nature research

Corresponding author(s): Eimear Kenny and Chris Carlson

Last updated by author(s): 04-15-2019

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 was used for data collection.
Data analysis	Software used for data analysis in this manuscript includes: Plink (v1.90), R, Illumina's GenomeStudio, SNPRelate, PC-Relate, METAL, GENESIS, SUGEN, and FINEMAP (v1.3). All software used in this analysis is described in detail within the Methods section (Pg 14) and Supplemental Information.

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

Data Availability: Individual-level phenotype and genotype data are available through dbGaP at https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs000356. Allele frequency data will be available for all genotyped sites on dbSNP (https://www.ncbi.nlm.nih.gov/projects/SNP/) and the University of Chicago Geography of Genetic Variants Browser (http://popgen.uchicago.edu/ggv/). Clinically relevant variant frequency data is also available through ClinGen.

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 conducted genome-wide association (GWAS) of 49,839 non-European individuals for 26 clinical and behavioral phenotypes. As far as we are aware, this is the largest study of multi-ethnic, non-European participants genotyped on a single array for GWAS. See Methods (Page 14) and Supplemental Information (Pages 2-3) for more information about the samples.
Data exclusions	Several rationale for data exclusions were deployed in this study. Participants were excluded from the analysis if they had missing or outlier phenotype values, confounding co-morbidities, medications, or otherwise non-standard measures of phenotype as described in the Supplementary Information. Genotyped or imputed variants were excluded if they failed quality control or technical filters described in the Methods section.
Replication	Over 95% of the findings of this study reproduced known, published genotype-phenotype associations in the GWAS catalog. In addition, 5% variants in this study (26 loci) reach the threshold of statistical significance, indicating novel findings. However, in the absence of independent large-scale studies for those in the same populations, these could not be independently reproduced.
Randomization	Models for variant association testing for each phenotype included covariates for gender and study center, to adjust for population group including self-identified race/ethnicity and the first 10 principal components of genetic ancestry, and phenotype-specific covariates (including, age of diagnosis/trait measurement, medications, smoking, and co-morbidities).
Blinding	All participating studies had population-base recruitment strategies, regardless of outcomes, therefore blinding was not necessary during data collection or analysis in this 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
☒	☐ 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	A total of 49,839 individuals of non-European ancestry were included in this study, comprising of 15,381 men and 34,458 women. Genotyped individuals self-identified as Hispanic/Latino (N=22,216), African American (N=17,299), Asian (N=4,680), Native Hawaiian (N=3,940), 112Native American (N=652), or Other (N=1,052). Detailed population characteristics by study and phenotype harmonization information can be found in Supplemental Information Sections 1 and 2, as well as Supplementary Table 1.
Recruitment	All participating studies had population-based recruitment strategies, regardless of outcomes. Details can be found in Supplemental Materials (Section 1).
Ethics oversight	Study protocols were approved for all studies by the appropriate boards at their respective institutions: Fred Hutchinson Cancer Research Center Institutional Review Board (WHI), University of North Carolina Office of Human Research Ethics/IRB (OHRE/IRB; HCHS/SOL), University of Southern California IRB (MEC), University of Hawaii IRB (MEC), Icahn School of Medicine at Mount Sinai IRB (BioMe), and the Stanford University IRB (GRP).

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