nature portfolio

Corresponding author(s): Gang Liu

Last updated by author(s): Feb 18, 2025

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	Touch-screen and online questionnaires, physical measurements, and clinical laboratory test were used to collect data.
Data analysis	Multiple Cox proportional hazards regression model was used to evaluate the associations between modifiable risk factors and common complications of type 2 diabetes. Multiple linear regression model and the least absolute shrinkage and selection operator (LASSO) regression were used to select representative proteins. The LASSO regression was performed using the "glmnet" R package. The mediation analyses were conducted using the %mediate SAS macro and "mediation" R package. The above statistical analyses were conducted in SAS version 9.4 (SAS Institute,Cary, NC) and R version 4.3.2 (R Core Team). Additionally, the Kyoto Encyclopedia of Genes and Genomes (KEGG) andGene Ontology (GO) biological processes (BP) enrichment analyses were performed using the online software Hiplot (https://hiplot.com.cn/).The analytic code used in this study will be made available from the corresponding author upon reasonable request.

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 data supporting the findings from this study are available within the manuscript and its supplementary information. Source data are provided with this paper. The raw UK Biobank data are protected and are not available due to data privacy laws. Researchers can apply to use the UK Biobank resource for health-related research and public interest via the UK Biobank Access Management System (https://ams.ukbiobank.ac.uk/ams/).

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 of participants was determined based on self-report. Among 14,102 participants with type 2 diabetes, 4,930 (35.0%) were female. In stratified analyses by sex (male, female), results were largely consistent, with no significant interactions (P>0.05, Supplementary Table 2-3).
Reporting on race, ethnicity, or other socially relevant groupings	Socially relevant categorization variables, including ethnicity, household income, educational attainment, employment status, and Townsend deprivation index, were used in your manuscript. All these variables served as covariates in multiple regression model and mediation analysis. Ethnicity, household income, educational attainment, and employment status were self-reported. Townsend deprivation index calculated prior to participant joining UK Biobank, based on the preceding national census output areas. Each participant is assigned a score corresponding to the output area in which their postcode is located. The multiple regression model was adjusted for age (continuous, years), sex (male, female), ethnicity (white, non-white), Townsend deprivation index (continuous), education attainment (high, intermediate, and low qualifications), employment status (employed, unemployed), household income before tax (<£18 000, £18 000-£30 999, £31 000-£51 999, £52 000-£100 000, >£100 000), moderate alcohol consumption (yes, no), diabetes duration (continuous, years), and diabetes medicine use (none, oral hypoglycemic drugs only, insulin therapy and others).
Population characteristics	Among 14,102 participants with type 2 diabetes, 4,930 (percentage = 35.0%) were female, with a mean age of 60.0 years. The baseline characteristics of participants according to cardiovascular health score are shown in Table 1. Overall, participants with higher cardiovascular health score were relatively older, were more likely to be female, non-white, employed, and have higher education qualifications, better economic status, and moderate drinking.
Recruitment	The UK Biobank is a large population-based cohort study, which recruited around half a million participants aged 37-73 years in 2006-2010 across England, Scotland, and Wales. The details of the study design have been shown elsewhere (PMID: 25826379)
Ethics oversight	The UK Biobank received ethical approval from the North West Multi-Centre Research Ethical Committee. All participants provided written informed consent.

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 sample size of main cohort (n=14,102) was determined based on the prevalence of type 2 diabetes and the availability of information related to modifiable risk factors, including diet, physical activity, nicotine exposure, sleep duration, body mass index, blood lipids, blood glucose, and blood pressure. The sample size of proteomic subset (n=1,287) was further determined based on the availability of plasma proteomics data.
Data exclusions	Individuals refused to participant in follow-up, who were free from type 2 diabetes at baseline, or had missing information on modifiable risk factors (i.e., diet, physical activity, nicotine exposure, sleep duration, body mass index, blood lipids, blood glucose, and blood pressure) were excluded from our study.
Replication	For associations of cardiovascular health score and degree of risk factor control with risks of outcomes, we performed several sensitivity analyses using competing risk model, excluding individuals who died or developed endpoints within the first year of follow-up, and including

those whose missing information on cardiovascular health metrics was imputed. The results of sensitivity analyses remain largely unchanged. For identifying representative proteins associated with cardiovascular health score, sensitivity analysis using elastic net regression was performed in addition to the least absolute shrinkage and selection operator (LASSO) regression. All proteins that have identified in LASSO regression are also identified in elastic net regression.

Randomization

Randomization was not directly relevant to this observational study. We have adjusted relevant covariates in statistical models, including age, sex, ethnicity, Townsend deprivation index, education attainment, employment status, household income before tax, moderate alcohol consumption, diabetes duration, and diabetes medicine use.

Blinding

Health-related outcomes (inpatients/mortality data) were blindly linked to baseline data. Data provided to the researchers did not contain any of personally identifiable variables.

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.