nature research

Corresponding author(s): Lillias Maguire

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	Software employed for the genotype analysis, quality control, phasing, imputation, and kinship analysis was performed for sequencing data of participants in the Michigan Genomics Initiative using the following programs. Genoype analysis: Illumina GenomeStudio (module 1.9.4, algorithm GenTrain 2.0). Variant quality control: PLINK (v1.90). Phasing: SHAPEIT2 (v2. r837). Pairwise kinship analysis: KING (v1.4.2). Imputation: Minimac3 (v1.0.13). Data management was performed in R, version 3.4.3. All software programs employed are available for public use and no custom code was employed.
Data analysis	The DEPICT software employed by the authors for gene prioritization, tissue, and pathway analysis is publically available for download from the Broad Institute (https://data.broadinstitute.org/mpg/depict/). For alternative reference files suitable for 1,000 genomes imputed data, contact Prof. Tune Pers at tune.pers@sund.ku.dk . Other statistics are performed in the language-and-environment R, version 3.4.3, with functions in the default libraries. Likewise, data management was also done in R.

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

The UK BioBank genomic and phenotypic data supporting this publication are publicly available from the Roslin Institute, University of Edinburgh (http://geneatlas.roslin.ed.ac.uk/). The Michigan Genomics Initiative (MGI) genomic and phenotypic data are not publicly available due to restrictions on participant privacy. MGI data can be made available on reasonable request to the corresponding author with permission of the University of Michigan Institutional Review Board.

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	The sample sizes were determined by enrollment of participants in the UKBiobank, European population (n=409,728) and Michigan Genomics Initiative (31,211). The UKBiobank represents one of the largest genomic data resources in the world and the Michigan Genomics Initiative is a large institutional resource which provides an appropriate cohort for replication.
Data exclusions	Only European ancestry populations were included, both to facilitate genomic analysis and due to distinct diverticular phenotypes known to exist in other ethnic groups. Genomic data was quality controlled with the exclusion of poor quality variants, contaminated variants, variants deviating from Hardy-Weinberg equilibrium, and variant call rate <99%. Related individuals were excluded.
Replication	The initial genome-wide association study, carried out in the UK Biobank was replicated in the Michigan Genomics Initiative with replication of several genetic loci of interest within this smaller population. Additionally, replication of genes of interest was assessed in the one pre-existing GWAS of diverticular disease.
Randomization	There was no randomization element to this study. Participants were enrolled and then assessed longitudinally for health outcomes.
Blinding	Patients were assigned to case/control groups based on their clinical development of diverticular disease as encoded in the medical record by treating clinicians not affiliated with the genomic banks. Treating heath care providers were unaware of patient participation in the biobanks and this specific study.

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 UK Biobank cohort includes 500,000 male and female participants ages 40-69 years of age, enrolled in the United Kingdom from 2006-2010.

For the purposes of this study, only participants of European ancestry were included, totaling 409,728, and data was controlled for age, gender, relatedness, and principal components. Participants filled out an extensive baseline questionnaire and provided samples for genome sequencing and basic hematologic and chemical laboratory testing. Subsequent diagnostic codes for inpatient hospital admissions were linked to the participant records.
Participants in the Michigan Genomics Initiative are adult patients of the University of Michigan Health System who provided consent for whole genome sequencing, chemical lab testing, and correlation of these results to their electronic medical records specifically for the purposes of research. In both cohorts, participant data was controlled for age, sex, and principal components.

Recruitment

Patients were recruited to the UKBiobank via mail communications sent to individuals aged 40-69 in the United Kingdom. Participants enrolled in 22 centers across the UK. Participants in the Michigan Genomics Initiative include patients of the University of Michigan Health System enrolled during a health care encounter. The populations are distinct in that only one population was enrolled in a health care setting. Additionally, the age criteria of the UKBiobank exclude the oldest and youngest patients afflicted with diverticular disease.