nature research

Corresponding author(s): Oliver Stegle,Inês Barroso

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 Authors & Referees and the Editorial Policy Checklist.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☒	☐	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	N/A
Data analysis	All association, interaction testing and downstream interpretation were performed using the methods that are available in the StructLMM software package, https://github.com/limix/struct-lmm. PLINK v1.90b3.32 was also used.

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/reviewers upon request. 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 BIOS RNA data can be obtained from the European Genome-phenome Archive (EGA; accession/EGAS00001001077). Genotype data are available from the respective biobanks.

Field-specific reporting

Please select the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences

For a reference copy of the document with all sections, see nature.com/authors/policies/ReportingSummary-flat.pdf

Life sciences

Study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	For the Biobank application, we use the maximal number of samples that passed QC and filtering processes as described in the Online Methods. For the eQTL analysis, we used all the samples that were used in the original analysis
Data exclusions	For the UK Biobank application, we excluded samples that were not of British ancestry and samples and variants that did not pass QC as detailed in the Online Methods. For the eQTL application, we excluded variants that did not pass QC as detailed in the Online Methods.
Replication	We have compared any associations identified with known findings, and we for some analyses, we have considered hold out validation within data from UK biobank. As the focus of the manuscript is on a new method, we have not attempted to replicate any associations using independent data.
Randomization	Not relevant because this study is not experimental
Blinding	Not relevant because this study is not experimental

Materials & experimental systems

Policy information about availability of materials

n/a	Involved in the study
☒	☐ Unique materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Research animals
☐	☒ Human research participants

Human research participants

Policy information about studies involving human research participants

Population characteristics	UK Biobank is a population based cohort, with unrelated individuals aged between 40 and 69 years at recruitment that are British. The four cohorts (CODAM, LLD, LLS and RS) used in our BIOS study are jointly summarized. The age range of the individuals differed for the different biobanks (Supplementary Fig. 7, Zhernakova et al Nature Genetics 2017). The number of samples per cohort used in our study can be found in Supplementary Table 1, Zhernakova et al Nature Genetics 2017.

Method-specific reporting

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ Magnetic resonance imaging