nature research

Corresponding author(s): Jacqueline M. Vink

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	For this study, we meta-analyzed genetic and phenotype data from 3 different cohorts. Data collection procedures differed per cohort and are summarized in the Methods section (for details on data collection in the ICC cohorts see Stringer et al., 2016). In general, DNA was extracted from blood or saliva samples, genotyped, and imputed using European reference data. Phenotype information was collected using paper-and-pencil or online surveys.
Data analysis	PLINK 2.0- genomewide association tool; MAGMA v1.06- gene-based tests; S-PrediXcan - gene expression analysis; LD score regression - genetic correlations and heritability; R qqman - visualisation of GWAS results; LocusZoom - creation of regional plots; GTEx Analysis Release V7 - eQTL; METASOFT - eQTL forest plot; MR-Base - mendelian randomization; R gsmr - mendelian randomization; METAL - meta-analysis *all software mentioned here is publicly available

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

Summary statistics based on the UK-Biobank and ICC samples and full results from the top 10,000 SNPs based on all three subsamples (i.e. including the 23andMe sample) will be available via LDhub (http://ldsc.broadinstitute.org/gwashare/). Codes and scripts are available upon reasonable request. Full summary statistics can only be provided after permission by 23andMe.

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	A previous GWAS of the International Cannabis Consortium with a sample size of ~33,000 had limited power to detect genomewide significant hits. We expanded this sample with all available data that we had access to, resulting in a 5-fold increase in sample size, providing sufficient power to detect genomewide signals.
Data exclusions	Extensive quality control procedures were used to select valid SNPs and individuals using pre-established criteria. These have been described in Supplementary Table S12 and include exclusion of related individuals and individuals with missing data, variants with a low HWE, a low minor allele frequency, a low imputation quality score, or high missingness rates, and variants whose alleles and allele frequency differ from those in reference panels. For secondary analysis, sometimes a subset of the genome-wide data was used (i.e., SNPs that could be mapped to a gene in gene-based tests, SNPs that were present in reference files that were used by LocusZoom, LDscore regression, or S-PrediXcan).
Replication	We did not divide our sample into a discovery and replication sample, so that we had one large sample with sufficient power to detect a genome-wide signal. We have been as transparant as possible about our methodology and summary statistics will be made available, so that replication can be attempted by other research groups.
Randomization	N/A; we did not use an experimental design.
Blinding	N/A; we did not use an experimental design. Analists were not blind to case-control status.

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	The sample included N=184,765 individuals, with 55,5% females and a mean age of 35.7 (range 16-87). Only individuals from

Population characteristics	European ancestry were included. Using principal components for population stratification, we controlled for systematic ancestry differences within this European cohort.
Recruitment	Recruitment strategies differed for the 3 cohorts (ICC, 23andMe, UKB) and within the different cohorts making up the ICC cohort (Stringer et al., 2016). 23andMe is a commercial platform where individuals can have their DNA genotyped at their own costs, and can provide permission to make their material available for research. UKB and ICC participants are volunteer samples. Information about recruitment can be found in our supplementary material and in the supplementary material of Stringer et al., 2016.

Stringer, S. et al. Genome-wide association study of lifetime cannabis use based on a large meta-analytic sample of 32 330 subjects from the International Cannabis Consortium. Translational psychiatry 6, e769, doi:10.1038/tp.2016.36 (2016).