nature research

Corresponding author(s): Jian Yang

Last updated by author(s): Aug 20, 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	No software was used for data collection.
Data analysis	We used fastGWA-GLMM (implemented in GCTA v1.93.3), SAIGE v0.42.1, PLINK2 v2.00a2.3, REGENIE v2.0.2, BOLT-LMM v2.3.2, GCTB v2.03, codes from the Neale Lab, R (package PHESANT v0.15), and flashPCA v2.0 to perform simulations and statistical analyses of the UK Biobank data. We have also mentioned the following software tools in the paper: fastGWA (implemented in GCTA v1.93.3), ACAT-V (implemented in ACAT v0.91), fastSPA (implemented in SPAtest v2.0.0), DISSECT v1.15.2c, GMMAT v1.3.1, and BGENIE v1.2.

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 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 individual-level genotype and phenotype data are available through formal application to the UK Biobank (http://www.ukbiobank.ac.uk). The summary-level statistics of the 2,989 binary traits from our real data analyses are available at our data portal (https://yanglab.westlake.edu.cn/resources/ukb_fastgwa/imp_binary/).

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 456,348 individuals of European ancestry from the UKB for real data analysis and subsets of the data for simulations. Given an effective sample size of >350,000, we have >99.8% power to detect a genetic variant explaining only 0.02% of the phenotypic variance at p < 5e-8.
Data exclusions	We excluded UK Biobank participants of non-European ancestries because the sample sizes were insufficient to compare the methods' computing performance or identify associated variants for most traits. Standard quality control criteria were applied to exclude genetic variants with minor allele frequencies < 0.0001, Hardy-Weinberg Equilibrium test P-value < 1e-6, imputation info score < 0.8, or missingness rate > 0.1 from the UK Biobank data to avoid the inclusion of low-quality variants in the association analyses.
Replication	We replicated the simulation for 100 times in each simulation scenario to investigate the performance of fastGWA-GLMM in comparison with existing methods. All replications were successfully performed. The real data analyses were not replicated because of the use of the whole UKB sample to maximize the statistical power.
Randomization	NA. We did not use any study design that required randonmization.
Blinding	NA. We did not use any study design that required 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
☐	☒ 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 the UK Biobank participants used in this study are of European ancestry, aged between 40 and 69 at recruitment. The female to male ratio is approximately 1.2 (see https://www.ukbiobank.ac.uk/key-documents/ for more details of the population characteristics of the UK Biobank sample).
Recruitment	This study used data from participants recruited by the UKB. Please see https://www.ukbiobank.ac.uk/key-documents/ for more details of the recruitment. In summary, during 2006-2010, the UKB conducted its recruitment phase of more than 500,000 participants who gave their consent, answered questions, had physical measurements and gave samples (blood, urine and saliva) at a baseline assessment visit. It should be noted that the UKB is not representative of the general population on a variety of sociodemographic, physical, lifestyle and health-related characteristics, with evidence of a 'healthy volunteer' selection bias. As a result, the UKB is not a suitable resource for deriving generalizable disease prevalence and incidence rates. The selection bias in the UKB would not affect our simulation results because all the phenotypes were simulated based on randomly generated genetic and environmental effects. However, our real data analyses results should be interpreted with caution as those from many other genome-wide association analyses of the UKB data.
Ethics oversight	This study was approved by the University of Queensland Human Research Ethics Committee B (2011001173) and the Ethics Committee of Westlake University (20200722YJ001).

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