

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 Phenotype data were collected via online surveys using the 23andMe Research platform. No additional software was used.

Data analysis The code used to run the analyses can be found at <https://github.com/23andMe/PD-microbiome>.

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](#)

Sequencing data and the study specific survey questions and responses are available through the Fox DEN tool (<https://foxden.michaeljfox.org/insight/explore/insight.jsp>). To access these data, researchers are asked to complete and e-sign a data use agreement (DUA) at <https://foxden.michaeljfox.org>.

Human research participants

Policy information about [studies involving human research participants and Sex and Gender in Research.](#)

Reporting on sex and gender

Participants self-reported biological sex. We report so statistically significant associations with any metrics assessed in the manuscript and biological sex. The percentages of self-reported biological sex for controls and cases are reported in main table 1.

Population characteristics

The mean age of the PD cohort was 61.7 years (interquartile range [IQR]: 58 - 66) with a slight male preponderance (54.9%). Mean age at diagnosis was 58.5 years (IQR: 55 - 63), and mean disease duration (time elapsed in years since self reported diagnosis) was 3.2 years (IQR: 2 - 4). The most frequently reported motor symptoms were tremor (66.2%), rigidity (40.7%), and bradykinesia (49.3%); 88.3% were treated with anti-parkinsonian medication, including dopamine precursors (70.7%), dopamine agonists (31.6%), and monoamine oxidase-B inhibitors (26.0%). The mean age of the controls was mean age 62.6 years (IQR 58.2 - 67.0)

Recruitment

Participants were 23andMe, Inc. customers, who were recruited through the 23andMe research platform. The inclusion criteria for participants with PD were: 1) a self-reported PD diagnosis; 2) disease duration < 5 years (amended to < 8 years to meet enrollment goals); and 3) current age between 50 and 69 years. Age and sex matched controls who did not report a diagnosis of PD nor family history of PD were randomly selected from 23andMe's Research cohort. The exclusion criteria for both groups included: 1) positive LRRK2 G2019S or GBA N370S carrier status; 2) antibiotic usage within the last 6 months; 3) self-reported vegetarian/vegan diet; 4) non-European genetic ancestry; and 5) mailing addresses outside the US.

Ethics oversight

The study was IRB approved (protocol No. US-IRB-26, AAHRPP-accredited IRB)

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

We targeted a roughly 2:1 case to control ratio. We recruited 470 cases and 304 controls. After QC the final sample size used for the primary analyses included 666 stool microbiome samples (445 PD cases and 221 controls) and 658 saliva microbiome samples (438 PD cases and 220 controls)

Data exclusions

The total participants on which the analyses were performed was reduced due to sequencing QC, missing age data or age outside of inclusion criteria (controls only), removing participants who had experienced gastroenteritis (after noting that their stool microbiome composition was quite different from all others), and after rarefaction of reads (in which we remove samples with total read counts fewer than 10,000).

Replication

In order to utilize the full statistical power of our sample size, we did not hold out samples for replication.

Randomization

To account for the lack of randomization between cases and controls we included specific covariates in statistical models, which were age, BMI, dessert consumption frequency, ever having smoked, flossing teeth frequency, laxative use in the past month, and number of weekly bowel movements.

Blinding

Biological samples were processed blindly, and batches were mixed with cases and controls. Blinding was not relevant to the rest of the analysis as the case vs. control status was needed to explore differences in the data.

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

Methods

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