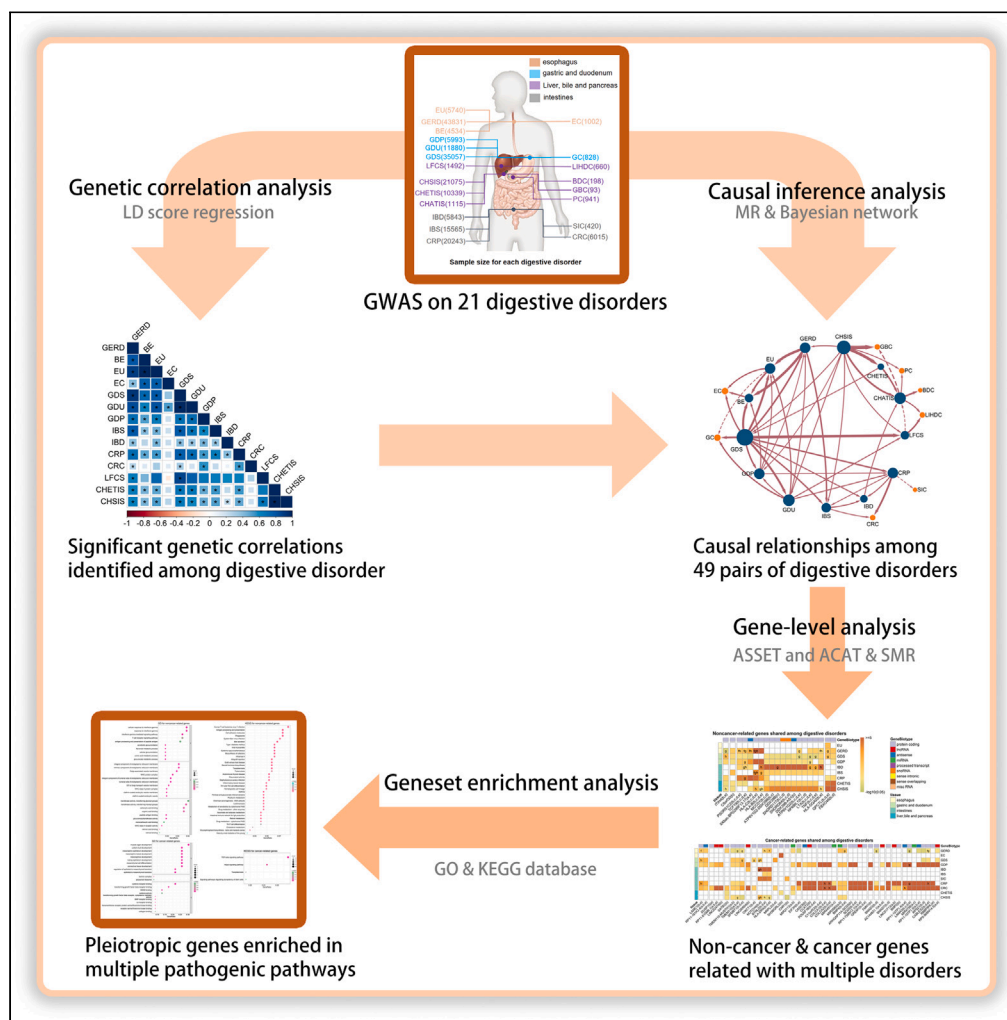


Article

A cross-disorder study to identify causal relationships, shared genetic variants, and genes across 21 digestive disorders



Yue Jiang, Yihong Zhang, Can Ju, ..., Yefei Zhu, Sipeng Shen, Yongyue Wei

ywei@pku.edu.cn (Y.W.)
sshenn@njmu.edu.cn (S.S.)
zhuyf@njmu.edu.cn (Y.Z.)

Highlights

This is a comprehensive study on shared genetic backgrounds of 21 digestive diseases

Genetic correlations and causal relationships among these diseases are revealed

Shared genetic variants and genes inform potential pathogenesis of these diseases



Article

A cross-disorder study to identify causal relationships, shared genetic variants, and genes across 21 digestive disorders

Yue Jiang,^{3,6} Yihong Zhang,^{4,6} Can Ju,² Ruyang Zhang,² Hui Li,⁵ Feng Chen,² Yefei Zhu,^{4,*} Sipeng Shen,^{2,*} and Yongyue Wei^{1,7,*}

SUMMARY

Digestive disorders are a significant contributor to the global burden of disease and seriously affect human quality of life. Research has already confirmed the presence of pleiotropic genetic loci among digestive disorders, and studies have explored shared genetic factors among pan-cancers, including various malignant digestive disorders. However, most cross-phenotype studies within the digestive tract system have been limited to a few traits, with no systematic coverage of common benign and malignant digestive disorders. Here, we analyzed data from the UK Biobank to investigate 21 digestive disorders, exploring the genetic correlations and causal relationships between diseases, as well as the common genetic factors and potential biological pathways driving these relationships. Our findings confirmed the extensive genetic correlation and causal relationship between digestive disorders, providing important insights into the genetic etiology, causality, disease prevention, and clinical treatment of diseases.

INTRODUCTION

Digestive disorders have significantly increased the years living with disability;¹ three digestive malignant neoplasms are ranked into top 10 according to the incidence worldwide, including colorectal cancer (CRC), gastric cancer (GC), and liver cancer (LC).² Among the top 10 cancers with the worst prognosis, half are digestive malignant neoplasms, including colon, gastric, liver, esophageal, and pancreatic cancers.³ Identifying causal factors of development of digestive disorders is crucial for disease prevention. Profound influences of genetic variations on the risk of a broad list of digestive disorders have been studied by large-scale genomic studies, and unveiled genetic loci for Barrett's esophagus (BE) and other digestive disorders.⁴ Notably, a substantial proportion of the heritability is contributed by common variants leading to susceptibility of multiple digestive disorders, emphasizing the complexity and highly polygenic nature of these conditions.^{5,6}

Illustrating the causal relationships of cross-disorders and their shared genes has considerable implications for disease prevention and mechanistic understanding.⁷ Simultaneously, deciphering the functional genomics of shared genetic factors across cross-traits aids in uncovering the biological mechanisms of pleiotropic loci, facilitating the identification of targets for clinical diagnosis, treatment, and drug intervention. Such study has been successfully implemented in psychiatric disorders⁸ and pan-cancer.⁹ For digestive disorders, previous genome-wide association studies have identified several pleiotropic loci that were shared among gastroesophageal reflux disease (GERD) and severe esophageal and colorectal diseases.^{10,11} However, these studies have not delved deeply into the underlying mechanisms and have been limited to a few traits, lacking comprehensive research that systematically covers common benign and malignant digestive disorders.

Here, we present a cross-trait analysis on a broad list of digestive disorders in UK Biobank (UKB). We address three major questions regarding the shared genetic basis of these disorders: 1) causal relationship among these digestive disorders; 2) novel susceptibility loci and annotated genes contributed to the risk of digestive disorders through multiple pathways; and 3) functional explorations of the shared genes.

¹Peking University Center for Public Health and Epidemic Preparedness and Response, Key Laboratory of Epidemiology of Major Diseases (Peking University), Ministry of Education, Beijing 100191, China

²Department of Biostatistics, School of Public Health, Center of Global Health, Nanjing Medical University, Nanjing 211166, China

³Clinical Stem Cell Center, The Affiliated Drum Tower Hospital of Nanjing University Medical School, Nanjing 210008, China

⁴Laboratory Medicine Center, The Second Affiliated Hospital of Nanjing Medical University, Nanjing 210011, China

⁵Department of Gastroenterology, The Affiliated Hospital of Integrated Traditional Chinese and Western Medicine, Nanjing University of Chinese Medicine, Nanjing 210028, China

⁶These authors contributed equally

⁷Lead contact

*Correspondence: ywei@pku.edu.cn (Y.W.), sshen@njmu.edu.cn (S.S.), zhuyf@njmu.edu.cn (Y.Z.)

<https://doi.org/10.1016/j.isci.2023.108238>



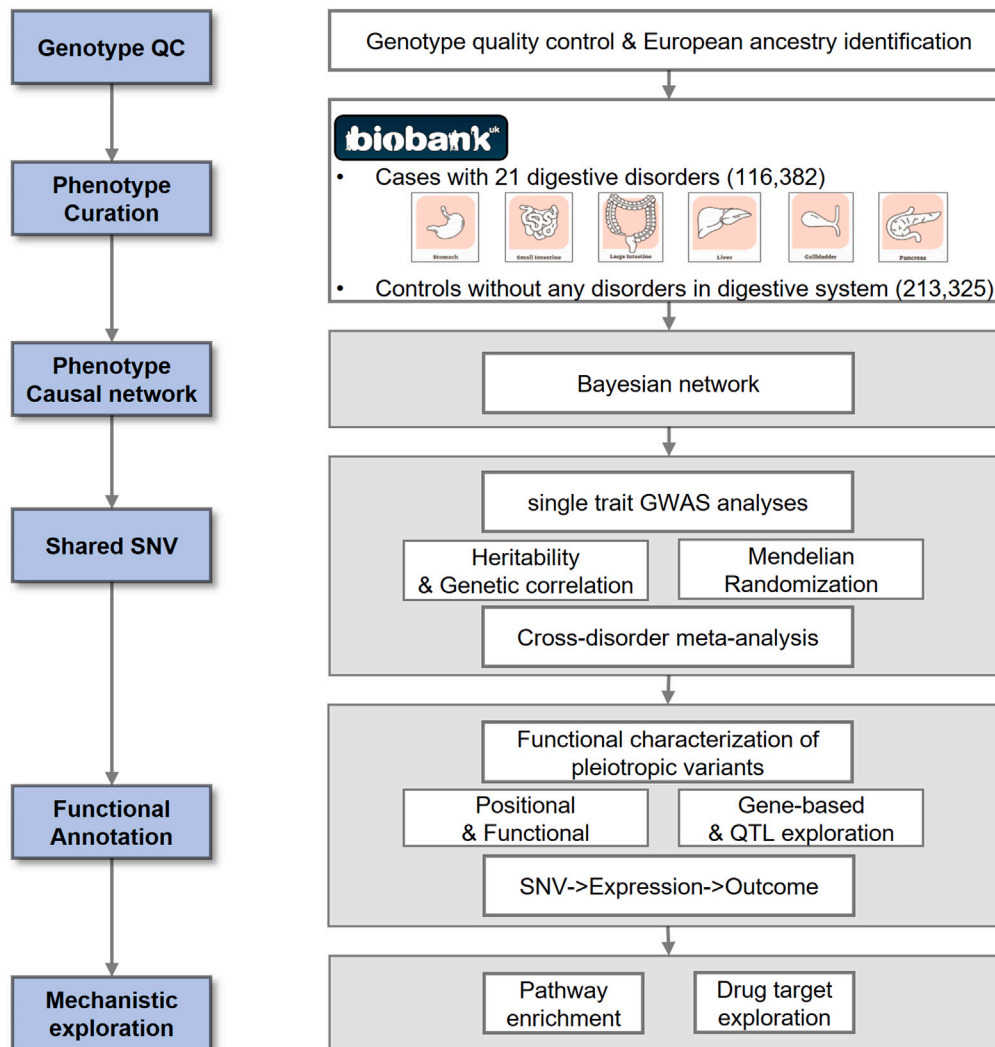


Figure 1. Flowchart of the study

In brief, delineation of the causal relationship of 21 digestive disorders in UKB allows for the identification of shared variants and genes to European populations from different level using the cross-trait approach.

RESULTS

The flowchart of our study is given in Figure 1. Based on the definition of disorders in the UKB (Table S1), 21 types of digestive disorders were included. The number of noncancer disorder cases varied from 1,115 for cholangitis (CHATIS) to 43,831 for GERD, while the number of cancer cases varied from 93 for gallbladder cancer (GBC) to 6,015 for CRC (Figure 2A; Table S2).

Genome-wide association studies of 21 digestive disorders

We conducted genome-wide association studies (GWASs) for 21 digestive disorders. A total of 204 independent variants reached genome-wide significant ($p \leq 5 \times 10^{-8}$) for individual disorders (Figure 2C), of which 13 were associated with two disorders. 113 variants overlapped or had $LD r^2 \geq 0.1$ with the previously identified SNPs, while the remaining 91 variants were novel ($r^2 < 0.1$) (Table S3). 69 novel variants were independent of previously reported variants but in known regions associated with digestive disorders. The top five disorders associated with them were cholelithiasis (CHSIS, 44 novel variants), cholecystitis (CHETIS, 10 novel variants), gastric and duodenal polyp (GDP, 3 novel variants), colorectal polyp (CRP, 3 novel variants), and GBC (3 novel variants). The remaining 22 novel variants were in the region that was not previously reported for any digestive disorder.

We estimated the SNP-based heritability (h^2_{SNP}) using linkage disequilibrium (LD) score regression on both the observed scale and liability scale, assuming the proportion of the cases in the sample as the disease lifetime risk estimates (Table S4). Among the 15 digestive disorders

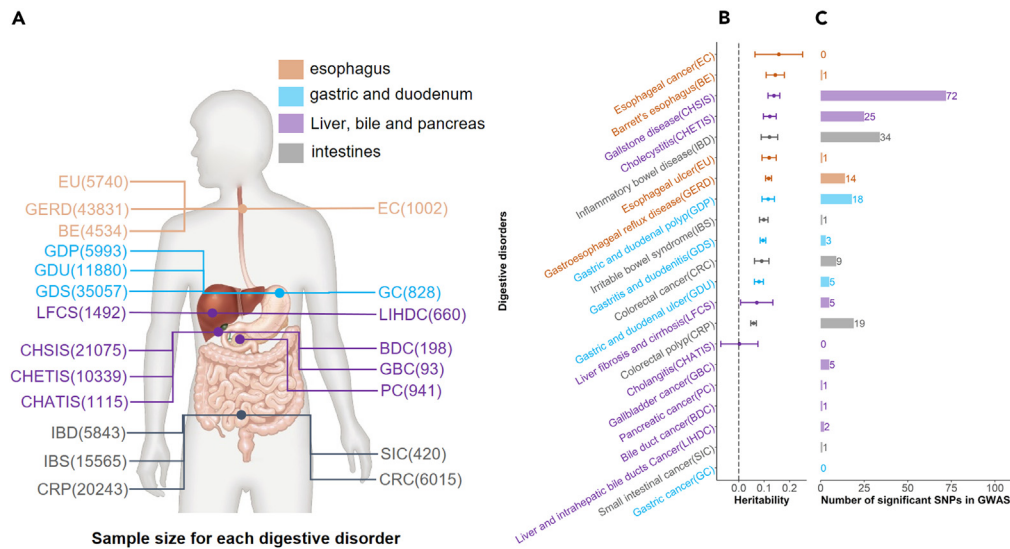


Figure 2. The 21 digestive disorders, heritability, and GWAS findings

(A) Digestive disorders presented by anatomical location and their sample size.
 (B) Dot plot indicates heritability estimates and 95% confidence interval for the digestive disorders having sample size over 1000.
 (C) Bar plot indicates the number of significant index SNPs ($p < 5 \times 10^{-8}$) from GWAS analyses.

with over 1,000 cases, 14 had significant genetic heritability. The h^2_{SNP} estimates ranged from 5.83% for CRP to 15.75% for esophageal cancer (EC), except for CHATIS (Figure 2B).

36 pleiotropic LD blocks were defined using LD clumping procedure according to the 204 SNPs identified in GWAS analyses (Figure S1; Table S8). 20 pleiotropic blocks had direct evidence (index SNPs were previously reported the association with risk of digestive disorders) or indirect evidence (index SNPs had high LD ($r^2 > 0.1$) with the SNPs previously reported the association with risk of digestive disorders) that were associated with the corresponding disorders.

Genetic causal relationships across 21 digestive disorders

Among 91 pairs of the 14 disorders with significant heritability, 64 pairs showed positive genetic correlation with Bonferroni correction ($p \leq 0.05/91$), and the other 18 pairs had nominally significant genetic correlations ($p \leq 0.05$), indicating considerable genetic basis of complex relationships among these disorders (Figure 3A; Table S5). Moreover, Bayesian network analysis obtained 53 high-confidence causal relationships among the digestive disorders (Table S6), of which 32 were positively correlated in pairwise genetic correlation analyses. Non-cancerous digestive disorders showed complex pathogenic interactions and, in turn, to multiple types of digestive cancers (Figure 3B).

Further, to validate the causal relationships inferred by Bayesian network, we carried out Mendelian randomization (MR) for all the relationships on the network (Table S7). The relationships, illustrated as 49 arcs on the network that were replicated in MR analysis with statistical significance ($p \leq 0.05$), were retained on the network (Figures 3B and 4, and Table S6). What's more, 48 retrained arcs relationships were additionally confirmed using generalized summary data-based MR method, of which 42 relationships were further confirmed using median-MR method and maximum-likelihood method¹² (Table S6).

Pleiotropic genetic variants of 21 digestive disorders

To provide further evidence for shared genetic factors, we performed a cross-disorder meta-analysis for the GWAS analyses on the 21 digestive disorders using ASSET. Of the 7,337 variants with $p \leq 1 \times 10^{-4}$ in the single-disorder GWASs, 539 variants were pleiotropic that passed the genome-wide significant threshold ($P_{meta} \leq 5 \times 10^{-8}$) (Figure 5; Tables S9 and S10); 176 of them were reported (Table S12). 75% (4404/539) and 74% (398/539) of the identified pleiotropic variants were related to gallbladder disorders including CHSIS and CHETIS, respectively. GERD, which had the largest number of cases, was associated with 62% (332/539) of the pleiotropic variants.

Among 539 pleiotropic SNPs, 498 exhibited effect with consistent direction, while 41 showed effect with different directions. Further, 114 LD blocks were shared between disorders using the same criteria in GWAS (Table S11).

Functional characterization of pleiotropic variants

We annotated SNPs by their physical location on genomic sequence. 28 of 539 SNPs (5%) were in exon region (Figure S2A), of which 22 were nonsynonymous. This result was consistent with previous findings that most (~93%) disease-associated SNPs in the GWAS Catalog were in non-coding regions.¹³

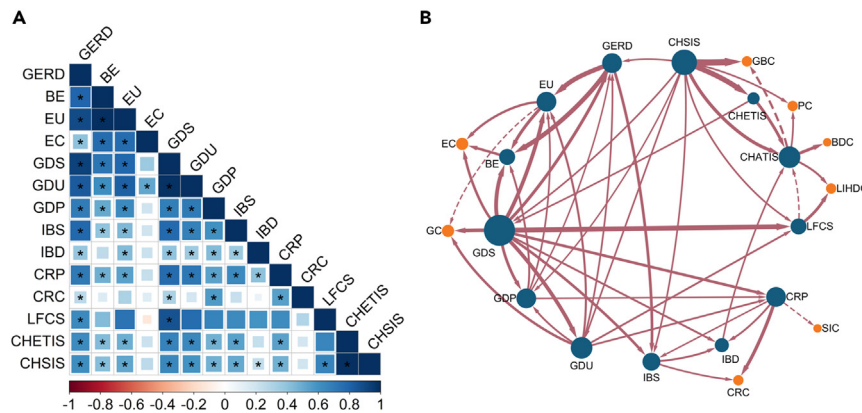


Figure 3. Genetic correlation and causal inference by Bayesian network

(A) Results for genetic correlation among the 14 digestive phenotypes with sample size over 1000. “*” represents genetic correlation significant after Bonferroni correction ($p < 0.05/91$). The color and size of the square scales with the correlation of pairwise of disorders.

(B) Causal network comprised 21 disorders constructed based on the intersectional results of Bayesian network and Mendelian randomization analysis to reveal complex genetic relationships. Orange nodes indicate cancers and dark blue nodes indicate noncancerous digestive disorders, with edges indicating the estimates of the IWW methods.

To comprehensively investigate the regulatory roles of the variants, we systematically annotated all the SNPs in aspect of functional basis (Figure S2B). Specifically, 128 SNPs interacted with the target genes through 3D chromatin loops, 209 SNPs located in or near super-enhancers/promoters, and 335 SNPs acted as eQTLs with the target genes. The 394 SNPs (73%) were annotated with at least one type of functional categories, and 71 SNPs (13%) had functional support from all three types of information. Detailed annotations were provided in Tables S9 and S10. The enrichment of functional annotations for these SNPs suggested that the pleiotropic SNPs might play essential roles in digestive disorders through functional regulation. Unique genes can be annotated from different sources (Figure S2C). Finally, we obtained 1,381 candidate genes for the following analysis.

Pleiotropic genes shared among digestive disorders

We tested 1,381 candidate genes that showed clues in the cross-trait meta-analysis according to the SNP annotation results (Table S13). 736 genes located at 146 independent genetic regions showed pleiotropic effects on 14 disorders after false discovery rate (FDR) correction for multiple testing ($FDR-q \leq 0.05$) (Figure S3A; Table S13). About half of these pleiotropic genes (369/736) were associated with at least three digestive disorders in ACAT analysis.

To better understand the pathogenic mechanism of these genes, we divided these 736 pleiotropic genes into two categories according to the results of ACAT: 690 noncancer-related pleiotropic genes and 46 cancer-related pleiotropic genes. The top noncancer-related pleiotropic genes were *MCCD1*, *ATP6V1G2*, and *LTA* that shared among seven digestive disorders in ACAT analysis, followed by 17 genes which were detected in six digestive disorders (Figure 6A). Notably, most of the top genes (*PSORS1C2*, *TCF19*, *XXbac-BPG299F13.17*, *HLA-C*, *HLA-B*, *MCCD1*, *ATP6V1G2-DDX39B*, *DDX39B*, *SNORD117*, *SNORD84*, *DDX39B-AS1*, *ATP6V1G2*, *NFKBIL1*, *LTA*, *TNF*, and *HLA-DQA2*) were located on 6p21.33 in major histocompatibility complex (MHC) region.

We highlighted 46 cancer-related pleiotropic genes which associated with three digestive cancers, including EC (1 gene), small intestinal cancer (SIC, 4 genes), and CRC (41 genes) (Figures 6C and 6D). Of these genes, 14 were novel, 39 were annotated from position, three were interacted with SNP through 3D chromatin loops, 16 were target genes for super-enhancer/promoter, and 27 were based on eQTLs.

Further, we explored whether the effect of genetic variants on risk of disorders was mediated by alteration of the corresponding genes' expression. The summary data-based MR test was performed on 14 disorders that shared genes identified in ACAT for 1,023 probes that had at least one *cis*-eQTL at $P_{eQTL} \leq 5 \times 10^{-8}$. After HEIDI test, we retained the results of gene-trait pairs that had been identified in the ACAT analysis (Table S14). We identified 65 pairs of association on gene expression and risk of disorder with $FDR-q \leq 0.05$ in the specific tissue, 299 pairs in the whole blood, and 184 pairs in the cross-tissue (Figures S3B–S3D). Of these, 130 genes' expressions were associated with two or more disorders. The analyses proved that the pleiotropic genetic variants and corresponding gene transcription contributed to the risk of multiple digestive disorders.

Functional enrichment analysis of pleiotropic genes

The Gene Ontology (GO) enrichment and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis performed by the noncancer-related genes and cancer-related genes were utilized to explore the shared biological functions and pathways related to digestive disorders (Tables S15, S16, S17, and S18). The top 10 significant GO results and all significant KEGG results are shown in Figure S4.

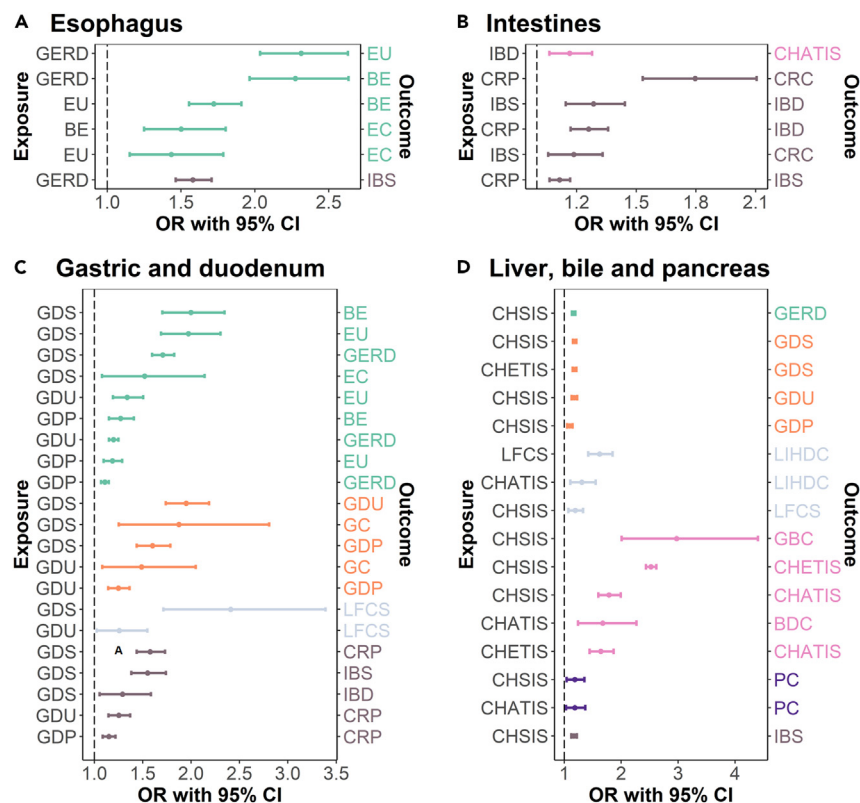


Figure 4. Causal relationships inferred by Mendelian randomization

The causal relationships from one causal disorder (on left y axis) to the other outcome disorder (on right y axis) were presented by the estimates (dot) and 95% confidence interval (horizontal line). A, B, C, and D showed the results of causal disorders at esophagus, gastric and duodenum, liver-bile-pancreas, and intestines, respectively.

For noncancer-related genes, the GO enrichment analysis showed that they were enriched in the biological process (BP) related to chronic inflammation and immune responses, such as cellular response to interferon-gamma. Meanwhile, noncancer-related genes were significantly enriched in cellular component (CC) related to intestinal inflammation,¹⁴ such as integral component of endoplasmic reticulum membrane. For the molecular function, these genes are enriched in MHC class II receptor activity, peptide antigen binding, and glucuronosyltransferase activity. For KEGG pathway, the genes were enriched in pathways which play the potential role in digestive system, such as antigen processing and presentation, bile secretion, and intestinal immune network for IgA production.

For cancer-related genes, the GO enrichment analysis results showed that the top significant BPs were related to epithelial-mesenchymal transition, which is known to be crucial for malignant progression, such as regulation of epithelial to mesenchymal transition.¹⁵ Interestingly, the top CC was laminin complex, which may act as regulators of cancer stem cells, and play an instrumental role in long-term cancer maintenance, metastasis development, and therapeutic resistance.¹⁶ For the KEGG pathway enrichment, the significant results were related to developmental pathways (TGF- β and Hippo) and signaling pathways regulating pluripotency of stem cells.

To summarize, these results mentioned previously suggested that the pleiotropic genes were closely related to digestive system and cancer.

Drug-gene interactions related to digestive disorders

As described previously, we detected 1,812 unique drugs that had drug-gene interactions with the target pleiotropic genes. The top ten genes related to drugs were *EHMT2*, *KCNH2*, *SMAD3*, *FEN1*, *ABCB1*, *MPHOSPH8*, *TNF*, *CNR1*, *UGT1A1*, and *CYP19A1*. Among these drugs, 66 drugs were indicated for digestive disorders (Table S19).

DISCUSSION

This is the first study that comprehensively investigated the causal relationships and shared genetic factors across 21 digestive disorders among 329,707 European individuals of UKB. Specifically, we explored 49 causal relationships among the digestive disorders and detected 539 pleiotropic SNPs enriched for regulatory functions, which mapped to 46 target genes shared across digestive cancers and noncancerous digestive disorders. Our findings provided new insights into the etiology and causality of digestive disorders.

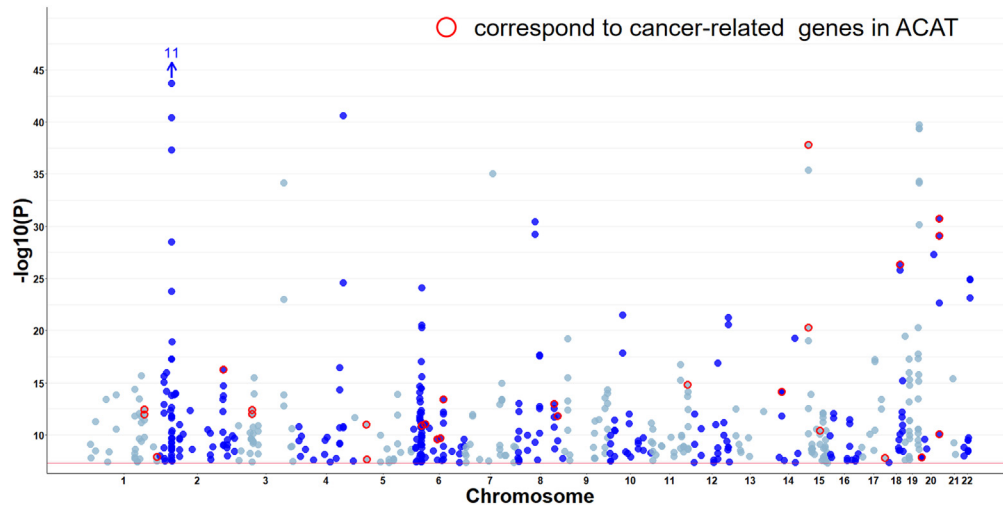


Figure 5. Manhattan plot of cross-disorder meta-analysis

The x axis represents genomic position (chromosomes 1–22), and y axis represents statistical significance in the scale of $-\log_{10}(P)$ for overall test). SNPs with genome-wide significance are shown above the horizontal red line which corresponding to the significant threshold at 5×10^{-8} . The highlighted SNPs in orange are SNPs associated with two subsets of disorders while having opposite association direction; those in yellow are novel SNPs. Red circles represent pleiotropic SNPs that associated with multiple disorders, and the minimum of GWAS p values was used for presentation.

The broad genetic overlap between pairwise disorders reflected the shared genetic across these digestive disorders, and prompting the exploration of phenotypic causal network, which further proved by MR, a sophisticated causal inference method. The 20 disorders were involved in the causal network except for SIC, partly due to insufficient number of cases. For some digestive disorders, we validated several relationships through the methodology of genetic studies that recognized in clinical and experimental studies, such as CHSIS and cancer,¹⁷ BE, and EC.¹⁸ In comparison, we discovered the existed significant genetic correlation between irritable bowel syndrome (IBS) and inflammatory bowel disease (IBD), differing from the previous study which removed IBD cases from IBS cases and may loss the potential overlap.¹⁰ Moreover, some relationships were evaluated by us for the first time, such as IBS and CRC, which provided new clue to genetic basis from both MR and Bayesian causal network. Most importantly, this study also indicates potential causal relationships among noncancerous disorders to cancers, such as from BE to EC, from IBS and CRP to CRC, which may provide evidence for the means forward for cancer prevention and warrants further investigation.

Our study identified pleiotropic LD blocks for digestive disorders, most of which were previously reported in GWAS on digestive disorders, indicating the reliability of our results. Interestingly, a considerable number of blocks were located at 2p21; the leading independent variant (rs56266464) is located at super enhancer of *ABCG5* and *ABCG8* which have role in cholesterol secretion and may contribute to sterol accumulation by mutation,¹⁹ and shared among GERD, gastritis and duodenitis (GDS), CHETIS, and CHSIS. Insights into these complex relationships may inform personalized treatment strategies, guide drug development, and facilitate early diagnosis and risk assessment, ultimately providing more accurate and individualized guidance for clinical decision-making.

Cross-trait meta-analysis detected more than double signals than that in GWAS, significantly increased statistical power, especially for the disorders with small sample size. The top 12 variants were all at 2p21, which was more tissue specific and congregated in hepatobiliary and pancreatic diseases (liver and intrahepatic bile ducts cancer, CHATIS, bile duct cancer, CHETIS, CHSIS, GBC, and pancreatic cancer). The top novel variant 13:29549405:AT:A which was associated with 20 digestive disorders, was an intronic variant located at 13q12.3 and ~50 kb upstream of microtubule-associated scaffold protein 2. This region was previously associated with non-alcoholic fatty liver disease (NAFLD)²⁰ and peptic ulcer.¹⁰ Moreover, this gene had been reported to be capable of regulating entotic cell-in-cell formation, which was described as a nonapoptotic cell death process that occurred in human tumors.²¹ Furthermore, in conjunction with the causal network, the pleiotropic variants could account for plentiful causal pathways. Meanwhile, rs760077, a missense variant of *MTX1* at 1q22 and in high LD ($r^2 = 0.72$) with rs2075570, which has been reported having association with susceptibility of CRC²² and gastric cancer²³ in European population, exhibits a significant association with risks of nine types of digestive disorders (esophageal ulcer, BE, gastric and duodenal ulcer, GC, IBD, CRP, liver fibrosis and cirrhosis, bile duct cancer, and GBC) which are also connected on the causal network.

It is noteworthy that we identified 41 variants that had heterogeneous effects on risk of digestive disorders although some of them had positive genetic correlation and causal relationships. Nine of them are located in the immune-mediated human leukocyte antigen region (6p21.3), which were highly polymorphic and had complex associations with digestive disorders with different pathological conditions.²⁴ Other regions also had several bidirectional variants. Notably, consistent with our results, rs1260326 in the exon of *GCKR* at 2p23.3 was reported to have the opposite effect in gallstone disease (risk allele: T, OR = 0.89)²⁵ compared to that with the other disorders, including NAFLD (risk allele: T, OR = 1.28),²⁶ IBD (risk allele: T, OR = 1.38),^{27,28} Crohn's disease, and ulcerative colitis (risk allele: T, OR = 1.046).²⁹ The heterogeneity of effects among digestive disorders could help the better understanding of cross-trait genetic relationships.

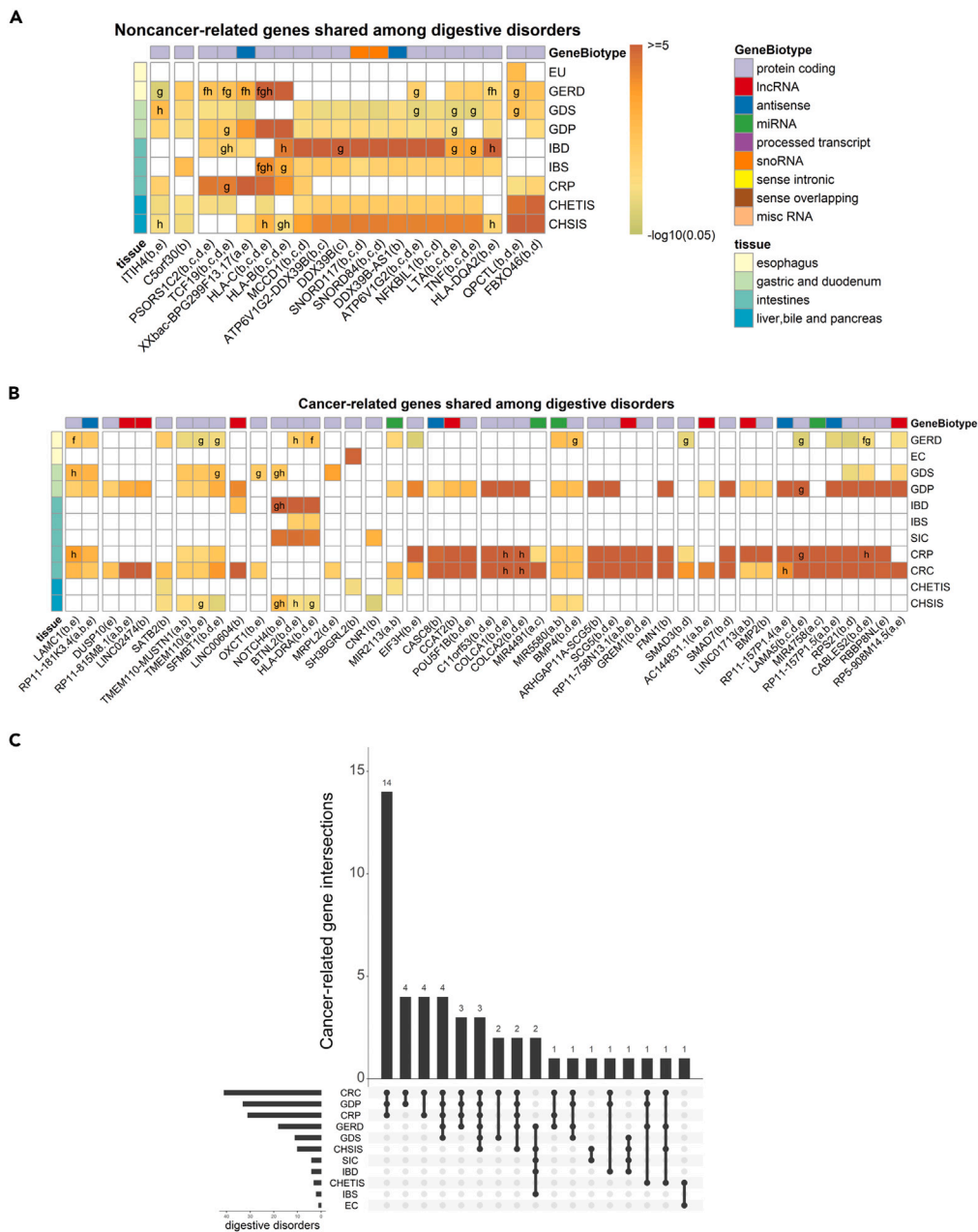


Figure 6. Pleiotropic genes annotation

(A) describes the noncancer-related genes shared among digestive disorders in ACAT.

(B) shows the cancer-related genes shared among digestive disorders in ACAT. Color of the labels in A and B indicates biological type and tissue of the gene.

(C) is the upset plot showing the overlap of pleiotropic cancer-related genes identified in the gene-based analysis for different digestive disorders.

Approach of the genes annotated.

- a : annotated as novel gene.
- b : positional mapped from SNPs.
- c : mapped through three-dimensional chromatin looping.
- d : mapped to super enhancer/promoter.
- e : mapped by eQTL relationship.
- f : identified by organ specific SMR analysis.
- g : identified by whole blood SMR analysis.
- h : identified by cross-organ SMR analysis.

Notably, the pleiotropic variants were annotated in both positional and functional aspects to maximize the list of potential genes involved in the risk of gastric disorders, especially for the noncoding variants.³⁰ In the gene-level analyses, abundant significant genes had been shown to play important roles in the digestive disorders' pathogenesis. *ATP6V1G2*, which is linked to seven types of digestive disorders including GERD, GDS, GDP, IBD, IBS, CHETIS, and CHSIS, plays a significant role in human energy metabolism and induces oxidative stress, and had been considered as the risk gene for CRC.³¹ *LTA* Lymphotoxin alpha, corresponding to the same list of seven disorders described previously, a member of the tumor necrosis factor family, is among the master regulators of intestinal lymphoid development³² and was suggested to play a bigger part in esophageal metaplasia.³³ Inter-alpha-trypsin inhibitor heavy chain 4, which was associated with six digestive disorders including GERD, GDS, GDP, CRP, CHETIS, and CHSIS, located on 3p21.1, has been reported related to growing early colorectal adenomas.³⁴

Moreover, we highlighted the 46 genes that could drive digestive disorders to cancers. Notably, the top three protein-coding genes (*TMEM110-MUSTN1*, *TMEM110*, and *SFMBT1*) shared among six disorders, including GERD, GDS, GDP, CRP, CHSIS, and CRC. *TMEM110-MUSTN1* is novel for digestive disorder, but has been identified as the putative marker for lung adenocarcinoma.³⁵ *TMEM110*, also known as *STIMATE*, was novel but was found to be a regulator of STIM1 activation, which could promote tumor growth and metastasis in a variety of cancer types.³⁶ Another reported gene, *SFMBT1*, had been verified the potential oncogenic function using *in vitro* functional assays in multiple CRC cells.³⁷ These findings provide a deeper understanding of the genetic mechanisms and pathogenesis underlying digestive disorders. Worth further exploration is whether these genes share similar genetic pathways among different digestive disorders, and whether their roles vary across different organs.

Functional enrichment analysis showed that genes related to benign digestive tract traits were mainly enriched in pathways related to chronic inflammation and immune response, which was basically consistent with the previous review.³⁸ However, genes related to malignant digestive tract traits are enriched in important signal pathways related to carcinogenesis, including TGF- β and Hippo signal pathways regulating stem cell pluripotency. Among them, several crosstalk modes between TGF- β family signal and Hippo signal have been proved to regulate the proliferation, invasion, and migration of cancer cells.³⁹ In this study, cancer targeting-related genes enriched in these pathways also affect other precancerous lesions, which may provide clues for the identification of cancer progression and metastasis. At the same time, the genes that jointly drive benign and malignant diseases may provide new insights into disease prevention and clinical treatment.

In this study, the cancer-related genes enriched in these pathways were found to impact other benign disorders, possibly contributing to a pro-oncogenic environment. This provides insights into potential clues for identifying cancer progression and metastasis, indicating shared biological mechanisms among diseases that might influence the development of cancer. By delving into the molecular mechanisms of these shared effects through gene and pathway analysis, we can infer and identify key factors that may influence cancer progression and metastasis. Simultaneously, the remarkable consistency of driver genes across different diseases in this study may offer novel insights for clinical treatment and disease prevention. If these driver genes maintain consistency across multiple diseases, they could play pivotal roles in the pathological processes of various conditions. This opens doors to opportunities for developing treatment methods and preventative strategies targeting these genes, introducing new possibilities for disease management and intervention. We also provided further evidence supporting existing drugs for the treatment of digestive disorders. For instance, several experimental researches had identified the importance of *EHMT2* (also known as *G9a*) in multiple digestive disorders including GC,⁴⁰ LC,⁴¹ and CRC.⁴²

Our study has several strengths. First, this is a comprehensive study to investigate the relationships among a broad list of digestive disorders, in both phenotypic and genetic aspects, from single disorder to multiple disorders, and from statistical association to medical causality. This comprehensiveness not only enriches our understanding of the interactions between diseases, but also provides insight into the development of more accurate prevention and treatment strategies. Second, our study uncovers several novel and crucial genetic variants and genes that contribute to the causal pathway of multiple digestive disorders, which provide new clues for further mechanistic and functional research. Third, we explain the relationship between chronic diseases and gastrointestinal tumors through gene-based analysis and identify pleiotropic genes with remarkable biological functions. As we know, gastrointestinal tumors may develop from chronic diseases. Thus, focusing on the shared genes among them may help identify the high-risk population carrying risk alleles that are more susceptible to cancers.

In summary, our study substantiates the extensive genetic correlations and causal relationships among 21 digestive disorders, identifying shared genetic factors and elucidating the underlying biological mechanisms among these conditions. These findings provide insights into the etiology, causal relationships, and potential drug targets for clinical interventions.

Limitations of the study

We also acknowledge the limitations of this study. First, number of cases for individual disorders varied from 93 for GDC to 43,831 for GERD. Sample size of part disorders was small, which limited power to detect pleiotropic effects. Additionally, the imbalance in sample sizes potentially results in an inflation of type I error rates. Second, we included only individuals of European ancestry to avoid potential confounding due to ancestral heterogeneity across distinct disorder studies. It is essential to evaluate the signals in non-European populations. Third, functional clues of this study were bioinformatics explorations using public databases which warrant well-designed experimental studies in future. Fourth, this study does not specifically explore the role of epigenetic factors, and requires more in-depth correlation analysis.

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- **KEY RESOURCES TABLE**
- **RESOURCE AVAILABILITY**
 - Lead contact
 - Materials availability
 - Data and code availability
- **EXPERIMENTAL MODEL AND SUBJECT DETAILS**
 - Study population
 - GWAS statistics
- **METHOD DETAILS**
- **QUANTIFICATION AND STATISTICAL ANALYSIS**
 - Genetic heritability and genetic correlation
 - Inference of Bayesian causal network
 - Mendelian randomization analysis
 - Cross-disorder GWAS meta-analysis
 - Functional annotation and gene-mapping of pleiotropic variants
 - Gene-based association analysis
 - SMR analysis for candidate genes
 - Pathway enrichment analysis and drug target exploration

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at <https://doi.org/10.1016/j.isci.2023.108238>.

ACKNOWLEDGMENTS

We thank all the participants of UK Biobank for their contribution of valuable biological samples and data.

This work was supported by the National Natural Science Foundation of China (81973142 to Y.W., 82103946 to S.S., 82220108002 to F.C., and 82273737 to R.Z.) and Social Development Project of Jiangsu Province (BE2022817 to H.L.).

AUTHOR CONTRIBUTIONS

Y.W. was responsible for the study conception; Y.J., Y. Zhang, S.S., Y. Zhu, and Y.W. performed study design, data analyses, functional interpretation, and manuscript draft; C.J., H.L., R.Z., and F.C. contributed to the discussion and critical revision of the manuscript. All authors approved the final version of the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: April 19, 2023

Revised: June 26, 2023

Accepted: October 13, 2023

Published: October 16, 2023

REFERENCES

1. GBD 2016 Disease and Injury Incidence and Prevalence Collaborators (2017). Global, regional, and national incidence, prevalence, and years lived with disability for 328 diseases and injuries for 195 countries, 1990–2016: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet* 390, 1211–1259.
2. Global Burden of Disease Cancer Collaboration, Fitzmaurice, C., Abate, D., Abbasi, N., Abbastabar, H., Abd-Allah, F., Abdel-Rahman, O., Abdelalim, A., Abdoli, A., Abdollahpour, I., et al. (2019). Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-Years for 29 Cancer Groups, 1990 to 2017: A Systematic Analysis for the Global Burden of Disease Study. *JAMA Oncol.* 5, 1749–1768.
3. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R.L., Torre, L.A., and Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA A Cancer J. Clin.* 68, 394–424.
4. Su, Z., Gay, L.J., Strange, A., Palles, C., Band, G., Whiteman, D.C., Lescai, F., Langford, C., Nanji, M., Edkins, S., et al. (2012). Common variants at the MHC locus and at chromosome 16q24.1 predispose to Barrett's esophagus. *Nat. Genet.* 44, 1131–1136.
5. Ek, W.E., Levine, D.M., D'Amato, M., Pedersen, N.L., Magnusson, P.K.E., Bresso, F., Onstad, L.E., Schmidt, P.T., Törnblom, H., Nordenstedt, H., et al. (2013). Germline genetic contributions to risk for esophageal adenocarcinoma, Barrett's esophagus, and gastroesophageal reflux. *J. Natl. Cancer Inst.* 105, 1711–1718.
6. Dai, J., Shen, W., Wen, W., Chang, J., Wang, T., Chen, H., Jin, G., Ma, H., Wu, C., Li, L., et al. (2017). Estimation of heritability for nine common cancers using data from genome-wide association studies in Chinese population. *Int. J. Cancer* 140, 329–336.

7. Nielsen, J.B., Rom, O., Surakka, I., Graham, S.E., Zhou, W., Roychowdhury, T., Fritsche, L.G., Gagliano Taliun, S.A., Sidore, C., Liu, Y., et al. (2020). Loss-of-function genomic variants highlight potential therapeutic targets for cardiovascular disease. *Nat. Commun.* **11**, 6417.
8. Cross-Disorder Group of the Psychiatric Genomics Consortium Electronic address plee0@mgh.harvard.edu (2019). Genomic Relationships, Novel Loci, and Pleiotropic Mechanisms across Eight Psychiatric Disorders. *Cell* **179**, 1469–1482.
9. Rashkin, S.R., Graff, R.E., Kachuri, L., Thai, K.K., Alexeeff, S.E., Blatchins, M.A., Cavazos, T.B., Corley, D.A., Emami, N.C., Hoffman, J.D., et al. (2020). Pan-cancer study detects genetic risk variants and shared genetic basis in two large cohorts. *Nat. Commun.* **11**, 4423.
10. Wu, Y., Murray, G.K., Byrne, E.M., Sidorenko, J., Visscher, P.M., and Wray, N.R. (2021). GWAS of peptic ulcer disease implicates *Helicobacter pylori* infection, other gastrointestinal disorders and depression. *Nat. Commun.* **12**, 1146.
11. An, J., Gharakhani, P., Law, M.H., Ong, J.S., Han, X., Olsen, C.M., Neale, R.E., Lai, J., Vaughan, T.L., Gockel, I., et al. (2019). Gastroesophageal reflux GWAS identifies risk loci that also associate with subsequent severe esophageal diseases. *Nat. Commun.* **10**, 4219.
12. Burgess, S., Butterworth, A., and Thompson, S.G. (2013). Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet. Epidemiol.* **37**, 658–665.
13. Maurano, M.T., Humbert, R., Rynes, E., Thurman, R.E., Haugen, E., Wang, H., Reynolds, A.P., Sandstrom, R., Qu, H., Brody, J., et al. (2012). Systematic localization of common disease-associated variation in regulatory DNA. *Science* **337**, 1190–1195.
14. Eugene, S.P., Reddy, V.S., and Trinath, J. (2020). Endoplasmic Reticulum Stress and Intestinal Inflammation: A Perilous Union. *Front. Immunol.* **11**, 543022.
15. Dongre, A., and Weinberg, R.A. (2019). New insights into the mechanisms of epithelial-mesenchymal transition and implications for cancer. *Nat. Rev. Mol. Cell Biol.* **20**, 69–84.
16. Qin, Y., Rodin, S., Simonson, O.E., and Hollande, F. (2017). Laminins and cancer stem cells: Partners in crime? *Semin. Cancer Biol.* **45**, 3–12.
17. Stinton, L.M., and Shaffer, E.A. (2012). Epidemiology of gallbladder disease: cholelithiasis and cancer. *Gut Liver* **6**, 172–187.
18. Evans, J.A., Carlotti, E., Lin, M.L., Hackett, R.J., Haughey, M.J., Passman, A.M., Dunn, L., Elia, G., Porter, R.J., McLean, M.H., et al. (2022). Clonal Transitions and Phenotypic Evolution in Barrett's Esophagus. *Gastroenterology* **162**, 1197–1209.
19. Sarkadi, B., Homolya, L., Szakács, G., and Várdi, A. (2006). Human multidrug resistance ABCB and ABCG transporters: participation in a chemotherapeutic defense system. *Physiol. Rev.* **86**, 1179–1236.
20. Chalasani, N., Guo, X., Loomba, R., Goodarzi, M.O., Haritunians, T., Kwon, S., Cui, J., Taylor, K.D., Wilson, L., Cummings, O.W., et al. (2010). Genome-wide association study identifies variants associated with histologic features of nonalcoholic fatty liver disease. *Gastroenterology* **139**, 1567–1576.
21. Overholzer, M., Mailleux, A.A., Mounieimne, G., Normand, G., Schnitt, S.J., King, R.W., Cibas, E.S., and Brugge, J.S. (2007). A nonapoptotic cell death process, entosis, that occurs by cell-in-cell invasion. *Cell* **131**, 966–979.
22. Fehringer, G., Kraft, P., Pharoah, P.D., Eeles, R.A., Chatterjee, N., Schumacher, F.R., Schildkraut, J.M., Lindström, S., Brennan, P., Bickeböller, H., et al. (2016). Cross-Cancer Genome-Wide Analysis of Lung, Ovary, Breast, Prostate, and Colorectal Cancer Reveals Novel Pleiotropic Associations. *Cancer Res.* **76**, 5103–5114.
23. Mocellin, S., Verdi, D., Pooley, K.A., and Nitti, D. (2015). Genetic variation and gastric cancer risk: a field synopsis and meta-analysis. *Gut* **64**, 1209–1219.
24. Cassinotti, A., Birindelli, S., Clerici, M., Trabattoni, D., Lazzaroni, M., Ardizzone, S., Colombo, R., Rossi, E., and Porro, G.B. (2009). HLA and autoimmune digestive disease: a clinically oriented review for gastroenterologists. *Am. J. Gastroenterol.* **104**, 195–217.
25. Joshi, A.D., Andersson, C., Buch, S., Stender, S., Noordam, R., Weng, L.C., Weeke, P.E., Auer, P.L., Boehm, B., Chen, C., et al. (2016). Four Susceptibility Loci for Gallstone Disease Identified in a Meta-analysis of Genome-Wide Association Studies. *Gastroenterology* **151**, 351–363.e28.
26. Anstee, Q.M., Darlay, R., Cockell, S., Meroni, M., Govaere, O., Tiniakos, D., Burt, A.D., Bedossa, P., Palmer, J., Liu, Y.L., et al. (2020). Genome-wide association study of non-alcoholic fatty liver and steatohepatitis in a histologically characterised cohort. *J. Hepatol.* **73**, 505–515.
27. Liu, J.Z., van Sommeren, S., Huang, H., Ng, S.C., Alberts, R., Takahashi, A., Ripke, S., Lee, J.C., Jostins, L., Shah, T., et al. (2015). Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat. Genet.* **47**, 979–986.
28. de Lange, K.M., Moutsianas, L., Lee, J.C., Lamb, C.A., Luo, Y., Kennedy, N.A., Jostins, L., Rice, D.L., Gutierrez-Achury, J., Ji, S.G., et al. (2017). Genome-wide association study implicates immune activation of multiple integrin genes in inflammatory bowel disease. *Nat. Genet.* **49**, 256–261.
29. Ellinghaus, D., Jostins, L., Spain, S.L., Cortes, A., Bethune, J., Han, B., Park, Y.R., Raychaudhuri, S., Pouget, J.G., Hübenthal, M., et al. (2016). Analysis of five chronic inflammatory diseases identifies 27 new associations and highlights disease-specific patterns at shared loci. *Nat. Genet.* **48**, 510–518.
30. Khurana, E., Fu, Y., Chakravarty, D., Demichelis, F., Rubin, M.A., and Gerstein, M. (2016). Role of non-coding sequence variants in cancer. *Nat. Rev. Genet.* **17**, 93–108.
31. Shao, Y., Jia, H., Huang, L., Li, S., Wang, C., Aikemu, B., Yang, G., Hong, H., Yang, X., Zhang, S., et al. (2021). An Original Ferroptosis-Related Gene Signature Effectively Predicts the Prognosis and Clinical Status for Colorectal Cancer Patients. *Front. Oncol.* **11**, 711776.
32. Spahn, T.W., and Kucharzik, T. (2004). Modulating the intestinal immune system: the role of lymphotoxin and GALT organs. *Gut* **53**, 456–465.
33. Dang, T., Meng, X., Modak, C., Wu, J., Chang, Z., Che, N., Narvaez, R., and Chai, J. (2019). Overexpression of CCN1 in Het1A cells attenuates bile-induced esophageal metaplasia through suppressing non-canonical NFκB activation. *Cytokine* **116**, 61–69.
34. Ivancic, M.M., Anson, L.W., Pickhardt, P.J., Megna, B., Pooler, B.D., Clipson, L., Reichelderfer, M., Sussman, M.R., and Dove, W.F. (2019). Conserved serum protein biomarkers associated with growing early colorectal adenomas **116**, 8471–8480.
35. Codreanu, S.G., Hoeksema, M.D., Slebos, R.J.C., Zimmerman, L.J., Rahman, S.M.J., Li, M., Chen, S.C., Chen, H., Eisenberg, R., Liebler, D.C., and Massion, P.P. (2017). Identification of Proteomic Features To Distinguish Benign Pulmonary Nodules from Lung Adenocarcinoma. *J. Proteome Res.* **16**, 3266–3276.
36. Villalobos, C., Hernández-Morales, M., Gutiérrez, L.G., and Núñez, L. (2019). TRPC1 and ORAI1 channels in colon cancer. *Cell Calcium* **81**, 59–66.
37. Yuan, Y., Bao, J., Chen, Z., Villanueva, A.D., Wen, W., Wang, F., Zhao, D., Fu, X., Cai, Q., Long, J., et al. (2021). Multi-omics analysis to identify susceptibility genes for colorectal cancer. *Hum. Mol. Genet.* **30**, 321–330.
38. Chiba, T., Marusawa, H., and Ushijima, T. (2012). Inflammation-associated cancer development in digestive organs: mechanisms and roles for genetic and epigenetic modulation. *Gastroenterology* **143**, 550–563.
39. Luo, K. (2017). Signaling Cross Talk between TGF-beta/Smad and Other Signaling Pathways. *Cold Spring Harb. Perspect. Biol.* **9**, a022137.
40. Hu, L., Zang, M.D., Wang, H.X., Zhang, B.G., Wang, Z.Q., Fan, Z.Y., Wu, H., Li, J.F., Su, L.P., Yan, M., et al. (2018). G9a promotes gastric cancer metastasis by upregulating ITGB3 in a SET domain-independent manner. *Cell Death Dis.* **9**, 278.
41. Colyn, L., Bárcena-Varela, M., Álvarez-Sola, G., Latasa, M.U., Uriarte, I., Santamaría, E., Herranz, J.M., Santos-Laso, A., Archederra, M., Ruiz de Gauna, M., et al. (2021). Dual Targeting of G9a and DNA Methyltransferase-1 for the Treatment of Experimental Cholangiocarcinoma. *Hepatology* **73**, 2380–2396.
42. Cha, S.T., Tan, C.T., Chang, C.C., Chu, C.Y., Lee, W.J., Lin, B.Z., Lin, M.T., and Kuo, M.L. (2016). G9a/RelB regulates self-renewal and function of colon-cancer-initiating cells by silencing *Let-7b* and activating the K-RAS/beta-catenin pathway. *Nat. Cell Biol.* **18**, 993–1005.
43. Sudlow, C., Gallacher, J., Allen, N., Beral, V., Burton, P., Danesh, J., Downey, P., Elliott, P., Green, J., Landray, M., et al. (2015). UK biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779.
44. Manichaikul, A., Mychaleckyj, J.C., Rich, S.S., Daly, K., Sale, M., and Chen, W.M. (2010). Robust relationship inference in genome-wide association studies. *Bioinformatics* **26**, 2867–2873.
45. Bycroft, C., Freeman, C., Petkova, D., Band, G., Elliott, L.T., Sharp, K., Motyer, A., Vukcevic, D., Delaneau, O., O'Connell, J., et al. (2018). The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209.
46. Chang, C.C., Chow, C.C., Tellier, L.C., Vattikuti, S., Purcell, S.M., and Lee, J.J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *GigaScience* **4**, 7.

47. Bulik-Sullivan, B.K., Loh, P.R., Finucane, H.K., Ripke, S., Yang, J., Patterson, N., Daly, M.J., Price, A.L., and Neale, B.M.; Schizophrenia Working Group of the Psychiatric Genomics Consortium (2015). LD Score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* *47*, 291–295.
48. Scutari, M., and Nagarajan, R. (2011). On Identifying Significant Edges in Graphical Models.
49. Heckerman, D., Geiger, D., and Chickering, D.M. (1994). Learning Bayesian Networks: The Combination of Knowledge and Statistical Data. In *Uncertainty Proceedings 1994*, R.L. de Mantaras and D. Poole, eds. (Morgan Kaufmann), pp. 293–301. <https://doi.org/10.1016/B978-1-55860-332-5.50042-0>.
50. Horton, R., Wilming, L., Rand, V., Lovering, R.C., Bruford, E.A., Khodiyar, V.K., Lush, M.J., Povey, S., Talbot, C.C., Jr., Wright, M.W., et al. (2004). Gene map of the extended human MHC. *Nat. Rev. Genet.* *5*, 889–899.
51. Zhu, Z., Zheng, Z., Zhang, F., Wu, Y., Trzaskowski, M., Maier, R., Robinson, M.R., McGrath, J.J., Visscher, P.M., Wray, N.R., and Yang, J. (2018). Causal associations between risk factors and common diseases inferred from GWAS summary data. *Nat. Commun.* *9*, 224.
52. Bowden, J., Davey Smith, G., Haycock, P.C., and Burgess, S. (2016). Consistent Estimation in Mendelian Randomization with Some Invalid Instruments Using a Weighted Median Estimator. *Genet. Epidemiol.* *40*, 304–314.
53. Yavorska, O.O., and Burgess, S. (2017). MendelianRandomization: an R package for performing Mendelian randomization analyses using summarized data. *Int. J. Epidemiol.* *46*, 1734–1739.
54. Bhattacharjee, S., Rajaraman, P., Jacobs, K.B., Wheeler, W.A., Melin, B.S., Hartge, P., Yeager, M., Chung, C.C., Chanock, S.J., and Chatterjee, N.; GliomaScan Consortium (2012). A subset-based approach improves power and interpretation for the combined analysis of genetic association studies of heterogeneous traits. *Am. J. Hum. Genet.* *90*, 821–835.
55. Yang, H., and Wang, K. (2015). Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. *Nat. Protoc.* *10*, 1556–1566.
56. Pan, Q., Liu, Y.J., Bai, X.F., Han, X.L., Jiang, Y., Ai, B., Shi, S.S., Wang, F., Xu, M.C., Wang, Y.Z., et al. (2021). VARAdb: a comprehensive variation annotation database for human. *Nucleic Acids Res.* *49*, D1431–D1444.
57. Lu, Y., Qian, C., Chen, H., Bo, X., and Zhang, C. (2017). 3DSNP: a database for linking human noncoding SNPs to their three-dimensional interacting genes. *Nucleic Acids Res.* *45*, D643–D649.
58. Consortium, G.T. (2015). Human genomics. The Genotype-Tissue Expression (GTEx) pilot analysis: multitissue gene regulation in humans. *Science* *348*, 648–660. <https://doi.org/10.1126/science.1262110>.
59. Vösa, U., Claringbould, A., Westra, H.-J., Bonder, M.J., Deelen, P., Zeng, B., Kirsten, H., Saha, A., Kreuzhuber, R., Kasela, S., et al. (2018). Unraveling the polygenic architecture of complex traits using blood eQTL metaanalysis. Preprint at bioRxiv, 447367. <https://doi.org/10.1101/447367>.
60. Quinlan, A.R., and Hall, I.M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* *26*, 841–842.
61. Rappaport, N., Twik, M., Plaschkes, I., Nudel, R., Iny Stein, T., Levitt, J., Gershoni, M., Morrey, C.P., Safran, M., and Lancet, D. (2017). MalaCards: an amalgamated human disease compendium with diverse clinical and genetic annotation and structured search. *Nucleic Acids Res.* *45*, D877–D887.
62. Liu, Y., and Xie, J. (2020). Cauchy combination test: a powerful test with analytic p-value calculation under arbitrary dependency structures. *J. Am. Stat. Assoc.* *115*, 393–402.
63. Mullins, N., Forstner, A.J., O’Connell, K.S., Coombes, B., Coleman, J.R.I., Qiao, Z., Als, T.D., Bigdeli, T.B., Børte, S., Bryois, J., et al. (2021). Genome-wide association study of more than 40,000 bipolar disorder cases provides new insights into the underlying biology. *Nat. Genet.* *53*, 817–829.
64. Guo, X., Lin, M., Rockowitz, S., Lachman, H.M., and Zheng, D. (2014). Characterization of human pseudogene-derived non-coding RNAs for functional potential. *PLoS One* *9*, e93972.
65. Zhu, Z., Zhang, F., Hu, H., Bakshi, A., Robinson, M.R., Powell, J.E., Montgomery, G.W., Goddard, M.E., Wray, N.R., Visscher, P.M., and Yang, J. (2016). Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat. Genet.* *48*, 481–487.
66. Freshour, S.L., Kiwala, S., Cotto, K.C., Coffman, A.C., McMichael, J.F., Song, J.J., Griffith, M., Griffith, O.L., and Wagner, A.H. (2021). Integration of the Drug-Gene Interaction Database (DGIdb 4.0) with open crowdsourced efforts. *Nucleic Acids Res.* *49*, D1144–D1151.
67. Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z., et al. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* *46*, D1074–D1082.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Biological samples		
UK Biobank: 57471	UK Biobank	https://www.ukbiobank.ac.uk/
Deposited data		
eQTL data from eQTLGen	eQTLGen	https://eqtlgen.org/
eQTL data from GTEx	GTEx	https://www.gtexportal.org/home
Software and algorithms		
PLINK v1.90		http://pngu.mgh.harvard.edu/purcell/plink/
LDSC		https://github.com/bulik/ldsc
ANNOVAR		https://annovar.openbioinformatics.org/en/latest/user-guide/download/
bedtool		https://code.google.com/archive/p/bedtools/
Summary-data-based Mendelian Randomization		https://yanglab.westlake.edu.cn/software/smr/
LDlink R package		https://github.com/CBIIT/LDlinkR
bnlearn R package		https://github.com/cran/bnlearn
MendelianRandomization R package		https://cran.r-project.org/web/packages/MendelianRandomization/
GSMR R package		https://yanglab.westlake.edu.cn/software/gsmr/
ASSET R package		https://dceg.cancer.gov/tools/analysis/asset
ACAT R package		https://github.com/yaowuliu/ACAT
clusterProfiler R package		https://bioconductor.org/packages/release/bioc/html/clusterProfiler.html
original code	This paper	Zenedo; https://doi.org/10.5281/zenodo.8405925

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources should be directed to and will be fulfilled by the lead contact, Yongyue Wei (ywei@pku.edu.cn).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The data used in this study is all from public databases. Data support the main findings in this study are accessible via the UK Biobank under application number 57471. Other data can be obtained from the GTEx and eQTLGen. Download URLs are listed in the [key resources table](#).

Original code has been deposited at Zenodo and is publicly available as of the date of publication. DOIs are listed in the [key resources table](#).

Any additional information required to reanalyze the data reported in this paper is available from the [lead contact](#) upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Study population

The data were obtained from the UKB cohort (Proposal ID: 57471). UKB is a population-based longitudinal cohort of ~500,000 individuals recruited at 22 centers across the United Kingdom.⁴³ The UKB phenotypes were derived from the following data field IDs: self-report (20001, cancer code; 20002, noncancer illness code), ICD10 (41270, diagnoses in ICD10; 40001, underlying (primary) cause of death in ICD10), ICD9 (41271, diagnoses in ICD9) ([Figure 2A](#) and [Table S1](#)). Individuals who have any other disorders of the digestive system were excluded according to the above data fields, and the rest of the individuals were defined as controls ([Table S2](#)). Analyses were limited to ‘Caucasian’ according to Field ID 22006 to reduce population stratification. The kinship relationship was inferred by KING software with

default parameters.⁴⁴ After filtering, 329,707 European individuals including 116,382 cases with at least one digestive disorder and 213,325 controls were retained.

GWAS statistics

Genotyping was conducted using either the UKB Axiom array or the UK BiLEVE array.⁴⁵ We excluded SNPs with imputation accuracy (Info) score < 0.8, minor allele frequency (MAF) < 0.01, Hardy-Weinberg equilibrium test P value < 1.0×10^{-6} , or missing genotype rate > 0.05 using PLINK 1.9,⁴⁶ leaving 8,573,123 variants for the following analyses.

We performed case-control GWAS analyses using a logistic regression model additively modeled the SNPs with genetic sex, age, and top 10 ancestry principal components (PCs) as covariates in PLINK. We randomly selected 20,000 European individuals and set SNPs of them as linkage disequilibrium (LD) reference. Independent trait-associated SNPs were generated using PLINK ($-clump-p1\ 5 \times 10^{-8}$ $-clump-r2\ 0.1$ $-clump-kb\ 500$).

The significant SNPs were searched in GWAS Catalog and were divided into two categories: previously reported SNPs related to digestive disorders and novel SNPs, via R package *LDlink* according to the published GWAS from GWAS Catalog. We determined that a SNP was potentially novel if GWAS Catalog SNPs had LD $r^2 \leq 0.1$ with the SNP.⁹ Based on the GWAS analysis, shared LD blocks were assessed for overlap among multiple digestive disorders.

METHOD DETAILS

Our study does not involve experiments, and the relevant statistical methods and analysis procedures will be discussed in the "Quantification and statistical analysis" section.

QUANTIFICATION AND STATISTICAL ANALYSIS

All statistical analyses using R packages were performed using R 4.2.1, unless otherwise stated. Information on specific statistical analyses are described below.

Genetic heritability and genetic correlation

We estimated the SNP-based heritability (h^2_{SNP}) using linkage disequilibrium score regression (LDSC).⁴⁷ To convert to liability-scale heritability, we adjusted for lifetime risks of each digestive disorder based on the proportion of the cases in the sample. The genomic inflation factor (λ_{GC}) was also reported for each disorder. Genetic correlations (r_g) for each pair of the 21 digestive disorders were calculated using bivariate LDSC.

Inference of Bayesian causal network

To understand the causal relationship among multiple digestive disorders, we conducted the Bayesian network using the score-based hill-climbing (HC) algorithm with a sufficiently large sample size to enable effective inference.⁴⁸ In our study, the network was bootstrapped 2000 times, using the 21 disorders as discrete variables, and arc directions were identified significantly which the probability is more than 85%. The strength of the probabilistic relationships expressed by the arcs was measured by the logarithm of the Bayesian Dirichlet equivalent score (bde).⁴⁹ For the undirected arc which probability of its direction is 0.5, we retained the direction which had the stronger strength. The Bayesian network was generated in R package *bnlearn*.

Mendelian randomization analysis

To explore the potential causal effect among all pairs of 21 digestive disorders, we used Mendelian randomization (MR) with exposure-significant SNPs as the instrument variables. Considering that some digestive disorders had insufficient number of cases, resulted in the limited genome-wide significant SNPs ($P \leq 5 \times 10^{-8}$), we relaxed the significance threshold to 5×10^{-6} to obtain sufficient genetic instrumental variables for those digestive disorders. Due to the complexity and strong linkage disequilibrium of the MHC region, only the most significant SNP within MHC region (chr6: 25-34 Mb) was reserved for MR analysis.⁵⁰ Based on the UKB reference panel, we used linkage disequilibrium $r^2 < 0.01$ as a clumping threshold and set the physical distance threshold to be 10 Mb to ensure uncorrelated genetic instruments.

We applied the inverse-variance weighted (IVW) method to estimate the causal relationship.¹² To ensure the robustness of the results, we additionally performed GSMR,⁵¹ median-MR,⁵² and mendelian randomization analysis using Maximum-likelihood¹² as sensitivity analysis to control for the influence of pleiotropic effects, instrumental outliers, and sample overlap. Statistical analyses were performed using the packages MendelianRandomization⁵³ and GSMR.

Cross-disorder GWAS meta-analysis

To identify the shared variation of the multiple digestive disorders, cross-disorder meta-analysis was carried out via association analysis based on subsets (ASSET).⁵⁴ We conducted ASSET analysis on the independent signals (index SNPs which $P_{GWAS} \leq 1 \times 10^{-4}$) for each digestive disorder. In the bidirectional pleiotropy analysis, P value for each direction is provided as well as an overall P for the total association signal for both directions combined. The pleiotropic independent variants were determined via LD clumping with overall $P \leq 5 \times 10^{-8}$, and other

SNPs were clumped with the lead variant if they had overall $P < 0.05$, were within 500kb of the index SNP and had $r^2 > 0.1$ with the index SNP. A SNP was determined to have effect with consistent direction if the overall $P \leq 5 \times 10^{-8}$ and the P for one direction was ≤ 0.05 . Similarly, a SNP was determined to have effect with different directions if the overall $P \leq 5 \times 10^{-8}$ and the P values for both directions were < 0.05 .

Functional annotation and gene-mapping of pleiotropic variants

In terms of assessing variant functions and mapping SNPs to genes, we first annotated SNPs based on ANNOVAR.⁵⁵ To perform a more comprehensive evaluation of functional genetic variations, we obtained candidate genes from other various resources including VARAdb⁵⁶ and 3DSNP⁵⁷ as supplement to ANNOVAR.

We also annotated SNPs to genes in which the SNPs located in or near the super-enhancers and promoters. 3DSNP was used to map SNPs to distal target genes through three-dimensional (3D) chromatin looping. Cis-eQTL mapping provides significant genes (FDR q value < 0.05) in nine specific digestive tissue types (esophagus muscularis, esophagus mucosa, esophagus gastroesophageal junction, stomach, small intestine terminal ileum, colon sigmoid, colon transverse, liver, pancreas) from GTEx v8⁵⁸ and whole blood from eQTLGen⁵⁹ database. The eQTLGen provides the largest existing eQTL summary statistics from 31,684 whole blood samples. All tissue or cell types corresponding to the data source above were digestive-specific or whole blood.

Annotation results of multiple means were merged to form a list of candidate genes. The function cluster from bedtools⁶⁰ was used to cluster these genes into independent 1-Mb regions. MalaCards,⁶¹ a database that provides the gene-disease relations from multiple data sources, was used to search for the existing evidence for the association between candidate genes and disorders.

Gene-based association analysis

We applied aggregated Cauchy association test (ACAT) to combine the statistical evidence from multiple SNPs within the corresponding gene to determine the association of target gene and individual digestive disorders via R package ACAT based on the GWAS summary results.⁶² Gene boundary relies on Ensembl database build GRCh37.3, extending 35 kb upstream and 10 kb downstream to include regulatory regions.⁶³ Genomic locations unavailable from Ensembl were manually annotated using NCBI's Gene online web resource. Pseudogenes were not included because of potential concerns of inaccurate calling.⁶⁴

SMR analysis for candidate genes

Summary-data-based Mendelian randomization (SMR)⁶⁵ was used to provide putative causal evidence between SNPs and disorders via gene expression. SMR was performed using the expression quantitative trait Loci (eQTL) summary statistics from the eQTLGen and GTEx v8 described in the supplementary methods. Only transcripts with at least one cis-eQTL ($P \leq 5 \times 10^{-8}$) were taken into consideration. The significant threshold for the Heterogeneity in Dependent Instrument (HEIDI) test was $PHEIDI \geq 0.01$.

Pathway enrichment analysis and drug target exploration

To explore functional discrepancy of the detected pleiotropic genes, the Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses were performed by the R package clusterProfiler with p -value cutoff = 0.05. To investigate the potential drugs related to digestive disorders, drug target genes and indications were obtained from the Drug-Gene Interaction Database (DGIdb),⁶⁶ and DrugBank version 5.1.9.⁶⁷