nature portfolio

Corresponding author(s): Uyenlinh L. Mishahi, Alisa Manning, Aaron Leong, Miriam Udler, Josep M. Mercader

Last updated by author(s): Jul 19, 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	No data collection software was used.
Data analysis	Shapeit v2 and v4, TOPMed Imputation Server v2, REGENIE v2 and v3.4.1, METAL v2011-03-25, GCTA v.1.94.1, R 3.5 and 4.3, PRS-CS v1, Plink v2, Bcftools v1.20.

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

Full statistics of the T2D case-control meta-analysis are available through the Common Metabolic Diseases Knowledge Portal (https://t2d.hugeamp.org/downloads.html) and through the GWAS catalog (https://www.ebi.ac.uk/gwas/, accession ID: GCST90444202).

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	All analyses were adjusted by sex using the "self-reported sex" as covariate.
Reporting on race, ethnicity, or other socially relevant groupings	The analyses were not stratified by race or genetic ancestry. Instead, analyses were performed in the whole cohorts and genetic principal components were used to correct for differences in ancestry.
Population characteristics	A full description of the demographics and population characteristic of each cohort is described in Supplementary Table 1 and Supplementary Note.
Recruitment	There was no recruitment performed for this work. Instead, we relied on previously recruited participants.
Ethics oversight	All human research approved was by the relevant Institutional Review Boards and conducted according to the Declaration of Helsinki. All participants provided written informed consent. The UKB has obtained ethical approval covering this study from the National Research Ethics Committee (REC reference 11/NW/0382). The approval for the analysis of MGB data was obtained from the MGB Institutional Review Board (study 2016P001018). The Genetic Epidemiology Research on Adult Health and Aging (GERA) cohort data was obtained through dbGaP under accession phs000674.v1.p1. The work described here was confirmed as meeting criteria for non-human subject research by the All of Us Institutional Review Board. Participation in MyCode is done through written consent under Geisinger Institutional Review Board Study #2016-0269. The results reported here were determined by the Geisinger IRB to meet the criteria for "Non-human subject research" as defined in 45CFR46.102(e). All research was performed in accordance with relevant guideline and regulations. I confirmed that all necessary patient/participant consent has been obtained and the appropiate institutional forms have been archived, and that any patient/participant/sample identifiers included were not known to anyone (e.g., hospital staff, patients or participants themselves) outside the research group so cannot be used to identify individuals.

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 conducted a meta-analysis comprising a total of 51,256 T2D cases and 380,487 controls with 12.2% cases of non-European ancestry. All individuals meeting the T2D case/control criteria from each cohort were included. For that sample size, we estimated to have >80% power to detect a variant with odds ratio >3 and minor allele frequency as low as 5x10-5. Specifically for the findings involving rare variants, we provide detailed estimations of the statistical power to detect association for each variant in Supplementary Tables 5 and 8.
Data exclusions	Controls younger than 45 years old were excluded. Individuals within the pre-diabetes range based on Hba1c or fasting glucose were excluded from the analysis.
Replication	The novel findings from the discovery meta-analysis were tested for replication in independent datasets, including the publicly available T2D GWAS meta-analysis (Vujkovic et al., 2021 and Kurki et al., 2023., R9). We also analyzed data from the Geisinger MyCode Community Health Initiative (GEISINGER), and the discovery non-overlapping individuals from the GERA and the AoU cohorts.is, which comprised 73,088 T2D cases and 79,827 controls.
Randomization	NA.
Blinding	NA.

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

Eukaryotic cell lines

Policy information about cell lines and Sex and Gender in Research

Cell line source(s)	3T3 L1 (RRID:CVCL_0123 / ATCC CCL 92.1): European Collection of Authenticated Cell Cultures (ECACC). All cell lines supplied by ECACC undergo comprehensive quality control and authentication procedures. These include testing for mycoplasma by culture isolation, Hoechst DNA staining and PCR, together with culture testing for contaminant bacteria, yeast and fungi.
Authentication	Authentication procedures used include species verification by DNA barcoding and identity verification by DNA profiling (STR profiling). The European Collection of Authenticated Cell Cultures (ECACC) Quality Control testing methods are accredited in accordance with the recognised International Standard to ISO/IEC 17025:2005 general requirements for the competence of testing and calibration laboratories.
Mycoplasma contamination	Negative.
Commonly misidentified lines (See ICLAC register)	None.

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.*