nature portfolio

Corresponding author(s): Chikashi Terao

Last updated by author(s): May 6, 2024

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 used
Data analysis	GATK(3.5.0, 3.6.0, 3.8.0), IMPUTE2 (2.3.2), BEAGLE(4.1), SHAPEIT2(2.837), EAGLE2(2.39), Minimac4(1.0.0), VEP(87), BOLT(2.3.4), PLINK2(2.0), LDSC(1.0.1), LDstore(2.0), FINEMAP(1.4), PAINTOR (3.0), susieR (version 0.12.19), PyMOL (2.5.2). Standalone software to create LD-matrix (LD-store) and conduct statistical finemapping (FINEMAP) are available at http://www.christianbenner.com/. We deposited custom analysis codes for association analysis and finemapping at https://doi.org/10.5281/zenodo.10934238. Further detailed scripts are available upon reasonable request to the corresponding author.

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 statistics and the results of statistical finemapping is available at JENGER website (http://jenger.riken.jp/en) and the National Bioscience Database Center (https://biosciencedbc.jp/en) under research ID hum0014 without any restriction. The imputation reference panel containing the 3,256 high-depth Japanese subjects will be available at the National Bioscience Database Center (https://biosciencedbc.jp/en) under research ID hum0014 and available to the

researchers after approval by the Human Data Review Board.
The protein 3D structure data was obtained from the Protein Data Bank (https://www.rcsb.org/). Human tissue expression data was obtained from Genotype-Tissue Expression (GTEx) Portal (https://www.gtexportal.org/home/). DNase1 hypersensitivity site and transcription factor footprints were obtained from public repositories (https://zenodo.org/records/3838751 and https://zenodo.org/records/3905306, respectively). Chromatin immunoprecipitation data was obtained from ENCODE website (https://www.encodeproject.org/). Allele frequency information for diverse human populations was obtained from gnomAD project website (https://gnomad.broadinstitute.org/). The list of clinically curated pathogenic variants was obtained from ClinVar database (https://www.ncbi.nlm.nih.gov/clinvar/).

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	BBJ first cohort n = 176,894, BBJ second cohort n = 12,098, NCGG cohort n = 14,224, ToMMo cohort n = 53,365. Specific sample sizes for each GWAS are described in Supplementary table 1. We excluded traits with total sample size < 4,000 to attain 80% power for 1 S.D. effect size at allele frequency 0.5%.
Data exclusions	For BBJ subjects, we excluded outliers from East Asian clusters in the first and second genetic principal component space in which we projected subjects in combination with 1000 Genomes Project samples. We also excluded samples with call rates less than 0.98, and samples whose reported sex information was not supported by genotypes in the X-chromosome. For ToMMo subjects, we excluded samples with call rate less than 0.97 and non-Japanese identified by principal component analysis (PCA) analyzed with combination of 1000 Genomes Project samples.
Replication	We perform replication analysis for 26 phenotypes which were available from both BBJ and ToMMo. We observed 85.4% (1,304/1,528) of lead signals were replicated in the ToMMo dataset with P < 0.05. Further information will be found in the supplementary information.
Randomization	Randomization is not applicable since genetic variants are inherently randomly distributed in the population and the study focuses on the association between the genetic variants and human phenotypic variation.
Blinding	Blinding is not applicable since the genetic and phenotype data were independently collected in advance.

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

Eukaryotic cell lines

Policy information about cell lines

Cell line source(s)	HEK293T cells and HeLa cells (#CCL-2) were purchased from the American Type Culture Collection (ATCC).
Authentication	HEK293T cells and HeLa cells (#CCL-2) were authenticated by the supplier (ATCC) using STR profiling.
Mycoplasma contamination	No contaminated cell lines were used
Commonly misidentified lines (See ICLAC register)	No misidentified cell lines were used.

Human research participants

Policy information about studies involving human research participants

Population characteristics	The BBJ is composed of ~200k participants who had one of the 47 diseases predefined. This cohort has trend of old age (mean age at enrollment 63.5 years and s.d. 14.0 years) and relatively high fraction of male subjects (54.0%). Detailed study design of BBJ is described in the following literatures. The NCGG is a hospital based cohort with elder participants (mean age at enrollment 70.7 years and s.d. 13.5 years) and relatively low fraction of male subjects (44.3%). ToMMo was established to contribute to the reconstruction of the Tohoku area which suffered from the Great East Japan Earthquake and contains large-scale adult general population cohort (mean age at enrollment 60.4 years and s.d. 11.22 years, 38.4% male). Nagai, A. et al. Overview of the BioBank Japan Project: Study design and profile. J Epidemiol 27, S2–S8 (2017). Hirata, M. et al. Cross-sectional analysis of BioBank Japan clinical data: A large cohort of 200,000 patients with 47 common diseases. J Epidemiol 27, S9–S21 (2017). Yasuda, Jun, et al. Genome analyses for the Tohoku Medical Megabank Project towards establishment of personalized healthcare. The journal of biochemistry 165.2 (2019): 139-158.
Recruitment	In this study, we included three different datasets constructed from the contemporary Japanese population [Biobank Japan (BBJ) 1st cohort, BBJ 2nd cohort, and National Center for Geriatrics and Gerontology (NCGG) cohort] and meta-analyzed the results. BBJ is a nationwide hospital-based biobank with 12 collaborating medical institutions. The first cohort targeted 47 diseases and recruited 200,000 people between 2003 and 2013, and the second cohort targeted 38 diseases and recruited 67,000 people between 2013 and 2018 (https://biobankjp.org/en/index.html). In this study, 12,098 people with available genotypes were included from BBJ 2nd cohort. The NCGG Biobank is a hospital-based biobank maintained by NCGG since 2012. The participants were recruited from NCGG hospital, Obu City, Aichi prefecture, and nearby medical institutes (https://www.ncgg.go.jp/english). The subjects in ToMMo were recruited from the health checkups conducted in two prefectures of Northeastern Japan: Miyagi and Iwate (https://www.megabank.tohoku.ac.jp/english/).
Ethics oversight	All participants provided written informed consent following the protocols approved by following institutional ethical committees, the ethics committees of RIKEN Center for Integrative Medical Sciences, the Institute of Medical Sciences, the University of Tokyo, and National Center for Geriatrics and Gerontology.

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