GP2: The Global Parkinson's Genetics Program

Author: The Global Parkinson's Genetics Program

Address for correspondence:

Cornelis Blauwendraat, PhD (cornelis.blauwendraat@nih.gov) and Andrew Singleton, PhD (singleta@mail.nih.gov)
Laboratory of Neurogenetics
NIA, NIH, Building 35, 35 Convent Drive
Bethesda, MD 20892, USA

Word Count: 5573

Running Title: GP2: The Global Parkinson's Genetics Program

Abstract

Supported by the Aligning Science Across Parkinson's (ASAP) Initiative, the Global Parkinson's Genetics Program (GP2) is an international collaborative effort aimed at making transformational progress in Parkinson's disease (PD) genetics.

The aim of GP2 is not simply to improve our understanding of the role genetics plays in PD across World populations, but to also make that understanding actionable. The path to achieving this vision involves the formation of a cohesive group of collaborators, data collection, production, and harmonization at scale, training of analysts around the world, and the development of a portal that democratizes data and analytical resources.

Bringing together collaborators from Africa, Asia, Central America, the Caribbean, Europe, the Middle East, North America, Oceania, and South America, GP2 will integrate and generate critical clinical and genetic data. Using these data, GP2 will accelerate the identification of novel risk loci and monogenic causes of disease, identify genetic modifiers of disease phenotype and monogenic penetrance, fine map risk loci, and understand population differences in PD genetics.

Introduction

Fundamentally, the identification of genetic causes and contributors of disease represents the first step in an etiology-based understanding of disease, which, in turn, is a required step in the development of therapeutics targeting the underlying disease process. Genetic understanding serves as a foundation for succeeding functional studies and as a central component of efforts to predict disease risk, onset, and progression, and to understand disease mechanisms in individual patients. Without a reliable and complete foundation of genetic understanding, we limit our ability to develop and deploy treatments.

A large number of risk loci and causative mutations for PD have been identified; however, it is clear that the majority of genetic risk remains to be found [1,2]. While much can be done with the existing knowledge, moving forward now to expand our genetic understanding will be the foundation that will support the development of a complete view of this network, providing an array of potential therapeutic opportunities. Increasing genetic information can only serve to improve our efforts to treat disease.

Notably, our understanding of the genetic basis of PD has thus far largely been centered on research in individuals of Northern European ancestry. While some genetic discoveries have been made outside of these populations, this work is the exception rather than the norm, and generally focuses on the identification of rare mutations; little has been done in the identification of more common genes or genetic risk discovery [3,4]. Thus, we do not know if our current understanding is generalizable to the rest of the world and how the basis of disease varies across populations. While it is tempting to argue that the genetic basis for the PD will generally be the same across populations, we know that differences in genetic risk exist, and further, there is evidence to suggest that genetic forms of disease can present differently across populations [3,5–8]. This fundamental limitation of current research creates an inequitable situation for patients.

To facilitate the rapid expansion of our understanding of the genetic architecture of PD, both in terms of the depth and global context of this knowledge, we have created the Global Parkinson's Genetics Program (GP2, www.gp2.org). GP2 is the first supported resource project of the Aligning Science Across Parkinson's (ASAP) initiative, an audacious effort supporting PD research [9]. GP2 is geared toward creating a worldwide collaborative effort that will first dramatically accelerate the identification of genetic contributors to disease and second establish a network of researchers that can best leverage this understanding to research, diagnose, and treat PD worldwide. Here we describe our mission, the path we have proposed to achieve this, and the core principles of data democratization, transparency, and diversity.

Mission and underlying principles

The mission of GP2 is to drive transformational progress in our understanding of the genetic architecture of PD and to serve as a useful and actionable resource for the research and therapeutic development community. To fully realize this mission, GP2 will need to engage and mobilize a worldwide community of researchers and participants, generate and analyze genetic data at an extremely large scale, create an infrastructure that removes obstacles to data access, and make data and results accessible and useful to the broader community (Figure 1).

There are several underlying principles of GP2 that we believe are central to the overall success of our work and the continued success of the research community.

1. Diversification: We will leverage the power of diversity across each dimension and in both researchers and participants.

2. Democratization: We will ensure that the data, its use, compute resources, and results are not only available but usefully accessible to the broader research community. We will democratize GP2 and the use of GP2 data through training for contributing sites to create clinical and data-analytical expertise locally and regionally, growing our own and collaborators' capabilities.

3. Transparency and reproducibility: The collection, data generation, data cleaning, and analysis will be performed in an open manner. Methods, data, code, and results will be available to the research community in order to facilitate reproducibility, reduce redundancy, and to foster community involvement in and improvement of approaches.

4. Collaboration and cooperation: We will promote a high degree of collaboration and cooperation across a global community. To be effective this must be centered on a shared vision, collective opportunities and responsibilities, and ensuring that each member has a voice in the organization.

5. Foundational, actionable resource generation: The data and results we produce will form the foundation of a wealth of scientific and clinical research; therefore they must be both easily available and in a form that is useful for and interpretable by the wider research community.

6. Safe, responsible data sharing: Data sharing with the research community is key; however this must be done in a manner that ensures participant privacy and is in line with local regulations.

Deliverables and path

Broadly, there are two scientific arms to *G*P2, one centered in genetically complex, typical PD, and the other in monogenic disease. Over the initial five-year span of the GP2 program, our path will lead us to a dramatic increase in the number of known genes, disease-causing mutations, and risk loci for both rare monogenic and typical complex PD. Further, this work will be extended, for the first time at scale, to underrepresented populations from around the World. We will generate dense genetic data in more than 150,000 participants, using a genotyping array specifically designed for this purpose. We will also generate whole-genome sequence data from more than 10,000 individuals, to determine the genetic cause in as yet unsolved monogenic cases and to generate much needed reference datasets. Furthermore, we will use long-read DNA sequencing to support the analysis of structural and repeat variability that is relatively resistant to interrogation using traditional genome sequencing methods.

In order to ensure a functional and efficient structure, we created a series of working groups and hubs that center on achieving specific aims and priorities within GP2. While these groups have clear aims and deliverables, they function as a continuum with shared members (Figure 2).

Complex disease genetics

The role of this group is to explore the genetic basis of typical, apparently sporadic, PD. The foundation of this work will be the genotyping of 150,000 participants using an array designed by us specifically for this purpose. The Neuro Booster Array is centered on the backbone of the global diversity array (GDA, 1.8+ million variants) (https://www.illumina.com/products/by-type/microarray-kits/infinium-global-diversity.html) but also includes more than 95,000 custom content variants that include neurological disease-oriented content, and population-specific boosters (Manuscript in prep. https://github.com/GP2code/Neuro_Booster_Array). Broadly, we expect to generate data on ~100,000 Northern European ancestry individuals and more than 50,000 subjects from underrepresented populations from around the world. We have established collaborations to collect and assess cases of Black American, East Asian, African, Indian, Caribbean and Central/South American provenance.

Data from underrepresented populations will be generated from a variety of sources. GP2 has already initiated partnerships with academic research centers in the US to improve representation of Black Americans within the project. Samples and data are being collected from East Asia by the International Parkinson’s Disease Genomics Consortium (IPDGC) East Asia group, with efforts ongoing in Taiwan, Japan, South Korea, and China. Likewise, IPDGC Africa has initiated collaborations across Africa beginning with patients from Nigeria, Egypt, Ethiopia, Ghana, Mali, Tanzania, Senegal, South Africa, Sudan, and Zambia. The Genetic Epidemiology of Parkinson’s disease (GEoPD) Consortium has developed collaborations across underrepresented populations from North and Sub-Saharan Africa, Australia, and Asia. The Luxembourgish-German Indian Alliance on Neurodegenerative diseases and Therapeutics (Lux-GIANT) has formed a collaborative group to investigate PD patients across India [10]. Lastly, the Latin American Research Consortium on the Genetics of Parkinson's Disease (LARGE-PD) group is a fully active collaborative group collecting and investigating patients from Argentina, Brazil, Chile, Costa Rica, Colombia, Ecuador, Honduras, Mexico, Peru, Puerto Rico, Uruguay, and West Indies [11]. As GP2 continues, there will be room to expand to other countries and populations underserved in our current research.

These data, collectively, will afford the opportunity to rapidly detect novel genetic risk for PD. Critically, the availability of similar data across ancestral groups will allow an assessment of the varied genetic contribution in different ancestries, including the identification of population-specific loci, an understanding of the differences in the heritable component of disease between groups, and the generation of population-specific genetic risk profiles. Notably, these data collectively provide the opportunity to refine association signals, with trans-ethnic fine mapping.

A crucial step will be integration of clinical phenotype data. We know that there are diverse outcomes of PD including rate of motor and cognitive deterioration, and medication side effects such as levodopa-induced dyskinesias [12]. We believe that these will be in part genetically determined [13,14] and that understanding the associated genes and pathways will lead to new biological insights and importantly personalised treatments. A barrier to this is the harmonisation of data, which is a major goal of the cohort integration group.

Monogenic disease genetics

While modern genetic methods provide tools for the rapid discovery of rare causal or high-risk mutations, several barriers exist that limit the efficiency of identifying novel causes of disease. First, multiplex families are overall rare and dispersed; second, the generation and analysis of genetic data are specialized and expensive; third, existing genetic data are not harmonized and often not available for sharing; and fourth, penetrance is reduced in dominantly inherited forms and, although high in recessive forms, age-dependent. The latter results in frequent absence of the most prominent red flag for the occurrence of a monogenic form of PD, i.e. positive family history, so that a significant proportion of patients miss out on genetic testing and research because they are not deemed good candidates. Collectively, this means that finding segregating mutations, or mutations in the same putative novel gene is difficult; this in turn has resulted in the publication of a growing number of potentially disease-associated mutations that are preliminary and can be quite misleading or confusing to the field. Importantly, and unlike findings from complex genetic studies of PD, these putative new monogenic causes are often readily implemented in PD gene panels for diagnostic testing by genetic testing companies, posing an additional challenge to patients, unaffected carriers, genetic counselors and physicians in terms of interpretation of the ensuing test results.

The monogenic disease arm of GP2 aims to address these obstacles and thereby create an efficient infrastructure to accelerate the identification of novel genetic causes of apparently monogenic PD. Leveraging the above-described global network of researchers contributing patient samples to GDA genotyping and including already existing resources from the monogenic field, such as IPDGC [15], the GEoPD and the MJFF Global Genetic Parkinson's Disease Study Group [16], the monogenic arm will collect >5,000 patients and families in whom a monogenic cause may be suspected. Particular emphasis will be placed on families from underrepresented populations. All currently known PD genes have been found in various populations around the globe, however, some occur at highly variable and population-specific frequencies, the most striking examples being the p.G2019S mutation in the *LRRK2* gene [17] and *GBA* mutations including p.N370S [18] and p.K198E [19]. In addition, it is conceivable that population-specific hereditary forms of PD may exist, as exemplified by X-linked dystonia-parkinsonism (XDP), a condition exclusively present in patients of Filipino ancestry (Pauly et al. Mov Disord, in press), for which the underlying genetic cause has been identified as well as genetic age-at-onset modifiers [20–22]. This condition has served as an important model for basal ganglia disease [23,24].

The monogenic arm will collect families, singleton cases, and patient-parent trios, and prioritize these for WGS or long-range sequencing based on a number of different criteria: family history and availability of samples from several affected (and unaffected) family members, age at onset, ethnicity (with a focus on underrepresented populations), and level of available genetic pre-screening. Importantly, all patients enrolled into the monogenic arm of the project will also undergo Neuro Booster Array genotyping which, based on its PD-related custom content, will result in the identification of a sizable number of patients with mutations in known PD genes. Assuming that an average of ~10-15% of all PD patients carry a mutation in a known PD gene (3-5%) or a high risk variant in *GBA* (8-10%) [25], we estimate a total of ~15,000 monogenic or high risk variant carriers to be detected in the total GP2 sample set. It is at this important interface, where the complex genetic and monogenic arms will interact most closely: While the complex genetics arm will identify carriers of known PD-causing mutations that can then be enrolled in various additional investigations, such as genetic modifier studies of age at onset, the monogenic hub will contact all submitters of patients to the monogenic hub to also enroll patient and control cohorts into the complex genetics hub. This interplay will create unprecedented opportunities not only for the discovery of novel genetic causes of PD but also for a better understanding of the known genetic forms of PD.

The Democratization of Data Resources

Over the last fifteen years, the genetics field has made great strides toward making data available to the research community. However, barriers still exist. Data is typically highly dispersed across silos/portals, there is often considerable administrative burden for data access especially when accessing multiple datasets and data use agreements can be restrictive. In addition, the analytical expertise to interpret results or analyze data can be high, and the cost of data analysis plus curation can be prohibitive. A key stated outcome of GP2 is the generation of data and analyses that can be readily accessed and interpreted by the broad research community. To achieve this, we are taking several steps to consolidate data in one place and provide all analysis scripts with the necessary context in GitHub (https://github.com/GP2code), a public domain.

We will place as much information as timely and responsibly possible in the public domain, without requiring an extensive data use agreement. While the protection of participant's data must be a priority, a large number of analyses only require summary results and many of these can be shared publicly. De-identified participant level data will be stored securely; however, we will only require a single data use agreement to access these cohorts, streamlining the data access process.

We will follow a model where researchers analyze data in place, rather than downloading data to local computers. This approach offers several advantages: it means that we can reduce redundancies by capturing standard quality control and analyses and sharing these as a common path. It means we can create collaborative and training opportunities and resources by working across a common, shared workspace. In GP2, we can be both standardized and transparent in our analyses, which can be easily shared with the research community, allowing independent testing, additive analyses, and crowd-sourced improvements to workflows. This approach democratizes the data, tools, and infrastructure, allowing individuals with the skillset but without the compute resources to access and analyze these data. The model of analysts coming to the data, rather than vice versa, is one being increasingly used or promoted by major national and global genetics and genomics initiatives such as the International Complex Disease Alliance (see whitepaper at https://www.icda.bio/) and the AllOfUs initiative (https://allofus.nih.gov/). To achieve these aims, our current model is centered around Terra (https://terra.bio/) with other platform options possible soon.

Terra is a cloud-based, scalable platform developed and actively maintained by the Broad Institute (https://www.broadinstitute.org/) specifically for biomedical research, supporting other initiatives similar to GP2. Terra supports direct access to data stored in Google Cloud Storage and all analysis can be performed on Terra, which eliminates the need for data to be downloaded to local systems and ensures that researchers work from the same data. This is especially important to comply with current GDPR standards and allowing participants to withdraw their information from future analyses. Currently, Terra supports two types of analysis: pipelining with workflows and real-time analysis with Jupyter Notebooks. Workflows allow researchers to perform whole pipelines, such as aligning sequence data per sample and joint-calling across populations. Notebooks follow an intuitive structure and allow for interactive analysis, text that provides appropriate context, and immediate visualization. Terra makes collaboration easy and transparent with the ability to share notebooks and workspaces selectively with other researchers who have similar permissions. Sharing notebooks on a public access workspace allows other researchers to easily use and reproduce your exact workflow. The Terra platform aligns with GP2 goals by creating a secure and easy way to analyze data in the cloud, and enabling collaborative, reproducible, and transparent science.

Training and networking

It is not enough to just make data available to the wider PD research community. A key part of GP2 is to develop training opportunities that will benefit clinicians and researchers around the world so that they may pursue their own questions using GP2 data.

These training opportunities aim to establish broad, foundation-level knowledge in genetics, bioinformatics, medical statistics and molecular biology, through a suite of new web-based materials. Our program of web-based training has been brought into sharp focus by the COVID-19 pandemic and we have accelerated our efforts to deliver educational material in spite of restrictions to travel. We also aim to support clinical training opportunities in regions that would benefit most from this. The increased training would lead to more accurate diagnosis and phenotyping of the participants donating samples for GP2.

For individuals that demonstrate exceptional drive and talent, whatever their background, we will support a range of individual, formal training opportunities in the form of taught courses (such as Masters degrees), bespoke training visits at centers of excellence, and full PhD studentships. We are committed to serving the needs of clinicians and researchers that have been underrepresented so far in research, and we will enable research training opportunities and build sustainable partnerships through a network of GP2 collaborators.

Promoting diversity

It is of both scientific and social imperative that research efforts be expanded to diverse populations. We believe that a critical component of expanding research to underserved and underrepresented populations is ensuring ownership of local studies by local researchers, an active voice and role in the global study, and training, development and other infrastructure support for local researchers to build on their expertise.

Within GP2, we aim to address this in several ways. All data are returned to contributing investigators immediately upon completion of harmonization and quality control; further, these data are available to the group within the GP2 Terra compute space, so the individual investigator can access the processing and quality control workspaces that have been used to generate these data. Local investigators can also use GP2 supported cloud computing to perform additional analyses on these data, removing the cost of providing their own infrastructure. A remaining challenge then is ensuring that sufficient expertise exists to perform these analyses. Here, GP2 is providing several layers of support. Within GP2, we have a data analysis core that not only serves to perform study-wide analyses but also to support members in the form of direct analytical services, partners in training, and by providing a data concierge service.

Uses of these data

The work planned for GP2 will provide foundational data and fundamental insights for PD research. We believe that this work touches almost every aspect of the path from basic biological understanding to therapeutic deployment. It has the potential to inform biology, improve modeling, identify and prioritize targets, inform trial design, increase trial efficiency, and match patients to therapies.

Most obviously, the work from GP2 will lead to a significant increase in the proportion of the known heritable component of disease. We estimate that our efforts will provide a significant increase in the knowledge that serves as the genesis for our path to functional understanding [1]. This will improve the genetic component of multi-modal risk prediction models and will provide us with greater power to determine whether genetic subtypes of the common disease are present.

Based on prior experience, the expansion of efforts into traditionally underrepresented populations will also lead to the identification of new loci that are either absent or of weaker effect in the Northern European population [3,5,6]. The underlying genetic data in the underrepresented groups will allow us to determine whether heritability differs substantially across ancestral groups and to construct ancestry-specific risk profiles. Comparisons across all of the ancestral groups will also be incredibly valuable, providing insights into the genetic basis for any potential differences in disease presentation and course, as well as highlighting which groups may be particularly suited to a distinct etiologic-based therapeutic approach. Further, the differences and similarities at individual loci allow us to use trans-ethnic fine-mapping to reduce the critical intervals in which functional risk alleles reside and to identify the functional effector gene.

Beyond risk, GP2 provides the opportunity to examine the genetic basis of variability in disease. Age at onset, progression, and the range of symptoms and comorbidities are each amenable to genetic discovery. Perhaps more importantly, this work paves the way for the creation of individual risk profiles for onset and trajectories of disease.

Creating the infrastructure and processes for the identification of mutations underlying rare monogenic forms of disease will accelerate our discovery of novel high-penetrant gene mutations. It will also aid in the efficiency of how the field promotes and tests candidate mutations; a large number of mutations in putative disease-linked genes have been published over the past five years, and the existence of reference variant data in a large and diverse series of PD cases will speed up validation and hopefully reduce the number of false-positives reports of disease-linked genes. Again, these fundamental insights provide the basis for functional investigation of disease and provide immediate targets for therapeutic investigation.

Combining complex genetic approaches with data produced in monogenic cases also allows the identification of genetic modifiers of disease penetrance. Past experience, while of limited power, has shown that this approach is feasible [25,26]. Further, these data suggest that such work may reveal differential effects of the same risk loci in primarily monogenic and complex disease, a potential functional basis for etiologic differences, or at least different etiologic weights between patients. This area is of stated interest to our colleagues in industry who are looking for potential modifiers of the disease process.

Broad Applicability

There are a wide variety of opportunities provided by the GP2 study that are not part of its most immediate scope. Subject cohorts for GP2 will come from a large range of sources, each with varying ancillary data and future workflows. There is significant potential for additional work made possible through the generation of GP2 data. Our partnership with the PPMI study (https://www.ppmi-info.org/) enables the use of genetic and inferred ancestry data in biomarker development work. This can involve a range of possibilities, including the integration of genetics within multi-modal disease predictions, an examination of the influence of ancestry on biomarkers, and the use of genetics to improve biomarker readouts by adjusting for genetic influence unrelated to disease.

Like PPMI, many studies have collected additional data (including imaging, phenotype and environmental data) that can be analyzed with genetics. The availability of imaging and neuropathology data provides an opportunity to understand the genetic basis of variance in these measures and may offer insights on disease subtyping. The availability of other data and tissue from samples genotyped in GP2 is also exciting. Combining data from largely unbiased methods such as transcriptomics, epigenomics, metabolomics, and proteomics with genetics provides the opportunity to map quantitative trait loci in both health and disease, an approach that has been used to understand basic etiology and improve biomarker accuracy.

The sponsorship of GP2 in collecting samples from underrepresented populations also opens up the opportunity to perform ancillary studies in these groups. At the most basic level this may include simple phenotypic characterization of patients from countries or ancestral groups where little has been previously done. It can also include work on attitudes and perspectives on disease and the use of genetics.

We believe these opportunities, and myriad others in the spaces of environment, epidemiology, epigenetics, and beyond, can and should be prioritized for further research. Our aim in GP2 is not to create an organization that tackles each of these, but rather to create a structure around which such studies can be organized. Our hope is that individual investigators within the GP2 network will propose new lines of investigation and form alliances and collaborations with other investigators to explore these questions. We believe that GP2 and our collective data and structure can serve as an initiator for such work.

Cellular context and function

The identification of novel PD-linked gene mutations and loci will add considerable insight into the underlying basis and biology of disease. There is excellent evidence from PD and other diseases that integrating these results with other genomic data can provide compelling evidence on the cellular context of genetic risk, a key step in ensuring the relevance of modeling efforts [2,27,28]. Likewise, the more genetic information we possess, the more complete a picture we can construct of the pathogenesis of disease. ASAP has supported a series of ambitious projects within the Collaborative Research Network (https://parkinsonsroadmap.org/research-network/) with a major focus on work to determine the biology of PD-associated genetics. In the relatively short time that we have reliably been able to identify genetic risk factors at scale, the translation to mechanistic understanding has been challenging. However, recent advances, particularly in the space of single-cell genomics, will elucidate the cellular context, effector gene, and immediate biological effect. Most immediately this will be achieved through the combination of genetic risk loci with human single-cell data on transcription, chromatin accessibility, and physical chromatin interactions. Once defined, the cellular context and disease-relevant effect on expression can be used in traditional reductionist or more recent systems approaches to define disease pathways and networks respectively. Regardless of the path taken, a fundamental understanding of the breadth and depth of genetic influence in PD will at a minimum enhance our ability to understand the disease process but more likely it will form the foundation of that knowledge.

Informing trial design

GP2's work can afford us insights into the idea of mechanistic subsetting of disease; whether there are many distinct networks involved in the disease process and the distribution of these mechanisms across a typical PD population. Critically, if we find enriched or divergent mechanisms in subsets of patients, it will be important to match patients with mechanistic personalized therapeutics. This is, in essence, an extension of the rational idea of testing (for example) *LRRK2* kinase inhibitors initially in *LRRK2* mutation-positive cases. Further, because an outcome of GP2 will be an improved prediction of disease risk and progression, this information will serve to facilitate improved clinical trial design, supporting trials in early (even pre-clinical) patients, and adjusting outcomes based on individualized predictions of progression.

The future

As discussed above, there are a large number of potential projects and opportunities associated with the GP2 data and collaborative framework. We are clearly at the beginning of our journey but it is certainly difficult to resist the temptation to speculate on what priorities and opportunities may emerge as we continue.

Clearly, the results of our current work plot the course forward. The efforts of the monogenic group are already providing a clearer understanding of what is left to be found and poised to address key questions in the field. Can we identify modifiers to monogenic disease? How close are we to identifying the majority of these factors? What critical features remain to determine which mutation carriers will express disease, and when? Likewise, data from the complex genetics group will address similar questions. We will be better able to answer the question of whether we have reached biological saturation in common risk loci - the point at which newly identified loci are failing to add new insight into disease biology - perhaps an appropriate stopping point for GWAS. We will also have a good idea of whether scaling up our sequencing efforts in complex disease makes sense; do we see significant additional discoveries or resolution using this method over genotyping?

At the simplest level, scale is immediately compelling and it is apparent even now that it would be particularly important to scale up our work in underrepresented groups. We know our current efforts in underrepresented groups represent a significant commitment and a major step forward for the field, however, they are just a start. As we show success in these groups it will be important to capitalize on this momentum and continue to invest in research within and by underserved communities so that we can realize the most potential from this work.

Lastly, we should also be considering the next immediate steps after genetic discovery. In many ways genetics has evolved into a systematic endeavor rather than a scientific one. We have a clear path forward and broad consensus on the road to take. This is also becoming true for the next immediate steps to translate from genetic maps to mechanism, and there may be some advantage to GP2 tackling some of these problems, particularly around the creation of resources to make this translation efficient. However, regardless of how we get to these next steps, we believe that GP2 will serve as the starting point on this journey and a foundation for the global PD research community. We are excited about the progress we've made so far, and thrilled to be making this journey with our partners.

Acknowledgements and funding

This research was 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. The Michael J Fox Foundation for Parkinson's Research.

Financial disclosure and conflict of interest

Dr. Singleton is supported by the Intramural Research Program of the National Institute on Aging and has received grant support from the Michael J Fox Foundation for Parkinson's Research and the Aligning Science Across Parkinson's Initiative. He has received royalty payments related to a diagnostic for Stroke. He has no conflicts of interest to report.

Dr. Blauwendraat is supported by the Intramural Research Program of the National Institute on Aging and has received grant support from the Michael J Fox Foundation for Parkinson's Research and the Aligning Science Across Parkinson's Initiative. He has no conflicts of interest to report.

References

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

2 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.

3 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* Published Online First: 20 April 2020. doi:10.1001/jamaneurol.2020.0428

4 Zhao Y, Qin L, Pan H, *et al.* The role of genetics in Parkinson's disease: a large cohort study in Chinese mainland population. *Brain* 2020;**143**:2220–34.

5 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.

6 Satake W, Nakabayashi Y, Mizuta I, *et al.* Genome-wide association study identifies common variants at four loci as genetic risk factors for Parkinson's disease. *Nat Genet* 2009;**41**:1303–7.

7 Gwinn-Hardy K, Chen JY, Liu HC, *et al.* Spinocerebellar ataxia type 2 with parkinsonism in

ethnic Chinese. *Neurology* 2000;**55**:800–5.

8 Subramony SH, Hernandez D, Adam A, *et al.* Ethnic differences in the expression of neurodegenerative disease: Machado-Joseph disease in Africans and Caucasians. *Mov Disord* 2002;**17**:1068–71.

9 Schekman R, Riley EA. Coordinating a new approach to basic research into Parkinson's disease. *Elife* 2019;**8**. doi:10.7554/eLife.51167

10 Rajan R, Divya KP, Kandadai RM, *et al.* Genetic Architecture of Parkinson's Disease in the Indian Population: Harnessing Genetic Diversity to Address Critical Gaps in Parkinson's Disease Research. *Front Neurol* 2020;**11**:524.

11 Zabetian CP, Mata IF, Latin American Research Consortium on the Genetics of PD (LARGE-PD). LARGE-PD: Examining the genetics of Parkinson's disease in Latin America. *Mov Disord* 2017;**32**:1330–1.

12 Lim S-Y, Tan AH, Ahmad-Annuar A, *et al.* Parkinson's disease in the Western Pacific Region. *Lancet Neurol* 2019;**18**:865–79.

13 Tan MMX, Lawton MA, Jabbari E, *et al.* Genome-Wide Association Studies of Cognitive and Motor Progression in Parkinson's Disease. *Mov Disord* Published Online First: 28 October 2020. doi:10.1002/mds.28342

14 Iwaki H, Blauwendraat C, Leonard HL, *et al.* Genomewide association study of Parkinson's disease clinical biomarkers in 12 longitudinal patients' cohorts. *Mov Disord* 2019;**34**:1839–50.

15 International Parkinson Disease Genomics Consortium (IPDGC). Ten Years of the International Parkinson Disease Genomics Consortium: Progress and Next Steps. *J Parkinsons Dis* 2020;**10**:19–30.

16 Vollstedt E-J, Kasten M, Klein C, *et al.* Using global team science to identify genetic parkinson's disease worldwide. *Ann Neurol* 2019;**86**:153–7.

17 Healy DG, Falchi M, O'Sullivan SS, *et al.* Phenotype, genotype, and worldwide genetic penetrance of LRRK2-associated Parkinson's disease: a case-control study. *Lancet Neurol* 2008;**7**:583–90.

18 Sidransky E, Nalls MA, Aasly JO, *et al.* Multicenter analysis of glucocerebrosidase mutations in Parkinson's disease. *N Engl J Med* 2009;**361**:1651–61.

19 Velez-Pardo C, Lorenzo-Betancor O, Jimenez-Del-Rio M, *et al.* The distribution and risk effect of GBA variants in a large cohort of PD patients from Colombia and Peru. *Parkinsonism Relat Disord* 2019;**63**:204–8.

20 Bragg DC, Mangkalaphiban K, Vaine CA, *et al.* Disease onset in X-linked dystonia-parkinsonism correlates with expansion of a hexameric repeat within an SVA retrotransposon in. *Proc Natl Acad Sci U S A* 2017;**114**:E11020–8.

21 Westenberger A, Reyes CJ, Saranza G, *et al.* A hexanucleotide repeat modifies expressivity of X-linked dystonia parkinsonism. *Ann Neurol* 2019;**85**:812–22.

22 Aneichyk T, Hendriks WT, Yadav R, *et al.* Dissecting the Causal Mechanism of X-Linked Dystonia-Parkinsonism by Integrating Genome and Transcriptome Assembly. *Cell* 2018;**172**:897–909.e21.

23 Weissbach A, Bäumer T, Rosales R, *et al.* Neurophysiological fingerprints of X-linked dystonia-parkinsonism: A model basal ganglia disease. *Mov Disord* 2015;**30**:873–5.

24 Hanssen H, Prasuhn J, Heldmann M, *et al.* Imaging gradual neurodegeneration in a basal ganglia model disease. *Ann Neurol* 2019;**86**:517–26.

25 Blauwendraat C, Reed X, Krohn L, *et al.* Genetic modifiers of risk and age at onset in GBA associated Parkinson’s disease and Lewy body dementia. *Brain* 2020;**143**:234–48.

26 Iwaki H, Blauwendraat C, Makarious MB, *et al.* Penetrance of Parkinson’s Disease in LRRK2 p.G2019S Carriers Is Modified by a Polygenic Risk Score. *Mov Disord* 2020;**35**:774–80.

27 Bryois J, Skene NG, Hansen TF, *et al.* Genetic identification of cell types underlying brain complex traits yields insights into the etiology of Parkinson’s disease. *Nat Genet* 2020;**52**:482–93.

28 Agarwal D, Sandor C, Volpato V, *et al.* A single-cell atlas of the human substantia nigra reveals cell-specific pathways associated with neurological disorders. *Nat Commun* 2020;**11**:4183.

Figures

Figure 1: The scientific outcomes and underlying structural priorities of GP2

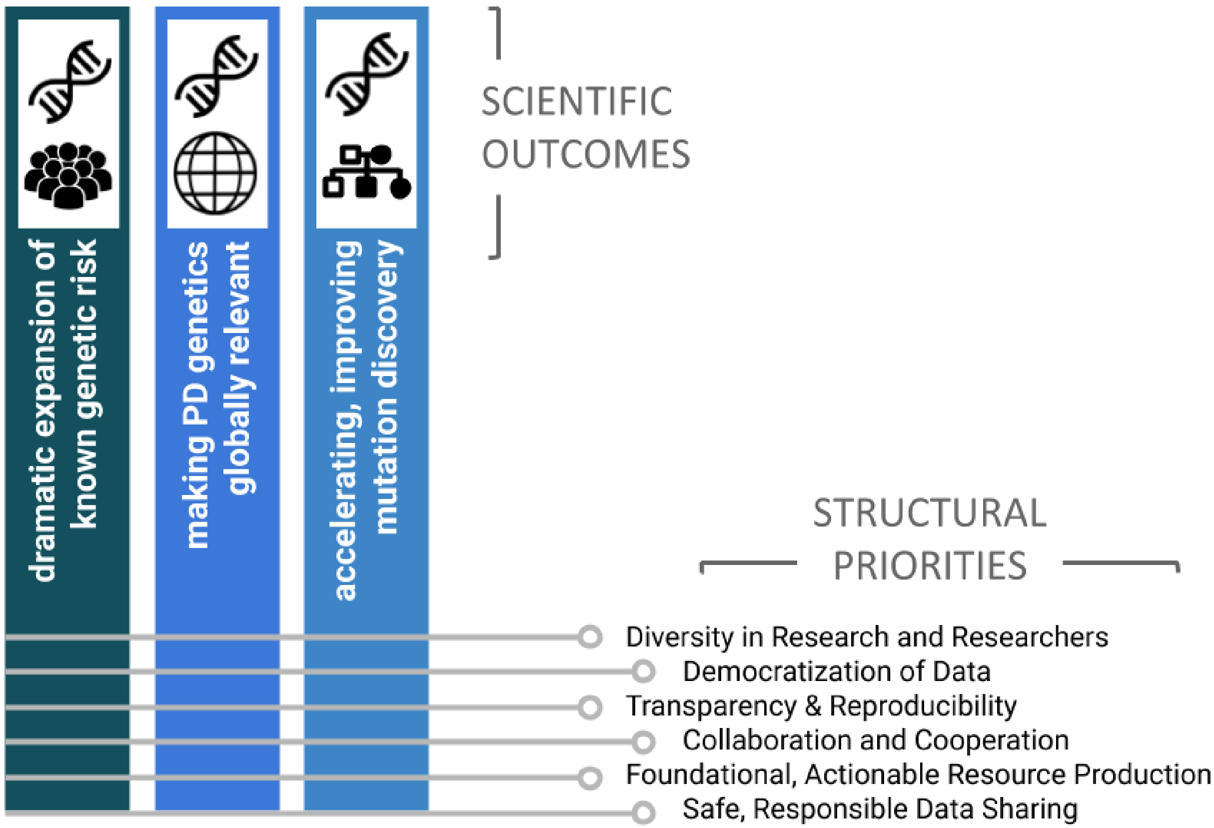
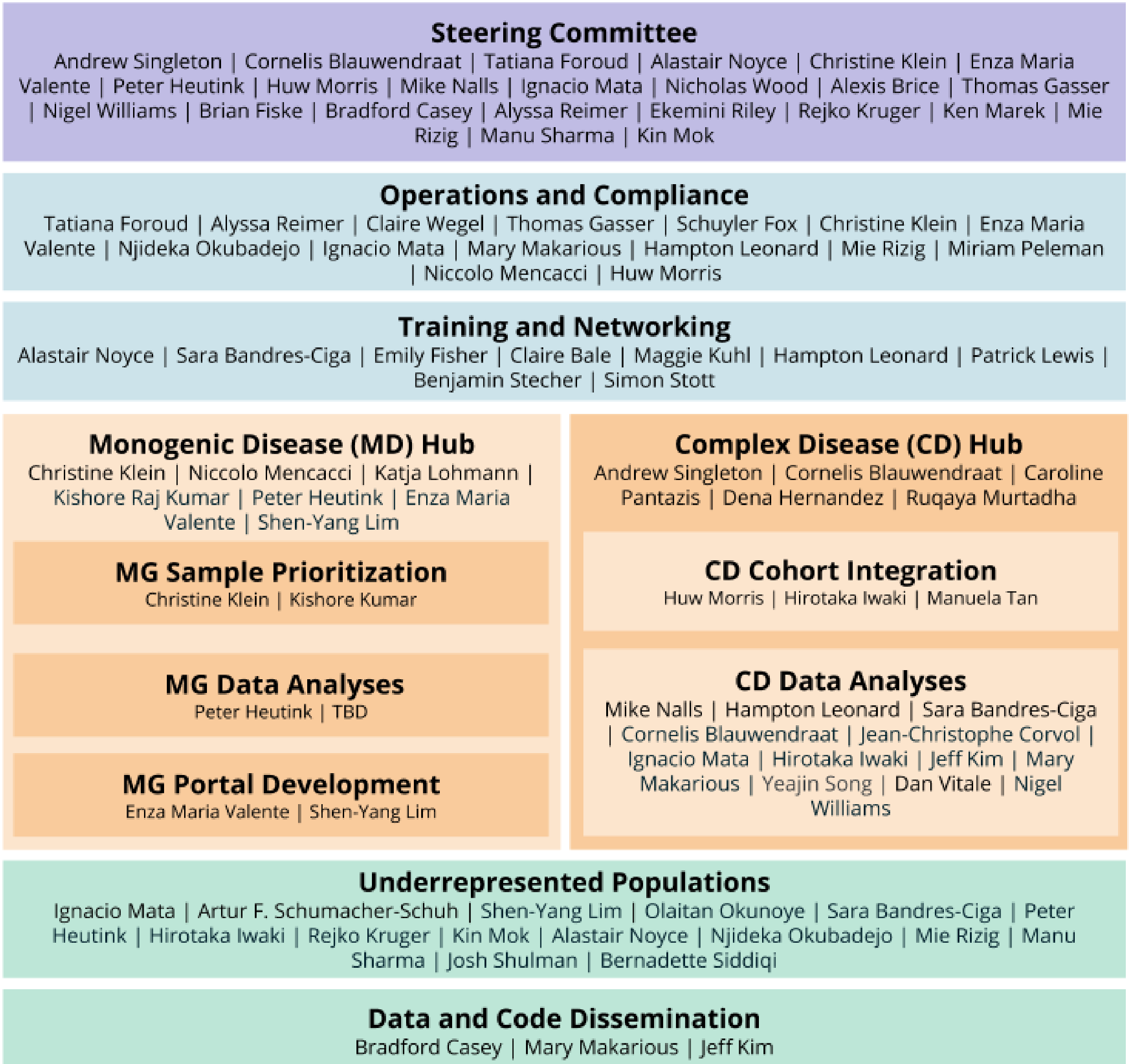


Figure 2: Organizational structure of GP2



Steering committee: This group oversees the general operation of GP2 and collectively crafts the strategies for achieving our mission and maintaining our core principles.

Operations and compliance: creates and manages the operational workflow for GP2 in addition to ensuring compliance with national and international regulations.

Monogenic hub: designs and manages the collection and prioritization of samples from patients with putative monogenic disease, organizes whole genome sequencing, and performs data analysis.

Sample prioritization: This group will use a scoring system to prioritize cases where a monogenic cause is strongly suspected, with the goal of increasing the likelihood of discovering new PD genes. Additionally, we hope to gather samples from across the world and include patients from underrepresented populations.

Data analyses: To execute and coordinate core analyses to identify the genetic basis of monogenic forms of PD

Portal development: To develop and manage a user-friendly web-based interface through which researchers can contribute data and samples, and obtain results of their patients, as well as aggregated data from other GP2 participating centers.

Complex hub
This group ensures that the complex disease effort operates efficiently, meeting the data generation and analysis milestones laid out in the proposal

Cohort integration: This group liaises with clinical/genetic cohorts to engage them in the overall project, define common clinical datasets, harmonize data for phenotype analysis and help to ensure collaborative clinical-genetic research with investigators working on clinical cohorts around the world.

Data analyses: To design and execute core analyses to identify the basis of genetically complex forms of PD

Underrepresented populations
This working group has the goal of increasing representation and deciphering the genetic factors that are associated with PD in non-European ancestry populations.

Data and Code Dissemination
To enable open science by sharing meaningful data, analytical code, and results, balancing participant privacy, regulations, and data security

Training and Networking
To establish a virtual center-of-excellence to promote training, networking and communication throughout the GP2 project.

Appendix

Global Parkinson's Genetics Program:

Name	Institute
Ai Huey Tan	Division of Neurology, Department of Medicine; and The Mah Pooi Soo & Tan Chin Nam Centre for Parkinson's & Related Disorders, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
Alastair Noyce	Preventive Neurology Unit, Wolfson Institute of Preventive Medicine, Barts and the London School of Medicine and Dentistry, Queen Mary University, London, United Kingdom
Alejandro Martinez Carrasco	Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, London, UK

Alexis Brice	Institute for Brain and Spinal Cord, Paris, France
Alyssa Reimer	The Michael J. Fox Foundation for Parkinson’s Research, New York, NY, USA
Anastasia Illarionova	German Center for Neurodegenerative Diseases (DZNE), Tuebingen, Germany
Andrew Singleton	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
Artur Schumacher-Schuh	Universidade Federal do Rio Grande do Sul, Hospital de Clínicas de Porto Alegre, Porto Alegre, Brazil
Benjamin Stecher	Independent patient advocate and consultant
Bernadette Siddiqi	The Michael J. Fox Foundation for Parkinson’s Research, New York, NY, USA
Bradford Casey	The Michael J. Fox Foundation for Parkinson’s Research, New York, NY, USA
Brian Fiske	The Michael J. Fox Foundation for Parkinson’s Research, New York, NY, USA
Caroline Pantazis	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
Christine Klein	Institute of Neurogenetics, University of Luebeck, Luebeck
Claire Bale	Parkinsons UK, London, UK
Claire Wegel	Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA
Cornelis Blauwendraat	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
Dan Vitale	Data Tecnica International, Glen Echo, MD, USA
Dena Hernandez	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
Ekemini Riley	Aligning Science Across Parkinson's, Chevy Chase, MD, USA
Emily Fisher	Preventive Neurology Unit, Wolfson Institute of Preventive Medicine, Barts and the London School of Medicine and Dentistry, Queen Mary University, London, United Kingdom
Enza Maria Valente	Departmento of Molecular Medicine, University of Pavia and IRCCS Mondino Foundation, Pavia, Italy
Eva Juliane Vollstedt	Institute of Neurogenetics, University of Luebeck, Luebeck
Hampton Leonard	Data Tecnica International, Glen Echo, MD, USA
Hirotaka Iwaki	Data Tecnica International, Glen Echo, MD, USA
Huw Morris	Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, London, UK
Ignacio Juan Keller Sarmiento	Northwestern University, Feinberg School of Medicine, Department of Neurology, Chicago, IL, USA
Ignacio Mata	Genomic Medicine Institute, Cleveland Clinic & Molecular Medicine, Cleveland Clinic Lerner College of Medicine of Case Western Reserve University

Jeff Kim	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
John Hardy	Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, London, UK
Kaileigh Murphy	The Michael J. Fox Foundation for Parkinson’s Research, New York, NY, USA
Katja Lohmann	Institute of Neurogenetics, University of Luebeck, Luebeck
Ken Marek	The Michael J. Fox Foundation for Parkinson’s Research, New York, NY, USA
Kin Mok	UK Dementia Research Institute at UCL and Department of Neurodegenerative Disease, UCL Institute of Neurology, University College London, London, UK
Kishore Kumar	Molecular Medicine Laboratory, Concord Hospital, University of Sydney, Garvan Institute, Australia
Lara Lange	Institute of Neurogenetics, University of Luebeck, Luebeck
Maggie Kuhl	The Michael J. Fox Foundation for Parkinson’s Research, New York, NY, USA
Manu Sharma	Centre for Genetic Epidemiology, Institute for Clinical Epidemiology and Applied Biometry, University of Tübingen, Tübingen, Germany
Manuela Tan	Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, London, UK
Mary Makarious	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
Michelle Durborow	The Michael J. Fox Foundation for Parkinson’s Research, New York, NY, USA
Micol Avenali	Department of Brain and Behavioural Sciences, University of Pavia and IRCCS Mondino Foundation, Pavia, Italy
Mie Rizig	Department of Molecular Neuroscience, UCL Institute of Neurology and National Hospital for Neurology and Neurosurgery, London, UK
Mike Nalls	Data Tecnica International, Glen Echo, MD, USA
Niccolo Mencacci	Northwestern University, Feinberg School of Medicine, Department of Neurology, Chicago, IL, USA
Nicholas Wood	Department of Clinical and Movement Neurosciences, UCL Queen Square Institute of Neurology, London, UK
Nigel Williams	MRC Centre for Neuropsychiatric Genetics and Genomics, Institute of Psychological Medicine and Clinical Neurosciences, Cardiff University, Cardiff, UK
Njideka Okubadejo	Neurology Unit, Department of Medicine, College of Medicine, University of Lagos, Idi Araba, Lagos State
Olaitan Okunoye	Department of Movement and Clinical Neurosciences, Queen Square Institute of Neurology, University College London.
Patrick Lewis	Comparative Biomedical Sciences, Royal Veterinary College, London, UK
Peter Heutink	German Center for Neurodegenerative Diseases (DZNE), Tuebingen, Germany

Rejko Kruger	Luxembourg Centre for Systems Biomedicine (LCSB), University of Luxembourg, Esch-sur-Alzette, Luxembourg
Ruqaya Murtadha	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
Sabina Adams	Preventive Neurology Unit, Wolfson Institute of Preventive Medicine, Barts and the London School of Medicine and Dentistry, Queen Mary University, London, United Kingdom
Sara Bandres-Ciga	Laboratory of Neurogenetics, National Institute on Aging, National Institutes of Health, Bethesda, MD, USA
Schuyler Fox	The Michael J. Fox Foundation for Parkinson’s Research, New York, NY, USA
Shen-Yang Lim	Division of Neurology, Department of Medicine; and The Mah Pooi Soo & Tan Chin Nam Centre for Parkinson's & Related Disorders, Faculty of Medicine, University of Malaya, Kuala Lumpur, Malaysia
Simon Stott	Cure Parkinsons Trust, London, UK
Soraya Bardien	Division of Molecular Biology and Human Genetics, Faculty of Medicine and Health Sciences, Stellenbosch University, Cape Town
Tatiana Foroud	Department of Medical and Molecular Genetics, Indiana University School of Medicine, Indianapolis, IN, USA
Thomas Gasser	Hertie Institute for Clinical Brain Research, University of Tübingen, Tübingen, Germany
Todd Sherer	The Michael J. Fox Foundation for Parkinson’s Research, New York, NY, USA
Yeajin Song	Data Tecnica International, Glen Echo, MD, USA