nature portfolio

Corresponding author(s): John E. Landers, Ph.D.

Last updated by author(s): Apr 16, 2024

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	All raw sequencing data was aligned to the GRCh38 reference genome using BWA-mem (v.2.2.1) according to the pipeline described by Regier et al., Nat. Commun. (2018) (implementation can be found at https://github.com/maarten-k/realignment and Zenodo: https://doi.org/10.5281/zenodo.10963076). Joint genotyping was performed using a uniform pipeline according to the GATK best practices (v. 4.2.6.1).
Data analysis	Handling and filtering of VCF files was performed using VCFtools (v. 0.1.16), BCFtools (v. 1.9) and PLINK (v. 1.90b6.21). Ancestry was estimated using LASER (v.2.04). Variants were annotated using Ensembl (GRCh38.105), snpEff (v.5.1d) and dbNSFP (v. 4.3a). Sample and variant quality control was performed using PLINK (v. 1.90b6.21) and RVAT (v. 0.2.0), while sample relatedness was inferred using KING (v. 2.2.7). Haplotypes were inferred using Beagle (v. 5.4). All downstream analyses were performed using custom R code (performed in R 3.6.3) that we made available in the RVAT R package (v. 0.2.0) (available on GitHub:https://github.com/kennalab/rvat and Zenodo: https://doi.org/10.5281/zenodo.10973472). Other R packages used either as dependencies of RVAT or in other analyses and visualizations are ggplot2 (v. 3.4.2), ggrepel (v. 0.9.1), dplyr (v. 1.0.7), readr (v. 2.1.1), stringr (v. 1.4.0), tidyr (v. 1.1.4), magrittr (v. 2.0.1), kinship2 (v. 1.9.6), logistf (v. 1.25.0), SKAT (v. 2.2.5), SummarizedExperiment (v. 1.16.1), S4Vectors (v. 0.24.4), GenomicRanges (v. 1.38.0), IRanges (v. 2.20.2), DBI(v. 1.1.3) and RSQLite (v. 2.3.1). Sequence analysis for Sanger sequencing was performed using the v.29.0 PHRED/PHRAP/Consed software suite (http://www.phrap.org/) and variations in the sequences were identified with the Polyphred v6.15 software (http://droog.gs.washington.edu/polyphred/).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Exome sequencing data from the case cohort is available under dbGaP study phs001004 or are available via the corresponding author subject on the data sharing terms of the respective cohorts. Project MinE data is available upon request (https://www.projectmine.com/research/data-sharing/). dbGAP datasets used are available under the following accession numbers: Alzheimer's Disease Sequencing Project (ADSP) (phs000572); Autism Sequencing Consortium (ASC) (phs000298); Sweden-Schizophrenia Population-Based Case-Control Exome Sequencing (phs000473); Inflammatory Bowel Disease Exome Sequencing Study (phs001076); Myocardial Infarction Genetics Exome Sequencing Consortium: Ottawa Heart Study (phs000806); Myocardial Infarction Genetics Exome Sequencing Consortium: Malmo Diet and Cancer Study (phs001101); Myocardial Infarction Genetics Exome Sequencing Consortium: U. of Leicester (phs001000); Myocardial Infarction Genetics Exome Sequencing Consortium: Italian Atherosclerosis Thrombosis and Vascular Biology (phs000814); NHLBI GO-ESP: Women's Health Initiative Exome Sequencing Project (WHI) - WHISP (phs000281); Building on GWAS for NHLBI diseases: The US CHARGE Consortium (CHARGE-S): CHS (phs000667); Building on GWAS for NHLBI Diseases: the US CHARGE Consortium (CHARGE-S): ARIC (phs000668); Building on GWAS for NHLBI diseases: the US CHARGE consortium (CHARGE-S): FHS (phs000651); NHLBI GO-ESP Family Studies: Idiopathic Bronchiectasis (phs000518); NHLBI GO-ESP: Family Studies (Hematological Cancers) (phs000632); NHLBI GO-ESP: Family Studies: (familial atrial fibrillation) (phs000362); NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (ARIC) (phs000398); NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (CHS) (phs000400); NHLBI GO-ESP: Heart Cohorts Exome Sequencing Project (FHS) (phs000401); NHLBI GO-ESP: Lung Cohorts Exome Sequencing Project (asthma) (phs000422); NHLBI GO-ESP: Lung Cohorts Exome Sequencing Project (COPDGene) (phs000296); GO-ESP: Family Studies (Thoracic aortic aneurysms leading to acute aortic dissections) (phs000347). NHLBI TOPMed: Genomic Activities such as Whole Genome Sequencing and Related Phenotypes in the Framingham Heart Study (phs000974); NHLBI TOPMed: Genetics of Cardiometabolic Health in the Amish (phs000956); NHLBI TOPMed: Genetic Epidemiology of COPD (COPDGene) (phs000951); NHLBI TOPMed: The Vanderbilt Atrial Fibrillation Registry (VU_AF) (phs001032); NHLBI TOPMed: Cleveland Clinic Atrial Fibrillation (CCAF) Study (phs001189); NHLBI TOPMed: Partners HealthCare Biobank (phs001024); NHLBI TOPMed - NHGRI CCDG: Massachusetts General Hospital (MGH) Atrial Fibrillation Study (phs001062); NHLBI TOPMed: Novel Risk Factors for the Development of Atrial Fibrillation in Women (phs001040); NHLBI TOPMed - NHGRI CCDG: The Vanderbilt AF Ablation Registry (phs000997); NHLBI TOPMed: Heart and Vascular Health Study (HVH) (phs000993); NHLBI TOPMed - NHGRI CCDG: Atherosclerosis Risk in Communities (ARIC) (phs001211); NHLBI TOPMed: The Genetics and Epidemiology of Asthma in Barbados (phs001143); NHLBI TOPMed: Women's Health Initiative (WHI) (phs001237); NHLBI TOPMed: Whole Genome Sequencing of Venous Thromboembolism (WGS of VTE) (phs001402); NHLBI TOPMed: Trans-Omics for Precision Medicine (TOPMed) Whole Genome Sequencing Project: Cardiovascular Health Study (phs001368).

Additional data used includes: GATK hg38 resource bundle (https://console.cloud.google.com/storage/browser/genomics-public-data/resources/broad/hg38/v0/), BWA-mem GRCh38 reference genome (run "bwa.kit/run-gen-ref hs38DH" from https://github.com/lh3/bwa/tree/master/bwakit), 1000genomes phase 3 (ftp://ftp-trace.ncbi.nih.gov/1000genomes/ftp/release/20110521), Ensembl v.105 (https://ftp.ensembl.org/pub/release-105/gtf/homo_sapiens/Homo_sapiens.GRCh38.105.gtf.gz), gnomAD v.2.1.1 (non-neuro) (https://gnomad.broadinstitute.org/variant/6-146865220-C-G?dataset=gnomad_r2_1_non_neuro), AMP-PD (https://amp-pd.org/).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Sex was included as covariate in the single variant and gene-based analyses. Sex was genetically inferred and confirmed when self-reported sex was available. No gender-based analyses were performed in this manuscript.
Reporting on race, ethnicity, or other socially relevant groupings	No categorizations were made based on socially constructed variables such as race or ethnicity. Ancestry was inferred using the LASER software (Extended Data Fig. 2). We retained individuals of predominantly European ancestry. Single variant and gene-based analyses were adjusted for ancestry by including 10 principal components derived from common variants (MAF > 0.01).
Population characteristics	The following cohorts of patients with PD and controls were included: Alzheimer's Disease Sequencing Project: 3,321 controls Autism Sequencing Consortium (ASC): 1,107 controls Sweden-Schizophrenia Population-Based Case-Control Exome Sequencing: 3,023 controls Inflammatory Bowel Disease Exome Sequencing Study: 1,893 controls Women's Health Initiative Exome Sequencing Project (WHI): 731 controls Myocardial Infarction Genetics Exome Sequencing Consortium (MIGen): 6,416 controls The US CHARGE Consortium (CHARGE-S): 1,741 controls NHLBI Trans-Omics for Precision Medicine (TOPMed): 7,323 controls UK Biobank: 49,981 controls NYGC ALS Consortium: 342 controls Project MinE: 2,805 controls PROGENI: 1,600 patients Parkinson's Institute: 441 patients Mayo clinic: 783 patients The following characteristics were employed as outcome and covariates in the analysis cohort (2,184 cases, 69,775 controls): - PD diagnosis was determined in the clinic as described in the Methods section. - Inferred genetic sex: 35,877 females (880 cases, 34,997 controls) and 36,082 males (1,304 cases, 34,778 controls). - Genetic principal components: Inferred from 41,421 common markers (MAF > 0.01) using plink2.

	- Total rare exome-wide synonymous variant count
Recruitment	1600 familial PD cases were from the PROGENI study, which recruited multiplex PD families primarily through a living affected sibling pair. Subsequently, ascertainment was loosened to include PD probands having a positive family history of PD in a first degree relative, who was not required to be part of the study. All individuals completed an in-person evaluation that included Parts II and III of the Unified Parkinson Disease Rating Scale (UPDRS; Lang and Fahn 1989) and a diagnostic checklist that implemented the UK PD Brain Bank inclusion and exclusion criteria. Responses on the diagnostic checklist were then used to classify study subjects as having verified PD or non-verified PD. 783 familial PD cases were recruited at the Mayo clinic, patients were included who fulfilled the clinical diagnostic criteria of PD. 441 familial PD cases were from the Parkinson Institute Biobank, diagnosis of PD was made according to the UK Brain Bank criteria.
Ethics oversight	We received approval for this study from the institutional review boards of the participating centers, and written informed consent was obtained from all patients (consent for research).

Note that full information on the approval of the study protocol must also be provided in the manuscript.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	The full dataset the included 2,824 patients with PD and 78,683 control subjects. After quality control, a total of 2,184 index familial PD cases (one affected individual per family) and 69,775 controls were included in the analysis cohort. All QC-passing, unrelated indivuals were included with the aim of maximizing statistical power for rare variant discovery. As such, the exact sample size was not predetermined. We show that this dataset is well-powered for the detection of known (ultra-)rare PD variants in LRRK2 and GBA.
Data exclusions	The exclusion of samples and variants is detailed in the Methods section. In brief, low-quality samples and variants were excluded. In addition, related samples were excluded, retaining one affected individual per family in the analysis cohort. Analyses were restricted to low-frequency variants (MAF < 0.05) that were predicted to be non-synonymous.
Replication	We did not perform a replication analysis of the genetic analyses in a separate cohort. We do, however, provide multiple lines of evidence that corroborate the main finding of this study: - The RAB32 S71R variant was confirmed in all PD cases using a different technique (Sanger sequencing). - The RAB32 S71R variant segregated within three families. - We confirmed that the RAB32 S71R variant is very rare in the general population based on 44,329 non-neurological European individuals in gnomAD (v2.1.1) - Functional analyses showed that RAB32 S71R increases LRRK2 kinase activity, which is a key pathological mechanism in PD. All biological replicates (n=3) of the functional experiments are presented in the manuscript.
Randomization	Not applicable, as this is an observational case-control study. All sequence data was uniformly processed (including alignment to the GRCh38 reference genome and joint variant calling) to minimize technical variation across cohorts. Subsequently, strict QC was applied at both the sample and variant level to minimize technical variation, outliers and batch effects as detailed in the Methods section. Finally, analyses were adjusted for sex, ten principal components and total synonymous variant counts to account for potential confounding due to sex, population stratification and technical artifacts.
Blinding	Since this is a case-control study, blinding is not applicable. Sample ascertainment, genotyping, phenotyping and analysis were all performed independently.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☐	☒ Antibodies
☐	☒ Eukaryotic cell lines
☒	☐ Palaeontology and archaeology
☒	☐ Animals and other organisms
☒	☐ Clinical data
☒	☐ Dual use research of concern
☒	☐ Plants

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Antibodies

Antibodies used: Antibodies used in the study were used as follows: Anti-HA, DSHB Cat#: anti-HA rRb-IgG; Anti-Myc, clone 9E10, Biolegend Cat#: 626802; Anti-Phos-LRRK2 S1292, clone MJFR-19-7-8, Abcam Cat#: 203181.

Validation: The Myc and HA antibodies are species independent and do not require validation for that aspect of their use. Regarding the application use, the manufacturer's website for each antibody provides data showing validation for both immunoprecipitation and western blotting using these antibodies. Regarding the phospho-LRRK2 antibody (S1292), the manufacturer's website states that it was validated in WB and tested in human samples.

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s): HEK293 cells were acquired from ATCC

Authentication: The cell lines used were authenticated by morphological and growth characteristics

Mycoplasma contamination: None of the cell lines used were tested for mycoplasma contamination

Commonly misidentified lines (See ICLAC register): Not applicable for this study

Plants

Seed stocks: *Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.*

Novel plant genotypes: *Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.*

Authentication: *Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.*