nature research

Corresponding author(s): Anders D. Børglum

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	This is fully described in the Online Methods and associated Supplemental Note. In brief: GenCall, Birdseed and zCall were used for genotype calling and the genotypes merged.
Data analysis	This is fully described in the Online Methods and associated Supplemental Note. In brief: Genotype calling was done using GenCall (1.6.2.2), Birdseed (1.6) and zCall version 1 (Autocall, https://github.com/jigold/zCall). Quality control, imputation, association analyses, and polygenic risk scoring was done using the Ricopili pipeline: https://github.com/Nealelab/ricopili, which relies on the following software: SHAPEIT v2, IMPUTE2, Eigensoft 6.0.1 (incl. smartPCA), Plink 1.9, METAL 2011-03-25. For gene-based and gene-set analyses we used MAGMA 1.06. Estimation of credible SNPs were done using CAVIAR v1, and to test whether credible SNPs are enriched in active marks in the fetal brain and in-house implementation of the GREAT method was used (see Won et al. Nature. 2016;538(7626):523-527). Functional annotation of credible SNPs was done using Ensemble Variant Effect Predictor (VEP). SNP heritability, partitioning of the heritability and genetic correlations were estimated using LD score regression (https://github.com/bulik/ldsc) and LD hub (http://ldsc.broadinstitute.org/) for the large samples. Genetic correlation between ASD subtypes was estimated using GCTA v1.26.0. Multitrait association analyzes were conducted using MTAG (https://github.com/omeed-maghzian/mtag). R v3.4 was used in general for statistical analyzes and plotting (https://www.Rproject.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/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

As stated in the manuscript, summary statistics has already been made available at http://ipsych.au.dk/downloads/ and https://www.med.unc.edu/pgc/results-and-downloads. For access to genotypes from the PGC samples and the iPSYCH sample, researchers should contact the lead PIs Mark J. Daly and Anders D. Børglum for PGC-ASD and iPSYCH-ASD respectively.

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 sample size calculation was made. Previous studies of psychiatric disorders that are very polygenic (e.g. schizophrenia) have demonstrated that high numbers of cases and controls (in line with the sample size analyzed in this study) yield enough power to detect common risk variants with low effect sizes.
Data exclusions	Within each analyzed cohort we aimed at analyzing genetically homogeneous samples. Genetic outliers were excluded based on principal component analyses and related individuals were removed.
Replication	Consistency was checked internally between the 23 batches in iPSYCH and between iPSYCH and PGC. The 88 strongest signals were followed up in independent samples from other Nordic countries and Eastern Europe. We evaluated our results in three ways: sign test, genetic correlation, and metaanalysis of the combined samples, and these generally support the results from our primary GWAS. In addition support for the reported loci were obtained by MTAG analysis of correlated traits.
Randomization	It is an observational study comparing everybody in the selected birth cohorts with and ASD diagnosis as cases, and a random sample from the complement in said cohort as controls.
Blinding	In iPSYCH, diagnoses are drawn from registries. These are administrative data bases populated by data from the clinicians long before the current study. The blood samples are pulled from a biobank. Hence, the study participants and diagnosing clinicians are blinded with respect to this study. Genotyping is done on a massive scale on 85.000 individuals on 500.000 variables (which by imputation is expanded to ~10 million variables), and the data is generated without a specif goal or effect in mind except for an overall goal of investigating the genetic and environmental effects on psychiatric disorders. So although it is in principle possible for analysts in the lab to look up crude diagnostic data for a sample, it will not change the genotyping. - In the meta analysis we include data from the Psychiatric Genetics Consortium (PGC) which has been reported in an earlier publication. There design was different, but analyses analogous.

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	In the meta-analysis we included samples from the Psychiatric Genetics Consortium (PGC) and the iPSYCH sample. The iPSYCH sample was processed in 23 batches (genotyping, qc and imputation was done separately for theses batches) of approximately 3,500 individuals each. All analyzes were adjusted for batch, and principal components included to control for population stratification. The PGC samples are trio samples, hence there was no need to adjust for populations stratification there.
Recruitment	In iPSYCH, diagnoses are drawn from national registries and the blood samples are pulled from a the Danish Neonatal Screening Biobank. Hence, it is a population sample and bias from self-selection is impossible.