nature portfolio

Corresponding author(s): Xiao Chang, Guangyong Zhang

Last updated by author(s): Dec 23, 2024

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	The GWAS summary statistics for ten digestive diseases/traits analyses used in this study are deposited in the GWAS Catalog (https://www.ebi.ac.uk/gwas/) and the accession codes are as follows: irritable bowel syndrome (GCST90016564), acute pancreatitis (GCST90255375), non-alcoholic fatty liver disease (GCST90091033), appendicitis (GCST90038695), colorectal cancer (GCST90255675), inflammatory bowel disease (GCST004131), crohn's disease (GCST004132), ulcerative colitis (GCST004133), stool frequency (GCST90002250) and diverticular disease (GCST008105). The GWAS summary statistics for the gastritis-duodenitis and cholelithiasis-cholecystitis are publicly available at https://www.leelabsg.org/resources. The GWAS summary statistics for peptic ulcer used in this study are available at https://cnsgenomics.com/content/data, and the GWAS summary statistics for Gastroesophageal reflux disease are available at 10.6084 / m9.figshare.8986589. The anxiety and neuroticism GWAS summary statistics used in this study are publicly available in the GWAS Catalog under accession code GCST90038651 and GCST007339. The summary statistics for depression are deposited in: https://ipsych.dk/en/research/downloads/. The GWAS summary statistics for seven other psychiatric disorders, including major depressive disorder, schizophrenia, attention-deficit hyperactivity disorder, autism spectrum disorder, bipolar disorder, post traumatic stress disorder, and anorexia nervosa, are available for download at https://pgc.unc.edu/for-researchers/download-results/. GTEx project v.8 data were publicly available at https://gtexportal.org/home/.
Data analysis	The study outlined the use of several publicly accessible software tools to conduct statistical analyses across different programming environments, ensuring the reproducibility and robustness of the research. In the Python 2.7.18 environment, LDSC version 1.0.1 was used to compute genetic correlations and facilitate cell type-specific investigations. For Multi-Trait Analysis of GWAS, MTAG version 1.0.8 was employed, while HESS version 0.5.4-beta was used for Heritability Estimation from Summary Statistics. Per l5.16.3 was utilized to generate Manhattan plots using the CIRCOS tool version 0.69-9. In R 4.2.3, the TCSC package version1.0.0 was applied for identifying causal tissues in disease and complex trait research, and the WGCNA package version 1.72-5 was used for gene co-expression network analysis. Java 17.0.5 facilitated PPI network analysis through Cytoscape version 3.10.1, alongside the string App plugin version 2.0.1. Additionally, MAGMA

software version 1.10 was used for gene analysis and generalized gene-set analysis of GWAS data, capable of processing both raw genotype information and summary SNP p-values from previous GWAS or meta-analyses. These tools collectively enhance the reproducibility of the research and highlight the significance of accessible and well-documented software in scientific studies. To address the issue of multiple testing, we applied the Bonferroni method in LDSC, TCSC, and WGCNA.

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 dataset used in this study was obtained from the public database.

Human research participants

Policy information about studies involving human research participants and Sex and Gender in Research.

Reporting on sex and gender	This study does not differentiate between genders.
Population characteristics	GWAS data and eQTL data for Europeans used in this study.
Recruitment	None.
Ethics oversight	GWAS Catalog.

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	GWAS summary statistics were collected from publicly available data sources. Data for irritable bowel syndrome (IBS) were obtained from a large meta-analysis with 486,601 individuals (53,400 cases and 433,201 controls) . For gastroesophageal reflux disease (GERD), data were collected from the UK Biobank (UKBB) study, which included 71,522 cases and 261,079 controls. This dataset was defined based on ICD10 codes, self-reported GERD, and the use of GERD medication. Additionally, data from the QSkin study, involving heartburn and GERD medication use as recorded in the Pharmaceutical Benefits Scheme (PBS) medical records, were included . An additional dataset for peptic ulcers (PU), involving 16,666 cases, were sourced from the same gastrointestinal tract GWAS conducted on a cohort of 456,327 individuals from the UKBB . GWAS summary statistics for inflammatory bowel disease (IBD, 25,042 cases, 34,915 controls), ulcerative colitis (UC, 12,366 cases, 33,609 controls), and crohn's disease (CD, 12,194 cases, 28,072 controls) were derived from the same IBD GWAS dataset. GWAS summary statistics for gastritis-duodenitis (GD) and “cholelithiasis and cholecystitis” (CL&CC) were derived from the full European data subset from the Lee Lab (https://www.leelabsg.org/resources). In addition, GWAS summary statistics for acute pancreatitis (AP, 10,630 cases, 844,679 controls), appendicitis (APP, 4,089 cases, 480,509 controls), colorectal cancer (CRC, 78,473 cases, 107,143 controls), diverticular disease (DD, 31,964 cases, 419,135 controls), stool frequency (SF, 167,875 participants) and Non-alcoholic fatty liver disease (NAFLD) (12,194 cases, 28,072 controls) were sourced from the GWAS catalog. GWAS summary statistics for neuroticism (NE, 523,783 participants) and ANX (6,514 cases, 478,084 controls) were obtained from the GWAS Catalog. GWAS data for eight additional psychiatric disorders were obtained from the Psychiatric Genomics Consortium (PGC), including DEP (294,322 cases, 741,438 controls) , major depressive disorder (MDD, 170,756 cases, 329,443 controls) , schizophrenia (SCZ, 76,755 cases, 234,649 controls) , bipolar disorder (BIP, 41,917 cases, 371,549 controls) , attention-deficit hyperactivity disorder (ADHD, 38,691 cases, 186,843 controls) , autism spectrum disorder (ASD, 18,381 cases, 27,969 controls) , post traumatic stress disorder (PTSD, 32,428 cases, 174,227 controls) and anorexia nervosa (AN, 16,992 cases, 55,525 controls) .
Data exclusions	None.

Replication	None.
Randomization	Grouping by disease and control.
Blinding	The investigators were blinded to group allocation during data collection.

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

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging