nature portfolio

Corresponding author(s): Kevin S. O'Connell and Ole A. Andreassen

Last updated by author(s): Oct 4, 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	No software was used to collect the data in this study
Data analysis	Genotype QC, imputation, association analysis, meta-analysis: RICOPILI (https://sites.google.com/a/broadinstitute.org/ricopili/) EIGENSTRAT v6.1.4 (https://www.hsph.harvard.edu/alkes-price/software/) Eagle v2.3.5 (https://alkesgroup.broadinstitute.org/Eagle/) Minimac3 (https://genome.sph.umich.edu/wiki/Minimac3) PLINK v1.90 (https://www.cog-genomics.org/plink2/) METAL (version 2011-03-25) (https://genome.sph.umich.edu/wiki/METAL_Documentation) DENTIST (https://github.com/Yves-CHEN/DENTIST) LDSC (https://github.com/bulik/ldsc) Popcorn (https://github.com/brielin/Popcorn) Downstream analysis: Complex Traits Genetics Virtual Lab web platform (CTG-VL; https://vl.genoma.io) multitrait-based conditional and joint analysis using GWAS summary data (mtCOJO) (https://yanglab.westlake.edu.cn/software/gcta/#mtCOJO) MiXeR v1.3 (https://github.com/precimed/mixer) PRS-CS (https://github.com/getian107/PRScs) GSA-MiXeR (https://github.com/precimed/gsa-mixer) MAGMA (https://ctg.cncr.nl/software/magma)

SynGO tool v1.2 (https://www.syngoportal.org/)
Variant Effect Predictor (VEP) (GRCh37) Ensembl release 75 (https://www.ensembl.org/info/docs/tools/vep/index.html)
Summary data-based Mendelian randomization (SMR) (v1.3) (https://yanglab.westlake.edu.cn/software/smr/)
TWAS and FOCUS (http://gusevlab.org/projects/fusion/)
isoTWAS (https://github.com/bhattacharya-a-bt/isotwas)
OpenTargets platform (https://genetics.opentargets.org/)

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

Genome-wide association summary statistics for these analyses are available at https://www.med.unc.edu/pgc/download-results/. The full GWAS summary statistics for the 23andMe datasets will be made available through 23andMe to qualified researchers under an agreement with 23andMe that protects the privacy of the 23andMe participants. Please visit https://research.23andme.com/collaborate/#dataset-access for more information and to apply to access the data. After applying with 23andMe, the full summary statistics including all analysed SNPs and samples in the GWAS meta-analyses will be accessible to the approved researchers. Individual-level data are accessible through collaborative analysis proposals to the Bipolar Disorder working group of the PGC (https://pgc.unc.edu/for-researchers/data-access-committee/data-access-information/). This study included some publicly available datasets accessed through dbGAP (PGC bundle phs001254.v1.p1).

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	Sex was included as a covariate in the analyses. Sex-based analyses were not considered in this manuscript due to the amount of data already being presented. Theses analyses are being considered for a separate manuscript in the future that can focus specifically on this aspect.
Reporting on race, ethnicity, or other socially relevant groupings	Participants were grouped into European, East Asian, Latino and African American ancestral groups. This was based on participant/recruiter reported information and confirmed with genetic information.
Population characteristics	The first five genetic principal components, and any others as required, were included as covariates in the GWAS of each cohort.
Recruitment	Participants were recruited using different approaches. These are outlined for each cohort in the supplementary materials (cohort descriptions) of the manuscript.
Ethics oversight	These are listed for each cohort in the supplementary materials.

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	The final multi-ancestry meta-analysis included up to 158,036 cases and 2,796,499 controls aggregated from 79 cohorts. No sample size calculations were performed. This sample size is hypothesized to be insufficient to explain the full extent of genetic liability to bipolar disorder however, it is the largest study of its kind to date.
Data exclusions	Technical quality control was performed separately on each cohort for which individual-level data were provided separately according to standards developed by the PGC including; SNP missingness<0.05 (before sample removal), subject missingness<0.02, autosomal heterozygosity deviation (Fhet<0.2), SNP missingness<0.02 (after sample removal), difference in SNP missingness between cases and controls<0.02, SNP Hardy–Weinberg equilibrium (P>1×10−10 in BD cases and P>1×10−6 in controls), and mismatches between pedigree and genetically-determined sex based on the F statistic of X chromosome homozygosity (female F < 0.2 and male F >0.8). In addition, relatedness

was calculated across cohorts using identity by descent and one of each pair of related individuals (pi_hat>0.2) was excluded, prioritising exclusion of individuals related to the most others, controls over cases, and individuals from larger cohorts.

These exclusions follow standard guidelines in the field.

Replication

All available cohorts of bipolar disorder cases and controls were included in the primary multi-ancestry meta-analysis and therefore we do not perform replication for significant loci identified from these analyses.
We did test replication of previously identified loci associated with bipolar disorder from European ancestry. These results are detailed in the manuscript. Briefly, 31/64 previous loci met genome-wide significant in the present analysis containing all samples, and of the 33 that did not, 25 met genome-wide significant in either the Clinical samples or in the meta-analysis that excluded Self-reported data. Moreover, the direction of association for all top SNPs (12,151 SNPs with $p < 1 \times 10–5$) from the previous GWAS was consistent with the direction of association in this multi-ancestry meta-analysis of all samples

Randomization

This is an observational, genetic epidemiological study, and as such no randomization was performed.

Blinding

No disease-status blinding was used in recruitment or analysis. In principle, participants were recruited blind to their genotype and we do not expect to observe bias in association between genotype and disease-status.

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

Plants

Seed stocks

NA

Novel plant genotypes

NA

Authentication

NA