo, H., et al. (2019). Detecting non-cognitive features of prodromal neurodegenerative diseases. *Curr. Aging Sci.* 11, 242–249. doi: 10.2174/1874609818666190204094233

Shipilina, D., Pal, A., Stankowski, S., Chan, Y. F., and Barton, N. H. (2023). On the origin and structure of haplotype blocks. *Mol. Ecol.* 32, 1441–1457. doi: 10.1111/mec.16793

Sirugo, G., Williams, S. M., and Tishkoff, S. A. (2019). The missing diversity in human genetic studies. *Cell* 177, 26–31. doi: 10.1016/j.cell.2019.02.048

So, H. C., Kwan, J. S. H., Cherny, S. S., and Sham, P. C. (2011). Risk prediction of complex diseases from family history and known susceptibility loci, with applications for cancer screening. *Am. J. Hum. Genet.* 88, 548–565. doi: 10.1016/j.ajhg.2011.04.001

Sollis, E., Mosaku, A., Abid, A., Buniello, A., Cerezo, M., Gil, L., et al. (2023). The NHGRI-EBI GWAS Catalog: knowledgebase and deposition resource. *Nucl. Acids Res.* 51, 977–985. doi: 10.1093/nar/gkac1010

Song, S., Jiang, W., Hou, L., and Zhao, H. (2020). Leveraging effect size distributions to improve polygenic risk scores derived from summary statistics of genome-wide association studies. *PLoS Comput. Biol.* 16:e1007565. doi: 10.1371/journal.pcbi.1007565

Swart, Y., Uren, C., van Helden, P. D., Hoal, E. G., and Möller, M. (2021). Local ancestry adjusted allelic association analysis robustly captures tuberculosis susceptibility loci. *Front. Genet.* 12:716558. doi: 10.3389/fgene.2021.716558

Swart, Y., van Eeden, G., Uren, C., van der Spuy, G., Tromp, G., Möller, M., et al. (2022). GWAS in the southern African context. *PLoS ONE* 17:e0264657. doi: 10.1371/journal.pone.0264657

Tan, M. S., Jiang, T., Tan, L., and Yu, J. T. (2014). Genome-wide association studies in neurology. *Ann. Transl. Med.* 2:124. doi: 10.3978/j.issn.2305-5839.2014.11.12

Tan, T., and Atkinson, E. G. (2023). Strategies for the genomic analysis of admixed populations. *Annu. Rev. Biomed. Data Sci.* 6, 105–127. doi: 10.1146/annurev-biodatasci-020722-014310

Tang, M. X., Cross, P., Andrews, H., Jacobs, D. M., Small, S., Bell, K., et al. (2001). Incidence of AD in African-Americans, Caribbean Hispanics, and Caucasians in northern Manhattan. *Neurology* 56, 49–56. doi: 10.1212/WNL.56.1.49

Towns, C., Richer, M., Jasaityte, S., Stafford, E. J., Joubert, J., Antar, T., et al. (2023). Defining the causes of sporadic Parkinson's disease in the global Parkinson's genetics program (GP2). *NPJ Parkinson's Dis.* 9:131. doi: 10.1038/s41531-023-00533-w

Truong, B., Hull, L. E., Ruan, Y., Huang, Q. Q., Hornsby, W., Martin, H., et al. (2023). Integrative polygenic risk score improves the prediction accuracy of complex traits and diseases. *medRxiv*. doi: 10.1101/2023.02.21.23286110

Uffelmann, E., Huang, Q. Q., Munung, N. S., de Vries, J., Okada, Y., Martin, A. R., et al. (2021). Genome-wide association studies. *Nat. Rev. Methods Prim.* 1:59. doi: 10.1038/s43586-021-00056-9

Vaswani, P. A., Tropea, T. F., and Dahodwala, N. (2020). Overcoming barriers to parkinson disease trial participation: increasing diversity and novel designs for recruitment and retention. *Neurotherapeutics* 17, 1724–1735. doi: 10.1007/s13311-020-00960-0

Wonkam, A. (2021). Sequence three million genomes across Africa. *Nature* 590, 209–211. doi: 10.1038/d41586-021-00313-7

Yang, Z., Cieza, B., Reyes-Dumeyer, D., Montesinos, R., Soto-Añari, M., Custodio, N., et al. (2023). A benchmark study on current GWAS models in admixed populations. *Brief Bioinform*. 25:bbad437. doi: 10.1093/bib/bbad437

Zabetian, C. P., and Mata, I. F. (2017). LARGE-PD: examining the genetics of Parkinson's disease in Latin America. *Mov. Disord*. 32, 1330–1331. doi: 10.1002/mds.27081