nature research

Corresponding author(s):   Dr Loic Yengo and Prof. Peter Visscher

Last updated by author(s):   Jul 16, 2019

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	
☐	☒	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 were used for data collection.
Data analysis	Statistical analyses were performed mostly with R versions 3.5.1. Mixture of exponential distribution were fitted using the R package Renext. Other analyses were performed using PLINK 1.07 and PLINK v1.90b6.7.

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

We used genotypic data and phenotypic data from the UK biobank under project number 12505. UKB data can be accessed upon request once a research project has been submitted and approved by the UKB reserach committee.

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 pre-existing data from the UK Biobank (N=456,426) and excluded 12 study participants who retracted consent. Otherwise, we used all data made available.
Data exclusions	We used SNP data to infer ancestry and restricted our analyses to study participants of European ancestry in order to minimize confounding due to population stratification. We also only considered in certain analyses genetically unrelated participants using a threshold of 0.05 of the SNP-based genetic relationship matrix. Details of our approach to determine ancestry homogeneity of the sample are also described in a previous study: Yengo et al. Hum. Mol. Genet. 15;27(20):3641-3649 (2018).
Replication	We quantified extreme inbreeding in the UK population. Therefore, our main analyses were performed in the UK Biobank, which is the best and largest dataset to address this question. Replication in other populations was not attempted because of lack of data on a similar scale.
Randomization	N/A
Blinding	N/A

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
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

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	Estimates were adjusted for age at recruitment, recruitment centre (treated as a categorical factor), sex, year of birth (treated as a continuous variable), genotyping batch (treated as a factor), socioeconomic status measured by the Townsend deprivation index and population structure measured by 10 genetic principal components.
Recruitment	UK Biobank investigators sent postal invitations to 9,238,453 individuals registered with the UK's National Health Service who were aged 40–69 years and lived within approximately 25 miles (40 km) of one of 22 assessment centers located throughout England, Wales, and Scotland. Overall, 503,317 participants consented to join the study cohort and visited an assessment center between 2006 and 2010, resulting in a participation rate of 5.45%. (Fry et al. Am J Epidemiol. 2017 Nov 1;186(9):1026-1034).
Ethics oversight	The National Information Governance Board for Health and Social Care and the North West Multicentre Research Ethics Committee provided approval for UK Biobank to obtain the contact details of people within the eligible age range from local National Health Service Primary Care Trusts.

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