PARKINSON'S PROGRESSION MARKERS INITIATIVE

Play a Part in Parkinson's Research

Genetic principal components and polygenic risk score for Parkinson's disease derived from Whole genome Sequencing data in AMP-PD

Hirotaka Iwaki, Andrew Singleton
Center 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

PPMI Project ID: 9001

Summary

Here, we provide the genetic principal components (PCs), population inference, and the polygenic risk scores (PRS) from PPMI participants. These values were calculated from the AMP-PD whole-genome sequencing data (release version 3) at AMP-PD. There were 10,418 AMP-PD participants in total, including 1,807 PPMI participants (1319 after excluding PPMI Genetic Registry participants). WGS data for PPMI included in AMP-PD was generated in **PPMI Project 118**, and the PRS scores here are derived from this data. These PRS scores from are provided to the PPMI research community for use and inclusion in ongoing analysis of the study.

Use Note: Project 9001 is provided for PPMI users' convenience. Investigators should carefully review the information provided on PRS derivation and confirm that the analytical methods applied meet their research objectives. Further, investigators interested in combining Project 9001 with other data from PPMI or other studies should review the guidance and code provided within the GitHub link to ensure that this methodology is appropriate for their use and the available data.

Background and notes on usage

Polygenic risk scores (PRS) are a tool for understanding the genetic basis of complex traits and diseases. PRS are calculated by combining the effects of multiple genetic variants associated with disease. These genetic variants are typically identified through genome-wide association studies (GWAS), and the effects of each genetic variant are weighted based on their strength of association with the trait or disease. The sum of these weighted effects is the PRS.

- Participant PRS scores are calculated using a weighted sum of the magnitudes of associations of genetic variants. The weights are based on the strength of association of each genetic variant with PD.
- PRS scores reflect limitations of the data used to generate them, including age-dependent sampling, platform limitations, potential bias for genetic risk factors used in recruitment, and limitations to ancestral diversity.
- PRS may not reflect all of the genetic variants that contribute to disease.
- 488 PPMI Genetic Registry participants are excluded from the PRS data table, as data from these participants is not available via LONI

PPMI data within AMP-PD

This analysis of genetic risk utilized the genomic data from multiple studies as represented within AMP-PD. At time of analysis, AMP-PD was utilizing data release version 3 (December 2022)

Methods

Please refer to https://github.com/GP2code/amp-pd-v3-pc-prs for the complete code used in processing of the AMP-PD data for PRS derivation, and Nalls et al., 2019 for the original methods used in this analysis. A detailed overview can be found in the supplemental methods of that publication.

In summary, the genetic principal components (PCs) were calculated after merging with the reference panel from the 1000 GENOME project and the population (EUR, AFR, EAS, and OTHER) were assigned based on the PCs with the threshold of mean +/- 6SD of PC1-5. The PRS (PRS88) is generated from the 90 risk SNPs identified in the recent GWAS study (Nalls et al., 2019, PMID: 31701892) but two SNPs were missing because of the low quality associated with these probes within the array.

PRSp90 is the version substituting these missing SNPs with proxy SNPs to address this. This version is recommended for the majority of users. Specifically, rs11578699 was substituted by rs11577197 (perfect LD) and rs3742785 was substituted by rs10134885 (R2=0.989). PRS were also calculated with/without SNPs in the GBA and LRRK2 loci,. The PPMI participants were subset, and the PCs and the PRSs were re-standardized (mean=0, SD=1) using only PPMI participants for ease of use.

When the PRSs are used in analysis, Authors recommend that the PCs (at least PC1, PC2 and PC3) are also included in the model to adjust for population stratification. Please note that the PRS weights were derived from the European GWAS results. Therefore, the PRSs may be not easily generalizable to non-European populations. The inferred population is derived from the WGS data to aid researchers in evaluating the cohort based on the calculated PRS. Care should be taken when evaluating these results for non-European participants.

PRS Data Structure

The PRS data has been prepared as a table, where each row corresponds to a PPMI participant ID (PATNO), and each column represents the corresponding value for each variable outlined below. To use the PRS data in analysis, the relevant principle components and PRS score field can be used to assess cumulative risk with or without the impact of annotated pathogenic variants in the known familial risk genes (GBA1 and LRRK2) as described above. For the PCs and PRS variables, these data are derived values and do not have associated units. Refer to the GitHub, publication, and overview provided in the 'Methods' section above for specific calculations used to derive the variables described in the below table.

The se data are available through the LONI IDA for use alongside data on other variables of interest in analyzing PPMI data across the study.

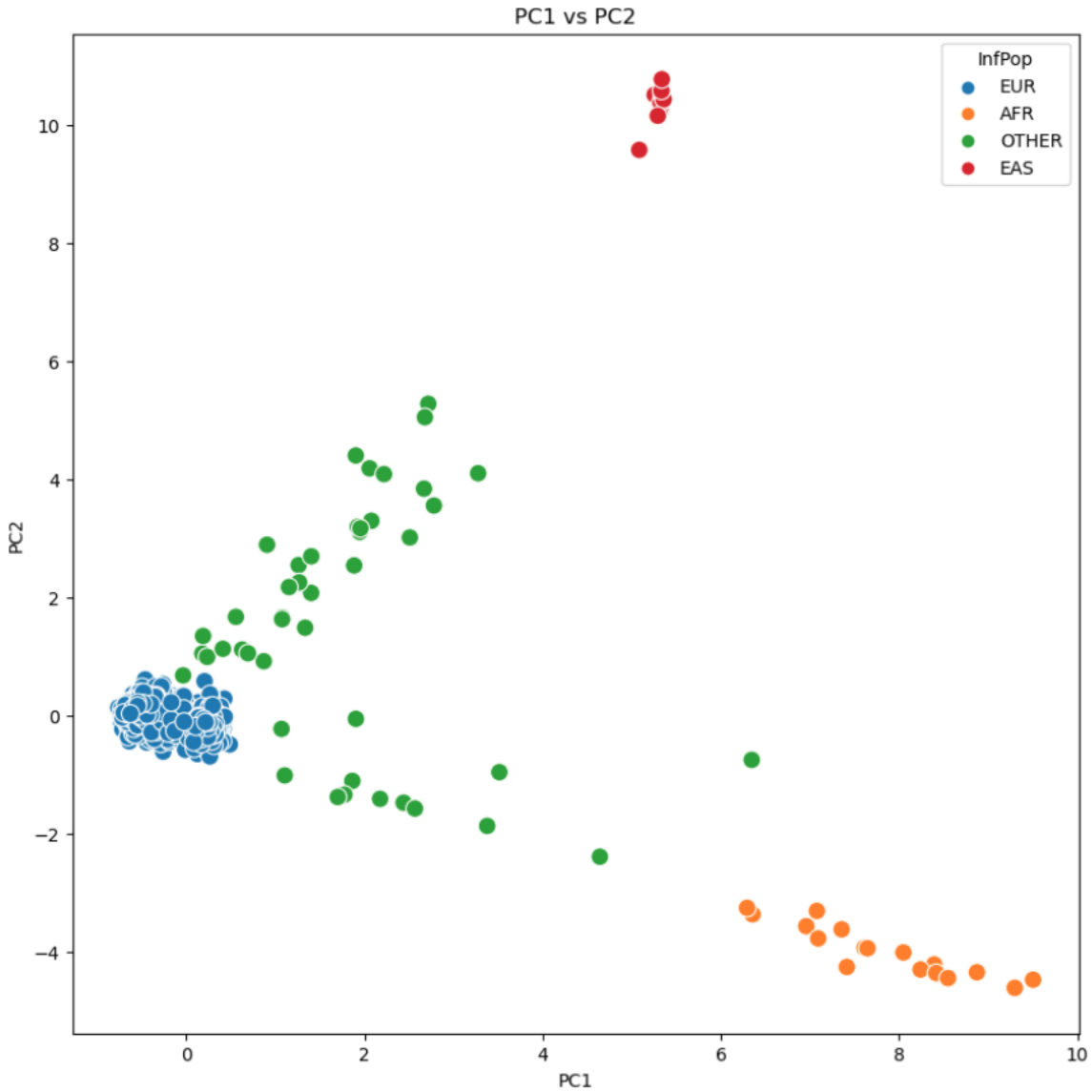
Variable Name	Data Type	Variable Description	Value	Value Description
PATNO	Numeric	Participant identifier	-	-
EVENT_ID	Text	Event identifier	SC	Screening visit[1]
Genetic_PRS_InfPop	Text	[Derived] Inferred ancestry	EUR AFR EAS OTHER	European African East Asian Not EUR, AFR or EAS
Genetic_PRS_PRS88	Numeric	[Derived] PRS composed of 88 PD risk associated SNPs (Original PD-GWAS article has 90 SNPs but 2 were missing)	-	-

[1] Since genetic data is considered stable over time, for purposes of this analysis, all results have been assigned a visit ID value of "SC" under EVENT_ID, to represent the earliest possible visit date. Users should be aware that the participants' biosample used under associated WGS Project 118 may have been taken at a visit other than SC

Genetic_PRS_PRS83	Numeric	[Derived] PRS composed of 83 PD risk associated SNPs (GBA and LRRK2 loci were excluded from the above 88 SNPs)	-	-
Genetic_PRS_PRSp90	Numeric	[Derived] PRS composed of 90 PD risk associated SNPs (The 2 missing SNPs were substituted by highly correlated SNPs in European population)	-	-
Genetic_PRS_PRSp88	Numeric	[Derived] PRS composed of 88 PD risk associated SNPs (LRRK2 locus were excluded)	-	-
Genetic_PRS_PRSp87	Numeric	[Derived] PRS composed of 87 PD risk associated SNPs (GBA locus were excluded)	-	-
Genetic_PRS_PRSp85	Numeric	[Derived] PRS composed of 85 PD risk associated SNPs (GBA and LRRK2 locus were excluded)	-	-

Genetic_PRS_PC1	Numeric	[Derived] Genetic principle components 1	-	-
Genetic_PRS_PC2	Numeric	[Derived] Genetic principle components 2	-	-
Genetic_PRS_PC3	Numeric	[Derived] Genetic principle components 3	-	-
Genetic_PRS_PC4	Numeric	[Derived] Genetic principle components 4	-	-
Genetic_PRS_PC5	Numeric	[Derived] Genetic principle components 5	-	-
Genetic_PRS_PC6	Numeric	[Derived] Genetic principle components 6	-	-
Genetic_PRS_PC7	Numeric	[Derived] Genetic principle components 7	-	-
Genetic_PRS_PC8	Numeric	[Derived] Genetic principle components 8	-	-
Genetic_PRS_PC9	Numeric	[Derived] Genetic principle components 9	-	-

Genetic_PRS_PC10	Numeric	[Derived] Genetic principle components 10	-	-



Scatter plots for genetic PCs colored by inferred ancestries. Colors indicate inferred ancestry based on available WGS data for each PPMI participant

PARKINSON'S PROGRESSION MARKERS INITIATIVE

Play a Part in Parkinson's Research

References

Nalls, Mike A., Cornelis Blauwendraat, Costanza L. Vallerga, Karl Heilbron, Sara Bandres-Ciga, Diana Chang, Manuela Tan, et al. 2019. "Identification of Novel Risk Loci, Causal Insights, and Heritable Risk for Parkinson's Disease: A Meta-Analysis of Genome-Wide Association Studies." Lancet Neurology 18 (12): 1091–1102.

About the Authors

This document was prepared by Hirotaka Iwaki, who is affiliated with the Center for Alzheimer's and Related Dementias (CARD), National Institute on Aging and National Institute of Neurological Disorders and Stroke, National Institutes of Health. For additional information, please reach out via email at hirotaka.iwaki@nih.gov.