

REVIEW

Open Access



# Transcriptome-wide association studies associated with Crohn's disease: challenges and perspectives

Keyu Jia<sup>1</sup> and Jun Shen<sup>1,2,3,4\*</sup>

## Abstract

Crohn's disease (CD) is regarded as a lifelong progressive disease affecting all segments of the intestinal tract and multiple organs. Based on genome-wide association studies (GWAS) and gene expression data, transcriptome-wide association studies (TWAS) can help identify susceptibility genes associated with pathogenesis and disease behavior. In this review, we overview seven reported TWASs of CD, summarize their study designs, and discuss the key methods and steps used in TWAS, which affect the prioritization of susceptibility genes. This article summarized the screening of tissue-specific susceptibility genes for CD, and discussed the reported potential pathological mechanisms of overlapping susceptibility genes related to CD in a certain tissue type. We observed that ileal lipid-related metabolism and colonic extracellular vesicles may be involved in the pathogenesis of CD by performing GO pathway enrichment analysis for susceptibility genes. We further pointed the low reproducibility of TWAS associated with CD and discussed the reasons for these issues, strategies for solving them. In the future, more TWAS are needed to be designed into large-scale, unified cohorts, unified analysis pipelines, and fully classified databases of expression trait loci.

**Keywords** Crohn's disease, Transcriptome-wide association studies, Susceptibility genes, GO functional analysis

## Introduction

Genome-wide association studies (GWAS) have been considerably successful in the past decade. From 2005 to 2022, approximately 400,000 single-nucleotide

polymorphisms (SNPs) associated with human traits were included in the NHGRI-EBI GWAS Catalog [1, 2]. However, a limitation of GWAS is that approximately 90% of these crucial signals are located in noncoding regions [3]. Transcriptome-wide association study (TWAS) is a bioinformatic approach that integrates large-scale GWAS, uses expression quantitative trait loci (eQTL) datasets to predict gene expression levels, and attempts to identify disease-related genes and verify associations of interest. This is important for exploring specimens that are not easily collected and phenotypes rarely collected with genetic data. TWAS can assess the association between variation in gene expression levels and phenotypic variation based on different population genotype and tissue-specific gene expression data—an additional analytical approach to GWAS data—and may be used to screen candidate pathogenic genes.

\*Correspondence:

Jun Shen

shenjun@renji.com

<sup>1</sup> Laboratory of Medicine, Baoshan Branch, Ren Ji Hospital, School of Medicine, Nephrology department, Shanghai Jiao Tong University, 1058 Huanzhen Northroad, Shanghai 200444, China

<sup>2</sup> Division of Gastroenterology and Hepatology, Key Laboratory of Gastroenterology and Hepatology, Ministry of Health, Inflammatory Bowel Research Center, Ren Ji Hospital, School of Medicine, Shanghai Institute of Digestive Disease, Shanghai Jiao Tong University, Shanghai, China

<sup>3</sup> NHC Key Laboratory of Digestive Diseases, Renji Hospital, Shanghai Jiaotong University School of Medicine, Shanghai, China

<sup>4</sup> Division of Gastroenterology and Hepatology, Baoshan Branch, Renji Hospital, School of Medicine, Shanghai Jiao Tong University, Shanghai, China



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

Crohn's disease (CD), a type of inflammatory bowel diseases (IBD), is a lifelong progressive disease with a tendency of symptoms to flare up or subside as the condition alternates between active and remission periods [4]. All segments of the gastrointestinal tract can be affected. The Montreal classification introduced subgroups of CD by considering the location of the disease, age of onset, and phenotype [5]. In recent decades, many researchers have believed that the pathogenesis of CD is an irrepressible immune response to luminal bacterial antigens. Immune cells participate in this process when infiltrating the gut of CD patients. Therefore, the onset of CD involves the immune system, segments of the digestive tract, intestinal tract contents, and multiple organs, which experience complications.

Over the past decades, GWAS have identified over 240 IBD susceptibility genes or loci outside the human leukocyte antigen region, and 37 of these genes are specific for CD [6, 7]. However, many of the polymorphic sites associated with CD are located in non-transcribed regions and do not cause amino acid substitutions or functional mutations, nor do they exhibit disease susceptibility. For instance, the GWAS results of Japanese patients with CD showed that among the 11 susceptibility gene loci associated with CD, only rs76418789 located in *IL23R* had an amino acid substitution [8]. The other 10 polymorphic sites may be located in these disease-susceptible regions, thereby affecting the expression of nearby genes and participating in the occurrence and development of CD [9]. However, the function of these SNPs in the occurrence and development of CD remains unclear. Thus, TWAS must be conducted to provide a biological context for interpreting disease risk loci by nominating candidate susceptibility genes not only in GWAS risk regions but also in other regions of potential that cannot be detected by current GWAS.

Therefore, we aim to provide an overview of previous TWAS on CD and summarize the databases and methods used in these studies. This review also discusses the overlapped susceptibility genes in different tissues and the potential pathway involved in CD by GO functional analysis, which may provide clues to explore the pathogenesis, diagnosis, and classification of CD.

### Heterogeneity across TWASs of CD

There were seven TWASs for CD overviewed in this review with details in Table 1. Four of these studies conducted TWA in single tissue types of Japanese [9], Korean [10], American [11], and British [12] populations. In particular, the Japanese study also used a cross-tissue eQTL database to explore susceptibility genes based on genotypic data from the Japanese population [9]. The other three studies used cross-tissue and multi-country

eQTL designs [13–15]. The selection of GWAS datasets, eQTLs, sample sizes, tissue types, and screening criteria varied among previous TWAS (Fig. 1).

### Selection of genotype data

Large-scale genotype data provide fundamental material for TWAS. Many multi-country GWASs reported to be associated with CD provide investigators with a variety of options for data selection [16]. As the genetic structure of disease-causing mutations varies in different populations, the effect sizes and risk prediction scores derived for SNPs in one population may not be directly generalizable to other populations [17, 18]. CD has a great ancestral dependency by comparing the GWAS data from East Asian and European ancestries [19]. As yet, most of the available genetic information is based on data from populations of European ancestry [20–22]. Key variants that are low in frequency or absent entirely in European populations are likely to be missed when studied in other populations, especially if the variant is ethnic-specific, leaving additional blind spots for future studies [23, 24]. Therefore, considering the study population during the study design stage, caution should be exercised in genotype dataset selection. Most of the 7 TWAS selected consistent races to establish this relationship. For example, the Japanese and Korean TWAS selected genotype and expression populations from their countries. Dai et al. chose genotyped data of European ancestry as the GWAS population and 24 CD patients and 23 healthy controls from the IBD–BIOM inception cohort from UK [12, 25]. Gettler et al. used both the genotype and expression populations of the Childhood CD Study derived from the RISK cohort [11]. However, the selection of a consistent population race may limit the sample size of the study.

### Selection of eQTL data

eQTL datasets are constructed using statistical models based on genotype and tissue expression data from the same population, which is an effective tool for fine-mapping GWAS that identifies SNPs associated with complex phenotypic traits and can be used to improve the heritability explained by identifiable genetic factors and to better understand the biological basis of complex traits [26].

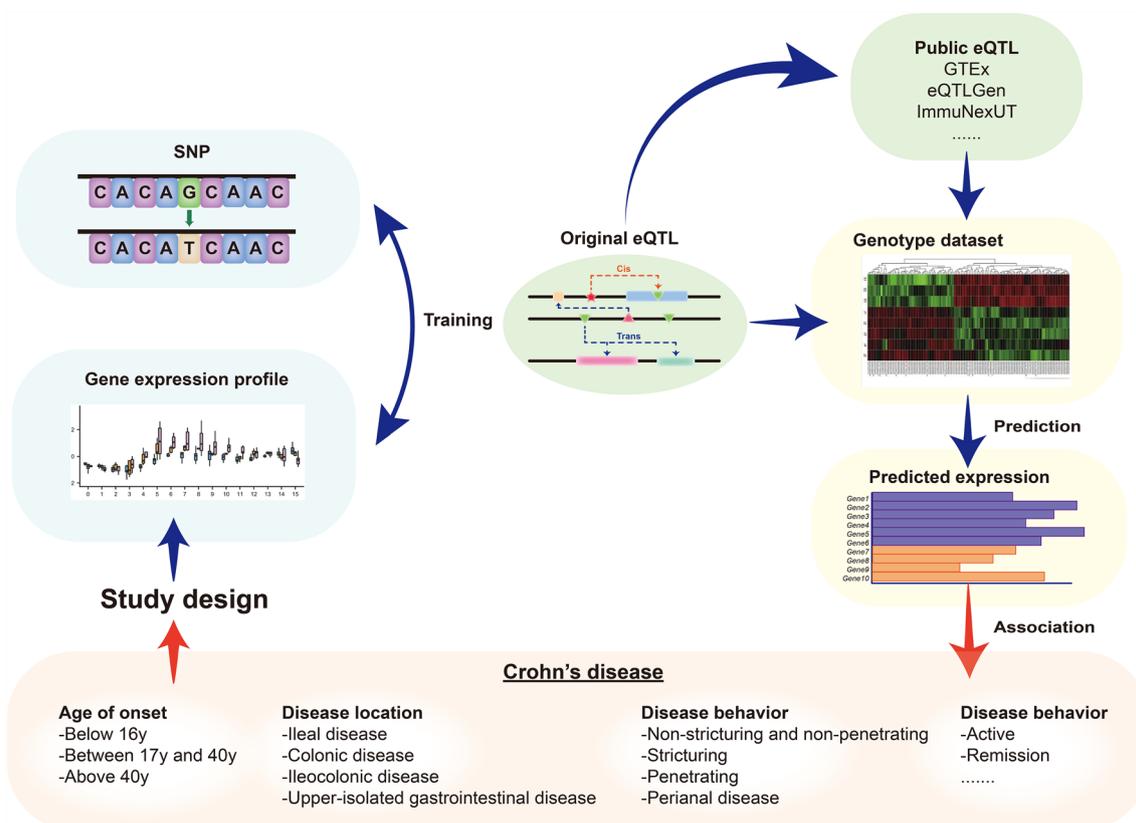
Previous TWAS exhibited acute instability when choosing various eQTL datasets based on different tissue types. Since eQTL limited the specific tissue to find differentially expressed genes, overlapping the same GWAS database with eQTLs of different tissues could result in different results of TWAS.

eQTLs are race-specific in several aspects. (1) For a specific phenotype, causal genes might be distinct across ethnicities. (2) Many polymorphisms were rare in some races but common in others, which could ignore some

**Table 1** Summary of the basic information of TWAS studies in CD

Year	GWAS data			Country	Tissue type of RNA-seq	eQTL N	Methods	Associated tissue type	Association screening criteria
	Country	Database	N						
Diez-Obrero et.al. [15]	Europe	IBD GWASs of European ancestries	IBD: 25,042 HC: 34,915	Spain (BarcUVA-Seq) Europe (CEDAR)	Ascending colon Transverse colon Descending colon	HC:138 HC:143 HC:164	S-PrediXcan	Ascending colon, Transverse colon, Descending colon	$P_{Bonferoni} < 0.05$
Uellendahl-Werth et. Al. [13]	15 countries cross Europe, North America, Australia	Ten case-control GWAS datasets	CD: 21,771 HC: 41,206	US (GTEx), Sweden (STARNET), UK (BLUE-PRINT)	Small intestine Sigmoid colon Transverse colon Whole blood CD14+ Monocytes, CD16+ Neutrophils, Naive T cell	non-IBD 77 non-IBD 124 non-IBD 169 non-IBD:338 non-IBD 338	UTMOST	Small intestine, Sigmoid colon, Transverse colon, Whole blood, CD14+ Monocytes, CD16+ Neutrophils, Naive T cell	$P_{Bonferoni} < 0.05$
Kakuta et.al. [9]	Japan	CD GWASs of Japanese	CD: 713 HC: 2063	Japan  US (GTEx)	Intestines  Sigmoid colon Transverse colon Small Intestine Whole blood EBV transformed lymphocytes	active CD: 15 active UC: 5  HC: 203 HC: 246 HC: 122 HC: 369 HC: 117	GWAS-eQTL analysis  FUSION	Effector memory T cells from inflammation sites  Sigmoid colon, Transverse colon, Small Intestine, Whole blood, EBV transformed lymphocytes	Susceptibility: $P_{FDR} < 0.05$ Candidate: $P_{FDR} < 0.1$
Jung et.al. [10]	Korea	CD GWAS of Korean	CD: 899 HC: 3805	Korea	Peripheral blood	CD:101	FUSION	Whole blood	$P_{Bonferoni} < 0.05$
Gettler et.al. [11]	Cross countries	15 GWAS of CD and/or UC	CD: 6299 HC: 15,148	America (GEO)	Terminal ileum	CD: 213, UC: 50, Unspecified IBD:4, HC: 35	COLOC	Terminal ileum	$P_{FDR} < 0.1$
Dai et.al. [12]	Europe	IIIBDGC	CD: 5956 HC: 21,770	British	Whole blood	CD: 24 HC: 23	TESA MetaXcan	Terminal ileum Whole blood Spleen	$FC \geq 1.5$ or $\leq 0.67$ $P_{Pascal} < 0.05$
Cheng et. al. [14]	Europe	IBD GWAS of European ancestries	CD: 18,405 UC: 14,308 HC: 34,241	Cross countries (GEO)	Intestinal  Whole blood  Peripheral blood	IBD: 134, non-IBD: 134  IBD: 75, HC:12  HC:42, UC:26, CD:59	FUSION	Sigmoid colon, Transverse colon, Whole blood/ Peripheral blood, Small Intestine terminal ileum EBV transformed lymphocytes	$P\text{-value} < 0.05$ & $FC > 1.5$

CD Crohn's disease, UC ulcerative colitis, IBD inflammatory bowel disease, HC healthy control, GWAS genome-wide association studies, TWAS transcriptome-wide association studies, eQTL expression quantitative trait loci; PGC Psychiatric Genomics Consortium, IIIBDGC International Inflammatory Bowel Disease Genetics Consortium, GEO Gene Expression Omnibus database, BarcUVA-Seq The University of Barcelona and University of Virginia RNA sequencing project; CEDAR correlated expression and disease association research, STARNET Stockholm-Tartu Atherosclerosis Reverse Networks Engineering Task study, S-PrediXcan Summary-PrediXcan, FC fold change



**Fig. 1** Workflow of TWAS in CD. The initial stage of TWAS design should consider characteristics of CD such as disease behavior, disease location, and disease status, and select GWAS datasets and eQTL populations with specific CD subtypes to ensure homogeneity within studies. TWAS could choose the public eQTL database or establish an original eQTL. The large-scale genotype dataset is used to predict expression data by eQTL to associated with CD related outcomes. *CD* Crohn's disease, *TWAS* transcriptome-wide association studies, *SNP* single nucleotide polymorphism, *eQTL* expression quantitative trait loci

associations between SNP and expression in the application. (3) The degree of association between a specific SNP and the expression level in one race might also differ from that in others. Owing to these diversities, when utilizing the same GWAS dataset, substantial discrepancies exist in the conclusions drawn among different studies conducted in different countries. As can be seen in Additional file 1: data S1, for the same TWAS methods, it tends to screen out more genes with a consistent ethnicity of GWAS population and eQTL population. For example, for whole blood, the Japanese study used GTEx (eQTL from the United States) to screen out only 1 gene, and the Korean study used their own eQTL to screen out a total of 21 genes, which also suggested that eQTLs are race-specific and the consistent ethnicity of GWAS and eQTL populations may increase the accuracy of expression prediction.

Whether eQTLs are disease-specific is uncertain [27]. Because epigenetic modifications differ according to the disease state [28], the relationship between expression mapping and genotype could also be affected. In our

review, most TWAS combined the expression data of healthy individuals and CD patients into the same eQTL establishment. Uniquely, the Japanese study selected data from CD and UC patients to construct its own eQTL dataset of IBD [9], which may interfere with the homogeneity of eQTLs in expression prediction and further association analysis between CD patients and healthy controls though reduce the test power concomitantly. The relationship between eQTL and GWAS associations at the same locus could be unpredictable for different disease types. A previous study of non-alcoholic fatty liver disease used the disease-specific eQTL to pinpoint individuals that harbor specific genotypes more or less susceptible to the disease [29]. Thus, using disease-specific eQTL to establish the relationship in patients is worth investigating.

The resolution is low and unreliable with a sample size of 100 [27], though the sample size of the original eQTLs in most CD TWASs did not exceed 50 pairs. Various public eQTL databases are available to explore CD, including the Genotype-Tissue Expression (GTEx, <https://www.>

[genome.gov/Funded-Programs-Projects/Genotype-Tissue-Expression-Project](https://www.genome.gov/Funded-Programs-Projects/Genotype-Tissue-Expression-Project)), eQTLGen (<https://www.eqtngen.org/phase1.html>), and the Blood eQTL browsers (<http://genenetwork.nl/bloodeqtlbrowser/>). Although GTEx has been in development for 10 years, it is still worth using because its data is relatively stable and still update yet with a wide range of tissue types and sample size. As the largest eQTL database, GTEx data included genotype data from 714 donors and 11,688 RNA-seq data from 53 tissue sites and two cell lines, with sufficient assay power to establish eQTLs in 48 tissue types/sites. Although the database only included the data of a healthy population, it could be combined with the expression database of CD patients for further analysis. As GTEx included data/patients solely from the United States, its utility is limited for other countries' populations. eQTLGen included 37 datasets from 31,684 individuals, including cis-eQTL, trans-eQTL, eQTS, and single-cell eQTLGen Consortium; however, only blood samples were used [30]. The Blood eQTL Browser, which has 5311 individuals' data, also included only blood samples [31], which limited the exploration of other tissue types more likely to be causally related to CD. Therefore, a database with more comprehensive and specific classifications across tissue types, diseases, and ethnicities is warranted to facilitate the use of multiple disease-targeting tissue types in future large-scale eQTL studies and to provide a unified platform for mining more robust associations in next-generation studies.

#### Selection of software or methods of TWAS

Integrating GWAS and expression data in TWAS could be performed by various tools, including PrediXcan for individual-level GWAS data, Fusion and S-PrediXcan for summary-level GWAS data, closely related methods, including SMR and HEIDI, based on Mendelian randomization (MR), and GWAS-eQTL colocalization methods, including Sherlock, coloc, QTLMatch, eCaviar, enloc, and RTC, for detecting genes whose expression regulated by the same GWAS hit [32, 33].

Of the seven CD TWAS, five TWAS used commonly used methods such as Fusion, PrediXcan or MetaXcan to integrate GWAS and expression data. MetaXcan is an expanded method calculating the results PrediXcan without using individual data [33].

Besides, Uellendahl-Werth et al. conducted cross-tissue TWAS associated with gut-brain-axis performed by UTMOST (multivariate-response penalized regression models) to predict cross-tissue gene expression [13]. And they observed that UTMOST could get more moderate associations and effectively select predictive cis eQTL variants compared with S-PrediXcan (logistic regression model) and FUSION (Bayesian linear mixed

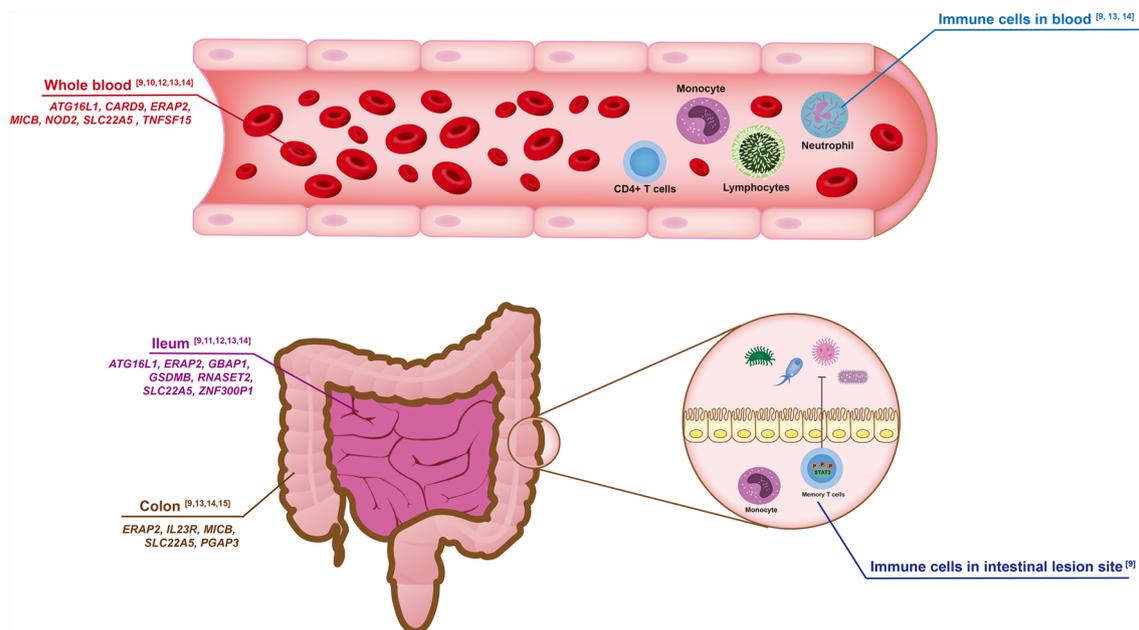
model) [13]. When comparing across studies, the genes screened by ULMOST and MetaXcan rarely overlap with those screened by traditional methods (Additional file 1: Data S1), which suggests that the methods bias was not the main result responsible for the poor reproducibility. Since Uellendahl-Werth et al. did not present the associated genes screened by S-PrediXcan and FUSION in their results, whether method bias contributed to the low repetition rate is uncertain.

Gettler et al. [11] used the coloc R package, which, as Wainberg et al. previously mentioned, is vulnerable because co-regulation bias makes it difficult to distinguish causality based on GWAS and expression data [32, 34]. Hukku et al. compared PrediXcan and GWAS-eQTL colocalization methods and found that the GWAS-eQTL colocalizations may have a higher specificity and limited sensitivity, and PrediXcan could be possible to report more results with difficult in biological interpreting [35]. Therefore, caution should be used when interpreting TWAS test results derived from controversial methods, as some of them may simply be false hits. [32].

#### Heterogeneity of various tissue types selected in eQTL

All segments of the digestive tract, including the systemic immune inflammatory response, could be involved in CD. Previous reviews have summarized the differences in the epidemiology, genetics, histology, microbiology, and immunology of ileal and colic celiac disease, suggesting that CD at different lesion sites should be regarded as distinct subtypes [36, 37]. Each segment of the digestive tract, immune cells, and other complex organs can be used as follow-up tissues to explore the unique pathology and etiology of CD. To date, only blood, immune cells, colon tissue, and ileum tissue have been reported in previous TWAS for CD. In this section, several biospecimen-related issues of utmost concern are discussed. Susceptibility genes associated with CD reported in TWASs in different tissue type were took union set and summarized in Additional file 2 Data S2. The overlapped susceptibility genes between TWASs are visualized in Fig. 2.

In this review, GO functional analyses of all susceptibility genes associated with CD in each tissue type (Shown in Additional file 2: Data S2) were conducted by "clusterProfiler" and "pathview" packages of R software [38, 39]. Since the logFC of these associated genes were not available, the GO function enrichment analysis was roughly conducted with a random assignment of foldchange (1 or -1), which can only suggest relevant functional enrichment and cannot indicate up-regulation or down-regulation. The top 10 results for significance (P value < 0.05) of Cellular Component (CC), Molecular Function (MF), and Biological



**Fig. 2** Summary of susceptibility genes associated with CD in different tissue types founding in TWAS. The bold gene names were the susceptibility genes at least found in two TWASs

Process (BP) are shown in Fig. 3. And the total significant results were shown in Additional file 3 data S3.

### Digestive tissue

Clinical inflammation in CD can infect the entire gastrointestinal tract from the mouth to the anus [40]. The intestinal epithelium is a single layer of columnar epithelium that produces mucus and antimicrobial factors and can establish a buffer zone between the luminal contents and itself [4]. Destruction of the intestinal epithelium may cause bacterial invasion and lead to the occurrence of CD. Thus, in CD studies, epithelial lesion tissue is commonly seen as the top-priority casual biospecimen [41].

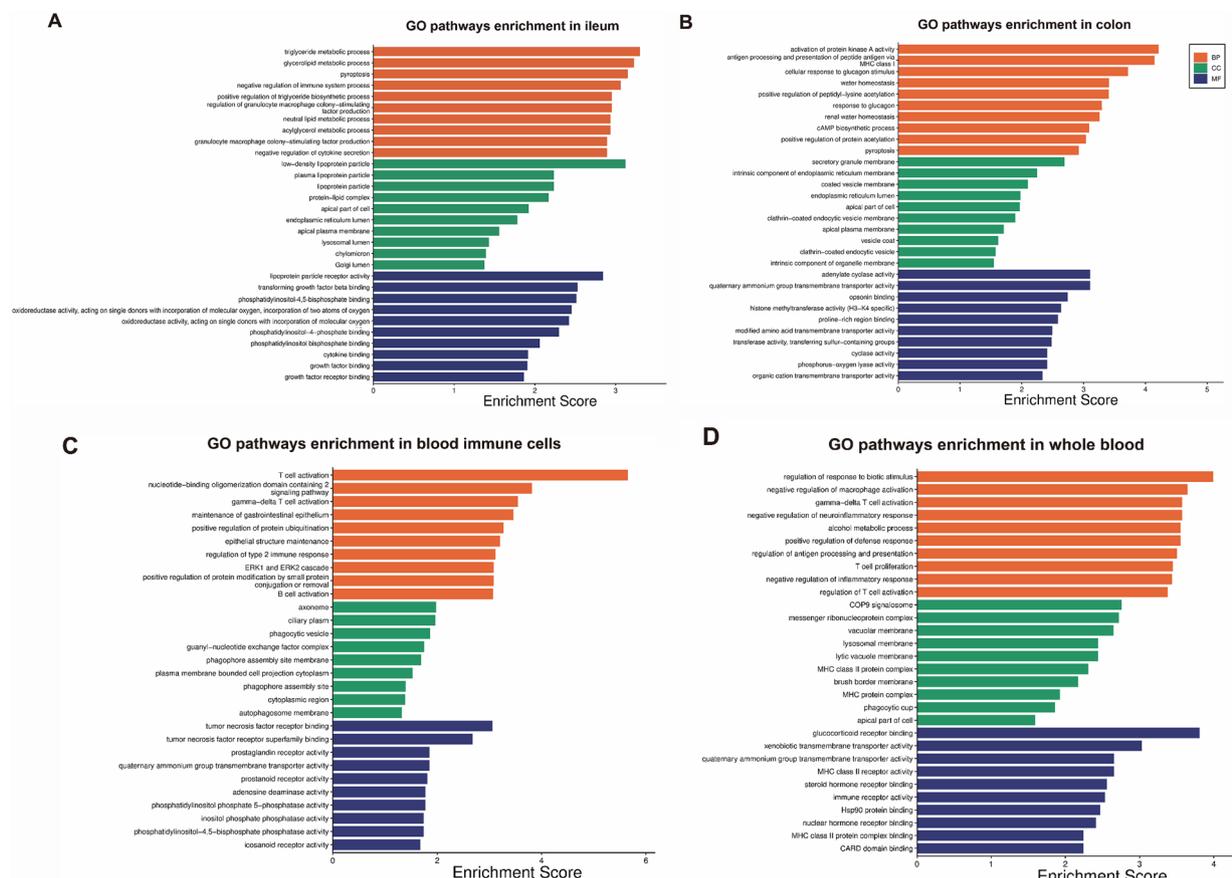
A comparative study observed that 19 hub genes were differentially expressed between the colon and ileum [42], suggesting that the tissues of different intestinal segments should be explored separately. Taking Th17 cells—identified in the pathogenesis of IBD—as an example, the genes related to Th17 pathways involved in CD were inconsistent between the colon and ileum, and the regulation (up or down) of mRNA expression levels of these genes in the colon and ileum lesion tissue from CD patients were also inconsistent [43–49]. These data suggest that gene expression patterns are significantly distinct at different intestinal sites.

### Ileum

The ileum, filled with abundant bile and digestive juice, is a relatively germ-free environment. A recent study found a trace amount of microbiota in the ileum, which differed radically from colonic microbiota [50]. A cultivable bacterial density of approximately 10<sup>4</sup> CFU/mL in the ileum was also much less than the 10<sup>11</sup>–10<sup>12</sup> CFU/mL in the colon [51]. This might indicate that the ileum and colon are situated in different external environments, which could stimulate differential gene expression.

Five TWASs selected ileum as etiological tissue of CD. *ATG16L1* [11, 13], *ERAP2* [11–13], *GBAP1* [11, 12], *GSDMB* [11, 12], *RNASET2* [11, 12], *SLC22A5* [11, 12], and *ZNF300P1* [11, 12, 14] in the ileum were susceptibility genes overlapped between at least two TWASs.

*ATG16L1* has been reported as an autophagy-related gene [6] and a major risk polymorphism in CD [52]. In a genotypic study in a U.K. population, the strongest association was observed for the *ATG16L1* risk variant with ileal disease [53]. And William J. Sandborn commented that *ATG16L1* genotype is associated with response to anti-TNF therapy. The *ATG16L1* mutation results in altered function and survival of highly specialized secretory epithelial Paneth cells located in small intestinal crypts, resulting in decreased secretion of antimicrobial proteins [54, 55]. The CD-associated *ATG16L1* mutation observed in animals is associated with low clearance of *Yersinia enterocolitica* and adherent-invasive *Escherichia*



**Fig. 3** The top 10 results obtained from GO functional analyses of all susceptibility genes associated with CD in each tissue type. *BP* biological process, *CC* cellular component, *MF* molecular function

*coli*, which further increase the production of inflammatory cytokines [56, 57]. And several studies have reported that the presence of the adherent-invasive *E. coli* abnormally colonize the ileal mucosa of CD patients [58]. Thus, the *ATG16L1* mutation may be involved in the pathological process of CD by causing intestinal barrier disruption. The expression of *ATG16L1* could be targeting regulated by a variety of micro-RNA, including miR-106b found in intestinal epithelial HCT116 cells, and four microRNA (miR-106a [59], miR-874 [60], miR-410 [61], and miR-223 [62]) found in cancer or other inflammation disease, which may provide pharmaceutical strategies for developing small molecule drugs for CD.

*ERAP2* shows limited polymorphism coding for Lys392Asn change, which affects the activity of aminopeptidases [63]. *ERAP2* forms a repertoire of ligands for HLA class I, involving in the processing of MHC-I ligands antigen presentation and the antigenic response of infection [64], which associated with various inflammation diseases, such as IBD, ankylosing spondylitis, birdshot chorioretinopathy, Behcet's disease and

psoriasis [65–67]. In this review, *ERAP2* is also observed an overlapping susceptibility gene in colon and blood, which has potential application in clinical screening and diagnosis of CD. Accumulating evidence suggests that *ERAP2* is tractable targets for the regulation of immune responses [68]. In pancreatic cancer cells, gemcitabine could increase the mRNA and protein levels of *ERAP2* [69]. However, no drugs targeted *ERAP2* in CD yet.

*GBAPI* is a pseudogene for the glucocerebrosidase (*GBA*) gene encodes for the enzyme glucocerebrosidase. Previous studies have demonstrated that *GBAPI* can act as a competitive endogenous RNA to competitively bind with microRNAs in gastric cancer [70, 85] and Parkinson's disease [71] through functional prediction, thereby promoting the expression of *GBA*. However, the role of glucocerebrosidase in CD has not been reported.

*GSDMB*, a member of the Gasdermins family, was originally known for its role in pyroptosis [72], and most prevalently expressed in gastrointestinal-associated organs, including stomach, small intestine and colon [73]. Studies have found that the expression of *GSDMB* is

increased in the inflammatory mucosa of ileum and colon of CD patients, and the related genes are enriched in cell proliferation, migration, and adhesion other than of pyroptosis [74]. As an inducer of GSDMB, methotrexate could induce upregulation of IEC-derived GSDMB-FL and translocation to the plasma membrane, but not lytic cell death in undifferentiated HT-29 cells. And the development of methotrexate in CD targeting GSDMB has entered phase III clinical trials (NCT00132899, Table 2).

*RNASET2* is the only human member of the Rh/T2/S family of acidic hydrolases [75]. An eQTL analysis observed an association between decreased *RNASET2* and *TNFSF15*-mediated IFN- $\gamma$  production, a key mediator of mucosal inflammation [76]. The circulating *RNASET2* protein levels was decreased in CD patients compared with healthy control [75]. In cancer studies, *RNASET2* has been found to be involved in recruitment, activation, and polarization of monocytes and macrophages [77, 78]. However, the role of *RNASET2* in CD needs to be further investigated.

*SLC22A5* code organic cation transporters (OCTN2), which was widely expressed and also the susceptibility genes observed in colon and blood (can be seen below). OCTN2 is mainly localized at the brush-border of apical membranes of intestinal epithelial cells and has a high transport capacity of L-carnitine in the small intestine, which is vital for  $\beta$ -oxidation of long-chain fatty acids in the mitochondria [79]. Several studies have observed the expression of OCTN2 downregulated in inflamed sites compared with non-inflamed sites both in patient intestinal tissue and mice model [80, 81]. And the PPAR $\alpha/\gamma$  may act as transcription factors in the expression of OCTN2 and further regulate inflammatory response [80]. OCTN2 also could transports drugs, such as TEA, ipratropium, prednisolone, and beta-lactam antibiotics [82–84].

*ZNF300P1* encode a long intergenic noncoding RNA, suggesting its primary function may be to regulate expression of other genes [85]. *ZNF300P1* was found

upregulated in ileum, rather than in colon or whole blood [86]. Besides, *ZNF300P1* may alter tissue-specific expression of TNF and a range of additional genes previously implicated in colitis and/or autophagy. Besides, *ZNF300P1* is reported to regulate polarity, proliferation, migration, and adhesion in ovarian epithelial cells [87], suggesting that it may similarly participant in intestinal epithelial functions.

The pathways of susceptibility genes associated with CD in ileum were enriched in lipid-related metabolism (Fig. 3A). Previous observational studies have reported a distinct lipid profile in CD patients compared with healthy population [88, 89]. And growing evidence showed emulsifying omega-3 fatty acids maybe a potential supplementary in maintaining remission of CD patients [90, 91]. An epidemiological study observed that lower total cholesterol levels, LDL-C, and HDL-C were associated with higher incidence of CD, but not UC [92]. Coincidentally, another study also observed that more lipid components significantly changed in CD patients than in UC patients compared with healthy population [93]. Considering that ileal lesions are present only in CD, the different association of lipid metabolism with these two types of IBD may be due to the location of the lesion in the ileum. However, A shotgun lipidomics study of noninflammatory ileal biopsy tissue identified only phosphatidylinositol 16:0/18:1 was different between healthy controls and CD patients, although the sample size was small [89]. Additional future exploration will be necessary to confirm this observation.

**Colon**

Unlike the ileum, the colon has a substantial bacterial load, which plays a crucial role in regulating gut health. Changes in the abundance of specific bacteria have been used as biomarkers for screening gastrointestinal disorders, including IBD, irritable bowel syndrome, adenomatous colonic polyps, and colorectal cancer. Changes in the abundance of some bacteria have been

**Table 2** Targeted drugs had been developed of overlapped genes

Genes	Target	Regulatory factor/drug	Disease	Developed process	Clinical trials. gov identifier
<i>GSDMB</i>	GSDMB	Methotrexate	CD	A Phase III randomized, placebo-controlled, double-blind	NCT00132899
<i>IL23R</i>	IL23 p19	Brazikumab	CD	The Phase 3 trial was terminated	NCT03961815
		Risankizumab	CD	In phase 3, randomized, placebo-controlled, double-blind trial	NCT06063967
		Mirikizumab	CD	In Phase 3 open-label trial	NCT04232553
		Guselkumab	CD	In Phase 3 randomized, placebo-controlled trial	NCT05347095
	IL23R	JNJ-67864238	CD	Study terminated early as futility criteria met	NCT04102111
<i>TNFSF15</i>	TNFSF15	PF-06480605	UC	Phase 2a single-arm trial	NCT02840721

CD Crohn’s disease, UC ulcerative colitis

used as biomarkers to screen for IBD and other gastrointestinal diseases [94–96]. A TWAS conducted in the gut microbiota has detected multiple tissue-specific candidate genes in the sigmoid and transverse colon, respectively, such as *TOB2P1* for Enterococcaceae in sigmoid colon, *WDR6* for Coprococcus in sigmoid colon, and *KCNIP3* for Veillonellaceae in transverse colon [97]. An association study using bioinformatic analysis in colorectal cancer also observed two overlapping pathways, the bile secretion and steroid hormone biosynthesis pathways, enriched by operational taxonomic units (OTUs) and gene expression patterns in colon tissue, respectively [98]. These results indicate a close cross-talk between the intestinal microbiota and the colon transcriptome. Thus, the results of colon TWAS could identify potential genes stimulated by the microbiota and provide hints to explore the colon-specific pathology of CD.

Four TWASs included colon as casual tissue, and each study included at least 2 segments of colon. In several studies, susceptibility genes observed in the sigmoid and transverse colon almost overlap. In Japanese population, among five genes observed in colon, four genes, including *ERV3-1*, *NPIP9*, *ZNF713*, and *WDR31* overlap between sigmoid and transverse colon [9]. However, none of these genes overlap with the results of other reports. For the cross-tissue TWAS, Uellendahl et al. found that 18 genes were differentially expressed in the sigmoid colon, and 31 genes were differentially expressed in the transverse colon [13]. Seven genes were overlapped between two segments of colon. In European ancestries, Cheng et al. found that *ZNF300P1* and *MICB* were significantly differentially expressed in both the sigmoid and transverse colon among the 3 susceptibility genes in colon [14]. In a meta-analysis, Virginia et al. conducted TWAS in three colon segments, including the ascending, transverse, and descending colon. And three colon segments had 11 overlapping genes, including *SLC22A5*, *GSDMB*, *ENTR1*, *ERAP2*, *C4A*, *FUT2*, *UBA7*, *GSDMA*, *FLRT3*, *RBM6*, and *HLA-C* [15]. In the comparison between Uellendahl's and Virginia's studies, *ERAP2* and *IL23R* in the transverse colon were observed in both. None of these susceptibility genes was replicated across studies in the sigmoid colon, because two of these studies found a relatively small number of genes (Additional file 2: data S2) [9, 14].

Among all susceptibility genes reported in colon regardless of which segments, *ERAP2* [13, 15], Interleukin-23 receptor (*IL23R*) [13, 15], Major histocompatibility complex class I chain-related gene B (*MICB*) [14, 15], Post-GPI attachment to the proteins 3 (*PGAP3*) [13, 15], and *SLC22A5* [14, 15] were overlapped between TWASs. *ERAP2* and *SLC22A5* were also overlapped with ileum discussed above.

*IL23R* is one of popular genes affects disease susceptibility and highly expressed on cell membrane of memory T cells and other immune cells, such as natural killer cells, monocytes, and dendritic cells [99]. *IL23R* interacts with IL-23, regulating the of immune activity and against infection by bacteria and viruses [99]. And the functional *IL23R* pathway polymorphisms play a role in modulating neonatal development of intestinal tolerance and bacterial colonization [100]. There were several of humanized monoclonal IgG, including Brazikumab, Risankizumab and Mirikizumab, could binds p19 of IL23 has entered clinical trial, and most of them has enter phase 3 clinical trials (Table 2). Due to the presence of protective or disease-associated variants in *IL23R* and related genes, only one locally acting oral peptide (JNJ-67864238) directly antagonizing IL-23R was found but was recently terminated after meeting criteria for futility [NCT04102111] [101].

*MICB* almost exclusively expressed in the intestinal epithelium [102]. *MICB* was reported in many human cancers via immune evasion [103–105]. And the immune cells, including natural killer (NK) cells and T cells, involved in *MICB* were also connected with CD. However, the functions of *MICB* in CD were still lack of evidence. In addition, *MICB* has only been reported as a CD susceptibility gene in whole blood, but not in blood immune cells, which needs further study.

*PGAP3* is ubiquitously expressed and code a Glycosylphosphatidylinositol (GPI)-specific phospholipase involving in lipid remodeling of GPI-anchored proteins [106]. The function of *PGAP3* was most reported in brain morphogenesis and mental development [107, 108]. However, the mechanism of *PGAP3* in CD was still under studied.

To be noted, the susceptibility genes associated with CD in colon enriched in the cell component of vesicle membrane (Fig. 3B), including exosomes, microvesicles and apoptotic bodies from endosomes, plasma membrane, plasma membrane/endoplasmic reticulum, respectively [109]. And the vesicle may be related with bacteria–host communication, which may involve in internalization of bacterial extracellular vesicles of epithelial cells [110]. Endocytic routes of intestinal epithelial cells, including macropinocytosis, clathrin-mediated endocytosis and lipid raft-mediated processes, may involve in CD pathogenesis [109]. Furthermore, extracellular vesicle (EV), mainly secreted by immune cells and intestinal epithelial cells, could package double-strand DNA (dsDNA), activating the STING pathway to provoke inflammatory responses [111]. Increasing evidence found EVs containing nucleotides have the potential to be biomarkers for the diagnosis of UC or general IBD [112–115]. And EVs may have therapeutic value for IBD [116].

However, the mechanism of EVs in the pathogenesis of CD remains to be further explored.

### Immune cells

The clinical symptoms of CD, including fever, diarrhea, and abdominal pain, mainly depend on the site of inflammation [117]. Various combinations of immune cell types and their locations may also indicate discrepant bio-metabolic pathways and pathogenesises.

### Intestinal immune cells

CD development involves a combination of environmental, microbial, and immune-mediated factors in individuals with susceptibility gene mutations [118]. Population studies have reported the highest incidence of CD activity in areas with high bacterial counts (colon) and relative retention of fecal material (terminal ileum and rectum) [119, 120]. Once bacteria destroy a single layer of columnar epithelium of the gut, the mucus and antimicrobial factors produced by the intestinal epithelium cannot defend against bacterial invasion [121], and the immune response occurs first in intestinal tissue. Thus, immune cells in the intestine could reflect the ultimate origin of CD.

At present, only Japanese studies have performed TWAS on immune cells in intestinal tissues using their own genome-wide and transcriptome data, and only *TNXA* was found to be significantly differentially expressed in CD4+ effector memory T cells (TEM cells) derived from lamina propria mononuclear cells (LPMCs) in the inflammatory sites of intestinal tissues [9]. Epstein–Barr virus (EBV) 3–1 in EBV-transformed lymphocytes was identified as a susceptibility gene for CD in Japanese patients using the GTEx database [9]. However, this gene was not screened in the populations of Western countries [14]. Owing to the small sample size, the study also broadened the significance level and defined *RAP1A* as a candidate gene associated with CD (FDR < 0.10) [9].

### Immune cell in blood

As the disease progresses, lesion locations may change or increase, and the risk of complicated diseases, such as rectal disease and perianal lesions, also increases. The metabolites of the microbiota associated with CD participate in immune progress, which can provoke the autoimmune response of the whole body [122]. Therefore, the genetics of immune cells in the circulatory system could also reflect the pathogenesis of CD.

There were three studies include blood immune cells as targeted tissue type. *ATG16L1*, *NOD2*, *ZGLP1*, *BRD7*, *CISD1*, and *SNX20* were significantly related to CD in multiple immune cells, including naïve CD4+ T cells, CD14+ monocytes, and CD16+ neutrophils, in the same

TWAS [13]. No gene was found to be significantly associated with CD using gene expression data from the Gene Expression Omnibus database to explore related genes in EBV-transformed lymphocytes, [14]. Because the diversity of immune cell types varies widely, no overlapped susceptibility gene in blood immune cells were observed across different TWASs.

Susceptibility genes in blood immune cell reported in TWASs are involved in the activation of immune cells and the maintenance of gastrointestinal epithelium (Fig. 3C), suggesting that immune cells may tend to function in the gut, where they may be more susceptible to CD.

### Blood

Whole blood is a heterogeneous tissue that includes a variety of immune cells, including lymphocytes, neutrophils, monocytes, and macrophages, with unique and disease-related roles in CD pathology. Extensive studies of whole blood or lymphoblasts are often used to maximize test power; however, they are mechanistically less relevant to disease. Owing to the relatively low cost of DNA and RNA extraction from whole blood, choosing whole blood for early exploration with a large sample size is feasible.

In a multi-tissue analysis using the GTEx in humans, compared with other tissue types, whole blood exhibited the fewest detected transcribed regions [123]. Whole blood seems to be a tissue type with less disease-specificity. However, it is also the most accessible biospecimen in clinical practice and could thus obtain sufficient test power with a large sample size. The differentially expressed gene is most likely to have the potential to be a biomarker and could be extensively used in clinical practice to help earlier diagnosis and disease classification.

Among seven TWASs associated with CD, five TWASs selected whole blood as targeted tissue. And among total 144 susceptibility genes in whole blood related with CD, *ATG16L1* [12, 13], Caspase recruitment domain 9 (*CARD9*) [12, 14], *ERAP2* [12, 13], *MICB* [12, 14], *NOD2* [10, 12, 13], *SLC22A5* [12, 14], and Tumor necrosis factor superfamily 15 (*TNFSF15*) [9, 10, 12] were overlapped between TWASs. *ATG16L1*, *ERAP2*, *MICB* and *SLC22A5* were overlapped with intestinal and discussed in the above.

Among the over 40 risk loci associated with CD identified to date, polymorphisms in *NOD2* account for the largest proportion of the genetic risk for this disease [124]. Experiments have demonstrated that *NOD2* recognizes bacterial muramyl dipeptides and recruits *ATG16L1* to bacterial entry sites on the plasma membrane, further regulating the intestinal barrier function and limiting transcellular permeability and bacterial

translocation [125, 126]. However, the differential expression of *NOD2* in circulatory—but not intestinal—tissues is puzzling. *NOD2* is widely expressed in macrophages and dendritic cells but to a lesser extent in intestinal epithelial cells [127] and T cells [128], which may explain this phenomenon. In addition, *NOD2* can act as a viral sensor protein and is activated by the orally bioavailable dinucleotide SB 9200 [129], but whether it is effective against CD is unknown.

*CARD9* was a member of CARD family and an adaptor molecule predominantly expressed in lymphoid tissues and immune cells. The expression of *CARD9* was observed significant reduced in CD patients compared with healthy controls [130]. *CARD9* is a central signaling molecule in the innate immune via mediating NF- $\kappa$ B signaling and against fungi, bacteria, virus and mycobacteria [131–134], which is closely related to the pathophysiological of CD development.

*TNFSF15* is a Th-1-polarized cytokine that participates in systemic inflammatory responses and functions in regulating immune cells, inducing apoptosis, inducing inflammation, and inhibiting tumorigenesis, which suggests that *TNFSF15* is possibly a susceptibility gene in the blood. However, two studies of IBD observed that *TNFSF15* was overexpressed in colonic tissues [135, 136]. The protective effect of *TNFSF15* polymorphisms on CD has been reviewed elsewhere [99, 137]. Notably, *TNFSF15* was only screened in two Asian populations, Japanese and Korean, and was not found in other populations, and it can be tentatively speculated that the association of *TNFSF15* with CD is stronger in Asian than in Western populations. Many previous studies have supported this hypothesis. A Japanese study reported a trend for a positive association between *TNFSF15* SNPs and the risk of anal lesions in CD [138]. Similar results were obtained in Chinese [139] and Korean population [140]. In European population, a protective effect of *TNFSF15* was observed in CD but fail to define a clinical subgroup of CD patients specifically associated with *TNFSF15* [19]. A most recent study compared the susceptibility genes associated with IBD between two population from East Asian and European ancestries, respectively. In this study, researchers found that the genetic basis of CD appears to be more ancestral than that of UC due to the allele frequency of *NOD2* and the influence of *TNFSF15* [19]. And a meta-analysis also observed that East Asians gene have unique SNPs of *TNFSF15* associated with IBD [141]. And PF-06480605, an inhibitor of *TNFSF15*, has been developed and enter the Phase 2a clinical trial as a treatment strategy of UC patients (NCT02840721, Table 2).

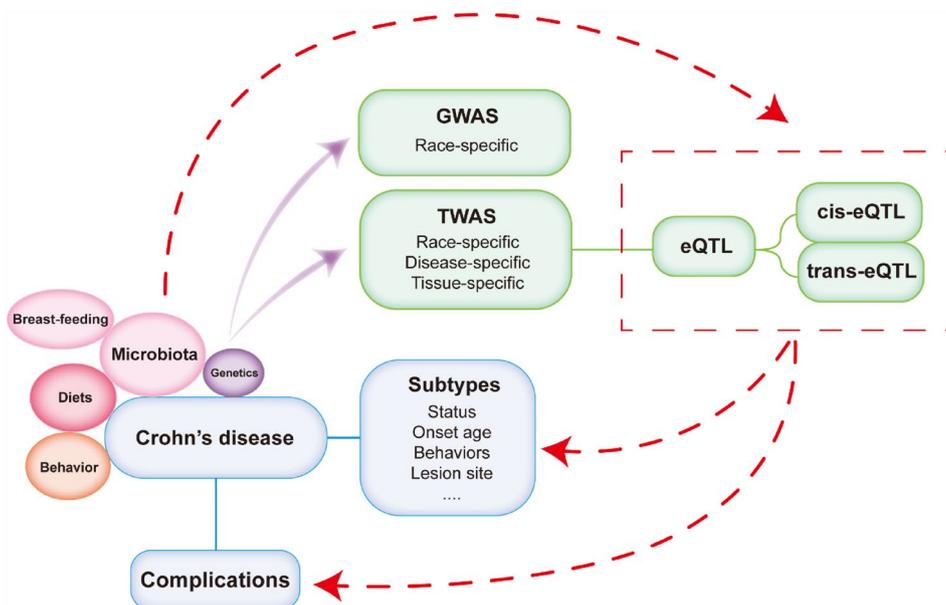
The rough GO pathways enriched in whole blood mainly involves in the regulation of responses to biotic stimulus, including the regulation of defends system and

the activation of immune system (Fig. 3D). This observation may suggest that the activity of immune cells and the process of inflammatory response in the whole blood may be affected by genetic background and reflect the disease status of CD.

### Future perspectives

In this review, we summarized the susceptibility genes and enriched pathways associated with CD found in TWAS. Most susceptibility genes replicated between different tissues can only be observed in the same TWAS. For a fixed tissue type, susceptibility genes were rarely replicated in different TWAS. For instance, Uellendahl et al. identified *ATG16L1* as a susceptibility gene in almost all tissue types [13] but absent in Cheng's multi-tissue TWAS [14]. Similarly, *MICB* was significantly associated with CD in Cheng's study [14] but was absent in the other TWAS [9, 11–13]. In this section, we discuss the reasons for these issues, strategies for solving them, and future directions, as shown in Fig. 4.

Epidemiologic studies have reported that the incidence of IBD has now leveled off in developed areas but continues to increase in developing regions [142]. This phenomenon suggests that CD is more likely to result from the interaction of dietary behaviors and environmental factors with host immune mechanisms. The onset of CD can even be traced back to infancy; breast milk, containing oligosaccharides, contributes to the establishment of intestinal flora in infants and has a longer duration of benefit in inhibiting the adhesion of enteropathogenic bacteria and protecting against the development of the disease [142]. A recent study reported a transcriptome-wide association with gut microbiota [97]. Since numerous studies have demonstrated that the gut microbiota has a close relationship with CD pathology, the TWAS in the gut microbiota may provide new insights into investigating novel pathological mechanisms. This new gut microbiota-based tool may be influenced by diet and medication use, and its applicability to CD research remains to be demonstrated. Thus, broader exploration is needed, and the results should be interpreted with caution. Although over 200 genetic loci associated with IBD have been identified by GWAS, these variants can only explain a small proportion of the heritability of IBD (approximately 26% for CD and 19% for UC) [143]. No genetic markers have been reported to be predictive of complications [37, 144]. Considering that total expression is affected by genetic and environmental factors and that predicted expression in TWAS is only a part of the total expression, gene expression data assessed by genotype data and eQTLs have strong limitations and biases. The predicted expression in TWAS was generally slightly higher than the total expression correlations. The analysis



**Fig. 4** Overview of future research direction of TWAS in CD. GWAS, genome-wide association studies; TWAS, transcriptome-wide association studies; eQTL, expression quantitative trait loci

of the correlation between predicted expression and CD may result in the significance of non-causal genes over causal genes due to linkage disequilibrium [32]. When a large proportion of genes are in linkage disequilibrium, the linkage disequilibrium region may also contain causal associations not related to the gene set. Even if there is no causal relationship between the gene set and the phenotype, it can still exhibit a significantly high rate [145]. Therefore, more work is needed to understand the genetic structure of CD.

Before the emergence of GWAS, most genotypic-phenotypic associations failed to replicate owing to small sample sizes, improper reliance on standard significance thresholds, failure to account for associations with low prior probabilities, and failure to assess the same SNPs across studies [1, 146]. We must acknowledge that the replication of TWAS results is not easy. The following reasons make the replication for CD TWAS even harder. (1) As a complex disease, the subtypes of CD patients, such as age of onset, lesion site, disease behaviors, disease process (active/remission), and medication used (hormone/biologicals), have different expression profile [42, 147], which should be included in GWAS datasets. Due to the lack of detailed information on large-scale genotype data sets, few studies have explored the relationship between predicted expression data and CD subtypes. This critically restricts further exploration of genetic factors in CD pathology. As a progressive disease, the onset time of symptoms and occurrence of complications are also important outcomes. (2) The establishment

of eQTLs, including the sample size, ethnic consistency with the GWAS population, and different control populations, will also have a great impact on the TWAS results. (3) The rapid improvement and update of TWAS methods also reduce the reproducibility of TWAS results. Therefore, replication of these findings still needs a good deal of work in the future.

Disease behavior of CD changes over time, and patients with inflammation as the main presenting behavior at diagnosis are highly likely to develop fistulas or stricture complications within 20 years [148]. Prolonged inflammatory responses during clinical remission can lead to complications (strictures, fistulas, and abscesses) and progressive intestinal damage [149]. The Montreal classification considers in detail whether a prescribed time point should be given before disease behavior classification [5]. The homogeneity of participants and tissue type could effectively ensure the credibility of the results. Even compared with healthy people, the heterogeneity within CD patients can cause a large bias. Most of the populations in the available studies were heterogeneous, including patients with various lesion sites or different disease states (active or remission). There are also other kinds of population heterogeneity. For instance, Gettler et al. selected a TWAS population from a RISK cohort that recruited children and adolescents under 17 years of age [11, 150]. The susceptibility genes in this study may indicate a different pathology compared to that in the adult study. This Japanese study investigated differentially expressed genes between patients diagnosed with active

CD and active UC, which may reflect different pathological processes in CD patients and healthy controls [9]. The heterogeneity of TWAS results in different tissue types is discussed at length above. Two IBD fine mapping studies published in 2017 found less than 30% concordance between eQTL and GWAS in identifying key genes [151, 152]. One study suggested a more significant overlap between eQTL and methylation QTL, and both studies suggested that related effects may be specific to cell type or disease status [151, 152]. During the past 100 years, the incidence of inflammatory bowel disease has sharply risen, then plateaued in the western world, whereas countries outside the western world seem to be in the first stage of this sequence [142]. The fact that CD patients are often treated for life and tend to be concentrated in a few well-known treatment centers, it is feasible to obtain sufficient numbers of patients and biological samples through recruitment in well-known treatment centers. And a certain tissue type within the same study design is recommended to be consistent, allowing for increased sample size on a limited budget. In the future, larger-scale, cell/tissue-specific, and status-specific studies will be vital to resolve this problem. With the advancement of technology, single-cell RNA sequencing and single-cell TWAS have already emerged, substantially improving the homogeneity of samples and further facilitating targeted interpretations of TWAS outcomes and disease mechanisms for individual cell types or specific disease states [153].

Among the most frequently mentioned TWAS genes, such as *ATGL16L1*, *NOD2* and *IL23R*, were most reported by coding risk variants in GWAS studies instead of replicating the results of RNA-seq. Among the seven studies, only one TWAS provided the gene list associated with CD obtained by RNA-seq data [12]. Of the 95 associated genes screened by TWAS and the 35 associated genes screened by RNA-seq, only two genes (*RPL9* and *STMN3*) overlapped. Since the most of RNA-seq data involved in the other six TWAS were either not associated with CD or did not include appropriate cases and controls (HC only, CD only, CD and UC), we found another two well-designed studies for comparison. Two Asina studies identified differentially expressed genes by RNA-seq in CD patients [154, 155]. Unfortunately, there was no susceptibility gene overlap between RNA-seq results and TWAS results neither in the ileum or colon. And the susceptibility genes screened by TWAS was less overlapped with the results of RNA-seq. There are several reasons for this phenomenon: (1) According to the distance of gene effect, eQTL includes cis-eQTLs (local) and trans-eQTLs (distal) [156]. A previous study observed that pervasive cis-eQTLs affect the majority of human genes (~75%) [157, 158], but a large twin study claimed

that only 10% of the variation in gene expression was explained by cis-eQTL [123]. However, cis-eQTLs remain the only reliable tool in the TWAS method [32], which was limited in assessing the allele-specific expression [159]. (2) Stretch enhancers are large chromatin-defined regulatory elements that regulate the expression of cell type-specific genes and are enriched in disease-associated genetic variants in disease-associated cell types. However, eQTL effect sizes for stretch enhancers may be smaller than for ubiquitous promoter regions, which may lead to prediction bias [34]. (3) The pleiotropy, including horizontal pleiotropy and vertical pleiotropy, is widely existed in genome but the exact extent is still unknown [160]. Most of the genes may be indirect causative genes for complex traits, and some of the GWAS gene expression predicted by eqtl may be amplified due to horizontal pleiotropy [35]. (4) Gene expression may be affected by heritable epigenetic variation, small signaling molecules or other environment factors [161]. For example, *NOD2* expression could be induced by bacterial lipopolysaccharide, short-chain fatty acids, hormonal vitamin D, and TNF- $\alpha$  [162], which make it harder to predict the real expression levels. Although rarely reported in RNA-seq studies, these gene expressions are involved in mucosal immunity as previous reported [99, 162, 163]. This suggests that RNA-seq and TWAS may have complementary roles in explaining genetic associations of complex traits.

Counterintuitively, a susceptibility gene, such as *NOD2*, identified in a tissue type is not always consistent with its function. This observation raises the question of whether differential expression results obtained by eQTLs can explain causal associations, and a growing body of data has raised this question. As Wainberg et al. pointed out, the TWAS method is merely a statistical test to predict expression and disease risk from genetic evidence, which can be used to screen candidate disease-causing genes but does not guarantee causality [32].

In conclusion, the following three considerations might benefit future TWAS for CD, facilitating a more rational study design. (1) Despite its generic nature, we require GWAS data from different countries and disease states with large sample sizes. (2) The demand for a comprehensive classification, including race, tissue, lesion site, status, and progressive time points, is increasing with the accumulation of eQTL data. (3) Transcriptome-wide data combined with new technologies, such as single-cell approaches, will provide novel insights into the pathological mechanisms of CD and progress in TWAS. (4) In future TWAS, the results and data of intermediate processes should also be provided to facilitate the integration of data from multiple studies, dig deeper into genetic information, and provide more predictions for drug discovery.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13578-024-01204-w>.

**Additional file 1.** The list of susceptibility genes in various tissue types screened out by different TWAS methods.

**Additional file 2.** All susceptibility genes associated with CD in various tissue types screened by TWAS.

**Additional file 3.** The total significant results of GO functional analyses.

### Acknowledgements

The authors are grateful to all funders who contributed to the manuscript.

### Author contributions

JS conceived the idea. KJ curated data. JS and KJ wrote the manuscript. All authors were involved in the interpretation of the results.

### Funding

Supported by grants from the Most Important Discipline, Baoshan Medical Key Discipline and Characteristic Brand Construction Project (BSZK-2023-Z06), 2022 Baoshan Medical and Health Project (2023-E-10, 2023-E-13) and Cultivation Funding for Clinical Scientific Research Innovation, Renji Hospital, School of Medicine, Shanghai Jiaotong University (RJPY-LX-004).

### Availability of data and materials

Data sharing is not applicable to this article as no datasets were generated or analysed during the current study.

### Declarations

#### Ethics approval and consent to participate

Not applicable.

#### Consent for publication

Not applicable.

#### Competing interests

The authors declare that they have no competing interests.

Received: 27 September 2023 Accepted: 4 February 2024

Published online: 25 February 2024

### References

- Kraft P, Zeggini E, Ioannidis JP. Replication in genome-wide association studies. *Stat Sci*. 2009;24(4):561–73. <https://doi.org/10.1214/09-STS290>.
- Sollis E, Mosaku A, Abid A, Buniello A, Cerezo M, Gil L, Groza T, Gunes O, Hall P, Hayhurst J, et al. The NHGRI-EBI GWAS catalog: knowledgebase and deposition resource. *Nucleic Acids Res*. 2023;51(D1):D977–85. <https://doi.org/10.1093/nar/gkac1010>.
- Maurano MT, Humbert R, Rynes E, Thurman RE, Haugen E, Wang H, Reynolds AP, Sandstrom R, Qu H, Brody J, et al. Systematic localization of common disease-associated variation in regulatory DNA. *Science*. 2012;337(6099):1190–5. <https://doi.org/10.1126/science.1222794>.
- Torres J, Mehandru S, Colombel JF, Peyrin-Biroulet L. Crohn's disease. *Lancet*. 2017;389(10080):1741–55. [https://doi.org/10.1016/S0140-6736\(16\)31711-1](https://doi.org/10.1016/S0140-6736(16)31711-1).
- Satsangi J, Silverberg MS, Vermeire S, Colombel JF. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. *Gut*. 2006;55(6):749–53. <https://doi.org/10.1136/gut.2005.082909>.
- Jostins L, Ripke S, Weersma RK, Duerr RH, McGovern DP, Hui KY, Lee JC, Schumm LP, Sharma Y, Anderson CA, et al. Host-microbe interactions have shaped the genetic architecture of inflammatory bowel disease. *Nature*. 2012;491(7422):119–24. <https://doi.org/10.1038/nature11582>.
- Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, Ripke S, Lee JC, Jostins L, Shah T, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet*. 2015;47(9):979–86. <https://doi.org/10.1038/ng.3359>.
- Kakuta Y, Kawai Y, Naito T, Hirano A, Umeno J, Fuyuno Y, Liu Z, Li D, Nakano T, Izumiya Y, et al. A genome-wide association study identifying RAP1A as a novel susceptibility gene for Crohn's disease in Japanese individuals. *J Crohns Colitis*. 2019;13(5):648–58. <https://doi.org/10.1093/ecco-jcc/jjy197>.
- Kakuta Y, Ichikawa R, Fuyuno Y, Hirano A, Umeno J, Torisu T, Watanabe K, Asakura A, Nakano T, Izumiya Y, et al. An integrated genomic and transcriptomic analysis reveals candidates of susceptibility genes for Crohn's disease in Japanese populations. *Sci Rep*. 2020;10(1):10236. <https://doi.org/10.1038/s41598-020-66951-5>.
- Jung S, Liu W, Baek J, Moon JW, Ye BD, Lee HS, Park SH, Yang SK, Han B, Liu J, et al. Expression quantitative trait Loci (eQTL) mapping in Korean patients with Crohn's disease and identification of potential causal genes through integration with disease associations. *Front Genet*. 2020;11:486. <https://doi.org/10.3389/fgene.2020.00486>.
- Gettler K, Giri M, Kenigsberg E, Martin J, Chuang LS, Hsu NY, Denson LA, Hyams JS, Griffiths A, Noe JD, et al. Prioritizing Crohn's disease genes by integrating association signals with gene expression implicates monocyte subsets. *Genes Immun*. 2019;20(7):577–88. <https://doi.org/10.1038/s41435-019-0059-y>.
- Dai Y, Pei G, Zhao Z, Jia P. A Convergent study of genetic variants associated with Crohn's disease: evidence from gwas, gene expression, methylation, eQTL and TWAS. *Front Genet*. 2019;10:318. <https://doi.org/10.3389/fgene.2019.00318>.
- Uellendahl-Werth F, Maj C, Borisov O, Juzenas S, Wacker EM, Jorgensen IF, Steiert TA, Bej S, Krawitz P, Hoffmann P, et al. Cross-tissue transcriptome-wide association studies identify susceptibility genes shared between schizophrenia and inflammatory bowel disease. *Commun Biol*. 2022;5(1):80. <https://doi.org/10.1038/s42003-022-03031-6>.
- Cheng B, Liang X, Wen Y, Li P, Zhang L, Ma M, Cheng S, Du Y, Liu L, Ding M, et al. Integrative analysis of transcriptome-wide association study data and messenger RNA expression profiles identified candidate genes and pathways for inflammatory bowel disease. *J Cell Biochem*. 2019;120(9):14831–7. <https://doi.org/10.1002/jcb.28744>.
- Diez-Obrero V, Moratalla-Navarro F, Ibanez-Sanz G, Guardiola J, Rodriguez-Moranta F, Obon-Santacana M, Diez-Villanueva A, Dampier CH, Devall M, Carreras-Torres R, et al. Transcriptome-wide association study for inflammatory bowel disease reveals novel candidate susceptibility genes in specific colon subsites and tissue categories. *J Crohns Colitis*. 2022;16(2):275–85. <https://doi.org/10.1093/ecco-jcc/jjab131>.
- Radstake TR, Gorlova O, Rueda B, Martin JE, Alizadeh BZ, Palomino-Morales R, Coenen MJ, Vonk MC, Voskuyl AE, Schuerwegh AJ, et al. Genome-wide association study of systemic sclerosis identifies CD247 as a new susceptibility locus. *Nat Genet*. 2010;42(5):426–9. <https://doi.org/10.1038/ng.565>.
- Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, Daly MJ, Bustamante CD, Kenny EE. Human demographic history impacts genetic risk prediction across diverse populations. *Am J Hum Genet*. 2017;100(4):635–49. <https://doi.org/10.1016/j.ajhg.2017.03.004>.
- Carlson CS, Matise TC, North KE, Haiman CA, Fesinmeyer MD, Buyske S, Schumacher FR, Peters U, Franceschini N, Ritchie MD, et al. Generalization and dilution of association results from European GWAS in populations of non-European ancestry: the PAGE study. *PLoS Biol*. 2013;11(9):e1001661. <https://doi.org/10.1371/journal.pbio.1001661>.
- Liu Z, Liu R, Gao H, Jung S, Gao X, Sun R, Liu X, Kim Y, Lee HS, Kawai Y, et al. Genetic architecture of the inflammatory bowel diseases across East Asian and European ancestries. *Nat Genet*. 2023;55(5):796–806. <https://doi.org/10.1038/s41588-023-01384-0>.
- Need AC, Goldstein DB. Next generation disparities in human genomics: concerns and remedies. *Trends Genet*. 2009;25(11):489–94. <https://doi.org/10.1016/j.tig.2009.09.012>.
- Bustamante CD, Burchard EG, De la Vega FM. Genomics for the world. *Nature*. 2011;475(7355):163–5. <https://doi.org/10.1038/475163a>.
- Popejoy AB, Fullerton SM. Genomics is failing on diversity. *Nature*. 2016;538(7624):161–4. <https://doi.org/10.1038/538161a>.

23. Manning A, Highland HM, Gasser J, Sim X, Tukiainen T, Fontanillas P, Grarup N, Rivas MA, Mahajan A, Locke AE, et al. A low-frequency inactivating AKT2 variant enriched in the finnish population is associated with fasting insulin levels and type 2 diabetes risk. *Diabetes*. 2017;66(7):2019–32. <https://doi.org/10.2337/db16-1329>.
24. Estrada K, Aukrust I, Bjørkhaug L, Burt NP, Mercader JM, García-Ortiz H, Huerta-Chagoya A, Moreno-Macías H, Walford G, Flannick J, et al. Association of a low-frequency variant in HNF1A with type 2 diabetes in a Latino population. *JAMA*. 2014;311(22):2305–14. <https://doi.org/10.1001/jama.2014.6511>.
25. Ventham NT, Kennedy NA, Adams AT, Kalla R, Heath S, O’Leary KR, Drummond H, Wilson DC, Gut IG, Nimmo ER, et al. Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease. *Nat Commun*. 2016;7:13507. <https://doi.org/10.1038/ncomms13507>.
26. Nicolae DL, Gamazon E, Zhang W, Duan S, Dolan ME, Cox NJ. Trait-associated SNPs are more likely to be eQTLs: an annotation to enhance discovery from GWAS. *PLoS Genet*. 2010;6(4): e1000888. <https://doi.org/10.1371/journal.pgen.1000888>.
27. Mo A, Marigorta UM, Arafat D, Chan LHK, Ponder L, Jang SR, Prince J, Kugathasan S, Prahalad S, Gibson G. Disease-specific regulation of gene expression in a comparative analysis of juvenile idiopathic arthritis and inflammatory bowel disease. *Genome Med*. 2018;10(1):48. <https://doi.org/10.1186/s13073-018-0558-x>.
28. Portela A, Esteller M. Epigenetic modifications and human disease. *Nat Biotechnol*. 2010;28(10):1057–68. <https://doi.org/10.1038/nbt.1685>.
29. Yoo T, Joo SK, Kim HJ, Kim HY, Sim H, Lee J, Kim HH, Jung S, Lee Y, Jamialahmadi O, et al. Disease-specific eQTL screening reveals an anti-fibrotic effect of AGXT2 in non-alcoholic fatty liver disease. *J Hepatol*. 2021;75(3):514–23. <https://doi.org/10.1016/j.jhep.2021.04.011>.
30. Vösa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, Kirsten H, Saha A, Kreuzhuber R, Yazar S, et al. Large-scale cis- and trans-eQTL analyses identify thousands of genetic loci and polygenic scores that regulate blood gene expression. *Nat Genet*. 2021;53(9):1300–10. <https://doi.org/10.1038/s41588-021-00913-z>.
31. Westra HJ, Peters MJ, Esko T, Yaghoobkar H, Schurmann C, Kettunen J, Christiansen MW, Fairfax BP, Schramm K, Powell JE, et al. Systematic identification of trans eQTLs as putative drivers of known disease associations. *Nat Genet*. 2013;45(10):1238–43. <https://doi.org/10.1038/ng.2756>.
32. Wainberg M, Sinnott-Armstrong N, Mancuso N, Barbeira AN, Knowles DA, Golan D, Ermel R, Ruusalepp A, Quertermous T, Hao K, et al. Opportunities and challenges for transcriptome-wide association studies. *Nat Genet*. 2019;51(4):592–9. <https://doi.org/10.1038/s41588-019-0385-z>.
33. Barbeira AN, Dickinson SP, Bonazzola R, Zheng J, Wheeler HE, Torres JM, Torstenson ES, Shah KP, Garcia T, Edwards TL, et al. Exploring the phenotypic consequences of tissue specific gene expression variation inferred from GWAS summary statistics. *Nat Commun*. 2018;9(1):1825. <https://doi.org/10.1038/s41467-018-03621-1>.
34. Viñuela A, Varshney A, van de Bunt M, Prasad RB, Asplund O, Bennett A, Boehnke M, Brown AA, Erdos MR, Fadista J, et al. Genetic variant effects on gene expression in human pancreatic islets and their implications for T2D. *Nat Commun*. 2020;11(1):4912. <https://doi.org/10.1038/s41467-020-18581-8>.
35. Hukku A, Sampson MG, Luca F, Pique-Regi R, Wen X. Analyzing and reconciling colocalization and transcriptome-wide association studies from the perspective of inferential reproducibility. *Am J Hum Genet*. 2022;109(5):825–37. <https://doi.org/10.1016/j.ajhg.2022.04.005>.
36. Atreya R, Siegmund B. Location is important: differentiation between ileal and colonic Crohn’s disease. *Nat Rev Gastroenterol Hepatol*. 2021;18(8):544–58. <https://doi.org/10.1038/s41575-021-00424-6>.
37. Cleyne I, Boucher G, Jostins L, Schumm LP, Zeissig S, Ahmad T, Andersen V, Andrews JM, Annese V, Brand S, et al. Inherited determinants of Crohn’s disease and ulcerative colitis phenotypes: a genetic association study. *Lancet*. 2016;387(10014):156–67. [https://doi.org/10.1016/S0140-6736\(15\)00465-1](https://doi.org/10.1016/S0140-6736(15)00465-1).
38. Yu G, Wang LG, Han Y, He QY. clusterProfiler: an R package for comparing biological themes among gene clusters. *OMICS*. 2012;16(5):284–7. <https://doi.org/10.1089/omi.2011.0118>.
39. Luo W, Brouwer C. Pathview: an R/Bioconductor package for pathway-based data integration and visualization. *Bioinformatics*. 2013;29(14):1830–1. <https://doi.org/10.1093/bioinformatics/btt285>.
40. Baumgart DC, Carding SR. Inflammatory bowel disease: cause and immunobiology. *Lancet*. 2007;369(9573):1627–40. [https://doi.org/10.1016/S0140-6736\(07\)60750-8](https://doi.org/10.1016/S0140-6736(07)60750-8).
41. Wu F, Dassopoulos T, Cope L, Maitra A, Brant SR, Harris ML, Bayless TM, Parmigiani G, Chakravarti S. Genome-wide gene expression differences in Crohn’s disease and ulcerative colitis from endoscopic pinch biopsies: insights into distinctive pathogenesis. *Inflamm Bowel Dis*. 2007;13(7):807–21. <https://doi.org/10.1002/ibd.20110>.
42. Zhang Y, Shen B, Zhuge L, Xie Y. Identification of differentially expressed genes between the colon and ileum of patients with inflammatory bowel disease by gene co-expression analysis. *J Int Med Res*. 2020;48(5):300060519887268. <https://doi.org/10.1177/0300060519887268>.
43. Brand S. Crohn’s disease: Th1, Th17 or both? The change of a paradigm: new immunological and genetic insights implicate Th17 cells in the pathogenesis of Crohn’s disease. *Gut*. 2009;58(8):1152–67. <https://doi.org/10.1136/gut.2008.163667>.
44. Bogaert S, Laukens D, Peeters H, Melis L, Olievier K, Boon N, Verbruggen G, Vandesompele J, Elewaut D, De Vos M. Differential mucosal expression of Th17-related genes between the inflamed colon and ileum of patients with inflammatory bowel disease. *BMC Immunol*. 2010;11:61. <https://doi.org/10.1186/1471-2172-11-61>.
45. Holtta V, Klemetti P, Sipponen T, Westerholm-Ormio M, Kociubinski G, Salo H, Rasanen L, Kolho KL, Farkkila M, Savilahti E, et al. IL-23/IL-17 immunity as a hallmark of Crohn’s disease. *Inflamm Bowel Dis*. 2008;14(9):1175–84. <https://doi.org/10.1002/ibd.20475>.
46. Andoh A, Zhang Z, Inatomi O, Fujino S, Deguchi Y, Araki Y, Tsujikawa T, Kitoh K, Kim-Mitsuyama S, Takayanagi A, et al. Interleukin-22, a member of the IL-10 subfamily, induces inflammatory responses in colonic subepithelial myofibroblasts. *Gastroenterology*. 2005;129(3):969–84. <https://doi.org/10.1053/j.gastro.2005.06.071>.
47. Seiderer J, Elben I, Diegelmann J, Glas J, Stallhofer J, Tillack C, Pfennig S, Jurgens M, Schmechel S, Konrad A, et al. Role of the novel Th17 cytokine IL-17F in inflammatory bowel disease (IBD): upregulated colonic IL-17F expression in active Crohn’s disease and analysis of the IL17F pHis161Arg polymorphism in IBD. *Inflamm Bowel Dis*. 2008;14(4):437–45. <https://doi.org/10.1002/ibd.20339>.
48. Schmechel S, Konrad A, Diegelmann J, Glas J, Wetzke M, Paschos E, Lohse P, Goke B, Brand S. Linking genetic susceptibility to Crohn’s disease with Th17 cell function: IL-22 serum levels are increased in Crohn’s disease and correlate with disease activity and IL23R genotype status. *Inflamm Bowel Dis*. 2008;14(2):204–12. <https://doi.org/10.1002/ibd.20315>.
49. Fujino S, Andoh A, Bamba S, Ogawa A, Hata K, Araki Y, Bamba T, Fujiyama Y. Increased expression of interleukin 17 in inflammatory bowel disease. *Gut*. 2003;52(1):65–70. <https://doi.org/10.1136/gut.52.1.65>.
50. Villmones HC, Halland A, Stenstad T, Ulvestad E, Weedon-Fekjær H, Kommedal Ø. The cultivable microbiota of the human distal ileum. *Clin Microbiol Infect*. 2021;27(6):912.e7–912.e13. <https://doi.org/10.1016/j.cmi.2020.08.021>.
51. Guarner F, Malagelada JR. Gut flora in health and disease. *Lancet*. 2003;361(9356):512–9. [https://doi.org/10.1016/S0140-6736\(03\)12489-0](https://doi.org/10.1016/S0140-6736(03)12489-0).
52. Kee BP, Ng JG, Ng CC, Hilmi I, Goh KL, Chua KH. Genetic polymorphisms of ATG16L1 and IRGM genes in Malaysian patients with Crohn’s disease. *J Dig Dis*. 2020;21(1):29–37. <https://doi.org/10.1111/1751-2980.12829>.
53. Prescott NJ, Fisher SA, Franke A, Hampe J, Onnie CM, Soars D, Bagnall R, Mirza MM, Sanderson J, Forbes A, et al. A nonsynonymous SNP in ATG16L1 predisposes to ileal Crohn’s disease and is independent of CARD15 and IBD5. *Gastroenterology*. 2007;132(5):1665–71. <https://doi.org/10.1053/j.gastro.2007.03.034>.
54. Ouellette AJ. Paneth cells and innate mucosal immunity. *Curr Opin Gastroenterol*. 2010;26(6):547–53. <https://doi.org/10.1097/MOG.0b013e32833dcccde>.
55. Cadwell K, Liu JY, Brown SL, Miyoshi H, Loh J, Lennerz JK, Kishi C, Kc W, Carrero JA, Hunt S, et al. A key role for autophagy and the autophagy gene Atg16l1 in mouse and human intestinal Paneth cells. *Nature*. 2008;456(7219):259–63. <https://doi.org/10.1038/nature07416>.

56. Murthy A, Li Y, Peng I, Reichelt M, Katakam AK, Noubade R, Roose-Girma M, DeVoss J, Diehl L, Graham RR, et al. A Crohn's disease variant in Atg16L1 enhances its degradation by caspase 3. *Nature*. 2014;506(7489):456–62. <https://doi.org/10.1038/nature13044>.
57. Sadaghian Sadabad M, Regeling A, de Goffau MC, Blokzijl T, Weersma RK, Penders J, Faber KN, Harmsen HJ. The ATG16L1-T300A allele impairs clearance of pathosymbionts in the inflamed ileal mucosa of Crohn's disease patients. *Gut*. 2015;64(10):1546–52. <https://doi.org/10.1136/gutjnl-2014-307289>.
58. Martínez-Medina M, Aldeguer X, Lopez-Siles M, González-Huix F, López-Oliu C, Dahbi G, Blanco JE, Blanco J, García-Gil LJ, A. Darfeuille-Michaud Molecular diversity of *Escherichia coli* in the human gut: new ecological evidence supporting the role of adherent-invasive *E coli* (AIEC) in Crohn's disease. *Inflamm Bowel Dis*. 2009;15(6):872–82. <https://doi.org/10.1002/ibd.20860>.
59. Liu K, Hong D, Zhang F, Li X, He M, Han X, Zhang G, Xu G, Stonehouse NJ, Jiang Z, et al. MicroRNA-106a inhibits autophagy process and antimicrobial responses by targeting ULK1, ATG7, and ATG16L1 during mycobacterial infection. *Front Immunol*. 2020;11: 610021. <https://doi.org/10.3389/fimmu.2020.610021>.
60. Huang H, Tang J, Zhang L, Bu Y, Zhang X. miR-874 regulates multiple-drug resistance in gastric cancer by targeting ATG16L1. *Int J Oncol*. 2018;53(6):2769–79. <https://doi.org/10.3892/ijo.2018.4593>.
61. Chen R, Li X, He B, Hu W. MicroRNA-410 regulates autophagy-related gene ATG16L1 expression and enhances chemosensitivity via autophagy inhibition in osteosarcoma. *Mol Med Rep*. 2017;15(3):1326–34. <https://doi.org/10.3892/mmr.2017.6149>.
62. Li Y, Zhou D, Ren Y, Zhang Z, Guo X, Ma M, Xue Z, Lv J, Liu H, Xi Q, et al. Mir223 restrains autophagy and promotes CNS inflammation by targeting ATG16L1. *Autophagy*. 2019;15(3):478–92. <https://doi.org/10.1080/15548627.2018.1522467>.
63. Evnouchidou I, Birtley J, Seregin S, Papakyriakou A, Zervoudi E, Samiotaki M, Panayotou G, Giastasi P, Petrakis O, Georgiadis D, et al. A common single nucleotide polymorphism in endoplasmic reticulum aminopeptidase 2 induces a specificity switch that leads to altered antigen processing. *J Immunol*. 2012;189(5):2383–92. <https://doi.org/10.4049/jimmunol.1200918>.
64. Hamilton F, Mentzer AJ, Parks T, Baillie JK, Smith GD, Ghazal P, Timpson NJ. Variation in ERAP2 has opposing effects on severe respiratory infection and autoimmune disease. *Am J Hum Genet*. 2023;110(4):691–702. <https://doi.org/10.1016/j.ajhg.2023.02.008>.
65. Goyette P, Boucher G, Mallon D, Ellinghaus E, Jostins L, Huang H, Ripke S, Gusareva ES, Annese V, Hauser SL, et al. High-density mapping of the MHC identifies a shared role for HLA-DRB1\*01:03 in inflammatory bowel diseases and heterozygous advantage in ulcerative colitis. *Nat Genet*. 2015;47(2):172–9. <https://doi.org/10.1038/ng.3176>.
66. López de Castro JA, Alvarez-Navarro C, Brito A, Guasp P, Martín-Esteban A, Sanz-Bravo A. Molecular and pathogenic effects of endoplasmic reticulum aminopeptidases ERAP1 and ERAP2 in MHC-I-associated inflammatory disorders: Towards a unifying view. *Mol Immunol*. 2016;77:193–204. <https://doi.org/10.1016/j.molimm.2016.08.005>.
67. Castro-Santos P, Moro-García MA, Marcos-Fernández R, Alonso-Arias R, Díaz-Peña R. ERAP1 and HLA-C interaction in inflammatory bowel disease in the Spanish population. *Innate Immun*. 2017;23(5):476–81. <https://doi.org/10.1177/1753425917716527>.
68. Georgiadis D, Mpakali A, Koumantou D, Stratikos E. Inhibitors of ER aminopeptidase 1 and 2: from design to clinical application. *Curr Med Chem*. 2019;26(15):2715–29. <https://doi.org/10.2174/0929867325666180214111849>.
69. Yu P, Luo S, Cai J, Li J, Peng C. ERAP2 as a potential biomarker for predicting gemcitabine response in patients with pancreatic cancer. *Aging (Albany NY)*. 2022;14(19):7941–58. <https://doi.org/10.18632/aging.204324>.
70. Ma G, Liu H, Du M, Zhang G, Lin Y, Ge Y, Wang M, Jin G, Zhao Q, Chu H, et al. A genetic variation in the CpG island of pseudogene GBAP1 promoter is associated with gastric cancer susceptibility. *Cancer*. 2019;125(14):2465–73. <https://doi.org/10.1002/cncr.32081>.
71. Straniero L, Rimoldi V, Samarani M, Goldwurm S, Di Fonzo A, Krüger R, Deleidi M, Aureli M, Soldà G, Duga S, et al. The GBAP1 pseudogene acts as a ceRNA for the glucocerebrosidase gene GBA by sponging miR-22-3p. *Sci Rep*. 2017;7(1):12702. <https://doi.org/10.1038/s41598-017-12973-5>.
72. Broz P, Pelegrín P, Shao F. The gasdermins, a protein family executing cell death and inflammation. *Nat Rev Immunol*. 2020;20(3):143–57. <https://doi.org/10.1038/s41577-019-0228-2>.
73. Fagerberg L, Hallström BM, Oksvold P, Kampf C, Djureinovic D, Odeberg J, Habuka M, Tahmasebpoor S, Danielsson A, Edlund K, et al. Analysis of the human tissue-specific expression by genome-wide integration of transcriptomics and antibody-based proteomics. *Mol Cell Proteomics*. 2014;13(2):397–406. <https://doi.org/10.1074/mcp.M113.035600>.
74. Rana N, Privitera G, Kondolf HC, Bulek K, Lechuga S, De Salvo C, Corridoni D, Antanaviciute A, Maywald RL, Hurtado AM, et al. GSDMB is increased in IBD and regulates epithelial restitution/repair independent of pyroptosis. *Cell*. 2022;185(2):283–298.e17. <https://doi.org/10.1016/j.cell.2021.12.024>.
75. Biener-Ramanujan E, Rosier F, Coetzee SG, McGovern DDP, Hazelett D, Targan SR, Gonsky R. Diagnostic and therapeutic potential of RNASET2 in Crohn's disease: Disease-risk polymorphism modulates allelic-imbalance in expression and circulating protein levels and recombinant-RNASET2 attenuates pro-inflammatory cytokine secretion. *Front Immunol*. 2022;13: 999155. <https://doi.org/10.3389/fimmu.2022.999155>.
76. Gonsky R, Fleshner P, Deem RL, Biener-Ramanujan E, Li D, Potdar AA, Bilsborough J, Yang S, McGovern DPB, Targan SR. Association of ribonuclease T2 gene polymorphisms With decreased expression and clinical characteristics of severity in Crohn's disease. *Gastroenterology*. 2017;153(1):219–32. <https://doi.org/10.1053/j.gastro.2017.04.002>.
77. Baranzini N, Pedrini E, Girardello R, Tettamanti G, de Eguileor M, Taramelli R, Acquati F, Grimaldi A. Human recombinant RNASET2-induced inflammatory response and connective tissue remodeling in the medicinal leech. *Cell Tissue Res*. 2017;368(2):337–51. <https://doi.org/10.1007/s00441-016-2557-9>.
78. De Vito A, Orecchia P, Balza E, Reverberi D, Scaldaferrri D, Taramelli R, Noonan DM, Acquati F, Mortara L. Overexpression of murine rnaset2 in a colon syngeneic mouse carcinoma model leads to rebalance of intratumor M1/M2 macrophage ratio, activation of t cells, delayed tumor growth, and rejection. *Cancers*. 2020. <https://doi.org/10.3390/cancers12030717>.
79. Tamai I, Ohashi R, Nezu J, Yabuuchi H, Oku A, Shimane M, Sai Y, Tsuji A. Molecular and functional identification of sodium ion-dependent, high affinity human carnitine transporter OCTN2. *J Biol Chem*. 1998;273(32):20378–82. <https://doi.org/10.1074/jbc.273.32.20378>.
80. Li P, Wang Y, Luo J, Zeng Q, Wang M, Bai M, Zhou H, Wang J, Jiang H. Downregulation of OCTN2 by cytokines plays an important role in the progression of inflammatory bowel disease. *Biochem Pharmacol*. 2020;178: 114115. <https://doi.org/10.1016/j.bcp.2020.114115>.
81. Palmieri O, Latiano A, Scimeca D, Bossa F, Corritore G, Latiano T, Andriulli A, Annese V. IL23R, ATG16L1, IRGM, OCTN1, and OCTN2 mRNA expression in inflamed and noninflamed mucosa of IBD patients. *Inflamm Bowel Dis*. 2011;17(8):1832–3. <https://doi.org/10.1002/ibd.21613>.
82. Mo J, Lim LY, Zhang ZR. L-Carnitine ester of prednisolone: pharmacokinetic and pharmacodynamic evaluation of a type I prodrug. *Int J Pharm*. 2014;475(1–2):123–9. <https://doi.org/10.1016/j.ijpharm.2014.08.049>.
83. Nakamura T, Nakanishi T, Haruta T, Shirasaka Y, Keogh JP, Tamai I. Transport of ipratropium, an anti-chronic obstructive pulmonary disease drug, is mediated by organic cation/carnitine transporters in human bronchial epithelial cells: implications for carrier-mediated pulmonary absorption. *Mol Pharm*. 2010;7(1):187–95. <https://doi.org/10.1021/mp900206j>.
84. Nakanishi T, Haruta T, Shirasaka Y, Tamai I. Organic cation transporter-mediated renal secretion of ipratropium and tiotropium in rats and humans. *Drug Metab Dispos*. 2011;39(1):117–22. <https://doi.org/10.1124/dmd.110.035402>.
85. Khalil AM, Guttman M, Huarte M, Garber M, Raj A, Rivea Morales D, Thomas K, Presser A, Bernstein BE, van Oudenaarden A, et al. Many human large intergenic noncoding RNAs associate with chromatin-modifying complexes and affect gene expression. *Proc Natl Acad Sci USA*. 2009;106(28):11667–72. <https://doi.org/10.1073/pnas.0904715106>.
86. Ajayi TA, Innes CL, Grimm SA, Rai P, Finethy R, Coers J, Wang X, Bell DA, McGrath JA, Schurman SH, et al. Crohn's disease IRGM risk alleles are associated with altered gene expression in human tissues. *Am J Physiol*

- Gastrointest Liver Physiol. 2019;316(1):G95–g105. <https://doi.org/10.1152/ajpgi.00196.2018>.
87. Gloss B, Moran-Jones K, Lin V, Gonzalez M, Scurry J, Hacker NF, Sutherland RL, Clark SJ, Samimi G. ZNF300P1 encodes a lincRNA that regulates cell polarity and is epigenetically silenced in type II epithelial ovarian cancer. *Mol Cancer*. 2014;13:3. <https://doi.org/10.1186/1476-4598-13-3>.
  88. Hrabovsky V, Zadak Z, Blaha V, Hyspler R, Ticha A, Karlik T. Lipid metabolism in active Crohn's disease: pre-results. *Biomed Pap Med Fac Univ Palacky Olomouc Czech Repub*. 2006;150(2):363–6. <https://doi.org/10.5507/bp.2006.056>.
  89. Sewell GW, Hannun YA, Han X, Koster G, Bielawski J, Goss V, Smith PJ, Rahman FZ, Vega R, Bloom SL, et al. Lipidomic profiling in Crohn's disease: abnormalities in phosphatidylinositols, with preservation of ceramide, phosphatidylcholine and phosphatidylserine composition. *Int J Biochem Cell Biol*. 2012;44(11):1839–46. <https://doi.org/10.1016/j.biocel.2012.06.016>.
  90. Yasueda A, Shinzaki S, Iijima H, Mizushima T, Nishimura J, Hiyama S, Ohno S, Ito T. Safety of emulsifying lipid formulation containing omega-3 polyunsaturated fatty acids for patients with Crohn's Disease. *Anticancer Res*. 2016;36(7):3753–9.
  91. Swan K, Allen PJ. Omega-3 fatty acid for the treatment and remission of Crohn's disease. *J Complement Integr Med*. 2013. <https://doi.org/10.1515/jcim-2012-0010>.
  92. Soh H, Im JP, Han K, Park S, Hong SW, Moon JM, Kang EA, Chun J, Lee HJ, Kim JS. Crohn's disease and ulcerative colitis are associated with different lipid profile disorders: a nationwide population-based study. *Aliment Pharmacol Ther*. 2020;51(4):446–56. <https://doi.org/10.1111/apt.15562>.
  93. Fan F, Mundra PA, Fang L, Galvin A, Moore XL, Weir JM, Wong G, White DA, Chin-Dusting J, Sparrow MP, et al. Lipidomic profiling in inflammatory bowel disease: comparison between ulcerative colitis and Crohn's Disease. *Inflamm Bowel Dis*. 2015;21(7):1511–8. <https://doi.org/10.1097/mib.0000000000000394>.
  94. Gao M, Zhong A, Patel N, Alur C, Vyas D. High throughput RNA sequencing utility for diagnosis and prognosis in colon diseases. *World J Gastroenterol*. 2017;23(16):2819–25. <https://doi.org/10.3748/wjg.v23.i16.2819>.
  95. Dulal S, Keku TO. Gut microbiome and colorectal adenomas. *Cancer J*. 2014;20(3):225–31. <https://doi.org/10.1097/ppo.0000000000000050>.
  96. Ni J, Wu GD, Albenberg L, Tomov VT. Gut microbiota and IBD: causation or correlation? *Nat Rev Gastroenterol Hepatol*. 2017;14(10):573–84. <https://doi.org/10.1038/nrgastro.2017.88>.
  97. Pan C, Ning Y, Jia Y, Cheng S, Wen Y, Yang X, Meng P, Li C, Zhang H, Chen Y, et al. Transcriptome-wide association study identified candidate genes associated with gut microbiota. *Gut Pathog*. 2021;13(1):74. <https://doi.org/10.1186/s13099-021-00474-w>.
  98. Zhang Q, Zhao H, Wu D, Cao D, Ma W. A comprehensive analysis of the microbiota composition and gene expression in colorectal cancer. *BMC Microbiol*. 2020;20(1):308. <https://doi.org/10.1186/s12866-020-01938-w>.
  99. Naser SA, Arce M, Khaja A, Fernandez M, Naser N, Elwasila S, Thanigachalam S. Role of ATG16L, NOD2 and IL23R in Crohn's disease pathogenesis. *World J Gastroenterol*. 2012;18(5):412–24. <https://doi.org/10.3748/wjg.v18.i5.412>.
  100. Libioulle C, Louis E, Hansoul S, Sandor C, Farnir F, Franchimont D, Vermeire S, Dewit O, de Vos M, Dixon A, et al. Novel Crohn disease locus identified by genome-wide association maps to a gene desert on 5p131 and modulates expression of PTGER4. *PLoS Genet*. 2007;3(4):58. <https://doi.org/10.1371/journal.pgen.0030058>.
  101. Sewell GW, Kaser A. Interleukin-23 in the pathogenesis of inflammatory bowel disease and implications for therapeutic intervention. *J Crohns Colitis*. 2022. <https://doi.org/10.1093/ecco-jcc/jjac034>.
  102. Groh V, Bahram S, Bauer S, Herman A, Beauchamp M, Spies T. Cell stress-regulated human major histocompatibility complex class I gene expressed in gastrointestinal epithelium. *Proc Natl Acad Sci USA*. 1996;93(22):12445–50. <https://doi.org/10.1073/pnas.93.22.12445>.
  103. Ferrari de Andrade L, Kumar S, Luoma AM, Ito Y, Alves da Silva PH, Pan D, Pyrdol JW, Yoon CH, Wucherpfennig KW. Inhibition of MICA and MICB shedding elicits NK-cell-mediated immunity against tumors resistant to cytotoxic T cells. *Cancer Immunol Res*. 2020;8(6):769–80. <https://doi.org/10.1158/2326-6066.Cir-19-0483>.
  104. Ferrari de Andrade L, Tay RE, Pan D, Luoma AM, Ito Y, Badrinath S, Tsoucas D, Franz B, May KF Jr, Harvey CJ, et al. Antibody-mediated inhibition of MICA and MICB shedding promotes NK cell-driven tumor immunity. *Science*. 2018;359(6383):1537–42. <https://doi.org/10.1126/science.aao0505>.
  105. Feng Q, Yu S, Mao Y, Ji M, Wei Y, He G, Chang W, Zhu D, Ren L, Xu J. High MICB expression as a biomarker for good prognosis of colorectal cancer. *J Cancer Res Clin Oncol*. 2020;146(6):1405–13. <https://doi.org/10.1007/s00432-020-03159-0>.
  106. Alhaidari AI, Albakri AS, Alhumaidi SS. A novel PGAP3 gene mutation-related megalocornea can be misdiagnosed as primary congenital glaucoma. *Cureus*. 2022;14(9):e29387. <https://doi.org/10.7759/cureus.29387>.
  107. Da's SI, Aamer W, Hasan W, Al-Maraghi A, Al-Kurbi A, Kilani H, AlRayahi J, Zamel K, Stotland MA, Fakhro KA. PGAP3 Associated with Hyperphosphatasia with mental retardation plays a novel role in brain morphogenesis and neuronal wiring at early development. *Cells*. 2020. <https://doi.org/10.3390/cells9081782>.
  108. Abdel-Hamid MS, Issa MY, Otaify GA, Abdel-Ghafar SF, Elbendary HM, Zaki MS. PGAP3-related hyperphosphatasia with mental retardation syndrome: Report of 10 new patients and a homozygous founder mutation. *Clin Genet*. 2018;93(1):84–91. <https://doi.org/10.1111/cge.13033>.
  109. Shen Q, Huang Z, Yao J, Jin Y. Extracellular vesicles-mediated interaction within intestinal microenvironment in inflammatory bowel disease. *J Adv Res*. 2022;37:221–33. <https://doi.org/10.1016/j.jare.2021.07.002>.
  110. Díaz-Garrido N, Badia J, Baldomà L. Microbiota-derived extracellular vesicles in interkingdom communication in the gut. *J Extracell Vesicles*. 2021;10(13):e12161. <https://doi.org/10.1002/jev2.12161>.
  111. Zhao F, Zheng T, Gong W, Wu J, Xie H, Li W, Zhang R, Liu P, Liu J, Wu X, et al. Extracellular vesicles package dsDNA to aggravate Crohn's disease by activating the STING pathway. *Cell Death Dis*. 2021;12(9):815. <https://doi.org/10.1038/s41419-021-04101-z>.
  112. Liu R, Tang A, Wang X, Chen X, Zhao L, Xiao Z, Shen S. Inhibition of lncRNA NEAT1 suppresses the inflammatory response in IBD by modulating the intestinal epithelial barrier and by exosome-mediated polarization of macrophages. *Int J Mol Med*. 2018;42(5):2903–13. <https://doi.org/10.3892/ijmm.2018.3829>.
  113. Polytarchou C, Hommes DW, Palumbo T, HatziaPOSTOLOU M, Koutsoumpa M, Koukos G, van der Meulen-de Jong AE, Oikonomopoulos A, van Deen WK, Vorvis C, et al. MicroRNA214 Is associated with progression of ulcerative colitis, and inhibition reduces development of colitis and colitis-associated cancer in mice. *Gastroenterology*. 2015;149(4):981–92. <https://doi.org/10.1053/j.gastro.2015.05.057>.
  114. Valter M, Verstockt S, Finalet Ferreira JA, Cleyneen I. Extracellular vesicles in inflammatory bowel disease: small particles big players. *J Crohns Colitis*. 2021;15(3):499–510. <https://doi.org/10.1093/ecco-jcc/jjaa179>.
  115. Chen P, Huang S, Yu Q, Chao K, Wang Y, Zhou G, Zhuang X, Zeng Z, Chen M, Zhang S. Serum exosomal microRNA-144–3p: a promising biomarker for monitoring Crohn's disease. *Gastroenterol Rep*. 2022. <https://doi.org/10.1093/gastro/goab056>.
  116. Nazari H, Alborzi F, Heirani-Tabasi A, Hadizadeh A, Asbagh RA, Behboudi B, Fazeli MS, Rahimi M, Keramati MR, Keshvari A, et al. Evaluating the safety and efficacy of mesenchymal stem cell-derived exosomes for treatment of refractory perianal fistula in IBD patients: clinical trial phase I. *Gastroenterol Rep*. 2022. <https://doi.org/10.1093/gastro/goac075>.
  117. Baumgart DC, Sandborn WJ. Inflammatory bowel disease: clinical aspects and established and evolving therapies. *Lancet*. 2007;369(9573):1641–57. [https://doi.org/10.1016/s0140-6736\(07\)60751-x](https://doi.org/10.1016/s0140-6736(07)60751-x).
  118. Glassner KL, Abraham BP, Quigley EMM. The microbiome and inflammatory bowel disease. *J Allergy Clin Immunol*. 2020;145(1):16–27. <https://doi.org/10.1016/j.jaci.2019.11.003>.
  119. Rutgeerts P, Goboos K, Peeters M, Hiele M, Penninckx F, Aerts R, Kerremans R, Vantrappen G. Effect of faecal stream diversion on recurrence of Crohn's disease in the neoterminal ileum. *Lancet*. 1991;338(8770):771–4. [https://doi.org/10.1016/0140-6736\(91\)90663-a](https://doi.org/10.1016/0140-6736(91)90663-a).
  120. Janowitz HD, Croen EC, Sachar DB. The role of the fecal stream in Crohn's disease: an historical and analytic review. *Inflamm Bowel Dis*. 1998;4(1):29–39. <https://doi.org/10.1097/00054725-199802000-00006>.

121. Vaishnav S, Yamamoto M, Severson KM, Ruhn KA, Yu X, Koren O, Ley R, Wakeland EK, Hooper LV. The antibacterial lectin RegIII $\gamma$  promotes the spatial segregation of microbiota and host in the intestine. *Science*. 2011;334(6053):255–8. <https://doi.org/10.1126/science.1209791>.
122. Rooks MG, Garrett WS. Gut microbiota, metabolites and host immunity. *Nat Rev Immunol*. 2016;16(6):341–52. <https://doi.org/10.1038/nri.2016.42>.
123. Grundberg E, Small KS, Hedman ÅK, Nica AC, Buil A, Keildson S, Bell JT, Yang TP, Meduri E, Barrett A, et al. Mapping cis- and trans-regulatory effects across multiple tissues in twins. *Nat Genet*. 2012;44(10):1084–9. <https://doi.org/10.1038/ng.2394>.
124. Fritz T, Niederreiter L, Adolph T, Blumberg RS, Kaser A. Crohn's disease: NOD2, autophagy and ER stress converge. *Gut*. 2011;60(11):1580–8. <https://doi.org/10.1136/gut.2009.206466>.
125. Shawkil A, McCole DF. Mechanisms of intestinal epithelial barrier dysfunction by adherent-invasive *Escherichia coli*. *Cell Mol Gastroenterol Hepatol*. 2017;3(1):41–50. <https://doi.org/10.1016/j.jcmgh.2016.10.004>.
126. Kennedy NA, Lamb CA, Berry SH, Walker AW, Mansfield J, Parkes M, Simpkins R, Tremelling M, Nutland S, Parkhill J, et al. The impact of NOD2 variants on fecal microbiota in Crohn's disease and controls without gastrointestinal disease. *Inflamm Bowel Dis*. 2018;24(3):583–92. <https://doi.org/10.1093/ibd/izx061>.
127. Gutierrez O, Pipaon C, Inohara N, Fontalba A, Ogura Y, Prosper F, Nunez G, Fernandez-Luna JL. Induction of Nod2 in myelomonocytic and intestinal epithelial cells via nuclear factor-kappa B activation. *J Biol Chem*. 2002;277(44):41701–5. <https://doi.org/10.1074/jbc.M206473200>.
128. Shaw MH, Reimer T, Sanchez-Valdepenas C, Warner N, Kim YG, Fresno M, Nunez G. T cell-intrinsic role of Nod2 in promoting type 1 immunity to *Toxoplasma gondii*. *Nat Immunol*. 2009;10(12):1267–74. <https://doi.org/10.1038/ni.1816>.
129. Suresh M, Korolowicz KE, Balarezo M, Iyer RP, Padmanabhan S, Cleary D, Gimi R, Sheri A, Yon C, Kallakury BV, et al. Antiviral efficacy and host immune response induction during sequential treatment with SB 9200 followed by Entecavir in Woodchucks. *PLoS ONE*. 2017;12(1):e0169631. <https://doi.org/10.1371/journal.pone.0169631>.
130. Yamamoto-Furusho JK, Fonseca-Camarillo G, Furuzawa-Carballeda J, Sarmiento-Aguilar A, Barreto-Zuñiga R, Martínez-Benitez B, Lara-Velazquez MA. Caspase recruitment domain (CARD) family (CARD9, CARD10, CARD11, CARD14 and CARD15) are increased during active inflammation in patients with inflammatory bowel disease. *J Inflamm (Lond)*. 2018;15:13. <https://doi.org/10.1186/s12950-018-0189-4>.
131. Wu W, Hsu YM, Bi L, Songyang Z, Lin X. CARD9 facilitates microbe-elicited production of reactive oxygen species by regulating the LyGDI-Rac1 complex. *Nat Immunol*. 2009;10(11):1208–14. <https://doi.org/10.1038/ni.1788>.
132. Hsu YM, Zhang Y, You Y, Wang D, Li H, Duramad O, Qin XF, Dong C, Lin X. The adaptor protein CARD9 is required for innate immune responses to intracellular pathogens. *Nat Immunol*. 2007;8(2):198–205. <https://doi.org/10.1038/ni1426>.
133. Iliev ID, Funari VA, Taylor KD, Nguyen Q, Reyes CN, Strom SP, Brown J, Becker CA, Fleshner PR, Dubinsky M, et al. Interactions between commensal fungi and the C-type lectin receptor Dectin-1 influence colitis. *Science*. 2012;336(6086):1314–7. <https://doi.org/10.1126/science.1221789>.
134. Hartjes L, Ruland J. CARD9 Signaling in Intestinal Immune Homeostasis and Oncogenesis. *Front Immunol*. 2019;10:419. <https://doi.org/10.3389/fimmu.2019.00419>.
135. Jin S, Chin J, Seeber S, Niewoehner J, Weiser B, Beaucamp N, Woods J, Murphy C, Fanning A, Shanahan F, et al. TLR1A/TNFSF15 directly induces proinflammatory cytokines, including TNF $\alpha$ , from CD3+CD161+ T cells to exacerbate gut inflammation. *Mucosal Immunol*. 2013;6(5):886–99. <https://doi.org/10.1038/mi.2012.124>.
136. Lin WW, Hsieh SL. Decoy receptor 3: a pleiotropic immunomodulator and biomarker for inflammatory diseases, autoimmune diseases and cancer. *Biochem Pharmacol*. 2011;81(7):838–47. <https://doi.org/10.1016/j.bcp.2011.01.011>.
137. He L, Chen J, Sun J, Peng J, He Q. Protective association of TNFSF15 polymorphisms with Crohn's disease and ulcerative colitis: a meta-analysis. *Saudi J Gastroenterol*. 2018;24(4):201–10. [https://doi.org/10.4103/sjg.SJG\\_5\\_18](https://doi.org/10.4103/sjg.SJG_5_18).
138. Kakuta Y, Kinouchi Y, Negoro K, Takahashi S, Shimosegawa T. Association study of TNFSF15 polymorphisms in Japanese patients with inflammatory bowel disease. *Gut*. 2006;55(10):1527–8. <https://doi.org/10.1136/gut.2006.100297>.
139. Zhang M, Wang X, Jiang X, Yang X, Wen C, Zhi M, Gao X, Hu P, Liu H. Polymorphisms of the TNF gene and three susceptibility loci are associated with Crohn's disease and Perianal Fistula Crohn's disease: a study among the Han population from South China. *Med Sci Monit*. 2019;25:9637–50. <https://doi.org/10.12659/msm.917244>.
140. Yang DH, Yang SK, Song K, Hong M, Park SH, Lee HS, Kim JB, Lee HJ, Park SK, Jung KW, et al. TNFSF15 is an independent predictor for the development of Crohn's disease-related complications in Koreans. *J Crohns Colitis*. 2014;8(10):1315–26. <https://doi.org/10.1016/j.crohns.2014.04.002>.
141. Ng SC, Tsoi KK, Kamm MA, Xia B, Wu J, Chan FK, Sung JJ. Genetics of inflammatory bowel disease in Asia: systematic review and meta-analysis. *Inflamm Bowel Dis*. 2012;18(6):1164–76. <https://doi.org/10.1002/ibd.21845>.
142. Ng SC, Shi HY, Hamidi N, Underwood FE, Tang W, Benchimol EI, Panaccione R, Ghosh S, Wu JCY, Chan FKL, et al. Worldwide incidence and prevalence of inflammatory bowel disease in the 21st century: a systematic review of population-based studies. *Lancet*. 2017;390(10114):2769–78. [https://doi.org/10.1016/s0140-6736\(17\)32448-0](https://doi.org/10.1016/s0140-6736(17)32448-0).
143. Chen GB, Lee SH, Brion MJ, Montgomery GW, Wray NR, Radford-Smith GL, Visscher PM. Estimation and partitioning of (co)heritability of inflammatory bowel disease from GWAS and immunochip data. *Hum Mol Genet*. 2014;23(17):4710–20. <https://doi.org/10.1093/hmg/ddu174>.
144. Huang C, Haritunians T, Okou DT, Cutler DJ, Zwick ME, Taylor KD, Datta LW, Maranville JC, Liu Z, Ellis S, et al. Characterization of genetic loci that affect susceptibility to inflammatory bowel diseases in African Americans. *Gastroenterology*. 2015;149(6):1575–86. <https://doi.org/10.1053/j.gastro.2015.07.065>.
145. de Leeuw CA, Neale BM, Heskes T, Posthuma D. The statistical properties of gene-set analysis. *Nat Rev Genet*. 2016;17(6):353–64. <https://doi.org/10.1038/nrg.2016.29>.
146. Ioannidis JP. Why most published research findings are false. *PLoS Med*. 2005;2(8):e124. <https://doi.org/10.1371/journal.pmed.0020124>.
147. Verstockt S, De Hertogh G, Van der Goten J, Verstockt B, Vancamelbeke M, Machiels K, Van Lommel L, Schuit F, Van Assche G, Rutgeerts P, et al. Gene and miRNA regulatory networks during different stages of Crohn's disease. *J Crohns Colitis*. 2019;13(7):916–30. <https://doi.org/10.1093/ecco-jcc/jjz007>.
148. Louis E, Collard A, Oger AF, Degroote E, Aboul Nasr FA, Belaiche, *Behavior of Crohn's disease according to the Vienna classification: changing pattern over the course of the disease*. *Gut*. 2001;49(6):777–82. <https://doi.org/10.1136/gut.49.6.777>.
149. Pariente B, Cosnes J, Danese S, Sandborn WJ, Lewin M, Fletcher JG, Chowers Y, D'Haens G, Feagan BG, Hibi T, et al. Development of the Crohn's disease digestive damage score, the Lemann score. *Inflamm Bowel Dis*. 2011;17(6):1415–22. <https://doi.org/10.1002/ibd.21506>.
150. Haberman Y, Tickle TL, Dexheimer PJ, Kim MO, Tang D, Karns R, Baldasano RN, Noe JD, Rosh J, Markowitz J, et al. Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J Clin Invest*. 2014;124(8):3617–33. <https://doi.org/10.1172/JCI75436>.
151. Huang H, Fang M, Jostins L, Umičević Mirkov M, Boucher G, Anderson CA, Andersen V, Cleynen I, Cortes A, Crins F, et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature*. 2017;547(7662):173–8. <https://doi.org/10.1038/nature22969>.
152. Chun S, Casparino A, Patsopoulos NA, Croteau-Chonka DC, Raby BA, De Jager PL, Sunyaev SR, Cotsapas C. Limited statistical evidence for shared genetic effects of eQTLs and autoimmune-disease-associated loci in three major immune-cell types. *Nat Genet*. 2017;49(4):600–5. <https://doi.org/10.1038/ng.3795>.
153. Regev A, Teichmann SA, Lander ES, Amit I, Benoist C, Birney E, Bodenmiller B, Campbell P, Carninci P, Clatworthy M, et al. The human cell atlas. *elife*. 2017. <https://doi.org/10.7554/eLife.27041>.
154. He B, Wang F, Shu J, Cheng Y, Zhou X, Huang T. Developing a non-invasive diagnostic model for pediatric Crohn's disease using RNA-seq analysis. *Front Genet*. 2023;14:1142326. <https://doi.org/10.3389/fgene.2023.1142326>.

155. Hong SN, Joung JG, Bae JS, Lee CS, Koo JS, Park SJ, Im JP, Kim YS, Kim JW, Park WY, et al. RNA-seq reveals transcriptomic differences in inflamed and noninflamed intestinal mucosa of crohn's disease patients compared with normal mucosa of healthy controls. *Inflamm Bowel Dis*. 2017;23(7):1098–108. <https://doi.org/10.1097/mib.0000000000001066>.
156. Luningham JM, Chen J, Tang S, De Jager PL, Bennett DA, Buchman AS, Yang J. Bayesian genome-wide TWAS method to leverage both cis- and trans-eqtl information through summary statistics. *Am J Hum Genet*. 2020;107(4):714–26. <https://doi.org/10.1016/j.ajhg.2020.08.022>.
157. Battle A, Brown CD, Engelhardt BE, Montgomery SB. Genetic effects on gene expression across human tissues. *Nature*. 2017;550(7675):204–13. <https://doi.org/10.1038/nature24277>.
158. Etheridge AS, Gallins PJ, Jima D, Broadaway KA, Ratain MJ, Schuetz E, Schadt E, Schroder A, Molony C, Zhou Y, et al. A New liver expression quantitative trait locus map from 1,183 individuals provides evidence for novel expression quantitative trait loci of drug response, metabolic, and sex-biased phenotypes. *Clin Pharmacol Ther*. 2020;107(6):1383–93. <https://doi.org/10.1002/cpt.1751>.
159. Hu YJ, Sun W, Tzeng JY, Perou CM. Proper use of allele-specific expression improves statistical power for cis-eQTL mapping with RNA-Seq data. *J Am Stat Assoc*. 2015;110(511):962–74. <https://doi.org/10.1080/01621459.2015.1038449>.
160. Watanabe K, Stringer S, Frei O, Umičević Mirkov M, de Leeuw C, Polderman TJC, van der Sluis S, Andreassen OA, Neale BM, Posthuma D. A global overview of pleiotropy and genetic architecture in complex traits. *Nat Genet*. 2019;51(9):1339–48. <https://doi.org/10.1038/s41588-019-0481-0>.
161. Li D, Liu Q, Schnable PS. TWAS results are complementary to and less affected by linkage disequilibrium than GWAS. *Plant Physiol*. 2021;186(4):1800–11. <https://doi.org/10.1093/plphys/kiab161>.
162. Sidiq T, Yoshihama S, Downs I, Kobayashi KS. Nod2: A critical regulator of ileal microbiota and crohn's disease. *Front Immunol*. 2016;7:367. <https://doi.org/10.3389/fimmu.2016.00367>.
163. Lapaquette P, Bringer MA, Darfeuille-Michaud A. Defects in autophagy favour adherent-invasive *Escherichia coli* persistence within macrophages leading to increased pro-inflammatory response. *Cell Microbiol*. 2012;14(6):791–807. <https://doi.org/10.1111/j.1462-5822.2012.01768.x>.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.