nature portfolio

Corresponding author(s): Steven Gazal

Last updated by author(s): Mar 14, 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	No software was used for data collection.
Data analysis	Our code is available at https://alkesgroup.broadinstitute.org/cS2G/code and https://doi.org/10.5281/zenodo.6353668. LD score regression software (ldsc) version 1.0.1 is available at https://github.com/bulik/ldsc PolyFun software version 1.0.0 is available at https://github.com/omerwe/polyfun Data analyses were performed using R version 3.6.1.

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

The List of 19,995 genes, summary statistics of the 63 independent traits, training and validation critical gene sets, S2G and cS2G strategies, SNP annotations, predicted causal SNP-disease pairs from UK Biobank fine-mapping analyses and from the NHGRI-EBI GWAS catalog, and SNP-heritability causally explained by SNPs linked to each gene have been made publicly available at https://alkesgroup.broadinstitute.org/cS2G and https://doi.org/10.5281/zenodo.6354007. Links for all datasets used to create S2G strategies are provided in Supplementary Table 26.

Access to the UK Biobank resource is available via application at http://www.ukbiobank.ac.uk/.
GWAS catalog https://www.ebi.ac.uk/gwas/api/search/downloads/full
Open Targets SNP-gene pairs https://raw.githubusercontent.com/opentargets/genetics-gold-standards/master/gold_standards/processed/gwas_gold_standards.191108.tsv
SNP-gene pairs from ref.48 https://urldefense.proofpoint.com/v2/url?u=https-3A__www.dropbox.com_s_kz2c49rpm2yanf5_all-5FbyCS-5Frev1.txt-3Fdl-3D0&d=DwMFaQ&c=WO-RGvefibhHBZq3fL85hQ&r=pj2hZETq-6Xv2-wuSquXm871XqnKfXGPV5duZ9gf88w&m=IWrDyJkE3HPLhLS3pLXW8e80amAyxaNTtTw2ULHrbLA&s=A8lWwTxhGFCV0avT0-G3s0X7Cs1TXlFqvPx8woQ6iiU&e=

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	We used the larger sample size that was available.
Data exclusions	We restrict our analyses on individuals from European ancestry. We excluded the MHC region from S-LDSC analyses and analyzed only autosomes, as recommended and developed by S-LDSC authors.
Replication	S2G evaluations were replicated on three curated disease-associated lists of SNP-gene pairs. We used training and validation critical gene sets to train and validate the combined S2G strategy. We compared the proportion of heritability linked to genes explained by genes ranked by top per-gene heritability by using per gene heritability that has been used for the ranking (using N=337K British UK Biobank training samples) and by running S-LDSC on summary statistics computed from N=122K European-ancestry UK Biobank validation samples that were distinct from the N=337K British UK Biobank training samples (to avoid winner's curse).
Randomization	Analyses were performed by maximizing the datasets and samples investigated; randomization is inapplicable here.
Blinding	Blinding is not possible in this study: we did not collect data for this study, but analyzed publicly available summary statistics and functional datasets.

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