Supplementary methods

Courage-PD international consortium

The COURAGE-PD (COmprehensive Unbiased Risk factor Assessment for Genetics and Environment in Parkinson‘s Disease) is an international consortium comprising of individual-level data from 35 cohorts on Parkinson’s disease (PD) from different populations worldwide. The GEo-PD (Genetic Epidemiology of Parkinson’s disease; https://geopd.biomedinfo.org/) consortium represents one of the major components of COURAGE-PD, which also aims at conducting collaborative studies on genetic susceptibility in PD. In addition, several other studies from Europe contributed to COURAGE-PD. One of the main features of COURAGE-PD is that participating sites are distributed in the five continents, therefore representing a highly diverse population, as shown in **eTable 1**. All studies were approved by local ethical committees following the procedures of each country, and material transfer agreements were set up between participating sites and the University of Tübingen (Germany).

Phenotyping

All the study sites collectively comprised of 27,538 subjects. PD was diagnosed according to standard criteria (United Kingdom Parkinson’s Disease Society Brain Bank - UKPDSBB, Gelb, Bower).[1-3] Phenotypic data on disease status, self-repored gender and ethnicity, age at the enrolment in the study, age at the onset of PD, age at diagnosis of PD, duration of PD, presence of *GBA* or *LRKK2* mutations in European-ancestry individuals, and family history of PD were collected. We could not find case-control status or gender status on 245 subjects, which were subsequently excluded from the study.

Genotyping

According to the study’s consortium agreement, participating sites contributed DNA and demographic/environmental data. DNAs (25µl of DNA at a concentration of 50 to 100 ng/µl) were shipped for quality control to the University of Tübingen (Germany) and genotyped in a

central laboratory in Munich (Institute of Human Genetics, Helmholtz Zentrum, Germany). The samples from two sites (Gasser, Morris/Wood) were genotyped at the Laboratory of Neurogenetics (National Institute on Aging, National Institutes of Health, Bethesda, USA). Demographic data were harmonized and collected using a standardized form and cleaned at Inserm U1018 (Villejuif, France). The COURAGE-PD consortium used the same array, Neurochip, to genotype its participants.[4] Briefly, Neurochip is a custom-designed array containing a tagging variant backbone with excellent genome-wide coverage of about 0.307 million variants, including a manually curated custom content comprised of 0.179 million variants implicated in diverse neurological diseases, including PD.

Pre-imputation Quality control

The genotyped dataset from the 35 study cohorts comprised 27,293 subjects. Genotypes were called using Illumina GenomeStudio (GenomeStudio v2011.1 with Genotyping module v1.9.4). After the exclusion of X and Y chromosomal variants, up to 0.485 million variants remained across all the cohorts. Genotyped data was exported from Genome Studio to PLINK format for subsequent downstream quality control (QC) procedures, implemented using an automated pipeline developed at the University of Tübingen. The pipeline uses PLINK (PLINK v1.07, https://www.cog-genomics.org/plink), R software (R version 3.6.2, R Foundation for Statistical Computing, Vienna, Austria), EIGEN software (EIGENSTRAT v4.2) and terminal-based commands on a Linux platform. Briefly, Individual and SNP level QC were implemented independently for each study cohort. Individuals with elevated missing genotyping rates (≥4%) or outlying heterozygosity (Mean ± 4SD), individuals with discordant sex information based on comparison of the homozygosity rate for X-linked SNPs for each individual to the expected rate, all except one of the duplicated or related individuals based on identity by descent (IBD)>0.185, and ancestral outliers based on comparison of first two principal components after merging with HapMap populations were excluded. Individual-level QC was followed by SNP level QC. SNPs with a high missing rate (>10%), a significant difference in the rate of missingness between PD cases and controls ($P<10^{-5}$), minor allele frequency (MAF) $<5\times10^{-8}$ and Hardy-Weinberg disequilibrium (HWE) $P<5\times10^{-8}$ were

excluded. Individual-level QC resulted in the exclusion of 2434 subjects, leaving 24,859 subjects comprising 14,380 PD cases and 10479 controls. SNP level QC resulted in the exclusion of on an average 0.158 million variants (32.5%) across all the study cohorts were excluded, leaving an average of 0.328 million variants.

Imputation

After QC, we used the HRC/1000G imputation preparation and checking tool (https://www.well.ox.ac.uk/~wrayner/tools/HRC-1000G-check-bim-v4.3.0.zip) to check for Ref/Alt allele assignments, incorrect strands, incorrect chromosomal positions, deviation from allele frequency, and conflicting palindromic SNPs. This was followed by imputation of autosomal variants, which was performed separately for each study cohort based on 0.271 to 0.374 million SNPs, through the Michigan Imputation server using the HRC reference panel and the GRCh37/hg19 assembly with an R^2 filter of 0.3. After imputation, the average number of SNPs available among European-ancestry and Asian-ancestry based studies were 9.38 million and 3.70 million, respectively.

Post-imputation quality control

Each study dataset was used for generating four principal components believed to explain the maximum proportion of variability in population substructure. Only PD cases were carried forward for analysis in the present study. We further excluded five study cohorts due to the following reasons: i) missing age at onset data (n=1), ii) overlap of study subjects with those present in IPDGC cohort (n=3), iii) a low number of PD cases, i.e. <25 (n=1). We also excluded some of the subjects from specific study cohorts due to the following reasons: i) missing age at onset data (n=48), ii) overlap of study subject with that present in IPDGC cohort (n=1237), iii) *GBA* or *LRRK2* mutant carriers or positive family history of PD (n=1285). In summary, a total of 8535 samples from 30 study cohorts was carried forward for the genotype-phenotype association analysis.

References

1. Gelb DJ, Oliver E, Gilman S. Diagnostic criteria for Parkinson disease. Arch Neurol 1999;56(1):33-9.

2. Gibb WR, Lees AJ. The relevance of the Lewy body to the pathogenesis of idiopathic Parkinson's disease. J Neurol Neurosurg Psychiatry 1988;51(6):745-52.

3. Bower JH, Maraganore DM, McDonnell SK, et al. Incidence and distribution of parkinsonism in Olmsted County, Minnesota, 1976-1990. Neurology 1999;52(6):1214-20.

4. Blauwendraat C, Faghri F, Pihlstrom L, et al. NeuroChip, an updated version of the NeuroX genotyping platform to rapidly screen for variants associated with neurological diseases. Neurobiol Aging 2017;57:247.e9-47.e13.