nature portfolio

Corresponding author(s): Franks PW, Coral DE

Last updated by author(s): Nov 1, 2023

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	We collected GWAS summary statistics for type 2 diabetes (T2D) and each of the three complications we considered in our study (coronary artery disease [CAD], acute ischaemic stroke [AIS] and chronic kidney disease [CKD]) from the GWAS Catalog. The links are provided in the data availiability statement. We downloaded data from multiple large consortia from the publicly available repository Open GWAS, hosted by the University of Bristol Medical Research Council Integrative Epidemiology Unit, which we queried through the R interface using the "ieugwasr" package. Summary statistics for tissue specific gene expression were downloaded from the GTEx online repository. Links are also provided in the data availability statement. The R version used was 4.1.2. Data collection in UK Biobank is described in Bycroft et al, Nature 2018. Data collection in the ORIGIN trial is described in ORIGIN Trial Investigators et al, Am Heart J 2008.
Data analysis	Data manipulation was done using mainly packages from the "tidyverse" collection. The association of polygenic scores to MACE were derived using the "survival" package in R. Pooled concordant and discordant estimates were calculated using the package "meta". To select traits where we find the most relevant differences between genetically determined obesity profiles we used the package "Boruta". For the SMR & HEIDI analysis, we adapted code from the original publication by Zhu et al. 2016 to be used within the R environment. For clumping genetic variants we used PLINK version 1.9. The R version used was 4.1.2.

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 GWAS summary analysed were accessed through the DIAGRAM Consortium data download website (http://diagram-consortium.org/downloads.html), the GWAS Catalog (https://www.ebi.ac.uk/gwas), the Open GWAS database (https://gwas.mrcieu.ac.uk) and the GTEx consortium website (https://gtexportal.org/home). UK Biobank data are available through a procedure described at http://www.ukbiobank.ac.uk/using-the-resource. Individual-level genetic and clinical data from the ORIGIN trial cannot be shared publicly due to patient confidentiality.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	We performed sex-stratified analyses due to 1) known differences in incidence rates between men and women (higher in men), and 2) available risk calculators are designed to be applied separately for each sex. Sex was determined by genotyping analysis; individuals whose genetic sex did not match reported sex were excluded, in order to have results relevant to biological sex and guard against distortion of estimates due to possible sex chromosome aneuploidies. These and other analyses that we perfomed in UK Biobank only included individuals from European descent. This is because our initial step to find concordant and discordant variants was done using GWAS summary statistics that were done in European populations, on the basis of achieving maximal power to detect discordant variants and guard against spurious associations due to population stratification. European descent was determined using genotyping data. The association of concordant and discordant polygenic scores with MACE were also tested on individuals of Latin American descent in ORIGIN. This was also determined by genotyping data.
Population characteristics	The mean age of individuals in UK Biobank at first assessment included in our study was 57 years, ranging between 39 to 72 years. 54% of participants were females. 5% had diabetes at baseline. 10% were currently smoking. 4% had experienced a MACE event prior to recruitment. The mean age of individuals in the ORIGIN trial was 64 years. 46% of individuals included in our study were of European descent, the remaining 54% were of Latin American descent. 36% of participants were females. Over 80% of participants had diabetes prior to entering the study. 12% were current smokers. 53% had prior MACE at baseline.
Recruitment	UK Biobank participants were assessed between 2006 and 2010 in 22 assessment centres throughout the UK, covering a variety of different settings to provide socioeconomic and ethnic heterogeneity and urban–rural mix. Invitations to participate were sent via mail to potential participants identified through the National Health Service. Participants that were included are not representative of the sampling population, as there is evidence for healthy volunteer selection bias. The ORIGIN trial is an international multicentre randomized controlled trial that recruited participants from multiple ancestries with impaired fasting glucose, impaired glucose tolerance, newly detected diabetes, or established diabetes, and high cardiovascular risk. The latter was ascertained by history of prior myocardial infarction or stroke or other documented manifestation of cardiovascular disease, such as coronary or peripheral stenosis detected in angiography, left ventricular hypertrophy or albuminuria
Ethics oversight	Ethics approval for the UK Biobank was obtained from the North West Centre for Research Ethics Committee. Analysis of individual-level data from UK Biobank participants in Lund University was approved by the Swedish Ethical Review Authority (2021-0317). The ORIGIN trial was approved by the Ethical Committee at each site and each participant provided written informed consent. Both studies conformed to the ethical principles for medical research involving human participants outlined in the Declaration of Helsinki. All participants provided written informed consent at enrolment.

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	To obtain genetic variants with significant effects on T2D and each of the three comorbidities, we selected studies with the largest sample

Sample size	sizes to date for both traits to ensure maximal statistical power to detect cross-trait associations. The sample size of T2D GWAS was over 900,000 individuals. The sample size of CAD and the AIS GWAS was over 1 million individuals. The sample size of the CKD GWAS was over 600,000 individuals. At the p-value threshold we applied to univariate GWAS in the context of cross-trait analysis ($p < 2.23e-4$, equal to the square root of the traditional genome-wide threshold 5e-8), these GWAS are well-powered to detect associations among common variants (over 80% for minor allele frequencies above 10% and an expected per-allele relative risk of 1.05 under an additive model, using the formulas by Skol et al Nat Genet 2006 [PMID:16415888], available at https://csg.sph.umich.edu/abecasis/cats/gas_power_calculator). In the phenome-wide scans, we selected studies in which we could ensured the allele frequencies were sufficiently similar to the reference panel (difference less that 20%). At an alpha value of 0.05, a sample size of at least 1000 individuals, and an expected variance explained by the polygenic score of 1%, we estimated a statistical power of 88% to detect an association, according to the formulas by Dudbridge et al Plos Genet 2013 (PMID:23555274). To estimate the statistical power to detect an effect of polygenic scores in survival analyses, we used the 'ssizeEpiCont.default' function from the R package 'powerSurvEpi' (version 0.1.3, last updated October 2022), maintained by Weiliang Qiu et al. Given an association where a one standard deviation unit increase in the polygenic score corresponds to a 1.05 hazard ratio of disease, and an overall incidence of disease of 5%, our calculations suggest a sample size of 66,000 individuals to achieve 80% statistical power. Our analysis of the UK Biobank data meets this threshold. In the ORIGIN study, due to eligibility criteria, we expected a higher incidence than in UK Biobank. Assuming a 50% incidence rate, the sample size required to reach 80% power exceeded 6,500 individuals, which was not met in ORIGIN.
Data exclusions	In the phenome-wide comparison between concordant and discordant SNPs, we excluded studies that were performed in population other than Europeans, to be able to compare effects across phenotypes. To test the association of concordant and discordant GRSs with MACE in ORIGIN we only included population from European and Latin American descent. To test the association with mortality in UK Biobank, we only included individuals of European ancestry. We excluded individuals with inconsistency between their reported and genetic sex, had sex chromosome aneuploidy or were outliers for heterozygosity or missingness.
Replication	To verify our results when comparing concordant and discordant diabetogenic profiles in GWAS summary data, we constructed genetic risk scores with each SNP set and assessed their associations with MACE in UK Biobank and in the ORIGIN trial.
Randomization	This study design does not involve randomization.
Blinding	This study design does not involve blinding.

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