

## 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](#).

### 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                                                                                                                                                                                                                                                                                      |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement                                                                                                                                    |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly                                                                                                                                    |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided<br><i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>                                                               |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of all covariates tested                                                                                                                                                                                                                     |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons                                                                                                                                        |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> 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) |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted<br><i>Give <math>P</math> values as exact values whenever suitable.</i>                            |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings                                                                                                                                                                      |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes                                                                                                                                                |
| <input type="checkbox"/>            | <input checked="" type="checkbox"/> 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   | We used publicly available software for the data analysis (R 3.3.0, plink 1.9 and 2.0, ADMIXTURE 1.3.0, TreeMix 1.13, selscan 1.1.0b and LDSC 1.0.0, Eagle (version 2.3), multicoreTSNE package of python, umap package of python, fineSTRUCTURE (version 2), Grimon software). We also used custom scripts to perform five dimensionality reduction methods on genotype data. We provided these scripts at <a href="https://github.com/saorisakaue/Genotype-dimensionality-reduction">https://github.com/saorisakaue/Genotype-dimensionality-reduction</a> . |

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. 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 genotype data of BioBank Japan used in this study are available from the Japanese Genotype-phenotype Archive (JGA; [http://trace.ddbj.nig.ac.jp/jga/index\\_e.html](http://trace.ddbj.nig.ac.jp/jga/index_e.html)) with accession code JGAD00000000123 [<https://ddbj.nig.ac.jp/jga/viewer/view/study/JGAS00000000114>]. The GWAS summary statistics of BioBank Japan are available at the National Bioscience Database Center (NBDC) Human Database with the accession code hum0014 [<https://humandbs.biosciencedbc.jp/en/hum0014-v18>]. UK Biobank analysis was conducted via the application 31063. All other data are contained in the article file and its supplementary information or available upon request.

## 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](https://www.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     | Clinical information and genotype data were obtained from BioBank Japan (BBJ) project, which is a prospective biobank that collaboratively collected DNA and serum samples from 12 medical institutions in Japan and recruited approximately 200,000 participants. Of them, we included individuals with genotype and basic phenotype data |
| Data exclusions | We excluded individuals with the age under 18, low call rate in genotyping (< 98%), closely related (PI_HAT < 0.125), and ancestry other than Japanese (based on PCA plot) as described elsewhere.                                                                                                                                         |
| Replication     | We randomly split the whole cohort in two groups, and used one group for discovery GWASs and the effect size derivation to be used in PRS construction. Using the other group with totally independent individuals, we calculated and validated PRSs.                                                                                      |
| Randomization   | In the above randomization, we assigned a random number to each participant and randomly split the whole participants into 2 groups.                                                                                                                                                                                                       |
| 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                                                         | Methods                                                                             |
|------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------|
| n/a                                                                                      | n/a                                                                                 |
| <input checked="" type="checkbox"/> <input type="checkbox"/> Involved in the study       | <input checked="" type="checkbox"/> <input type="checkbox"/> Involved in the study  |
| <input checked="" type="checkbox"/> <input type="checkbox"/> Antibodies                  | <input checked="" type="checkbox"/> <input type="checkbox"/> ChIP-seq               |
| <input checked="" type="checkbox"/> <input type="checkbox"/> Eukaryotic cell lines       | <input checked="" type="checkbox"/> <input type="checkbox"/> Flow cytometry         |
| <input checked="" type="checkbox"/> <input type="checkbox"/> Palaeontology               | <input checked="" type="checkbox"/> <input type="checkbox"/> MRI-based neuroimaging |
| <input checked="" type="checkbox"/> <input type="checkbox"/> Animals and other organisms |                                                                                     |
| <input type="checkbox"/> <input checked="" type="checkbox"/> Human research participants |                                                                                     |
| <input checked="" type="checkbox"/> <input type="checkbox"/> Clinical data               |                                                                                     |

## Human research participants

Policy information about [studies involving human research participants](#)

|                            |                                                                                                                                                                                                                                                    |
|----------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Population characteristics | The detailed information of the participants in BBJ is described in the previous literatures [Hirata, M. et al. J. Epidemiol. 27, S9–S21 (2017), and Nagai, A. et al. J. Epidemiol. 27, S2–S8 (2017)].                                             |
| Recruitment                | BioBank Japan (BBJ) project collaboratively collected DNA and serum samples from 12 medical institutions in Japan and recruited approximately 200,000 participants with the diagnosis of at least one of 47 diseases, mainly of Japanese ancestry. |
| Ethics oversight           | All the participants provided written informed consent approved from ethics committees of RIKEN Center for Integrative Medical Sciences, and the Institute of Medical Sciences, the University of Tokyo.                                           |

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