nature research

Corresponding author(s): Schizophrenia Working Group Psychiatric Genomics Consortium

Last updated by author(s): October 21, 2021

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 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 and risk scoring framework: RICOPILI (https://github.com/Ripkelab/ricopili/wiki) RICOPILI: QC, frequency estimation, association analysis and risk scoring: PLINK1.9 (https://www.cog-genomics.org/plink2/) Principal Components Analysis: Eigenstrat (https://github.com/DReichLab/EIG/tree/master/EIGENSTRAT) Genotype phasing: EAGLE (https://github.com/poruloh/Eagle), SHAPEIT https://mathgen.stats.ox.ac.uk/genetics_software/shapeit/shapeit.html Genotype imputation: IMPUTE (https://mathgen.stats.ox.ac.uk/impute/impute_v2.html) QC and heritability: LDSC (https://github.com/bulik/ldsc) Meta-analysis: METAL (https://github.com/statgen/METAL) Downstream analysis: Set-based association analysis: MAGMA (https://ctg.cncr.nl/software/magma) Fine mapping: FINEMAP(http://christianbenner.com/#sss) Gene ontology annotation: BioMart R package (https://bioconductor.org/packages/release/bioc/html/biomaRt.html) Mendelian randomization and pleiotropy: SMR + HEIDI (https://cnsgenomics.com/software/smr/) Conditional analysis: GCTA-COJO (https://cnsgenomics.com/software/gcta/index.html#Overview) Transcriptome imputation: FUSION (http://gusevlab.org/projects/fusion/) Transcriptome imputation and epigenetics: EpiXcan (https://zhangw17.u.hpc.mssm.edu/epixcan/about.php)

Heritability and polygenicity estimation: GCTB-SBayesS (https://cnsgenomics.com/software/gctb/#Overview)

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 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

Genome-wide summary statistics from the discovery meta-analysis, and ancestry specific meta-analyses will be made available upon publication at the following site: https://www.med.unc.edu/pgc/results-and-downloads/.
See "Supplementary Cohort Descriptions" for information for participating cohorts.

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 discovery sample consisted of 74,776 samples with Schizophrenia and 101,023 controls, aggregated from 99 distinct cohorts. While this sample size is hypothesized to be insufficient to explain the full extent of genetic liability to Schizophrenia, it is the largest study of its kind to date.
Data exclusions	We excluded individual samples and genetic markers using the standard RICOPILI pipeline (Lam. et. al. 2019).This consisted of, 1) Excluding variants with call rate below 95%; 2) Excluding subjects with call rate below 98%; 3) Excluding monomorphic variants; 4) Excluding subjects with inbred coefficient above 0.2 and below -0.2; 5) Excluding subjects with mismatch in reported gender and chromosome X computed gender; 6) Excluding variants with missing rate differences greater than 2% between cases and controls; 7) Subsequent to step 6, exclude variants with call rate below 98%; and 8) Exclude variants in violation of Hardy-Weinberg equilibrium ($P < 10\text{-}6$ for controls or $P < 10\text{-}10$ for cases). Using imputed data to estimate identity-by-descent, we were able to identify duplicate samples within and between cohorts. Samples with pi-hat > 0.2 were extracted, followed by Fisher-Yates shuffle on all samples. When deciding which samples to retain, trio were preferred, followed by cases, and thereafter a random sample for each related pair was removed. To identify population outliers, we computed Principal Components within each cohort and inspected PCA plots for the top four PCs manually, identifying outliers and removing them These exclusions follow standard guidelines for genome-wide association studies.
Replication	We replicated results from the discovery meta-analysis of 89 cohorts using summary statistics provided by deCODE Genetics (1,979 cases and 142,627 controls of European ancestry).
Randomization	This is an observational, genetic-epidemiological study, and as such no randomization was performed. Based on the principles of mendelian inheritance, it is hypothesized that such study designs are protected against typical forms of confounding. See section "Mendelian Randomization" in the methods texts. One known confounder in genome-wide association studies is population structure, which we adjust for by computing Principal Components (PCs) within each cohort, and adding the top four PCs — plus any in the top twenty that are marginally associated with disease-status — as covariates in the marker-level logistic regression. This is a standard form of adjustment for GWAS study designs.
Blinding	No disease-status-blinding was used in recruitment or analysis. In principle, samples were recruited blind with respect to their genotype, and we do not expect to observe bias in the 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
☐	☒ Human research participants
☒	☐ Clinical data
☒	☐ Dual use research of concern

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	This study aggregates data from 100 different cohorts, for a total of 75,703 cases and 242,251 controls and a replication sample. Of these, 76 in the discovery sample were drawn from populations of European descent (53,386 cases, 77,258 controls), and 14 in populations of East Asian descent (14,004 cases, 16,757 controls). Additionally, of the 100 cohorts, 4 European cohorts (1,369 trios) and 2 East Asian cohorts consisted of parent-proband trios (1,009 trios).
Recruitment	See "Supplementary Cohort Descriptions" for descriptions of recruitment strategies for participating cohorts. Additionally, see "Supplementary Note" section "Potential Heterogeneity due to Ascertainment" for evidence that differing ascertainment did not lead to heterogeneity in results.
Ethics oversight	The study protocols were approved by the institutional review board at each center involved with recruitment. Informed consent and permission to share the data were obtained from all subjects, in compliance with the guidelines specified by the recruiting center's institutional review board. Genotyping of samples recruited in mainland China were processed and analysed by Chinese groups on Chinese local servers, to comply with the Human Genetic Resources Administrative Regulations. Only summary statistics, with no individual-level data, were included in the final study from samples recruited from mainland China. See "Supplementary Cohort Descriptions" for descriptions of ethics approval protocols for participating cohorts.

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