nature portfolio

Corresponding author(s): *Double-anonymous peer review submissions: write DAPR and your manuscript number here instead of author names.*

Last updated by author(s): *YYYY-MM-DD*

# 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<br>*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<br>*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 for data collection |
| Data analysis | Genotyping platforms used for each cohort are described in the "Clinical databases: sample description, clinical assessment and genotyping" section of the manuscript.<br>Imputation and quality control of the clinical datasets were performed using PLINK 2.0 and the McCarthy Group tools, whereas imputation of the EstBB dataset was performed using Eagle v2.4.1 and Beagle v5.4 (beagle.22Jul22.46e.jar).<br>SNP-based heritability ($h^2$SNP) and genetic correlations were estimated using Linkage Disequilibrium Score Regression (LDSC v1.0)[32].<br>MiXeR (https://github.com/precimed/mixer) was used to estimate the number of trait-influencing variants and the genetic overlap between traits.<br>Genome-wide association study (GWAS)-by-subtraction analyses were performed using Genomic Structural Equation Modeling (GenomicSEM v0.0.2) in R v3.4.3, using the publicly available code (https://github.com/PerlineDemange/non-cognitive).<br>Independent lead variants were identified through linkage disequilibrium-based clumping using PLINK v1.09 and the 1000 Genomes Project Phase 3 European reference panel (http://www.internationalgenome.org/).<br>Conditional association analyses were performed using the COJO method implemented in Genome-wide Complex Trait Analysis (GCTA)[36].<br>Functional annotation of genome-wide significant variants was performed using FUMA (https://fuma.ctglab.nl/).<br>Genome-wide polygenic scores were computed using PRS-CS, with the 1000 Genomes European reference panel used for allele weighting and PLINK 2.0 used for allele scoring. Association analyses were performed in R.<br>Causal mediation analyses were conducted using the R package mediation. |

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

GWAS summary data for adhd and non_adhd have been deposited in the GWAS catalog with accession no. GCP001881. ADHD GWAS summary statistics from the Psychiatric Genomics Consortium (PGC) are available for download at https://www.med.unc.edu/pgc/download-results/. CUD GWAS summary statistics are available through the Gelernter Lab website (https://medicine.yale.edu/lab/gelernter/). GWAS summary statistics used in the genetic correlation analyses are accessible through the links provided in Supplementary Table 9. Individual-level EstBB data are not publicly available due to data protection regulations. Access is subject to ethical approval and managed through the EstBB. Further information is available at https://urldefense.com/v3/__https://genomics.ut.ee/en/content/estonian-biobank__;!!D9dNQwwGXtA!S-7Mtf6li1bfrGxbW_d0XB7Tmq8Vbz04CXalaxEFYS3muWStRNU7z0RBeDGhyP11cKAlPXAstCtYbW4YJm-b$. The clinical datasets analyzed in the current study are not publicly available because participants did not provide consent for public data sharing and the relevant ethical approvals do not permit public release.

## 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 | Biological sex (as determined by the sex chromosomes) was used as a covariate in the analysis.<br>All included GWAS summary statistics included both male and female subjects. |
| Reporting on race, ethnicity, or other socially relevant groupings | The study includes data on European ancestry individuals. |
| Population characteristics | To achieve adequate statistical power for the GWAS analyses, the ADHD and CUD datasets used in this study combined data from multiple cohorts, each with potentially different population characteristics. Detailed information on the cohorts included in the respective univariate GWAS analyses is provided in the supplementary materials of the corresponding GWAS publications. The current manuscript reports the total sample sizes (cases and controls) and diagnostic definitions for ADHD and/or substance use disorder (SUD).<br>The Estonian Biobank (EstBB) comprises approximately 20% of the adult population of Estonia (N = 212,000; 66% female; mean year of birth, 1971) and has been described previously (https://doi.org/10.1038/s41467-025-58465-3). |
| Recruitment | The present study used secondary data provided by the Estonian Biobank. Participant recruitment has been previously described (https://doi.org/10.1038/s41467-025-58465-3).<br>The clinical datasets were recruited as previously described (https://doi.org/10.1038/s41398-024-02923-x and https://doi.org/10.1038/s41386-020-0664-5. |
| Ethics oversight | Individual level data analysis in the EstBB was carried out under ethical approval 1.1-12/3455 from the Estonian Committee on Bioethics and Human Research (Estonian Ministry of Social Affairs), using data according to release application 6-7/GI/17033 from the EstBB.<br>The study in the clinical datasets was approved by the Clinical Research Ethics Committee (CREC) of Hospital Universitari Vall d'Hebron and methods were performed in accordance with the relevant guidelines and regulations. Written informed consent was obtained from all participants or their parents/caregivers before inclusion in the study. |

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 | For the adhd and non_adhd GWAS, the effective sample sizes were estimated, following Demange et al 2021 (https://doi.org/10.1038/ |

| | |
|---|---|
| Sample size | s41588-020-00754-2)<br>For GWAS summary statistics used in the study sample sizes are available in Supplementary Table 1 and Supplementary Table 9.<br>For the clinical datasets sample sizes are avaialble in the "Clinical datasets: sample description, clinical assessment and genotyping" section of the manuscript. |
| Data exclusions | GWAS are based only on participants with European ancestry. For the PGS analyses, all available data was used, participants with a non-European ancestry were excluded, due to the low portability of PGS between different ancestry populations. |
| Replication | No replication was done. Polygenic score predictions were performed in two clinical cohorts. oChort-specific and meta-analytic estimates are reported. |
| Randomization | NA |
| Blinding | NA |

# 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 | *Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.* |
| Novel plant genotypes | *Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.* |
| Authentication | *Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosiacism, off-target gene editing) were examined.* |