

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
<i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i> |
| ☐ | ☒ 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
<i>Give P values as exact values whenever suitable.</i> |
| ☒ | ☐ 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	No software was used for data collection
Data analysis	The code used in this study is publicly available on GitHub at: https://github.com/Nastyagodneva/HPP_resource_code and is provided under the BSD-2-Clause license, allowing modification and reuse. For long-term accessibility and reproducibility, the code has also been deposited in Zenodo and is available under DOI: 10.5281/zenodo.14855716.

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](#)

The raw individual-level gut microbiome metagenomic data, and basic phenotypes (age, sex, and BMI) used in this study are available at the European Genome-phenome Archive (<https://ega-archive.org/>) under accession EGAS00001007204.

Individual level genetics data used in this study are available EGA at the European Genome-phenome Archive (<https://ega-archive.org/>) under accession number EGAS00001008040.

Due to privacy and ethical considerations set forth by the IRB committee of the study, all other datasets that are part of the Human Phenotype Project (HPP) - such as blood test results and anthropometrics measurements - are accessible to researchers from universities and other research institutions at: <https://humanphenotypeproject.org/data-access> using the following procedure. Interested bona fide researchers should contact info@pheno.ai who will then provide an agreement that allows accessing the data for non commercial purposes. Data access will then be provided through an account in a secure cloud based environment.

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

This study explicitly considers sex/gender as a biological and sociocultural variable where relevant. Data collection and analysis were conducted in accordance with established guidelines to ensure appropriate representation and stratification by sex/gender. Results are presented stratified by sex/gender where possible to highlight potential differences.

Reporting on race, ethnicity, or other socially relevant groupings

This study considers ethnicity as a demographic variable, determined through a combination of self-reported data and genetic information. Where feasible, results are stratified by ethnicity to highlight population-specific findings.

Population characteristics

The study cohort comprised 13,018 adult participants (46.4% male, 53.6% female), with a mean age of 51.4 years (SD 8.5). Relevant covariates included body mass index (BMI), which averaged 26.1 kg/m² (SD 4.2)

Recruitment

Volunteers are self-assigned at the trial website (<https://www.project10k.org.il/en/>), where they encounter electronic consent modules and an initial screening survey. Self-selection bias may be present in this study, as participation was voluntary and likely attracted individuals who are more health-conscious, technologically engaged, or motivated to undergo detailed phenotyping. This could result in a study population that is healthier and more socioeconomically advantaged than the general population, potentially limiting generalizability.

Ethics oversight

All participants sign an informed consent form upon arrival to the research site. All identifying details of the participants are removed prior to the computational analysis. The HPP cohort study is conducted according to the principles of the Declaration of Helsinki and was approved by the Institutional Review Board (IRB) of the Weizmann Institute of Science (IRB approval number 2392-4).

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

No formal sample size calculation was performed prior to data collection. Instead, sample size was determined based on participant availability within the larger cohort from which the data were drawn. All individuals who met the inclusion criteria and had complete data for the relevant phenotypic and clinical measurements were included in each analysis.

The final sample size of 13,018 participants provides substantial statistical power to detect meaningful associations, particularly for multivariable modeling and stratified analyses. This large cohort ensures sufficient variability across key covariates (e.g., age, BMI), enhances the robustness of findings, and supports generalizability within similar populations.

Data exclusions

Participants are screened for the following exclusion criteria based on a designated questionnaire filled online: 1. Pregnancy or currently undergoing fertility treatments. 2. More than 3 hospitalizations in the previous year. 3. Cardiovascular disease such as myocardial infarction,

congestive heart failure, valve dysfunction, cerebro-vascular accident (CVA) or transient ischemic attack (TIA). 4. Chronic neurologic or psychiatric diagnosis such as dementia, parkinson's disease and amyotrophic lateral sclerosis (ALS). 5. Chronic respiratory disease such as cystic fibrosis, interstitial lung diseases, ventilation need. 6. Chronic kidney disease such as renal tubular acidosis, urinary catheter usage or dialysis. 7. Chronic gastrointestinal disease such as inflammatory bowel disease (Crohn's disease or Ulcerative Colitis) and liver cirrhosis. 8. Chronic metabolic diseases such as diabetes or adrenal insufficiency. 9. Antibiotic usage in the last 3 months. 10. Chronic antibiotic usage. 11. Weight loss > 5 % in the recent year without deliberate diet. 12. Previous or active malignancy.

Replication	Replication was not applicable to this study as the primary objective was to perform an exploratory analysis using a large, well-characterized cohort.
Randomization	Randomization was not applicable to this study, as it was an observational, cross-sectional analysis of naturally occurring variation within a large human cohort. Participants were not assigned to experimental groups or interventions; instead, existing phenotypic and clinical data were analyzed to explore associations between variables. As such, randomization was neither required nor feasible.
Blinding	Blinding was not applicable to this study, as it was a non-interventional, observational analysis using pre-existing clinical and phenotypic data. No experimental treatment or group assignment was performed, and data analysis was conducted using objective, quantitative measures

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

Clinical data

Policy information about [clinical studies](#)

All manuscripts should comply with the ICMJE [guidelines for publication of clinical research](#) and a completed [CONSORT checklist](#) must be included with all submissions.

Clinical trial registration	Ethics approval was granted by the Institutional Review Board (IRB) of the Weizmann Institute of Science.
Study protocol	https:// www. project10k. org. il/ en
Data collection	Established in 2018. Since the study's launch, it has grown to include 27,916 participants who signed up to the study with 13,072 who already completed their first visit as of December 2024. The data collection process of the HPP cohort along with the timepoints in which they are collected is presented in Figure 1A in the manuscript
Outcomes	not relevant

Plants

Seed stocks	<i>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.</i>
Novel plant genotypes	<i>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.</i>
Authentication	<i>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, mosaicism, off-target gene editing) were examined.</i>