nature portfolio

Corresponding author(s): Ditte Demontis

Last updated by author(s): Jun 6, 2022

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	This is fully described in the Online Methods a. In brief: (1.6.2.2) and Birdseed (1.6) and were used for genotype calling of variants with minor allele freqeuncy > 0.01 i iPSYCH1. iPSYCH2 genotypes were called using GenTrain V3
Data analysis	Common variants: imputation of variants were done using Eagle v2.3.555 and Minimac3. Quality control and association analyses were done using Plink 1.9 and Eigensoft 6.1.3. SNP heritability was estimated using GCTA v1.93.1. Genetic correlations were calculated using LD score regression v.1.0.1 (https://github.com/bulik/ldsc). Polygenic score analyses were done using Plink 1.9 and R v. 3.6.0 and the packages 'multcomp' and 'fmsb'. Rare variants: alignment of sequence reads to the reference genome and and calling of genotypes were done using BWA and GATK v.3.4. QC was done in Hail v. 0.1 (https://github.com/hail-is/hail), annotation of variants was done using SnpEff v. 4.3t and SnpSift v. 4.3t. Test for difference i categories of rare variants were done in R v. 3.6.0 using the R packages foreign, nnet, ggplot2, reshape2 .

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

Summary statistics from GWAS of childhood, persistent and late-diagnosed ADHD are available at the iPSYCH website (https://ipsych.dk/en/research/downloads/). All relevant iPSYCH data are available from the authors after approval by the iPSYCH Data Access Committee and can only be accessed on the secured Danish server (GenomeDK, https://genome.au.dk) as the data are protected by Danish legislation.
For data access please contact: Ditte Demontis or Anders D. Børglum.

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	No sample size calculation was made. This study included all individuals diagnosed with ADHD born in Denmark between 1981 and 2008, so no the sample size number was fixed with respect to number of cases. Additionally, previous studies of ADHD have demonstrated that the disorder is very polygenic, but also that the 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 and perform analyses of the polygenic architecture (e.g. Demontis et al. Nature Communications, 12, 576(2021), Demontis et al; Nature Genetics 51, 63–75(2019)).
Data exclusions	We aimed at analyzing a genetically homogeneous sample. Genetic outliers were excluded based on principal component analyses and related individuals were removed.
Replication	Replicated the ADHD-polygenic score results one time in an independent smaller Spanish sample consisting of 453 individuals with childhood ADHD, 270 with persistent ADHD, 889 with late-diagnosed ADHD and 3,440 controls.
Randomization	Allocation into groups was not random. Individuals were allocated into three ADHD case groups (childhood, persistent and late-diagnosed ADHD) based on age at first ADHD diagnosis (ICD10 diagnosis code F90.0). The controls should not have a diagnosis of ADHD and were selected randomly from the same birth cohort as the cases.
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 overall for the full iPSYCH cohort (130,000 individuals are array genotyped), 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.

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	Common variants analyses, covariates used: 10 PCs derived from component analysis (on a set pruned variants), sex and covariate for genotyping round (iPSYCH1 or iPSYCH2). Rare variants analyses, covariates used: birth year, sex, first ten principal components from ancestry principal component analysis, number rSYN, percentage of target with coverage > 20x, mean read depth at sites within the exome target passing VQSR, total number of variants, sequencing wave.
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 not possible.
Ethics oversight	The study was approved by the Danish Data Protection Agency and the Scientific Ethics Committee in Denmark.

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