nature research

Corresponding author(s): Handan Melike Donertas
Janet M Thornton

Last updated by author(s): Feb 23, 2021

Reporting Summary

Nature Research 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 Research 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	The primary data used in this study is the UK Biobank. We did not collect additional data for the UK Biobank participants and used the data downloaded using bash (v4.2) following the guidelines by the UK Biobank. To access other publicly available, programmatically accessible data resources we used R (v3.5.0).
Data analysis	- BOLT-LMM v2.3.2 (https://data.broadinstitute.org/alkesgroup/BOLT-LMM/) - PLINK v1.90b6.4 - VarMap (https://www.ebi.ac.uk/thornton-srv/databases/VarMap) - All remaining analysis was performed using R 3.5.0, run using RStudio v1.1.453. The following R packages were used: TSclust v1.2.4, PAM (partition around medioids) algorithm implemented in the "cluster" package v2.0.7.1, HDL (High-definition likelihood) v1.3.8 (https://github.com/zhenin/HDL/), VariantAnnotation v1.28.13, TxDb.Hsapiens.UCSC.hg19.knownGene v3.2.2, GenomicRanges v1.32.3, biomaRt v2.36.1, RCurl v1.98.1.2, jsonlite v1.7.1, rtracklayer v1.40.3, liftOver v1.12.0, goseq v1.40.0, preprocessCore v1.50.0 - LCV method (github.com/lukejoconnor/LCV) implemented in R. - The following packages were used for data handling: tidyverse v1.3.0, data.table 1.12.4 - The following packages were used for data visualisation: ggforce v0.2.2.9000, ggpubr v0.4.0, ggrepel v0.8.2, GGally v2.0.0, RColorBrewer v1.1.2, scales v1.1.1, igraph v1.2.1, ggthemes v4.2.0, pheatmap v1.0.12, ggnetwork v0.5.8 - All other analysis is done using custom codes written in bash (v4.2) or R and are available in GitHub: https://github.com/mdonertas/ukbb_ageonset

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 Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

The primary data used in the study is the UK Biobank resource, which requires an application for access (https://www.ukbiobank.ac.uk/). This study is conducted under application number 30688. The UK Biobank GWAS summary statistics provided by Neale Lab were downloaded for Townsend Deprivation Index and diet regimes (http://www.nealelab.is/uk-biobank). GTEx v8 eQTL and expression data were accessed on 20.10.2020 via the GTEx data portal (https://www.gtexportal.org/home/datasets). GWAS Catalog v1.0.2 e96 dataset was accessed on 30-07-2019 via https://www.ebi.ac.uk/gwas/docs/file-downloads. The gene lists available in "Human Ageing Genomic Resources" were downloaded using https://genomics.senescence.info/download.html and CellAge data was kindly made available before the data release on Oct 02, 2019 by Avelar et al. We access ChEMBL (https://www.ebi.ac.uk/chembl/) and PubChem (https://pubchem.ncbi.nlm.nih.gov/) using their APIs, and UniChem (https://www.ebi.ac.uk/unichem/) mappings are used to map PubChem CIDs to ChEMBL IDs. DGIdb (https://www.dgidb.org/) was used to compile drug-target gene interactions. Results of Adelman et al. (2019) and Marttila et al. (2015) age-related methylation studies were downloaded as article supplementary files. We accessed 1000 Genomes allele frequencies using the vcf file provided on the 1000 Genomes project website (https://www.internationalgenome.org/data).

The full set of GWAS results from this study can be accessed using BioStudies (S-BSST407) and all other results generated in the analysis are provided as Supplementary Datasets and Tables.

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 used UK Biobank dataset to perform GWAS and used all available data. However, we limited the number of diseases we analyzed based on the number of cases. Since we analyzed the same set of SNPs that have MAF>=0.01 across multiple diseases, to decrease the false positive rate in GWAS, we limited the diseases to a subset with at least 2,000 cases - which would correspond to allele count of 20 if equally distributed.
Data exclusions	We excluded all participants who withdrawn their data from the UK Biobank after data access. We also excluded samples without genotypes, discordant sex information, high missingness and heterozygosity. We have conducted the whole analysis on 484,598 participants. The details of the exclusion criteria is given in the methods section.
Replication	We did not repeat the analysis in an independent cohort, instead we confirmed the results using GWAS-Catalog database. Using all studies available in GWAS-Catalog (excluding the ones using the UKBB data), we confirmed the results are reproducible. The details are available in the methods and the results are given as a supplementary table.
Randomization	In this study we analyze a cohort of individuals with or without disease. Thus no randomization was performed. However, we consider sex, age, BMI, recruitment center, ethnicity, batch, and the population structure (PCs) as covariates in our analysis. Moreover, we use linear mixed models.
Blinding	The study is observational and thus blinding does not apply

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
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	All analysis was performed using the UK Biobank. The details are available elsewhere (ukbiobank.ac.uk). We provide the relevant details in the supplementary materials. In short, the dataset is composed of both males (n=221,840) and females (n=262,758), spanning an age range between 37 to 73, with a median value of 58.
Recruitment	UK Biobank recruited ~500,000 participants aged between 40-69 years in 2006-2010 across the country. They were not selected for certain traits or diseases. However, it is noteworthy that similar to other population cohorts, it is subject to healthy participant bias, i.e. the participants have reduced disease rates.
Ethics oversight	UK Biobank has obtained Research Tissue Bank (RTB) approval from its ethics committee that covers the majority of proposed uses of the Resource, so researchers do not typically need to obtain separate ethics approval. This covers the presented study as we do not re-contact the participants or collect additional data, and the research use is within the RTB approval. The REC reference for UK Biobank is 16/NW/0274. All necessary patient/participant consent has been obtained and the appropriate institutional forms have been archived. The details of the ethics oversight protocols are available elsewhere (ukbiobank.ac.uk). This study is conducted under the application number 30688

Note that full information on the approval of the study protocol must also be provided in the manuscript.