nature portfolio

Corresponding author(s): Lina Jin, Yanan Ma, Xinyao Zhang

Last updated by author(s): Feb 10, 2026

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	No software was involved in data collection (data used is all directly available from UK Biobank., as described in detail in the paper).
Data analysis	The following software and packages were used for data analysis: Python version 3.11.10 packages: The statsmodels package was used to perform Logistic regression models, and scikit-learn regression models, and linear regression and ordinal logistic regression. The lifelines package was used to perform COX regression models. the lightgbm package was used to perform the Light Gradient Boosting Machine (LightGBM) classifier R version 4.4.1 packages: MR analyses were conducted using the TwoSampleMR package.

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

All detailed results of metabolite-health related trait associations, and metabolite-disease outcomes associations, and metabolites-disease Clusters, and predictive power of metabolites for disease outcomes, and potential causal metabolites of diseases have been deposited through an interactive portal and are publicly No available (accessible at https://net-matebolic.vercel.app/). Access to UK Biobank data can be requested through a standard data access procedure. All requests should be directed to the following link: https://www.ukbiobank.ac.uk/enable-your-research/apply-for-access. This study was conducted using the UK Biobank under approved application number 769193.

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	We took sex into considerations in our study and our findings could apply to both man and woman. Sex (Field ID 31) in the UK Biobank was determined based on self-reporting data via questionnaire, and all included participants gave written informed consent for sharing of individual-level data.
Reporting on race, ethnicity, or other socially relevant groupings	We took ethnicity into considerations as one population characteristic and the discovery cohort was mainly based on white populations. Ethnicity (Field ID 21000) in the UK Biobank was determined based on self-reporting data via questionaire, and all included participants gave written informed consent for sharing of individual-level data. The external validation cohorts also took ethnicity into considerations.
Population characteristics	This study included 212,751 participants from the UK Biobank with Nuclear Magnetic Resonance (NMR) metabolic measures.The average age was 56.6 years, with a sex distribution of 52.6% female and 47.4% male. The vast majority of participants were of White ethnicity (90.3%)
Recruitment	This study adopted participants recruited from the UKB, a community-based cohort comprising adults aged between 40 and 69 years. Over 500,000 participants were recruited from 22 assessment centers across the UK between 2006 and 2010, and they were all registered with the UK National Health Service. This study included 212,751 participants who underwent metabolic profiling of their blood plasma samples collected during the baseline visits, The assessment visits comprised interviews and questionnaires covering lifestyles and health conditions, physical measures, biological samples, imaging, and genotype. The database is linked to national health datasets, including primary care, hospital inpatient, death, and cancer registration data.
Ethics oversight	The UK Biobank has received ethical approval from the North West Multi-Centre Research Ethics Committee (MREC, https://www.ukbiobank.ac.uk/learn-more-about-uk-biobank/about-us/ethics), and informed consent through electronic signature was obtained from study participants. This study utilized the UK Biobank Resource under application number 769193.

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 statistical methods were used to predetermine sample sizes. This study included 212,751 participants who underwent metabolic profiling of their blood plasma samples collected during the baseline visits.
Data exclusions	Participants without metabolite measurement data and metabolite outliers (defined as values ≥4 interquartile ranges from the median) were excluded.
Replication	All available UKB data were used to maximize statistical power of the analysis therefore we did not repeat the analysis.
Randomization	There is nothing in this study that pertains to randomization. We are using existing data released by UK Biobank. Uk Biobank is anobservational prospective epidemiological study, and our study use all available subjects that fulfillthe criteria described above. Hence

thereis no equivalent process of randomization that comes into this analysis (this is not a controlled randomized study).

Blinding: Blinding was not applicable to this study.

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

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.*

Magnetic resonance imaging

Experimental design

Design type: Linear regression models using phenotypes derived from T1-MRI and Diffusion MRI

Design specifications: The UK Biobank designed the imaging acquisition protocals including 6 modalities, covering structual, diffusion and functional iamging. In the current study. T1-weighted structural image provides a high-resolution depiction of brain anatomy. Diffusion MRI (voxel-wise) was used to reflect the integrity of microstructural tissue compartments.

Behavioral performance measures: Behavioral performance in the MRI scanner was not used in this study

Acquisition

Imaging type(s): T1-weighted structural imaging

Field strength: 3T

Sequence & imaging parameters:
Resolution: 1×1×1 mm
Field-of-view: 208×256×256 matrix
Duration: 5 minutes
3D MPRAGE, sagittal, in-plane acceleration iPAT=2, prescan-normalise

Area of acquisition: The whole brain were acquired.

Diffusion MRI ☒ Used ☐ Not used

Parameters:
Resolution: 2x2x2 mm
Field-of-view: 104x104x72 matrix
Duraion: 7 minutes (including 36 seconds phase-encoding reversed data)
5x b=0 (+3x b=0 blip-reversed), 50x b=1000 s/mm2, 50x b=2000 s/mm2

Gradient timings: δ=21.4 ms, Δ=45.5 ms; Spoiler b-value = 3.3 s/mm2
SE-EPI with x3 multislice acceleration, no iPAT, fat saturation
For the two diffusion-weighted shells, 50 distinct diffusion-encoding directions were acquired.

Preprocessing

Preprocessing software: T1-MRI was processed and analyzed by the UK Biobank imaging team using Freesurfer to generate macrostructural phenotypes. Diffusion MRI was used to generate microstructural phenotypes using the DTI Imaging Fitting Tool (DTFIT). The full details of the imaging processing and quality control pipeline are available in an open-access article 10.1016/j.neuroimage.2017.10.034.

Normalization: see above

Normalization template: see above

Noise and artifact removal: see above

Volume censoring: see above

Statistical modeling & inference

Model type and settings: Linear regression models.

Effect(s) tested: The regression coefficient was obtained.

Specify type of analysis: ☒ Whole brain ☐ ROI-based ☐ Both

Statistic type for inference (See Eklund et al. 2016): Voxel-wise

Correction: Bonferroni correction was applied for multiple tests comparison (P<0.05/[251*909]).

Models & analysis

n/a	Involved in the study
☒	☐ Functional and/or effective connectivity
☒	☐ Graph analysis
☐	☒ Multivariate modeling or predictive analysis

Multivariate modeling and predictive analysis: The associations between metabolites and imaging traits were assessed using various regression models tailored to the specific types of traits, Linear regressions were adopted to test continuous and binary traits, while proportional odds logistic regression was applied for ordered categorical traits. All models adjusted the same set of covariates as aforementioned.