nature portfolio

Corresponding author(s): Yukinori Okada

Last updated by author(s): Apr 25, 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	We did not collect data for this study.
Data analysis	We used publicly available software for the data analysis (R3.6.1, plink 1.9 and 2.0, PRS-CS v1.0.0, PRS-CSx v1.0.0, LDSC v1.0.1, METAL 2020-05-05, Minimac3).

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 genotype data of BBJ are available at the NBDC Human Database (research ID: hum0014). UK Biobank analysis was conducted via the application 47821 (http://www.ukbiobank.ac.uk). Individual genotyping results and other cohort data used for the polygenic prediction are stored in Tohoku Medical Megabank Organization. In response to reasonable requests for these data (contact at dist@megabank.tohoku.ac.jp), we will share the stored data after assembling the data set after

approval of the Ethics Committee and Materials and Information Distribution Review Committee of Tohoku Medical Megabank Organization. Summary LD information used by PRS-CS or PRS-CSx is available at: https://github.com/getian107/PRScs. The PRS weights developed for TMM and BBJ-2nd have been released through the PGS Catalog (https://www.pgscatalog.org) with publication ID PGP000593 and score IDs PGS004615-PGS004620. The PRS weights can also be accessed through application at the NBDC Human Database with the accession code hum0197 (https://humandbs.biosciencedbc.jp/en/hum0197-latest).

Research involving human participants, their data, or biological material

Policy information about studies with human participants or human data. See also policy information about sex, gender (identity/presentation), and sexual orientation and race, ethnicity and racism.

Reporting on sex and gender	Since biological sex influences disease onset and trait differences, we used biological sex as in other medical and genetic studies. We obtained sex information for BBJ and BBJ-2nd participants from their medical records. We obtained sex information for UKBB and TMM participants by matching self-reported and genetic sex.
Reporting on race, ethnicity, or other socially relevant groupings	Genetic differences between populations influence polygenic prediction performance. PRS-CSx requires the input dataset to define the sample population. For the UK Biobank participants, we defined the population by grouping individuals based on genetic PCA and self-reported ethnicity provided by UK Biobank. For the BioBank Japan and Tohoku Medical Megabank Project participants, we defined the population by genetic PCA.
Population characteristics	We analyzed individual-level data from Biobank Japan, UK Biobank and Tohoku Medical Megabank Project. These datasets have already been throughly characterized in the following publications: 1. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203 (2018). 2. Nagai, A. et al. Overview of the BioBank Japan Project: Study design and profile. J Epidemiol 27, S2-S8 (2017). 3. Kuriyama et al. The Tohoku Medical Megabank Project: Design and Mission. J Epidemiol 26, 493-511 (2016).
Recruitment	We analyzed large population cohorts data that were recruited in previous studies
Ethics oversight	Biobank Japan and the second cohort of Biobank Japan: All the participants provided written informed consent approved by the ethics committees of RIKEN Center for Integrative Medical Sciences and the Institute of Medical Sciences at the University of Tokyo. UK Biobank: Collection of the UK Biobank (UKBB) data was approved by the UKBB's Research Ethics Committee. Approval to use UKBB individual-level in this work was obtained under application 47821. Tohoku Medical Megabank Project: Written informed consent was obtained from each participant when they were enrolled in the cohort studies. All the studies related to the Tohoku Medical Megabank Project were approved by the ethics committee of the Tohoku University School of Medicine and the ethics committee of the Tohoku Medical Megabank Organization.

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	We used samples from Biobank Japan, UK Biobank and Tohoku Medical Megabank Project obtained by stratified random sampling based on BMI, preserving case-control ratios. In polygenic prediction of type 2 diabetes, we used 97,884 samples (Ncase=27,642, Ncontrol=70,242) from BioBank Japan and UK Biobank, 26,000 samples (Ncase=6,000, Ncontrol=20,000) from Tohoku Medical Megabank Project, and 33,096 samples (Ncase=11,236, Ncontrol=21,860) from the second cohort of BioBank Japan. In coronary artery disease, we used 97,860 samples (Ncase=22,650, Ncontrol=75,030) from each Biobank Japan and UK Biobank.
Data exclusions	We excluded individuals who wished to be omitted from biobank dataset. We excluded samples and variants based on the standard quality control for GWAS and PRS analysis described in our manuscript.
Replication	We performed the polygenic prediction of type 2 diabetes in two independent Japanese cohorts, the Tohoku Medical Megabank Project and the second cohort of BioBank Japan, using PRS generated from GWAS summary statistics in BioBank Japan and UK Biobank.
Randomization	In sample stratification based on BMI values, we performed a stratified random sampling to maintain case-control ratios between BMI groups. In LOGO meta-analysis method, we randomly split participants into 5 groups while maintaining case-control ratios.
Blinding	We did not apply blinding of the samples because no intervention was conducted in our study.

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
☒	☐ Plants

Methods

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