

BRIEF COMMUNICATION OPEN



Defining the causes of sporadic Parkinson's disease in the global Parkinson's genetics program (GP2)

Clodagh Towns¹, Madeleine Richer¹, Simona Jasaityte¹, Eleanor J. Stafford^{1,2}, Julie Joubert¹, Tarek Antar³, Alejandro Martinez-Carrasco^{1,2}, Mary B. Makarios^{1,3,4}, Bradford Casey^{5,6}, Dan Vitale^{4,7,8}, Kristin Levine^{7,8}, Hampton Leonard^{3,7,8,9}, Caroline B. Pantazis^{4,7}, Laurel A. Screven¹⁰, Dena G. Hernandez³, Claire E. Wegel¹⁰, Justin Solle⁵, Mike A. Nalls^{3,4,7,8}, Cornelis Blauwendraat^{4,7,11}, Andrew B. Singleton^{3,7,12}, Manuela M. X. Tan¹³, Hirotaka Iwaki^{3,7,8}, Huw R. Morris^{1,2}✉ and the Global Parkinson's Genetics Program (GP2)*

The Global Parkinson's Genetics Program (GP2) will genotype over 150,000 participants from around the world, and integrate genetic and clinical data for use in large-scale analyses to dramatically expand our understanding of the genetic architecture of PD. This report details the workflow for cohort integration into the complex arm of GP2, and together with our outline of the monogenic hub in a companion paper, provides a generalizable blueprint for establishing large scale collaborative research consortia.

npj Parkinson's Disease (2023)9:131; <https://doi.org/10.1038/s41531-023-00533-w>

INTRODUCTION

Parkinson's disease (PD) is a multifactorial disorder with complex etiology. The largest genome-wide association study (GWAS) to-date included 37,688 cases, 18,618 proxy cases (unaffected first-degree relatives), and 1.4 million controls from European ancestry, and identified 90 independent risk signals across 78 genomic regions; 38 of which were novel signals¹. Despite these advances, PD GWAS are currently limited by scale, a focus on European populations, and limited integration with clinical phenotype data.

A power calculation based on the 2019 GWAS data indicates that inclusion of an additional ~99,000 cases would enable variants of smaller effect size that contribute to polygenic risk (p -value cut off: 1.35×10^{-3}) to reach genome-wide significance. Therefore, expanding PD GWAS to at least this size will result in identification of additional risk loci and improve genetic prediction of PD occurrence¹. The heritability of PD can be estimated using twin studies or statistical genetic methods and is thought to lie between 22% and 40% in European populations. Known genome wide significant loci currently explain ~16% of the heritability of PD¹. The use of polygenic risk score analysis (including loci that do not reach genome wide significance) indicates that there are likely to be a substantial number of loci contributing to PD risk that have not yet been defined. Our power analysis indicates that 99,000 PD cases will be needed to define loci with 80% power, a minor allele frequency of 0.21 and similar effects to the current state-of-the-art analysis. The variability of phenotypes observed in PD are likely to have a genetic basis²⁻⁴. Knowledge of associations between genotype and clinical outcomes will enable clinicians to provide patients with a more accurate prognosis. Understanding the gene-to-phenotype pathways responsible for specific PD features would provide an opportunity to develop treatments targeting phenotype-specific

disease pathways, resulting in more efficient and personalized treatment with fewer side effects. We aim to capture the diversity of PD outcomes, including Parkinson's itself but also related conditions such as prodromal Parkinson's, dementia with Lewy bodies and other Parkinson's plus syndromes, and to perform large-scale analyses of clinical-genetic data with sufficient power for gene discovery. This will comprise regression and time-to-event analysis for the phenotype of interest (e.g., dementia, dyskinesias, motor progression). It is likely that this will be limited to around 25% of samples included in this study with in depth longitudinal data, and further large scale longitudinal cohorts will be needed to explore the biology of progression and diverse phenotypes. The focus of PD GWAS on individuals of European ancestry has left gaps in our knowledge of PD-associated genetic variants in underrepresented populations and limited our ability to resolve GWAS loci⁵. To advance understanding of the genetic determinants of Parkinsonism on a global scale, we need to ensure representation of diverse ancestries with sample sizes sufficient to detect ancestry-specific signals. We aim to include at least 15,000 participants of African, South Asian, and East Asian ancestry, respectively.

The Global Parkinson's Genetics Program (GP2, <http://gp2.org/>), funded by the Aligning Science Across Parkinson's initiative (ASAP, <https://parkinsonsroadmap.org/>) will recruit PD, Prodromal Parkinson's and Parkinson's Plus (including Dementia with Lewy bodies, Progressive Supranuclear Palsy, Multiple System Atrophy and Cortico-basal syndrome) cohorts from across the world⁶. We will genotype >150,000 participants and integrate genetic data with harmonized clinical data for use in case-control and genotype-phenotype association studies. This figure will include a minimum of 50,000 individuals from ancestries currently underrepresented in Parkinson's research: including Black

¹Department of Clinical and Movement Neurosciences, Queen Square Institute of Neurology, University College London, London, UK. ²University College London, London, UK. ³Molecular Genetics Section, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. ⁴National Institutes of Health, Bethesda, MD, USA. ⁵Department of Clinical Research, Michael J. Fox Foundation for Parkinson's Research, New York City, NY, USA. ⁶The Michael J. Fox Foundation for Parkinson's Research, New York, NY, USA. ⁷Center for Alzheimer's and Related Dementias (CARD), National Institute on Aging and National Institute of Neurological Disorders and Stroke, National Institutes of Health, Bethesda, MD, USA. ⁸Data Tecnica International, Washington, DC, USA. ⁹National Institute on Aging/National Institutes of Health, Bethesda, MD, USA. ¹⁰Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA. ¹¹Integrative Genomics Unit, Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA. ¹²National Institute on Aging, Bethesda, MD, USA. ¹³Department of Neurology, Oslo University Hospital, Oslo, Norway. ✉email: h.morris@ucl.ac.uk

American, African, Middle Eastern, Central/ East/South Asian, Indian, Caribbean and Central/South American. The remaining participants are expected to be of European ancestry⁶. Cleaned genetic and clinical data will be harmonized across cohorts and made available to Parkinson's, Prodromal and Parkinson's Plus researchers via a controlled-access online repository with a unified user agreement. Contributing investigators are encouraged to play an active role in the project by proposing and leading analyses, and will receive authorship on GP2 publications as well as support for their analyses. GP2 also provides comprehensive training opportunities for researchers from contributing institutions via an online learning management system. The outcome of GP2 will be an open access resource which will integrate clinical and genetic data from a very large, diverse sample of cases, and facilitate discovery of new genetic determinants of Parkinson's, Prodromal and Parkinson's Plus occurrence and phenotypic variation in multiple ancestries. A global network of PD researchers will be established, facilitating future collaboration and advancement of PD research. The project began in January 2020 and current funding extends to 2029.

Coordinating cohorts around the world to include 150,000 participants is a considerable logistical challenge. As of May 2022, we have established a workflow for cohort integration, a standardized set of clinical data elements, a recommended consent template, a protocol for evaluating cohorts joining the study, and a process for integrating and harmonizing the incoming clinical data. More broadly, we have established the complex hub framework, which created a code of conduct and publication policy, created a DNA quality and shipping protocol, and created the NeuroBooster Array (NBA) genotyping workflow.

The size and depth of the dataset which is being generated by GP2 will provide a major opportunity to discover new genotype-phenotype associations. Core analyses, analyses that will be continuously updated by the analysis teams at GP2 with the inclusion of new samples, will include case-control, age at onset, and progression GWAS, as well as within-ancestry analyses of previously underrepresented populations. Beyond this, GP2 is supporting additional analytical projects proposed by contributing collaborators. The GP2 network will also facilitate collaboration on auxiliary studies between investigators sharing the same area of interest (e.g., biomarkers of different modalities) to address outstanding questions in PD research.

GP2 is conducted according to overarching principles of democratization of data, collaboration and cooperation, safe and responsible data sharing, commitment to diversity in research, transparency and reproducibility, and production of an actionable resource in accordance with the Findable, Accessible, Interoperable and Reusable (FAIR) scientific data management principles⁷. It is GP2 policy that local researchers are included in publications that use their data/samples. Here, we report the specific steps undertaken by the Cohort Integration Working Group (CIWG) to identify, recruit, and harmonize clinical cohorts within the complex disease arm of GP2. GP2 also has a monogenic arm (<https://gp2.org/working-groups/>) which focuses on cases with potential monogenic causes of PD, i.e., those with early age at onset or a family history of PD (Reference: PMID:37369645).

METHODS

PD and Parkinson's Plus investigators and cohorts are identified through relevant publications, or through existing consortia. We welcome contact from interested investigators around the world (Email: cohort@gp2.org). Prospective collaborators complete a Site Interest Form (SIF) (Supplementary Material) to provide a brief overview of their cohort [Supplementary material]. Data availability, type of cohort (e.g., brain bank, drug study etc.), ethnic diversity and sample/data sharing restrictions are all considered by the CIWG when reviewing SIF submissions. Cohorts' consent

documents are reviewed by the Operations and Compliance Working Group (OCWG) to ensure consent for sample/data sharing was obtained, and compliance with regional and cohort-specific data sharing restrictions. The OCWG assists the investigator if revisions are needed, and template consent language is provided online (<https://gp2.org/resources/consent-guidelines/> [Supplementary Material]). We have an inclusive approach; to-date >95% of cohorts evaluated have been accepted. Following cohort approval, a concise Collaboration Agreement is signed by the collaborating PI and the Michael J. Fox Foundation (MJFF), and the necessary material and data transfer agreements are completed. The cohort's samples are then transferred to one of GP2's genotyping centers, and data are transferred through secure upload to a cloud server.

Once the cohort's clinical data have been transferred, harmonization and quality control (QC) are performed by the CIWG, following a standard framework⁸. Core clinical data were decided on the basis of clinical scales available in existing PD cohorts (primarily the Tracking Parkinson's⁹, PPMI¹⁰, PDBP¹¹, NET PD-LS1¹², and GEPD¹³ cohorts) as well as on the basis of recent proposals for a modular set of assessment criteria to characterize longitudinal PD cohorts^{14,15}. The core clinical data elements are the clinical data categories of interest which will be harmonized across cohorts and used for analysis (Table 1). We have generated a common dataset for brain banks, and have added data elements for time-to-event analysis (Supplementary Table 1). We are also collecting information on availability of other data and samples, including, plasma, serum, RNA, fibroblasts, CSF, skin biopsy, PBL, iPSC to expand awareness of the resources that are available at contributing sites (Supplementary Material). The common dataset is defined in a common data dictionary (Supplementary Table 2). The harmonization of raw data uses a set of coding rules for re-coding and handling of missing data, using a modified custom script for each cohort.

DNA samples from each cohort are genotyped on the Neuro Booster Array (NBA; https://github.com/GP2code/Neuro_Booster_Array), developed in collaboration between Data Tecnica International, NIA, NINDS, Illumina Inc and GP2. The NBA consists of the Illumina Infinium Global Diversity Array (GDA; <https://www.illumina.com/products/by-type/microarray-kits/infinium-global-diversity.html>), a high-density (1.9 M total variants) global backbone optimized for cross-population imputation coverage of the human genome, and additional custom content (>95,000 variants) which includes known causal variants for various neurodegenerative diseases and imputation boosters for underrepresented populations. QC of genotype data by the Data Analysis Working Group (DAWG) follows a standard pipeline. A custom genotype clustering file is used to ensure representation of the diverse genetic ancestries within the GP2 data. The clustering file used for the latest data release is based on 2793 samples across 6 ancestry groups, and includes 420 Gaucher disease cases to capture variants of interest in the *GBA1* risk gene. This file is available for download via the GP2 Github repository (<https://github.com/GP2code>). Cleaned genetic data are returned to the contributing investigator to use as they wish, and are uploaded to the GP2 repository for use in combined analyses (Fig. 1). GP2 covers the costs of genotyping and sample shipment for all contributing cohorts, and can assist investigators with the analysis of their cohort's data.

The dataset which we are currently aggregating and harmonizing contains a wide variety of demographic and clinical factors, thanks to the diversity of contributing cohorts. As of March 2023, the CIWG has approved 145 cohorts for inclusion in GP2, of which 128 have been approved by the OCWG, and 74 have completed all necessary agreements and are in the process of transferring samples and clinical data. Samples that have been transferred are currently being genotyped and passed through the data QC pipeline. So far, the approved cohorts span over 50 different countries and territories. A map showing the geographic distribution of these cohorts, as well as current expected and

Table 1. Recommended core clinical data.

	MINIMUM	MINIMUM PLUS	CORE	EXTENDED
Demographics (e.g., NINDS CDE General Core) ¹⁶	✓	✓	✓	✓
Recruitment category (Case/Control)	✓	✓	✓	✓
Family History	✓	✓	✓	✓
Diagnostic checklist (MDS clinical diagnosis criteria ¹⁷ or UK Parkinson's Disease Society Brain Bank Diagnostic Criteria) ¹⁸		✓	✓	✓
Primary diagnosis and PD certainty (PPMI Primary Clinical Diagnosis) ¹⁰		✓	✓	✓
PD history (e.g., NINDS CDE for Parkinson's Disease) ¹⁶		✓	✓	✓
Global PD severity (CISI-PD) ¹⁹		✓	✓	✓
Behavioral and Environmental History (PD-RFQ-U) ¹⁶			✓	✓
Medical History			✓	✓
Current Medication Status			✓	✓
nM-EDL (MDS-UPDRS Part I [or UPDRS]) ^{20,21}			✓	✓
M-EDL (MDS-UPDRS Part II [or UPDRS]) ^{20,21}			✓	✓
Motor Assessment (MDS-UPDRS Part III [or UPDRS]) ^{20,21}			✓	✓
Complications (MDS-UPDRS Part IV [or UPDRS]) ^{20,21}			✓	✓
Cognitive Assessment (MMSE, MoCA) ^{22,23}			✓	✓
Motor (Hoehn and Yahr stage) ²⁴			✓	✓
Autonomic function assessment (SCOPA-AUT) ²⁵			✓	✓
pRBD (RBDSQ) ²⁶			✓	✓
Day time sleepiness (ESS) ²⁷				✓
Depression (GDS-15) ²⁸				✓
Orthostatic hypotension (Standing/sitting/supine HR and BP)				✓
Olfactory function (UPSIT or Sniffin-stick or BSIT) ^{29–31}				✓
General ADL (Schwab & England ADL) ³²				✓
PD EQL (PDQ-8) ³³				✓
Pain (Kings PD Pain Scale) ³⁴				✓

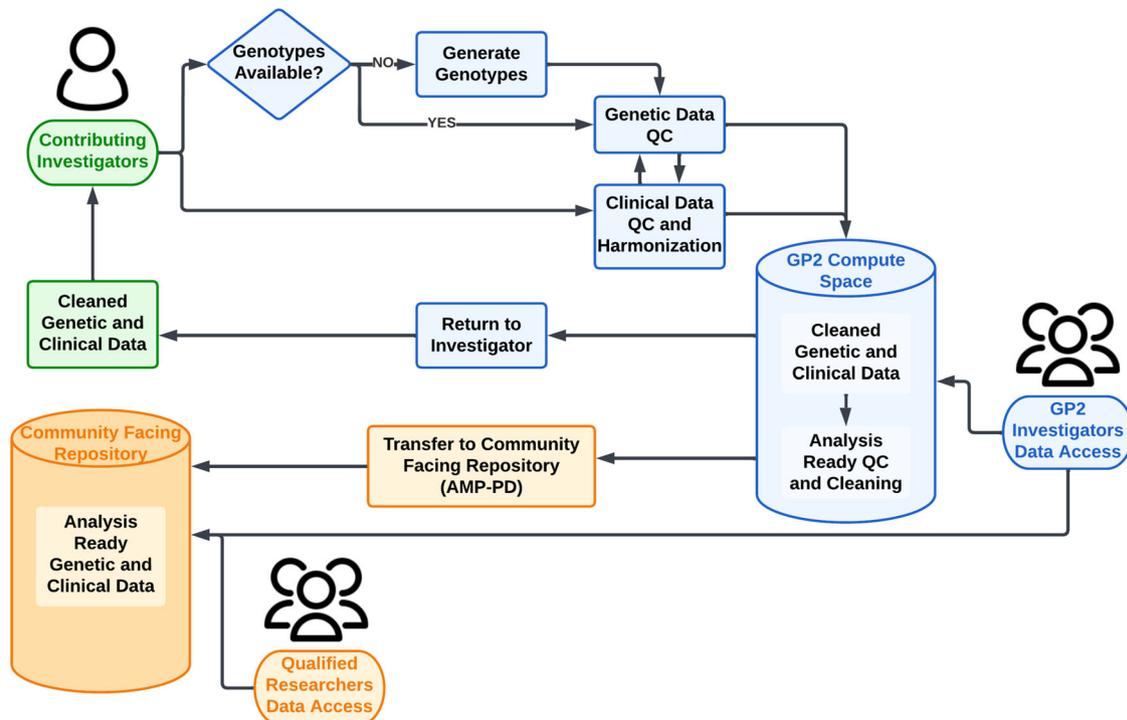


Fig. 1 Cohort integration workflow. Data are contributed and returned to the local PI (green); new genetic data, data cleaning and harmonization are carried out by GP2 and made available for analysis within the GP2 consortium (blue); and data are released to qualified investigators via AMP-PD (orange). AMP-PD Accelerating Medicines Partnership - Parkinson's Disease (<https://amp-pd.org>).

completed sample numbers, can be found on the GP2 website (<https://gp2.org/cohort-dashboard/> [Supplementary Material]).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this article.

DATA AVAILABILITY

GP2 has partnered with the Accelerating Medicines Partnership - Parkinson's Disease (AMP-PD; <https://amp-pd.org>) to share data generated by GP2, and in December 2021 the first GP2 genotyping data were released on the AMP-PD platform. As of 2023, the data consist of 14,902 samples (8190 PD cases), representing a broad range of diverse ancestries defined directly from the genotyping data from these cohorts. Genotyping and data QC is ongoing, and there will be regular data releases (2–4 times per year) as the project progresses. All contributing investigators have access to GP2 data, and external researchers can also gain access by following instructions on the GP2 website (<https://gp2.org/applying-for-gp2-data-access-on-the-amp-pd-platform/>). There are two tiers of data access. Tier 1 consists of summary statistics and any researcher can gain access by completing an online application. Tier 2 access includes deidentified, individual level genetic data, and to gain access researchers must sign a Data Usage Agreement co-signed by their institution.

CODE AVAILABILITY

Code used in GP2 core analyses and the custom clustering files are deposited in the GP2 GitHub repository (<https://github.com/GP2code>). Code created and used by the CIWG can be found on the "GP2-WorkingGroups/CD-Cohort-Integration" subpage of the GP2 GitHub repository (<https://github.com/GP2code/GP2-WorkingGroups/tree/main/CD-Cohort-Integration>). The Data and Code Dissemination Working Group (DCD-WG) at GP2 is committed to making these data, resources, and training materials accessible to the scientific community. The DCD-WG actively supports efforts to consolidate analysis scripts with the necessary analytical tools and decisions via GitHub (<https://github.com/GP2code>), consolidate training materials with several translations in collaboration with the Training and Networking Working Group (<https://gp2.org/training/>), and consolidate additional analytical pipelines on Terra (<https://amp-pd.org/tools>).

Received: 25 November 2022; Accepted: 22 May 2023;

Published online: 12 September 2023

REFERENCES

- Nalls, M. A. 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.* **18**, 1091–1102 (2019).
- Iwaki, H. et al. Genetic risk of Parkinson disease and progression: an analysis of 13 longitudinal cohorts. *Neurol. Genet.* **5**, 1–14 (2019).
- Latourelle, J. C. et al. Large-scale identification of clinical and genetic predictors of Parkinson's disease motor progression in newly-diagnosed patients: a longitudinal cohort study and validation. *Lancet Neurol.* **16**, 908–916 (2017).
- Tan, M. M. X. et al. Genome-wide association studies of cognitive and motor progression in Parkinson's disease. *Mov. Disord.* **36**, 424–433 (2021).
- Heinzel, S., Lerche, S., Maetzler, W. & Berg, D. Global, yet incomplete overview of cohort studies in Parkinson's disease. *J. Parkinsons. Dis.* **7**, 423–432 (2017).
- The Global Parkinson's Genetics Program (GP2). GP2: the global parkinson's genetics program. *Mov. Disord.* **36**, 842–851 (2021).
- Wilkinson, M. D. et al. The FAIR Guiding Principles for scientific data management and stewardship. *Sci. Data* **3**, 160018 (2016).
- Bennett, S. N. et al. Phenotype harmonization and cross-study collaboration in GWAS consortia: the GENEVA experience. *Genet. Epidemiol.* **35**, 159–173 (2011).
- Malek, N. et al. Tracking Parkinson's: study design and baseline patient data. *J. Parkinsons. Dis.* **5**, 947–959 (2015).
- Parkinson Progression Marker Initiative. The Parkinson Progression Marker Initiative (PPMI). *Prog. Neurobiol.* **95**, 629–635 (2011).
- Rosenthal, L. S. et al. The NINDS Parkinson's disease biomarkers program. *Mov. Disord.* **31**, 915–923 (2016).
- Kiebertz, K. et al. Effect of creatine monohydrate on clinical progression in patients with Parkinson disease. *JAMA* **313**, 584–593 (2015).
- Bardien, S., Mellick, G. D., Hattori, N., Ross, O. A. & Lesage, S. *Celebrating the Diversity of Genetic Research to Dissect the Pathogenesis of Parkinson's Disease* (Frontiers Media SA, Lausanne, 2021).
- Lerche, S. et al. Aiming for study comparability in Parkinson's disease: proposal for a modular set of biomarker assessments to be used in longitudinal studies. *Front. Aging Neurosci.* **8**, 1–4 (2016).
- Lerche, S. et al. Methods in neuroepidemiology characterization of European longitudinal cohort studies in Parkinson's disease-report of the JPND working group BioLoC-PD. *Neuroepidemiology* **45**, 282–297 (2015).
- Grinnon, S. T. et al. National institute of neurological disorders and stroke common data element project - approach and methods. *Clin. Trials* **9**, 322–329 (2012).
- Postuma, R. B. et al. MDS clinical diagnostic criteria for Parkinson's disease. *Mov. Disord.* **30**, 1591–1601 (2015).
- Gibb, W. R. G. & Lees, A. J. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. *J. Neurol. Neurosurg. Psychiatry* **51**, 745–752 (1988).
- Martinez-Martin, P. et al. Global versus factor-related impression of severity in Parkinson's disease: a new clinimetric index (CISI-PD). *Mov. Disord.* **21**, 208–214 (2006).
- Goetz, C. G. et al. Movement disorder society-sponsored revision of the unified Parkinson's disease rating scale (MDS-UPDRS): process, format, and clinimetric testing plan. *Mov. Disord.* **22**, 41–47 (2007).
- Fahn, S., Elton R. Members of the UPDRS Development Committee. in *Recent Developments in Parkinson's Disease*, Vol 2, (eds. Fahn S., Marsden C. D., Calne D. B., & Goldstein M.) 153–163, 293–304 (Macmillan Health Care Information, 1987).
- Folstein, M. F., Folstein, S. E. & McHugh, P. R. Mini-mental state'. A practical method for grading the cognitive state of patients for the clinician. *J. Psychiatr. Res.* **12**, 189–198 (1975).
- Nasreddine, Z. S. et al. The Montreal Cognitive Assessment, MoCA: A brief screening tool for mild cognitive impairment. *J. Am. Geriatr. Soc.* **53**, 695–699 (2005).
- Hoehn, M. M. & Yahr, M. D. Parkinsonism: onset, progression and mortality. *Neurology* **17**, 427–442 (1967).
- Visser, M., Marinus, J., Stiggelbout, A. M. & van Hilten, J. J. Assessment of autonomic dysfunction in Parkinson's disease: the SCOPA-AUT. *Mov. Disord.* **19**, 1306–1312 (2004).
- Stiasny-Kolster, K. et al. The REM sleep behavior disorder screening questionnaire - a new diagnostic instrument. *Mov. Disord.* **22**, 2386–2393 (2007).
- Johns, M. W. A new method for measuring daytime sleepiness: the Epworth sleepiness scale. *Sleep* **14**, 540–545 (1991).
- Yesavage, J. A. & Sheikh, J. I. Geriatric Depression Scale (GDS) recent evidence and development of a shorter version. *Clin. Gerontol.* **5**, 165–173 (1986).
- Doty, R. L., Shaman, P. & Dann, M. Development of the University of Pennsylvania Smell Identification Test: a standardized microencapsulated test of olfactory function. *Physiol. Behav.* **32**, 489–502 (1984).
- Wolfensberger, M. Sniffin'sticks: a new olfactory test battery. *Acta Otolaryngol.* **120**, 303–306 (2000).
- Doty, R. L., Marcus, A. & Lee, W. W. Development of the 12-item Cross-Cultural Smell Identification Test (CC-SIT). *Laryngoscope* **106**, 353–356 (1996).
- Schwab, R. S. & England, A. C. Projection technique for evaluating surgery in Parkinson's disease. in *Third Symposium on Parkinson's Disease* (eds Billingham, F. H. & Donaldson, M. C.) p. 152–157 (Churchill Livingstone, 1969).
- Jenkinson, C., Fitzpatrick, R., Peto, V., Greenhall, R. & Hyman, N. The PDQ-8: development and validation of a short-form Parkinson's disease questionnaire. *Psychol. Health* **12**, 805–814 (1997).
- Chaudhuri, K. R. et al. King's Parkinson's disease pain scale, the first scale for pain in PD: an international validation. *Mov. Disord.* **30**, 1623–1631 (2015).

ACKNOWLEDGEMENTS

This research is supported by the Aligning Science Across Parkinson's Initiative, the Intramural Research Program, National Institute on Aging, National Institutes of Health, Department of Health and Human Services, project ZO1 AG000949, and the Michael J. Fox Foundation for Parkinson's Research. Data used in the preparation of this article were obtained from Global Parkinson's Genetics Program (GP2). GP2 is funded by the Aligning Science Across Parkinson's (ASAP) initiative and implemented by The Michael J. Fox Foundation for Parkinson's Research (<https://gp2.org>). For a complete list of GP2 members see <https://gp2.org>.

AUTHOR CONTRIBUTIONS

C.T., M.R., S.J., E.J.S., J.J., T.A., A.C.M., M.M.X.T., H.I., and H.R.M. are members of the GP2 Cohort Integration Working Group (CIWG), of which H.R.M. is the lead, and M.M.X.T. and H.I. are co-leads. C.T. was the primary contributor in the drafting of this manuscript, assisted by M.R. and J.J. The draft was reviewed by all CIWG members prior to circulation to other contributing authors for review. M.B.M. is the co-lead of the GP2 Data and Code Dissemination Working Group and drafted the data availability section of this manuscript, together with B.C. who is senior associate

director at the Michael J. Fox Foundation's research division. D.V., K.L., H.L., and M.A.N. are members of the GP2 Complex Disease Data Analysis Working Group (DAWG), of which M.A.N. is the lead and H.L. is a co-lead. DAWG are responsible for the data cleaning and analysis described in this manuscript. DGH directs the Genomic Technologies Group within the Laboratory of Neurogenetics at NIH, which conducts the majority of sample genotyping for the Complex Network. C.E.W. and J.S. are members of the GP2 Operations and Compliance Working Group of which J.S. is a co-lead. C.B.P. and L.A.S. are scientific program managers for the GP2 Complex Disease Network. A.B.S. is the lead of the GP2 Complex Disease Network and C.B. is the co-lead. All listed authors reviewed the manuscript and provided comments and revisions prior to submission.

COMPETING INTERESTS

A.B.S. and C.B. are supported by the Intramural Research Program of the National Institute on Aging and have received grant support from the Michael J. Fox Foundation for Parkinson's Research and the Aligning Science Across Parkinson's Initiative. A.B.S. has received royalty payments related to a diagnostic for stroke. A.B.S. is an editor for npj Parkinson's Disease. A.B.S. was not involved in the journal's review of, or decisions related to, this manuscript. H.L., H.J., D.V., K.L., and M.A.N. are consultants employed by Data Tecnica International. Data Tecnica is engaged in a consulting agreement with the US National Institutes of Health. H.R.M. is employed by UCL. In the last 24 months, he reports paid consultancy from Biogen, Biohaven, Lundbeck; lecture fees/honoraria from Wellcome Trust, Movement Disorders Society; Research Grants from Parkinson's UK, Cure Parkinson's Trust, PSP Association, CBD Solutions, Drake Foundation, Medical Research Council, Michael J. Fox Foundation. H.R.M. is also a co-applicant on a patent application related to C9ORF72—Method for diagnosing a neurodegenerative disease (PCT/GB2012/052140). BC and JS are employed by the Michael J. Fox Foundation for Parkinson's Research. All other authors declare no financial or non-financial competing interests.

THE GLOBAL PARKINSON'S GENETICS PROGRAM (GP2)

Emilia M. Gatto¹⁴, Marcelo Kauffman¹⁵, Samson Khachatryan¹⁶, Zaruhi Tavadyan¹⁶, Claire E. Shepherd¹⁷, Julie Hunter¹⁸, Kishore Kumar¹⁹, Melina Ellis²⁰, Miguel E. Rentería²¹, Sulev Koks²², Alexander Zimprich²³, Artur F. Schumacher-Schuh²⁴, Carlos Rieder²⁵, Paula Saffie Awad²⁶, Vitor Tumas²⁷, Sarah Camargos²⁸, Edward A. Fon²⁹, Oury Monchi³⁰, Ted Fon³¹, Benjamin Pizarro Galleguillos³², Marcelo Miranda³³, Maria Leonor Bustamante³⁴, Patricio Olguin³², Pedro Chana³⁵, Beisha Tang³⁶, Huifang Shang³⁷, Jifeng Guo³⁸, Piu Chan³⁹, Wei Luo⁴⁰, Gonzalo Arboleda⁴¹, Jorge Orozco⁴², Marlene Jimenez del Rio⁴³, Alvaro Hernandez⁴⁴, Mohamed Salama⁴⁵, Walaa A. Kamel⁴⁶, Yared Z. Zewde⁴⁷, Alexis Brice⁴⁸, Jean-Christophe Corvol⁴⁹, Ana Westenberger⁵⁰, Anastasia Illarionova⁵¹, Brit Mollenhauer⁵², Christine Klein⁵⁰, Eva-Juliane Vollstedt⁵⁰, Franziska Hopfner⁵³, Günter Höglinger⁵³, Harutyun Madoev⁵⁰, Joanne Trinh⁵⁰, Johanna Junke⁵⁰, Katja Lohmann⁵⁰, Lara M. Lange⁵⁴, Manu Sharma⁵⁵, Sergiu Groppa⁵⁶, Thomas Gasser⁵⁵, Zih-Hua Fang⁵⁷, Albert Akpalu⁵⁸, Georgia Xiromerisiou⁵⁹, Georgios Hadjigorgiou⁵⁹, Ioannis Dagklis⁶⁰, Ioannis Tarnanas⁶¹, Leonidas Stefanis⁶², Maria Stamelou⁶³, Efthymios Dadiotis⁵⁹, Alex Medina⁶⁴, Germaine Hiu-Fai Chan⁶⁵, Nancy Ip⁶⁶, Nelson Yuk-Fai Cheung⁶⁵, Phillip Chan⁶⁶, Xiaopu Zhou⁶⁶, Asha Kishore⁶⁷, Divya KP⁶⁸, Pramod Pal⁶⁹, Prashanth Lingappa Kukkle⁷⁰, Roopa Rajan⁷¹, Rupam Borgohain⁷², Mehri Salari⁷³, Andrea Quattrone⁷⁴, Enza Maria Valente⁷⁵, Lucilla Parnetti⁷⁵, Micol Avenali⁷⁵, Tommaso Schirinzi⁷⁷, Manabu Funayama⁷⁸, Nobutaka Hattori⁷⁹, Tomotaka Shiraishi⁸⁰, Altnay Karimova⁸¹, Gulnaz Kaishibayeva⁸¹, Cholpon Shambetova⁸², Rejko Krüger⁸³, Ai Huey Tan⁸⁴, Azlina Ahmad-Annur⁸⁴, Mohamed Ibrahim Norlinah⁸⁵, Nor Azian Abdul Murad⁸⁶, Shahrul Azmin⁸⁷, Shen-Yang Lim⁸⁴, Wael Mohamed⁸⁸, Yi Wen Tay⁸⁴, Daniel Martinez-Ramirez⁸⁹, Mayela Rodriguez-Violante⁹⁰, Paula Reyes-Pérez⁹¹, Bayasgalan Tserensodnom⁹², Rajeev Ojha⁹³, Tim J. Anderson⁹⁴, Toni L. Pitcher⁹⁴, Arinola Sanyaolu⁹⁵, Njideka Okubadejo⁹⁵, Oluwadamilola Ojo⁹⁶, Jan O. Aasly⁹⁷, Lasse Pihlstrøm⁹⁸, Manuela Tan⁹⁸, Shoaib Ur-Rehman⁹⁹, Mario Cornejo-Olivas¹⁰⁰, Maria Leila Doquenía¹⁰¹, Raymond Rosales¹⁰¹, Angel Vinuela¹⁰², Elena Iakovenko¹⁰³, Bashayer Al Mubarak¹⁰⁴, Muhammad Umair¹⁰⁵, Eng-King Tan¹⁰⁶, Jia Nee Foo¹⁰⁷, Ferzana Amod¹⁰⁸, Jonathan Carr¹⁰⁹, Soraya Bardien¹¹⁰, Beomseok Jeon¹¹¹, Yun Joong Kim¹¹², Esther Cubo¹¹³, Ignacio Alvarez¹¹⁴, Janet Hoenicka¹¹⁵, Katrin Beyer¹¹⁶, Maria Teresa Periñan¹¹⁷, Pau Pastor¹¹⁸, Sarah El-Sadig¹¹⁹, Christiane Zweier¹²⁰, Paul Krack¹²⁰, Chin-Hsien Lin¹²¹, Hsiu-Chuan Wu¹²², Pin-Jui Kung¹²³, Ruey-Meei Wu¹²¹, Yihru Wu¹²², Rim Amouri¹²⁴, Samia Ben Sassi¹²⁵, A. Nazlı Başak¹²⁶, Gencer Genc¹²⁷, Özgür Öztıp Çakmak¹²⁶, Sibel Ertan¹²⁶, Alastair Noyce¹²⁸, Alejandro Martinez-Carrasco^{1,2}, Anette Schrag², Anthony Schapira², Camille Carroll¹²⁹, Claire Bale¹³⁰, Donald Grosset¹³¹, Eleanor J. Stafford^{1,2}, Henry Houlden², Huw R. Morris^{1,2}, John Hardy², Kin Ying Mok², Mie Rizig², Nicholas Wood², Nigel Williams¹³², Olaitan Okunoye², Patrick Alfray Lewis¹³³, Rauan Kaiyrzhanov², Rimona Weil², Seth Love¹³⁴, Simon Stott¹³⁵, Simona Jasaitye², Sumit Dey¹²⁸, Vida Obese², Alberto Espay¹³⁶, Alyssa O'Grady⁶, Andrew B. Singleton^{1,2,3,7,12}, Andrew K. Sobering¹³⁷, Bernadette Siddiqi⁶, Bradford Casey^{1,2,3,4,5,6}, Brian Fiske⁶, Cabell Jonas¹³⁸, Carlos Cruchaga¹³⁹, Caroline B. Pantazis^{1,2,3,4,7}, Charisse Comart⁶, Claire Wegel¹⁴⁰, Cornelis Blauwendraat^{1,2,3,4,7,11}, Dan Vitale^{4,7,8}, Deborah Hall¹⁴¹, Dena Hernandez⁴, Ejaz Shiamim¹⁴², Ekemini Riley¹⁴³, Faraz Faghri⁴, Geidy E. Serrano¹⁴⁴, Hampton Leonard^{3,7,8,9}, Hirotaka Iwaki^{3,7,8}, Honglei Chen¹⁴⁵, Ignacio F. Mata¹⁴⁶, Ignacio Juan Keller Sarmiento¹⁴⁷, Jared Williamson¹⁴², Jonggeol Jeff Kim⁴, Joseph Jankovic¹⁴⁸, Joshua Shulman¹⁴⁹, Justin C. Solle⁶, Kaileigh Murphy⁶, Karen Nuytemans¹⁵⁰, Karl Kiebert¹⁵¹, Katerina Markopoulou¹⁵², Kenneth Marek¹⁵³, Kristin S. Levine⁸, Lana M. Chahine¹⁵⁴, Laura Ibanez¹³⁹, Laurel Screven¹², Lauren Ruffrage¹⁵⁵, Lisa Shulman¹⁵⁶, Luca Marsili¹³⁶, Maggie Kuhl⁶, Marissa Dean¹⁵⁵, Mary B. Makarios^{1,3,4}, Mathew Koretsky⁴, Megan J. Puckelwartz¹⁵⁷,

ADDITIONAL INFORMATION

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41531-023-00533-w>.

Correspondence and requests for materials should be addressed to Huw R. Morris.

Reprints and permission information is available at <http://www.nature.com/reprints>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2023

Miguel Inca-Martinez¹⁴⁶, Mike A. Nalls^{3,4,7,8}, Naomi Louie⁶, Niccolò Emanuele Mencacci¹⁴⁷, Roger Albin¹⁵⁸, Roy Alcalay¹⁵⁹, Ruth Walker¹⁶⁰, Sara Bandres-Ciga⁴, Sohini Chowdhury⁶, Sonya Dumanis¹⁶¹, Steven Lubbe¹⁵⁷, Tao Xie¹⁶², Tatiana Foroud¹⁶³, Thomas Beach¹⁶⁴, Todd Sherer⁶, Yeajin Song⁴, Duan Nguyen¹⁶⁵, Toan Nguyen¹⁶⁵ and Masharip Atadzhanov¹⁶⁶

¹⁴Sanatorio de la Trinidad Mitre- INEBA, Buenos Aires, Argentina. ¹⁵Hospital JM Ramos Mejia, Buenos Aires, Argentina. ¹⁶Somnus Neurology Clinic, Yerevan, Armenia. ¹⁷Neuroscience Research Australia, Sydney, NSW, Australia. ¹⁸ANZAC Research Institute, Concord, NSW, Australia. ¹⁹Garvan Institute of Medical Research and Concord Repatriation General Hospital, Darlinghurst, NSW, Australia. ²⁰Concord Hospital, Concord, NSW, Australia. ²¹QIMR Berghofer Medical Research Institute, Herston, QLD, Australia. ²²Murdoch University, Perth, Australia. ²³Medical University Vienna Austria, Vienna, Austria. ²⁴Universidade Federal do Rio Grande do Sul / Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil. ²⁵Federal University of Health Sciences of Porto Alegre, Porto Alegre, Brazil. ²⁶Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil. ²⁷University of São Paulo, São Paulo, Brazil. ²⁸Universidade Federal de Minas Gerais, Belo Horizonte, Brazil. ²⁹Montreal Neurological Institute, Montreal, QC, Canada. ³⁰Institut universitaire de gériatrie de Montréal, Montreal, QC, Canada. ³¹McGill University, Montreal, QC, Canada. ³²Universidad de Chile, Santiago, Chile. ³³Fundación Diagnosis, Santiago, Chile. ³⁴Faculty of Medicine Universidad de Chile, Santiago, Chile. ³⁵CETRAM, Santiago, Chile. ³⁶Central South University, Changsha, China. ³⁷West China Hospital Sichuan University, Chengdu, China. ³⁸Xiangya Hospital, Changsha, China. ³⁹Capital Medical University, Beijing, China. ⁴⁰Zhejiang University, Hangzhou, China. ⁴¹Universidad Nacional de Colombia, Bogotá, Colombia. ⁴²Fundación Valle del Lili, Santiago De Cali, Colombia. ⁴³University of Antioquia, Medellín, Colombia. ⁴⁴University of Costa Rica, San Jose, Costa Rica. ⁴⁵The American University in Cairo, Cairo, Egypt. ⁴⁶Beni-Suef University, Beni Suef, Egypt. ⁴⁷Addis Ababa University, Addis Ababa, Ethiopia. ⁴⁸Paris Brain Institute, Paris, France. ⁴⁹Sorbonne Université, Paris, France. ⁵⁰University of Lübeck, Lübeck, Germany. ⁵¹Deutsches Zentrum für Neurodegenerative Erkrankungen, Göttingen, Germany. ⁵²University Medical Center Göttingen, Göttingen, Germany. ⁵³Department of Neurology, University Hospital, LMU Munich, Munich, Germany. ⁵⁴University of Lübeck and University Medical Center Schleswig-Holstein, Lübeck, Germany. ⁵⁵University of Tübingen, Tübingen, Germany. ⁵⁶University of Mainz, Mainz, Germany. ⁵⁷The German Center for Neurodegenerative Diseases, Göttingen, Germany. ⁵⁸University of Ghana Medical School, Accra, Ghana. ⁵⁹University of Thessaly, Volos, Greece. ⁶⁰Aristotle University of Thessaloniki, Thessaloniki, Greece. ⁶¹Ionian University, Corfu, Greece. ⁶²Biomedical research Foundation of the Academy of Athens, Athens, Greece. ⁶³Diagnostic and Therapeutic Centre HYGIEA Hospital, Marousi, Greece. ⁶⁴Hospital San Felipe, Tegucigalpa, Honduras. ⁶⁵Queen Elizabeth Hospital, Kowloon, Hong Kong. ⁶⁶The Hong Kong University of Science and Technology, Kowloon, Hong Kong. ⁶⁷Aster Medcity, Kochi, India. ⁶⁸Sree Chitra Tirunal Institute for Medical Sciences and Technology, Thiruvananthapuram, India. ⁶⁹National Institute of Mental Health & Neurosciences, Bengaluru, India. ⁷⁰Manipal Hospital, Delhi, India. ⁷¹All India Institute of Medical Sciences, Delhi, India. ⁷²Nizam's Institute Of Medical Sciences, Hyderabad, India. ⁷³Shahid Beheshti University of Medical Science, Tehran, Iran. ⁷⁴Magna Græcia University of Catanzaro, Catanzaro, Italy. ⁷⁵University of Pavia, Pavia, Italy. ⁷⁶University of Perugia, Perugia, Italy. ⁷⁷University of Rome Tor Vergata, Rome, Italy. ⁷⁸Juntendo University, Tokyo, Japan. ⁷⁹Juntendo University faculty of medicine, Tokyo, Japan. ⁸⁰Jikei University School of Medicine, Tokyo, Japan. ⁸¹Institute of Neurology and Neurorehabilitation, Almaty, Kazakhstan. ⁸²Kyrgyz State Medical Academy, Bishkek, Kyrgyzstan. ⁸³University of Luxembourg, Luxembourg, Luxembourg. ⁸⁴University of Malaya, Kuala Lumpur, Malaysia. ⁸⁵Universiti Kebangsaan Malaysia, Selangor, Malaysia. ⁸⁶UKM Medical Molecular Biology Institute, Kuala Lumpur, Malaysia. ⁸⁷Universiti Kebangsaan Malaysia Medical Centre, Kuala Lumpur, Malaysia. ⁸⁸International Islamic University, Kuala Lumpur, Malaysia. ⁸⁹Tecnologico de Monterrey, Monterrey, Mexico. ⁹⁰Instituto Nacional de Neurologia y Neurocirugia, Mexico City, Mexico. ⁹¹Universidad Nacional Autónoma de México, Mexico City, Mexico. ⁹²Mongolian National University of Medical Sciences, Ulaanbaatar, Mongolia. ⁹³Tribhuvan University, Kirtipur, Nepal. ⁹⁴University of Otago, Dunedin, New Zealand. ⁹⁵University of Lagos, Lagos, Nigeria. ⁹⁶College of Medicine of the University of Lagos, Lagos, Nigeria. ⁹⁷Norwegian University of Science and Technology, Trondheim, Norway. ⁹⁸Oslo University Hospital, Oslo, Norway. ⁹⁹University of Science and Technology Bannu, Bannu, Pakistan. ¹⁰⁰Universidad Científica del Sur, Lima, Peru. ¹⁰¹Metropolitan Medical Center, Manila, Philippines. ¹⁰²University of Puerto Rico, San Juan, Puerto Rico. ¹⁰³Research Center of Neurology, Moscow, Russia. ¹⁰⁴King Faisal Specialist Hospital and Research Center, Riyadh, Saudi Arabia. ¹⁰⁵King Abdullah International Medical Research Center, Jeddah, Saudi Arabia. ¹⁰⁶National Neuroscience Institute, Singapore, Singapore. ¹⁰⁷Nanyang Technological University, Singapore, Singapore. ¹⁰⁸University of KwaZulu-Natal, Durban, South Africa. ¹⁰⁹University of Stellenbosch, Stellenbosch, South Africa. ¹¹⁰Stellenbosch University, Stellenbosch, South Africa. ¹¹¹Seoul National University Hospital, Seoul, South Korea. ¹¹²Yongin Severance Hospital, Seoul, South Korea. ¹¹³Hospital Universitario Burgos, Burgos, Spain. ¹¹⁴University Hospital Mutua Terrassa, Barcelona, Spain. ¹¹⁵Institut de Recerca Sant Joan de Deu, Barcelona, Spain. ¹¹⁶Research Institute Germans Trias i Pujol, Barcelona, Spain. ¹¹⁷Instituto de Biomedicina de Sevilla, Seville, Spain. ¹¹⁸University Hospital Germans Trias i Pujol, Barcelona, Spain. ¹¹⁹Faculty of medicine university of Khartoum, Khartoum, Sudan. ¹²⁰Inselspital Bern, University of Bern, Bern, Switzerland. ¹²¹National Taiwan University Hospital, Taipei City, Taiwan. ¹²²Chang Gung Memorial Hospital, Taoyuan City, Taiwan. ¹²³National Taiwan University, Taipei City, Taiwan. ¹²⁴National Institute Mongi Ben Hamida of Neurology, Tunis, Tunisia. ¹²⁵Mongi Ben Hmida National Institute of Neurology, Tunis, Tunisia. ¹²⁶Koç University, Istanbul, Turkey. ¹²⁷Sisli Etfal Training and Research Hospital, Istanbul, Turkey. ¹²⁸Queen Mary University of London, London, UK. ¹²⁹University of Plymouth, Plymouth, UK. ¹³⁰Parkinson's UK, London, UK. ¹³¹University of Glasgow, Glasgow, UK. ¹³²Cardiff University, Cardiff, UK. ¹³³Royal Veterinary College University of London, London, UK. ¹³⁴University of Bristol, Bristol, UK. ¹³⁵Cure Parkinson's, London, UK. ¹³⁶University of Cincinnati, Cincinnati, OH, USA. ¹³⁷Augusta University / University of Georgia Medical Partnership, Augusta, GA, USA. ¹³⁸Mid-Atlantic Permanente Medical Group, Bethesda, MD, USA. ¹³⁹Washington University, St. Louis, MO, USA. ¹⁴⁰Indiana University, Bloomington, IN, USA. ¹⁴¹Rush University, Chicago, IL, USA. ¹⁴²Kaiser Permanente, Oakland, CA, USA. ¹⁴³Coalition for Aligning Science, Washington, WA, USA. ¹⁴⁴Banner Sun Health Research Institute, Sun City, AZ, USA. ¹⁴⁵Michigan State University, East Lansing, MI, USA. ¹⁴⁶Cleveland Clinic, Cleveland, OH, USA. ¹⁴⁷Northwestern University, Evanston, IL, USA. ¹⁴⁸Baylor College of Medicine, Houston, TX, USA. ¹⁴⁹Baylor College of Medicine, Texas Children's Hospital, Houston, TX, USA. ¹⁵⁰University of Miami Miller School of Medicine, Miami, FL, USA. ¹⁵¹Beth Israel Deaconess Medical Center, Boston, MA, USA. ¹⁵²North Shore University Health System, Chicago, IL, USA. ¹⁵³Institute for Neurodegenerative Disorders, New Haven, CT, USA. ¹⁵⁴University of Pittsburgh, Pittsburgh, PA, USA. ¹⁵⁵University of Alabama at Birmingham, Birmingham, AL, USA. ¹⁵⁶University of Maryland, Baltimore, MD, USA. ¹⁵⁷Northwestern University, Chicago, IL, USA. ¹⁵⁸University of Michigan, Ann Arbor, MI, USA. ¹⁵⁹Columbia University, New York, NY, USA. ¹⁶⁰James J. Peters Veterans Affairs Medical Center, New York, NY, USA. ¹⁶¹Aligning Science Across Parkinson's, Washington, WA, USA. ¹⁶²University of Chicago, Chicago, IL, USA. ¹⁶³Indiana University School of Medicine, Indianapolis, IN, USA. ¹⁶⁴Sun Health Research Institution, Sun City, AZ, USA. ¹⁶⁵Hue University, Huế, Vietnam. ¹⁶⁶University of Zambia, Lusaka, Zambia.