nature research

Corresponding author(s): David Howard

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.

Statistical parameters

When statistical analyses are reported, confirm that the following items are present in the relevant location (e.g. 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
☐	☒	An indication of 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 statistics 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
☐	☒	Clearly defined error bars *State explicitly what error bars represent (e.g. SD, SE, CI)*

Our web collection on statistics for biologists may be useful.

Software and code

Policy information about availability of computer code

Data collection	No software used for data collection.
Data analysis	BGENIE v1.1 METAL v.2011-03-25 Linkage Disequilbrium Score Regression v1.0.0 LD Hub v1.9 MR-Base (TwoSampleMR) v.0.4.9 MAGMA v1.06 Plink v1.90b4

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 upon request. 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

Summary statistics for 10,000 genetic variants from the meta-analysis of 23andMe_307k, UK Biobank, and PGC_139k and the summary statistics for all assessed genetic variants for the meta-analysis of depression in UK Biobank and PGC_139k are available from: http://dx.doi.org/10.7488/ds/2458.
To access the summary statistics for all assessed genetic variants for the meta-analysis of depression in 23andMe_307k, UK Biobank, and PGC_139k a data transfer agreement is required from 23andMe (dataset-request@23andMe.com) before a request should be made to the corresponding author (D.Howard@ed.ac.uk).
The raw genetic and phenotypic UK Biobank data used in this study, which were used under license, are available from: http://www.ukbiobank.ac.uk/.
The genome-wide summary statistics for the Hyde et al. analysis of 23andMe, Inc. data were obtained under a data transfer agreement. Further information about obtaining access to the 23and Me, Inc. summary statistics are available from: https://research.23andme.com/collaborate/
The genome-wide summary statistics for the Wray et al. analysis of PGC data were obtained under secondary analysis proposals #60 and #76. Further information about obtaining access to the PGC summary statistics are available from: http://www.med.unc.edu/pgc/statgen

Field-specific reporting

Please select 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/authors/policies/ReportingSummary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	No statistical methods were used to pre-determine sample sizes but the sample size in current analysis (n = 807,553 individuals) was greater than the published studies that contributed to the current meta-analysis (n = 459,481, 322,580, and 461,134 individuals)
Data exclusions	UK Biobank data We removed 131,790 individuals that had a shared relatedness up to the third degree and then calculated a genomic relationship matrix and identified one individual to be reinstated from within each related group that had a genetic relatedness less than 0.025 with all other participants which allowed us to add an additional 55,745 individuals back into our sample. We then used a checksum based approach 93 to identify and exclude 954 individuals from within the UK Biobank cohort that overlapped with the Major Depressive Disorder Working Group of the Psychiatric Genomics Consortium (PGC) cohorts analysed by Wray, et al. We also removed those UK Biobank individuals with a variant call rate < 90% or that were outliers based on heterozygosity, or variants with a call rate < 98%, a minor allele frequency < 0.005, those that deviated from Hardy-Weinberg equilibrium (P < 10-6), or had an imputation accuracy score < 0.1. 23andMe data We excluded variants with an imputation accuracy threshold < 0.6 and with a minor allele frequency < 0.005. PGC data This meta-analysis included the 23andMe_307k discovery cohort and a previous release of the UK Biobank data. We therefore obtained the summary statistics from their meta-analysis of major depression with the 23andMe_307k and the previous UK Biobank cohorts removed. We excluded variants with an imputation accuracy threshold < 0.6 or a minor allele frequency < 0.005.
Replication	We used a large independent cohort (~1.5M individuals) to assess the reproducibility of the significant variants associated with depression.
Randomization	UK Biobank and 23andMe cohorts were allocated in to cases and controls based on self-reported measures regarding depression. PGC was based on a clinically-diagnosed phenotype of major depressive disorder.
Blinding	The data was collected independently of the analysis and was typically from population-based studies. Therefore there was no need to use blinding or randomization.

Reporting for specific materials, systems and methods

Materials & experimental systems

n/a	Involved in the study
☒	☐ Unique biological materials
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants

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	UK Biobank and 23andMe cohorts were allocated in to cases and controls based on self-reported measures regarding depression. PGC was based on a clinically-diagnosed phenotype of major depressive disorder. UK Biobank consisted of 361,315 individuals with a trait prevalence of 35%. The age range of this cohort was 38 - 74 years of age with a mean age of 57.2 years and standard deviation of 8.0 years. 46% of this cohort were female. 23andMe consisted of 307,354 individuals with a trait prevalence of 25%. 52% of this cohort are female and the age of these individuals has not been reported to us. PGC consisted of 138,884 individuals with a trait prevalence of 31%. Age and sex have not been reported to us.
Recruitment	Patients were recruited differently in each of the cohorts used, with some cohorts collected from hospitals and others from population-based samples. The results were relatively consistent across cohorts suggesting that there was limited bias due to sample recruitment