nature portfolio

Corresponding author(s): Hailiang Huang

Last updated by author(s): Jan 20, 2023

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	Computer code relating to this study includes: RICOPILI v2019_Jun_25.001: https://sites.google.com/a/broadinstitute.org/ricopili EIGENSTRAT v6.1.4: https://github.com/DReichLab/EIG/tree/master/EIGENSTRAT bcftools v1.11: http://samtools.github.io/bcftools/bcftools.html LDSC v1.0.1: https://github.com/bulik/ldsc S-LDXR v0.3-beta: https://huwenboshi.github.io/s-ldxr HRC-1000G-check-bim v4.3.0: https://www.well.ox.ac.uk/~wrayner/tools/HRC-1000G-check-bim-v4.3.0.zip VcfCooker v1.1.1: https://genome.sph.umich.edu/wiki/VcfCooker Eagle2 v2.4.1: https://alkesgroup.broadinstitute.org/Eagle/ Minimac4 v1.0.0: https://genome.sph.umich.edu/wiki/Minimac4 apt software v2.10.2.2: https://www.thermofisher.com/us/en/home/life-science/microarray-analysis/microarray-analysis-partners-programs/affymetrix-developers-network/affymetrix-power-tools.html SNPolisher v3.0: https://downloads.thermofisher.com/SNPolisher_3.0.zip BEAGLE v5.1: https://faculty.washington.edu/browning/beagle/b5_1.html PLINK2 v2.00a3.6: https://www.cog-genomics.org/plink/2.0 METAL v2011-03-25: https://genome.sph.umich.edu/wiki/METAL MANTRA v1: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3460225/ COJO v1.92.2beta: https://yanglab.westlake.edu.cn/software/gcta/#COJO PRS-CS v1.0.0: https://github.com/getian107/PRScs PRS-CSx v1.0.0: https://github.com/getian107/PRScsx Python implementation for SuSiE: https://github.com/getian107/SuSiEx

FinnGen QC and Association analysis: https://finngen.gitbook.io/documentation/methods/genotype-imputation/genotype-data
FinnGen GWAS: https://finngen.gitbook.io/documentation/methods/phewas
IEU open GWAS project: https://gwas.mrcieu.ac.uk/phewas/
VEP v104.3: https://useast.ensembl.org/info/docs/tools/vep/index.html
Cytoscape v3.9.1: https://cytoscape.org/

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

CaVEMaN and DAP-G GTEx v8 Fine-Mapping cis-eQTL Data was retrieved from https://gtexportal.org/home/datasets#filesetFilesDiv15.
1000 Genomes Project Phase 3 is available from: https://www.internationalgenome.org/category/phase-3/
TOPMed reference panel R2 is available from: https://imputation.biodatacatalyst.nhlbi.nih.gov/#!
Human Genome Diversity Project is available from: https://www.internationalgenome.org/data-portal/data-collection/hgdp
Simons Genome Diversity Project is available from: https://www.simonsfoundation.org/simons-genome-diversity-project/
Korean Personal Genome Diversity Project is available from: http://opengenome.net/Main_Page
NBDC human database (Accession ID: JGAS000114) is available from: https://humandbs.biosciencedbc.jp/en/
STRING network dataset is available from: https://string-db.org/
NFE summary statistics is from the published research (de Lange et al., Nature Genetics 2017, 10.1038/ng.3760): ftp://ftp.sanger.ac.uk/pub/project/humgen/summary_statistics/human/2016-11-07/
FIN summary statistics is from FinnGen R7: https://www.finngen.fi/en/access_results
PRS weights and genome-wide summary statistics for the meta-analyzed EAS samples, and across all study samples (EAS and EUR) can be downloaded from https://www.ibdgenetics.org.

Individual-level genotype data for EAS samples are available upon request: SHA1: Z.L. (zhanjuliu@tongji.edu.cn), KOR1: K.S. (kysong@amc.seoul.kr), JPN1: Y.Kakuta (ykakuta@med.tohoku.ac.jp) and ICH1: IIBDGC (ibdgc-dcc@mssm.edu). Access to individual-level genotypes from samples recruited within mainland China are subject to the policies and approvals from the Human Genetic Resource Administration, Ministry of Science and Technology of the People's Republic of China.

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	Several studies suggest a sample size of over 10,000 individuals will be needed to properly power this study. All available samples from collaborators were used.
Data exclusions	We used standard quality control procedures for genome-wide association studies. SHA1, KOR1 and ICH1: QC was performed using the RICOPILI pipeline. We first excluded subjects with a mismatch in their reported sex and sex imputed from chromosome X, and updated 98 subjects with no sex reported with the imputed sex. We then performed QC using the following steps. Autosomes: excluding 1) variants with a call rate below 95%; 2) subjects with a call rate below 98%; 3) monomorphic variants; 4) subjects with an inbred coefficient above 0.2 and below -0.2; 5) variants with missing rate differences > 2% between cases (IBD) and controls; 6) variants with a call rate < 98%; and 7) variants in violation of Hardy-Weinberg equilibrium (two-sided) with P<10^−6 in controls or P<10^−10 in case (IBD). Chromosome X: we started with samples that passed the QC using autosomes. Variants in chromosome X from these samples were QC'ed by excluding: 1) monomorphic variants; 2) variants with a call rate below 98% in either male or female; 3) variants with missing rate differences > 2% between cases (IBD) and controls in either males or females; and 4) variants in violation of Hardy–Weinberg equilibrium with P<10^−6 in controls or P<10^−10 in case (IBD) in female. JPN1: Only autosome data was available. QC was performed using the following steps: 1) subjects with a sample call rate < 98% were excluded; 3) monomorphic variants were removed; 4) Variants that were not classified as Recommendations by SNPolisher (Affymetrix, Inc.) were removed; 5) variants with P < 10^-5 in Hardy-Weinberg equilibrium test (two-sided) were removed; 6) Variants with P < 10^-6 in association analysis were visually checked for clustering results, and seven variants which showed poor cluster resolution were excluded. FIN: We used the summary statistics from the FinnGen. QC was performed by the FinnGen (https://finngen.gitbook.io/documentation/methods/genotype-imputation): In sample-wise quality control steps, individuals with ambiguous gender, high genotype missingness (>5%), excess heterozygosity (+-4SD) and non-Finnish ancestry were excluded. In variant-wise quality control steps, variants with high missingness (>2%), low HWE P-value (<1e-6) and low minor allele count (MAC<3) were excluded. We did not further exclude any data.

NFE: We used the summary statistics from a published study (de Lange et al., Nature Genetics 2017). QC for NFE subjects is described in the paper including: researchers removed variants that were not present on either the Human Core Exome v12.1 chip nor the Human Core Exome v12.0 chip, had missingness >5%, had a significant difference in call rate between cases and controls ($P < 1 \times 10-5$), deviated from Hardy–Weinberg equilibrium in controls ($P < 1 \times 10-5$) or were affected by a genotyping batch effect (significant association ($P < 1 \times 10-5$) between an outlier group of cases discovered using principal-component analysis ($PC1 < -0.005$) and the remainder of the samples). Researchers then removed samples who had missingness >1%, heterozygosity ±3 s.d. from the mean, or mismatch between the reported and genotypic sex, samples who were first-degree relatives or closer (kinship coefficient >0.177) and non-European samples identified through principal-component analysis with HapMap 3 populations. We did not further exclude any data.

Replication

The analysis was performed on six independent cohorts that serve the purpose of replication with each other.

Randomization

No randomization was conducted. The current study is a large-scale genetics study. Randomness is achieved from the nature of how alleles are distributed in the population. Alleles are randomly passed from parent to offspring during meiosis, hence the nature of these types of studies tend to be shielded from extraneous confounds of standard epidemiological studies. One exception of this is the population structure, which can confound the random passing of genetic alleles. To resolve the problem, we performed the principal component (PC) analysis to capture the population structure as PCs. We then included PCs in GWAS as covariates to control the confounding driven by the population structure.

Blinding

The nature of the study is in of itself blind to study recruiters. No one would beforehand know the genotype make up of the cases and controls recruited in the current study.

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

Data was obtained from subjects living in East Asia (China, Japan, Korea), North America and Europe. All cases were clinically diagnosed to have IBD. Controls were either unaffected healthy controls from the IBD studies or population based controls from various public resources.

SHA1: All subject are from China. 60.8% of subjects are male, 64.4% of IBD cases are male, 58% of controls are male; 45% of subjects are IBD cases, 22% of subjects are CD cases and 21% of subjects are UC cases; 55% of subjects are controls.

KOR1: All subjects are from Korea. 50% of subjects are male, 65% of IBD cases are male, 40% of control are male; 42% of subjects are IBD cases, 21% of subject are CD cases and 14% of subject are UC cases (Jung et al., J. Crohns. Colitis 2021).

JPN1: All subjects are from Japan. 59% of subjects are male, 63% of IBD cases are male, and 48% of controls are male. 76% of subjects are IBD cases, 36% of subjects are CD cases and 40% of subjects are UC cases.

ICH1: All subjects are recruited from Japan, Korea, and Hong Kong SAR China. 57% of subjects are male, 61% of cases are male, 55% of controls are male; 42% of subjects are IBD cases, 25% of subjects are CD, 24% of subjects are UC (Liu et al., Nature Genetics 2015).

FIN: All subjects are from FinnGen. 44% of subjects in FinnGen biobank R7 are male; 47% IBD case are male; 1.8% of subjects are IBD case, 0.4% of subjects are CD cases, and 1.3% of subjects are UC cases (https://risteys.finngen.fi/).

NFE: 42% of subjects are IBD cases, 20% of subjects are CD cases, and 21% of subjects are UC cases (de Lange et al., Nature Genetics 2017). The published study did not report the distribution of sex of participants.

We do not have age, treatment or past diagnosis information for some cohorts. Those information usually have minimal impact on the quality of large-scale IBD GWAS. They are not commonly included as covariates for IBD GWAS.

Recruitment

SHA1: Subjects were recruited patients from multiple inpatient and outpatient IBD centers in China. The diagnosis of IBD followed either the Second European Evidence-based Consensus or the ECCO-ESGAR Guideline for Diagnostic Assessment in

IBD. We used clinical characteristics, radiological and endoscopic examination, and histological features in the diagnosis. We excluded patients with infectious diseases, other autoimmune diseases, tumors, and indeterminate colitis. A small number of patients were found to have inconsistent CD/UC diagnosis in our later analysis and were reassigned to and treated as IBD-U. Control subjects were recruited from the outpatient service at each recruitment site, who typically visited for routine physical examinations during the period of study. DNA samples were purified from blood using RelaxGene Blood DNA System (Tiangen Biotech, Beijing Co., Ltd., Beijing, China), and genotyped at Beijing CapitalBio Technology Co., Ltd. (Beijing, China) using the Illumina Asian Screening Array (ASA).

KOR1: Recruitment described in Jung et al., J. Crohns. Colitis 2021

JPN1:Recruitment described in Kakuta et al., J. Gastroenterol. 2018; Kakuta et al., J. Crohns. Colitis 2019; Okamoto et al., Inflamm. Bowel Dis. 2020; and Kakuta et al., J. Crohns. Colitis 2021

ICH1: Recruitment described in Liu et al., Nature Genetics 2015

FIN: Recruitment described in Kurki et al., medRxiv 2022

NFE: Recruitment described in de Lange et al., Nature Genetics 2017

Ethics oversight

Written Informed consent and permission to share the data were obtained from all subjects, in compliance with the guidelines specified by the recruiting center's institutional review board.

SHA1 was approved by the Institutional Review Board for Clinical Research of the Shanghai Tenth People's Hospital of Tongji University (SHSY-IEC-4.0/18-33/01). Samples recruited in mainland China were processed and analyzed in a Chinese server by the Chinese co-authors to comply with the Administrative Regulations on Human Genetic Resources (a regulation from the Ministry of Science and Technology of the People's Republic of China). Results from the analyses as aggregated information (e.g., summary statistics), which contain no individual-level nor identifiable data, were used in this study.

KOR1 was approved by the Institutional Review Board of Asan Medical Center (2017-0456).

JPN1 was approved by the Ethics Committee of Tohoku University School of Medicine (2020-1-608).

MGH Institutional Review Board reviewed and approved this study (2013P002634), including the use of ICH1.

Patients and control subjects in FinnGen provided informed consent for biobank research, based on the Finnish Biobank Act. Alternatively, separate research cohorts, collected prior the Finnish Biobank Act came into effect (in September 2013) and start of FinnGen (August 2017), were collected based on study-specific consents and later transferred to the Finnish biobanks after approval by Fimea (Finnish Medicines Agency), the National Supervisory Authority for Welfare and Health. Recruitment protocols followed the biobank protocols approved by Fimea. The Coordinating Ethics Committee of the Hospital District of Helsinki and Uusimaa (HUS) statement number for the FinnGen study is Nr HUS/990/2017. The FinnGen study is approved by Finnish Institute for Health and Welfare (permit numbers: THL/2031/6.02.00/2017, THL/1101/5.05.00/2017, THL/341/6.02.00/2018, THL/2222/6.02.00/2018, THL/283/6.02.00/2019, THL/1721/5.05.00/2019, THL/1524/5.05.00/2020, and THL/2364/14.02/2020), Digital and population data service agency (permit numbers: VRK43431/2017-3, VRK/6909/2018-3, VRK/4415/2019-3), the Social Insurance Institution (permit numbers: KELA 58/522/2017, KELA 131/522/2018, KELA 70/522/2019, KELA 98/522/2019, KELA 138/522/2019, KELA 2/522/2020, KELA 16/522/2020, Findata THL/2364/14.02/2020 and Statistics Finland (permit numbers: TK-53-1041-17 and TK/143/07.03.00/2020 (earlier TK-53-90-20). The Biobank Access Decisions for FinnGen samples and data utilized in FinnGen Data Freeze 7 include: THL Biobank BB2017_55, BB2017_111, BB2018_19, BB_2018_34, BB_2018_67, BB2018_71, BB2019_7, BB2019_8, BB2019_26, BB2020_1, Finnish Red Cross Blood Service Biobank 7.12.2017, Helsinki Biobank HUS/359/2017, Auria Biobank AB17-5154 and amendment #1 (August 17 2020), Biobank Borealis of Northern Finland_2017_1013, Biobank of Eastern Finland 1186/2018 and amendment 22 § /2020, Finnish Clinical Biobank Tampere MH0004 and amendments (21.02.2020 & 06.10.2020), Central Finland Biobank 1-2017, and Terveystalo Biobank STB 2018001.

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