nature portfolio

Corresponding author(s): Sijia Wang

Last updated by author(s): Jul 28, 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	No software was sued for data collection.
Data analysis	Code for the analysis is available at GitHub (https://github.com/Fun-Gene/fastQTLmapping) and Zenodo (https://doi.org/10.5281/zenodo.8084877)71. Most operations are carried out by R (https://cran.r-project.org/), and the plots are mainly made by ggplot2 v3.4.2 R package (https://cran.r-project.org/web/packages/ggplot2/index.html). mQTL mapping is performed by FastQTLmapping fastQTLmapping (https://github.com/Fun-Gene/fastQTLmapping) and R package MatrixEQTL v2.3 (https://cran.r-project.org/web/packages/MatrixEQTL/index.html). Heritability is estimated by GCTA (https://yanglab.westlake.edu.cn/software/gcta/). MKL is available at https://software.intel.com/tools/onemkl. GSL is available at http://www.gnu.org/software/gsl/. Annotation of SNP and CpG is based on ANNOVAR (https://annovar.openbioinformatics.org/en/latest/, date: 2020.11.2) and annotation of CpG is based on the manufacturer's manifest files (date: 2020.10.21). Genotype calling is based on GenomeStudio (https://support.illumina.com/array/array_software/genomestudio/downloads.html). Imputation of SNP chip is based on SHAPEIT2 (https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html) and IMPUTE2 (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html). Enrichment analysis of mQTLs and mCpGs is performed by R package clusterProfiler v4.8.1 (https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html). DNAm processing is based on R package minfi Bioconductor package v1.46.0 (https://bioconductor.org/packages/release/bioc/html/minfi.html) and CHAMP Bioconductor package v2.30.0 (https://bioconductor.org/packages/release/bioc/html/ChAMP.html). Cell-type mQTLs are estimated by CellDMC which is available as part of the EpiDISH v2.8 Bioconductor R-package (http://bioconductor.org/packages/devel/EpiDISH. eFORGE is run with the webserver at eFORGE2.0 (https://eforge.altiusinstitute.org/). Sharing Effect of cell-type mQTLs is estimated by R package mashr (https://cran.r-project.org/web/packages/mashr/index.html).The GO and KEGG pathway enrichment analyses of mCpGs are conducted using R package missMethyl v1.34.0 (https://bioconductor.org/packages/3.13/bioc/html/missMethyl.html). Genes enrichment for diseases/traits analysis is performed by the R package disgenet2r v0.99.3 (https://www.disgenet.org/disgenet2r) based on the DisGeNET knowledgebase (date: 2021.6.9). The Twotwo-Sample sample MR analysis is conducted using the R package TwoSampleMR v0.4.26 (https://mrcieu.github.io/TwoSampleMR/). The

HiChIP loops are processed by HiCCUPS implemented in the Juicer Tools (v0.7.5) with default parameter settings. The influence of SNPs on regulatory elements is calculated using the tool OpenCausal (https://github.com/liwenran/OpenCausal). Colocalization is performed by SMR v1.3.1 (https://yanglab.westlake.edu.cn/software/smr/#Download). Enrichment of mQTL CpGs enrichment in for TF motifs is performed by TFmotifView (http://bardet.u-strasbg.fr/tfmotifview/) and R package PWMEnrich v4.28.1 (https://bioconductor.org/packages/release/bioc/html/PWMEnrich.html). Phenome-wide association analysis is carried out by PheWAS (https://gwas.mrcieu.ac.uk/phewas).

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

Data Availability: Our mQTL database is available for download at https://www.biosino.org/panmqtl/, which incorporate mQTLs not only in East Asian (NSPT), but also in published European and South Asian data. The database also supports searching and visualization of genomic, functional and downstream disease/trait hits of mQTLs and mCpGs. The statistics of mQTLs in NSPT and CGZ cohort are available for download at NODE https://www.biosino.org/node under accession number OEP002902, or directly accessed at https://www.biosino.org/node/project/detail/OEP002902. The statistics of mQTLs replicated in CAS is available for download at OMIX https://ngdc.cncb.ac.cn/omix under accession number OMIX004116, or directly accessed at https://ngdc.cncb.ac.cn/omix/release/OMIX004116. The individual-level genotype data is not available because of IRB restriction due to privacy concern. The individual-level DNA methylation data can be requested at https://ngdc.cncb.ac.cn/omix/release/OMIX004363 (NSPT), https://ngdc.cncb.ac.cn/omix/release/OMIX004333 (CAS) and https://www.biosino.org/node/project/detail/OEP002902 (CGZ). Requests are normally processed within 1–3months. Data usage shall be in full compliance with the Regulations on Management of Human Genetic Resources in China. The DNAm dataset in buccal cells is available by submitting data requests to mrclha.enquiries@ucl.ac.uk; see full policy at http://www.nshd.mrc.ac.uk/data.aspx. Managed access is in place for this 69-year-old NSHD study to ensure that use of the data is within the bounds of consent given previously by participants, and to safeguard any potential threat to anonymity since the participants are all born in the same week. The mQTL results of the EUR cohort (GoDMC) were downloaded from http://mqtldb.godmc.org.uk/downloads. The mQTL results of the EUR cohort (FHS) were downloaded from https://ftp.ncbi.nlm.nih.gov/eqtl/original_submissions/FHS_meQTLs/ (date: 2020.9.14). The annotation of CpG probes was downloaded from https://zwdzwd.github.io/InfiniumAnnotation (date: 2019.11.25). Significant GWAS results were downloaded from GWAS Catalog (https://www.ebi.ac.uk/gwas/docs/file-downloads, date: 2020.12.25) and significant EWAS results was downloaded from EWAS Atlas (https://ngdc.cncb.ac.cn/ewas/downloads, date: 2020.12.25). The cis-eQTL results in whole blood were downloaded from GTEx V8 database (https://www.gtexportal.org/home/datasets, date: 2020.6.17) and HGDV (http://www.genome.med.kyoto-u.ac.jp/SnpDB/). The human gene information (Ensembl release v104) was downloaded from GENCODE (https://www.gencodegenes.org/human/release_37lift37.html) (date: 2021.4.26), the list of human transcription factors were from http://humantfs.ccbr.utoronto.ca/download.php (date: 2020.4.3), and the list of House-Keeping genes was downloaded from (https://www.tau.ac.il/~elieis/HKG/). Motifs information of TFs was obtained from JASPAR 2020 database (http://jaspar.genereg.net/) (date: 2021.7.2) and JASPAR 2022 (date: 2022.8.22). ChIP-seq signals of TFs were downloaded from ChIP-Atlas database (http://chip-atlas.org/) (date: 2021.6.2). Other data sources used in this study include: BLUEPRINT mQTLs summary statistics (https://ega-archive.org/datasets/EGAD00001005200); Phenoscanner GWAS summary statistics (http://www.phenoscanner.medschl.cam.ac.uk/); Functional genomic regions from the Functional Annotation of ANimal Genomes (FAANG) Project (https://www.faang.org); PCHi-C data (https://osf.io/u8tzp); H3K27ac HiChIP data (https://www.ncbi.nlm.nih.gov/geo/, GSE101498); The DNase-seq data for B cells and T cells and the H3K27ac ChIP-seq data of neutrophil cells (https://www.encodeproject.org); GO terms, KEGG pathways, and Reactome pathways were downloaded from the Molecular Signatures Database (https://www.gsea-msigdb.org/gsea/msigdb/index.jsp); FANTOM5 (https://fantom.gsc.riken.jp/data/). Experimental Factor Ontology (EFO) (https://www.ebi.ac.uk/ols/ontologies/efo).GWASs in BBJ (https://pheweb.jp/); GWASs in UKBB (https://pan.ukbb.broadinstitute.org/); super enhancer databases (http://www.licpathway.net/sedb/; http://www.asntech.org/dbsuper/; http://www.licpathway.net/SEanalysis/); Segmented functional regions from GM12878 cell line (http://genome.ucsc.edu/cgi-bin/hgTrackUi?db=hg19&g=wgEncodeAwgSegmentation); 15 chromatin states (https://egg2.wustl.edu/roadmap/data/byFileType/chromhmmSegmentations/ChmmModels/coreMarks/jointModel/final/).

Human research participants

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

Reporting on sex and gender	The term sex was consistently used in this study to indicate biological attribute. Sex information for participant was based on self-reported. Sex was applied as a covariate in mQTL mapping along with age and other factors. As we used SNPs and CpGs from the auto chromosomes, there are no obvious difference between females and males at least for trans-mQTLs.
Population characteristics	Samples in the discovery dataset included 1,310 males and 2,213 females, aged from 18 to 83 years old (mean ± SD = 50.21 ± 12.75). Samples in the validation dataset CAS included 634 males and 426 females, aged from 22 to 64 years old (mean ± SD = 40.87 ± 9.41). Samples in the validation dataset CGZ included 492 males and 306 females (aged from 24 to 70 years old, (mean ± SD = 51.0 ± 9.7).
Recruitment	The samples in the discovery set (NSPT, N=3,523) were recruited from three different regions of China. There may have been differences in ancestry within the samples. The samples in validation set (CAS, N=1,060) were recruited from CAS employees, who were characterized by a high level of education and a young to middle age. There may be bias in sample selection. The samples in validation set (CGZ, N=798) were recruited from a multicenter, double-blind, parallel-group, and placebo-controlled phase 3 trial conducted in 26 centers in China. There may be bias from sample selection, the samples may be prone to diabetes, and also diabetes and drug use of chiglitazar may have an influence on DNA methylation of the samples.
Ethics oversight	The discovery cohort is a sub project of The National Science & Technology Basic Research Project which was approved by the Ethics Committee of Human Genetic Resources of School of Life Sciences, Fudan University, Shanghai (14117). The Declaration of Helsinki Principles was followed and all participants provided written informed consent. CAS study protocol

was approved by the Institutional Review Board of Beijing Institute of Genomics and Zhongguancun Hospital (No.2020H020, No.2021H001, and No.20201229). All the participants had provided written informed consents. CGZ was a multicenter, double-blind, parallel-group, and placebo-controlled phase 3 trial conducted in 26 centers in China. The trial was registered with ClinicalTrials.gov (NCT02121717). Ethical approvals were obtained from Ethical Committees of the 26 study centers. All procedures performed in the study involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the Declaration of Helsinki and its later amendments or comparable ethical standards. 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	Samples in the discovery dataset included 3,523 Han Chinese individuals. No statistical method was used to predetermine sample size. We have sufficient power to detect mQTLs at minor allele frequency of 0.01 when DNA methylation variance explained is larger than 2% (Note that in the study we found that the median DNA methylation variance explained by a mQTL was 3.1%, with an interquartile range of 1.9%-6.3%).
Data exclusions	No data were excluded in analysis.
Replication	Replication was done in two independent Han Chinese samples (CAS, N=1,060 and CGZ, N=798) and was overwhelmingly successful.
Randomization	All of the included samples are volunteers. Due to random allocation of genetic variants during gamete production genetic association studies of germline association are not expected to be subject to the confounding and reverse causation typically seen in traditional obervational epidemiology studies.
Blinding	Blinding is not relevant to this study as both the genetic variants and methylation levels of each samples were obtained by high-throughput techniques and these information would not change with analyst.

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