frontiers | Research Topics

# Understanding molecular mechanisms to facilitate the development of biomarkers for therapeutic intervention in gastrointestinal diseases and sepsis

**Edited by**
Dipak Kumar Sahoo, Romy Monika Heilmann and Ashish Patel

**Published in**
Frontiers in Genetics

ISSN 1664-8714
ISBN 978-2-8325-6037-2
DOI 10.3389/978-2-8325-6037-2

## About Frontiers

Frontiers is more than just an open access publisher of scholarly articles: it is a pioneering approach to the world of academia, radically improving the way scholarly research is managed. The grand vision of Frontiers is a world where all people have an equal opportunity to seek, share and generate knowledge. Frontiers provides immediate and permanent online open access to all its publications, but this alone is not enough to realize our grand goals.

## Frontiers journal series

The Frontiers journal series is a multi-tier and interdisciplinary set of open-access, online journals, promising a paradigm shift from the current review, selection and dissemination processes in academic publishing. All Frontiers journals are driven by researchers for researchers; therefore, they constitute a service to the scholarly community. At the same time, the *Frontiers journal series* operates on a revolutionary invention, the tiered publishing system, initially addressing specific communities of scholars, and gradually climbing up to broader public understanding, thus serving the interests of the lay society, too.

## Dedication to quality

Each Frontiers article is a landmark of the highest quality, thanks to genuinely collaborative interactions between authors and review editors, who include some of the world's best academicians. Research must be certified by peers before entering a stream of knowledge that may eventually reach the public - and shape society; therefore, Frontiers only applies the most rigorous and unbiased reviews. Frontiers revolutionizes research publishing by freely delivering the most outstanding research, evaluated with no bias from both the academic and social point of view. By applying the most advanced information technologies, Frontiers is catapulting scholarly publishing into a new generation.

## What are Frontiers Research Topics?

Frontiers Research Topics are very popular trademarks of the *Frontiers journals series*: they are collections of at least ten articles, all centered on a particular subject. With their unique mix of varied contributions from Original Research to Review Articles, Frontiers Research Topics unify the most influential researchers, the latest key findings and historical advances in a hot research area.

Find out more on how to host your own Frontiers Research Topic or contribute to one as an author by contacting the Frontiers editorial office: frontiersin.org/about/contact

# Understanding molecular mechanisms to facilitate the development of biomarkers for therapeutic intervention in gastrointestinal diseases and sepsis

**Topic editors**

Dipak Kumar Sahoo — Iowa State University, United States

Romy Monika Heilmann — Leipzig University, Germany

Ashish Patel — Hemchandracharya North Gujarat University, India

**Citation**

Sahoo, D. K., Heilmann, R. M., Patel, A., eds. (2025). *Understanding molecular mechanisms to facilitate the development of biomarkers for therapeutic intervention in gastrointestinal diseases and sepsis*. Lausanne: Frontiers Media SA. doi: 10.3389/978-2-8325-6037-2

# Table of contents

frontiers | Frontiers in Genetics

TYPE Editorial
PUBLISHED 11 March 2025
DOI 10.3389/fgene.2025.1581299



OPEN ACCESS

EDITED AND REVIEWED BY
Jared C. Roach,
Institute for Systems Biology (ISB), United States

*CORRESPONDENCE
Dipak Kumar Sahoo,
dsahoo@iastate.edu,
dipaksahoo11@gmail.com
Romy M. Heilmann,
romy.heilmann@kleintierklinik.uni-leipzig.de
Ashish Patel,
uni.ashish@gmail.com

RECEIVED 21 February 2025
ACCEPTED 25 February 2025
PUBLISHED 11 March 2025

CITATION
Sahoo DK, Heilmann RM and Patel A (2025) Editorial: Understanding molecular mechanisms to facilitate the development of biomarkers for therapeutic intervention in gastrointestinal diseases and sepsis.
*Front. Genet.* 16:1581299.
doi: 10.3389/fgene.2025.1581299



# Editorial: Understanding molecular mechanisms to facilitate the development of biomarkers for therapeutic intervention in gastrointestinal diseases and sepsis

Dipak Kumar Sahoo[1]*, Romy M. Heilmann[2]* and Ashish Patel[3]*

[1]Department of Veterinary Clinical Sciences, College of Veterinary Medicine, Iowa State University, Ames, IA, United States, [2]Department for Small Animals, Veterinary Teaching Hospital, College of Veterinary Medicine, University of Leipzig, Leipzig, Saxony, Germany, [3]Department of Life Sciences, Hemchandracharya North Gujarat University, Patan, Gujarat, India

KEYWORDS

**gastrointestinal disorders, inflammatory bowel diseases, ulcerative colitis, molecular biomarkers, sepsis, colorectal cancer, disease diagnosis**

Editorial on the Research Topic
Understanding molecular mechanisms to facilitate the development of biomarkers for therapeutic intervention in gastrointestinal diseases and sepsis

Gastrointestinal (GI) disorders include a range of pathological conditions with varying severities and outcomes that impact the integrity and function of the GI tract. These conditions include indigestion, the inflammatory bowel diseases (IBDs) ulcerative colitis (UC) and Crohn's disease (CD), and malignant tumors. The resulting dysfunction of the intestinal barrier, leading to impaired permeability, allows for the translocation of luminal contents, including intact microbes and microbial products. This situation can cause severe sepsis and potentially fatal outcomes if timely intervention is not provided. Advancements in omics technology have facilitated the identification and evaluation of molecular biomarkers for disease diagnosis. These encompass genomic (e.g., single nucleotide polymorphisms), transcriptomic (including non-coding RNAs), epigenetic (e.g., DNA methylation), proteomic, metabolomic, and microbiome biomarkers (Dalal et al., 2020; Sahoo et al., 2024a; Sahoo et al., 2022b; Figure 1). These biomarkers hold significant clinical potential for improving diagnosis, prognosis, and treatment strategies in patients with GI disorders and sepsis.

For diagnosing sepsis, in addition to lactate as a widely utilized biomarker, other surrogate markers, such as C-reactive protein (CRP) and procalcitonin, which are produced in response to infection and inflammation, may assist in identifying patients at risk of developing severe sepsis before significant organ dysfunction occurs (Faix, 2013). As oxidative stress (OS) plays a key role in the progression of sepsis and septic shock to multiple organ failure (Sahoo et al., 2024b; Wong et al., 2025), OS markers associated with sepsis, specifically, superoxide dismutase, soluble endoglin, asymmetric dimethylarginine,

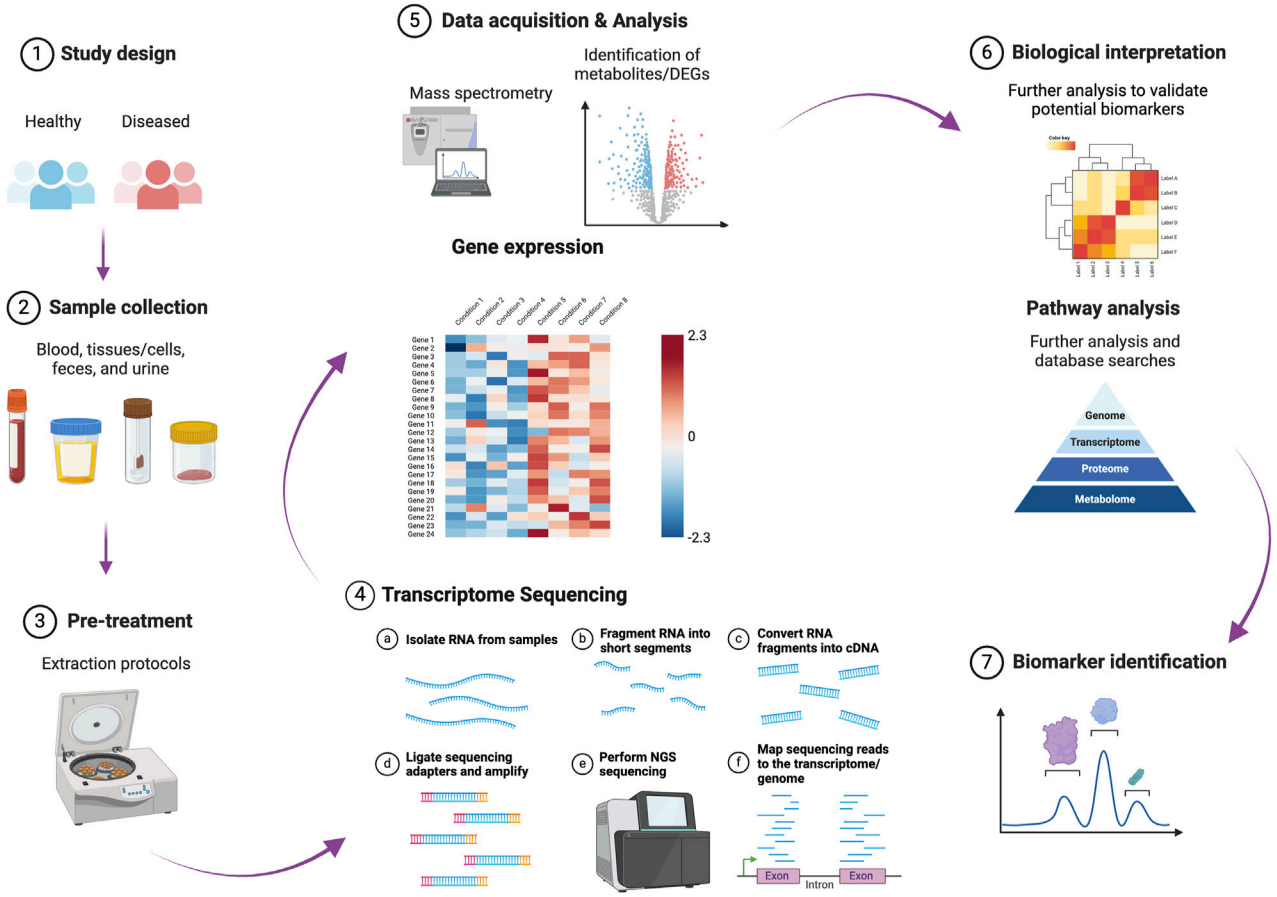



FIGURE 1
A schematic representation of the different phases involved in identifying and validating molecular biomarkers for disease diagnosis and prognosis. This figure was created in BioRender (https://BioRender.com/d22b503). DEGs: differentially expressed genes.

and neopterin, merit further clinical investigation (Helan et al., 2022). Several emerging biomarkers such as microRNA-486-5p, circular RNAs (circRNAs), HOXA distal transcript antisense RNA (a lncRNA located on chromosome 7q15.2), protein C (a vitamin K-dependent glycoprotein), prokineticin 2, and triiodothyronine also hold potential for enhancing early detection and prognostic assessment of sepsis with high sensitivity and specificity (He et al., 2024).

Mendelian randomization (MR) examines the causal effects of modifiable exposures, such as potential risk factors, on health by utilizing genetic variants linked to those specific exposures (Richmond and Smith, 2022). MR analysis by Zhang et al. showed that CMPF (3-carboxy-4-methyl-5-propyl-2-furanpropanoate) has an association with 28-day all-cause mortality in clinical cases of sepsis. The metabolic pathway of alpha-linolenic acid and linoleic acid was identified as a crucial factor in the development and progression of sepsis. The study by Ye et al. demonstrated that levels of soluble suppression of tumorigenicity 2 (sST2) in the blood have significant clinical diagnostic and prognostic implications in sepsis. Moreover, sST2 showed a comparable predictive capability to the SOFA (Sequential Organ Failure Assessment) and APACHE II (Acute Physiology and Chronic Health Evaluation II) scores and had a greater predictive capability than lactic acid levels in assessing the prognosis of patients with sepsis. Jin et al. used single-cell RNA sequencing (scRNA-seq) and identified several diagnostic markers for sepsis, such as PIM1 (proviral integration site for Moloney murine leukemia virus kinase 1), HIST1H1C (Histone Cluster 1 H1 Family Member C), and IGSF6 (Immunoglobulin Superfamily Member 6). The involvement of PIM1 in modulating the immune-inflammatory response during sepsis was verified through experimental validation, indicating that PIM1 is a promising novel therapeutic target.

The endoscopic evaluation of patients suspected of IBD with the collection of mucosal biopsies for histopathological confirmation continues to be the gold standard for establishing an IBD diagnosis, assessing treatment efficacy, and identifying post-operative recurrence; however, it is associated with high costs and invasiveness. Biomarkers enable non-invasive disease evaluation, with C-reactive protein and fecal calprotectin being the most frequently utilized biomarkers in current clinical practice. The T>C substitution SNPs that affect the functionality of the *DLG5* (discs large homolog 5) protein, along with the lack of *CARD15*/*NOD2* (caspase recruitment domain family number 15/nucleotide-binding oligomerization domain-containing protein 2) SNPs associated with CD pathogenesis, could serve as genomic

biomarkers (Dudzińska et al., 2018). Several potential microbiome biomarkers (microbial markers) have been reported. For example, *Faecalibacterium prausnitzii* and its phylogroups and elevated *Escherichia coli* counts serve as potential biomarkers for CD diagnosis. *Akkermansia muciniphila* is identified for pediatric CD diagnosis. Additionally, reductions in *Firmicutes* (*Clostridiales*) levels correlate with IBD severity, while increased abundance of *Lactobacilllaceae* and *Enterococcaceae* families, as well as the genera *Lactobacillus*, *Enterococcus*, and *Eggerthella*, are noted in UC patients (Wang et al., 2024). In dogs with CIE, there is a notable dysbiotic profile in both luminal and mucosal intestinal bacteria, marked by a reduction in *Clostridium* and *Bacteroides* and an increase in *Enterobacteriaceae* (Sahoo et al., 2022a).

The regulation of gene expression mediated by microRNA (miRNA) is essential for the appropriate development and functioning of the intestine (Shanahan et al., 2021). Numerous studies have effectively identified unique miRNA profiles that indicate the upregulation or downregulation of one or more miRNAs in intestinal biopsy samples from patients with IBD (James, 2020) and dogs affected with chronic inflammatory enteropathy (CIE) (Sahoo et al., 2024a). The colonic mucosa of patients with active UC was shown to overexpress specific miRNAs, including miR-21, miR-150, and miR-155, and have a reduction in miRNAs like miR-143 and miR-145 when compared to healthy controls (Alghoul et al., 2022). Comparing the colonic mucosa of patients with active CD and healthy controls, there was an upregulation of miR-196 and a downregulation of miR-7 (Alghoul et al., 2022). The differential expression of miRNAs in saliva, blood, and colon tissue samples was analyzed in UC and CD patients (Schaefer et al., 2015). This research highlighted multiple miRNAs (specifically, miR-21, miR-31, miR-142-3p, miR-142-5p) with expression levels exhibiting significant changes across all three sample types when comparing IBD patients to non-IBD controls (Schaefer et al., 2015). Recent studies indicate that DNA methylation of specific genes contributes to the pathogenesis of IBD, implying their potential utility as clinical biomarkers (Cooke et al., 2012; Nimmo et al., 2012). A comprehensive analysis of methylation patterns across the genome, performed on rectal mucosal biopsies, revealed specific differential gene signatures, including Fanconi anemia complementation group (FANCC), thyroid hormone receptor-associated protein 2 (THRAP2), globoside alpha-1,3-N-acetylgalactosaminyltransferase 1 (GBGT1), tumor necrosis factor ligand superfamily member 4 (TNFSF4), TNF superfamily member 12 (TNFSF12), docking protein 2 (DOK2), and fucosyltransferase 7 (FUT7). These genes exhibited notable differences in methylation levels in specimens from patients with CD or UC compared to healthy individuals (Cooke et al., 2012). Response to infliximab treatment in patients with IBD resulted in notable decreases in macrophage-derived cluster of differentiation 14 (CD14) and CD86 levels, along with the chemokine CCL2 (Magnusson et al., 2015), highlighting their potential as surrogate biomarkers to monitor IBD patients during treatment.

The research by Song et al. employing MR suggests that interleukin-13 (IL-13) contributes to the pathophysiology of IBD (CD and UC). Whereas macrophage migration inhibitory factor appears to be specifically related to CD, stem cell factor is more likely to play a role in the progression of IBD (CD and UC). Another MR study by Qian et al. showed that patients with gastric cancer exhibit decreased blood levels of tryptophan, nonadecanoate (19:0), and erythritol. Zhu et al. employed MR and demonstrated that the genes *GPBAR1* (G protein-coupled bile acid receptor 1), *IL1RL1* (Interleukin 1 receptor-like 1), *PRKCB* (Protein Kinase C Beta), and *PNMT* (Phenylethanolamine N-Methyltransferase) are linked to an increased risk of IBD. Whereas *IL1RL1* was shown to have a protective effect against the risk of CD, *GPX1* (Glutathione peroxidase 1), *GPBAR1*, and *PNMT* are implicated in the risk of UC. In a scRNA-seq study by Keever-Keigher et al., common expression patterns were observed in GI disorders, including an extensive upregulation of MTRNR2L8 (MT-RNR2 Like 8) across various cell types. The increase of XIST (X Inactive Specific Transcript) expression across different cell types in individuals with UC and an elevated expression of Th2 (T helper 2)-associated genes in eosinophilic disorders is also noteworthy.

Colorectal cancer (CRC) is a common GI neoplasia. To facilitate the early detection of CRC, a minimally invasive and reproducible technique known as liquid biopsy (LB) has been established. This method isolates cancer-derived components from the patients' peripheral blood and/or other body fluids, including circulating tumor cells (CTC), miRNA, long non-coding RNAs (lncRNAs), and circulating tumor DNA (ctDNA) (Zhang et al., 2023). Heat-shock protein 27 (Hsp27) has been identified as expressed explicitly in well-differentiated CRC and is linked to other significant CRC biomarkers, such as epidermal growth factor receptor (EGFR), tumor necrosis factors, protein kinase B (AKT), and human epidermal growth factor receptor 2 (ERBB2) (Gan et al., 2014). Glutathione S-transferase pi1 (GSTP1) and KTR8 were overexpressed in both well-differentiated and poorly differentiated CRCs. In contrast, triosephosphate isomerase (TPI), tubulin beta chain (TUBB), and fatty acid-binding protein (FABP1) were upregulated exclusively in well-differentiated CRCs. Human leukocyte antigen A (HLA-A) was observed to be increased in poorly differentiated CRC (Gan et al., 2014). Circular RNAs (circRNAs) such as hsa_circ_001978, hsa_circ_103627, hsa_circ_105039, and circ_0124554 may also serve as indicators for a diagnosis of CRC. Levels of serum miR-21, miR-29a, and miR-125b can potentially differentiate patients with early colorectal neoplasia from healthy ones (Yamada et al., 2015). Serum miRNAs, such as miR-21, miR-92a, miR-182S, and miR-223, along with other miRNAs like miR-17–5p, miR-18a–5p, miR-18b–5p, miR-103a–3p, miR-127–3p, miR-151a–5p, and miR-181a–5p may have clinical utility as biomarkers in the non-invasive screening for CRC (Zhang et al., 2023). CircRNAs have the potential to synergize with various proteins or RNAs, demonstrated by circ_0000523 and methyltransferase-like 3 (METTL3), to enhance the accuracy of a CRC diagnosis (Wang et al., 2022).

Carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9) currently serve as the primary serum tumor markers for assessing the prognosis of CRC. The research conducted by Dai et al. indicates that both the overall survival rate and the disease-free survival rate in patients with CRC progressively decline with an increasing number of positive tumor markers before and after surgical intervention. Nomograms utilizing pre-and postoperative CEA and CA19-9 demonstrate high accuracy in predicting survival and recurrence for stage I-III CRC patients following radical surgery, significantly outperforming the American Joint Committee on Cancer (AJCC) 8th Tumor-Node-Metastasis (TNM) stage. The

research by Rad et al. highlights the significant impact of machine learning algorithms on predicting CRC recurrence, specifically by examining the least number of serial CEA measurements necessary for accurate prediction of recurrence. The research conducted by Pan et al. involved a retrospective analysis of clinical data from 36,708 patients who underwent gastroscopy and colonoscopy between 2005 and 2022. Conventional adenomas (CAs), serrated polyps (SPs), non-adenomatous polyps (NAPs), and CRC were all associated with an increased risk of *Helicobacter pylori* infection and older age. The presence of moderate to severe intestinal metaplasia was associated with an increased risk of NAP and CAs. The risk of CRC was found to be increased with low-grade intraepithelial neoplasia, whereas gastric cancer was linked to high-grade intraepithelial neoplasia. A correlation was also observed between advanced gastric pathology and an increased risk of CRC.

In individual patients without clinical signs, biomarkers or biomarker panels have the potential to serve as a significant resource for screening to identify cancer at an early stage or recognize precancerous conditions. For symptomatic patients, these biomarkers can help differentiate cancerous from benign states. Furthermore, in cancer patients undergoing treatment such as surgical procedures, radiation therapy, and/or chemotherapy, surrogate disease biomarkers are valuable tools for evaluating the success of tumor elimination (complete resection and remission) and the potential for disease recurrence. Identifying and validating optimal biomarkers and biomarker panels is a crucial step, as it offers considerable potential for enhancing personalized medicine and overall clinical outcomes. However, further research into the specificity of molecular biomarkers for sepsis, IBD, and CRC is necessary before they can be utilized as diagnostic tools in clinical practice.

## Author contributions

DS: Conceptualization, Data curation, Formal Analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing–original draft, Writing–review and editing. RH: Conceptualization, Resources, Writing–original draft, Writing–review and editing. AP: Conceptualization, Resources, Writing–original draft, Writing–review and editing.

## Funding

The author(s) declare that no financial support was received for the research and/or publication of this article.

## Acknowledgments

DKS acknowledges the support of the Department of Veterinary Clinical Sciences (VCS) Core Lab, College of Veterinary Medicine, Iowa State University, Ames, Iowa, United States.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers at the time of submission. This had no impact on the peer review process and the final decision.

## Generative AI statement

The author(s) declare that no Generative AI was used in the creation of this manuscript.

## Publisher's note



## References

Alghoul, Z., Yang, C., and Merlin, D. (2022). The current status of molecular biomarkers for inflammatory bowel disease. *Biomedicines* 10, 1492. doi:10.3390/BIOMEDICINES10071492

Cooke, J., Zhang, H., Greger, L., Silva, A. L., Massey, D., Dawson, C., et al. (2012). Mucosal genome-wide methylation changes in inflammatory bowel disease. *Inflamm. Bowel Dis.* 18, 2128–2137. doi:10.1002/IBD.22942

Dalal, N., Jalandra, R., Sharma, M., Prakash, H., Makharia, G. K., Solanki, P. R., et al. (2020). Omics technologies for improved diagnosis and treatment of colorectal cancer: technical advancement and major perspectives. *Biomed. and Pharmacother.* 131, 110648. doi:10.1016/J.BIOPHA.2020.110648

Dudzińska, E., Gryzinska, M., and Kocki, J. (2018). Single nucleotide polymorphisms in selected genes in inflammatory bowel disease. *Biomed. Res. Int.* 2018, 6914346. doi:10.1155/2018/6914346

Faix, J. D. (2013). Biomarkers of sepsis. *Crit. Rev. Clin. Lab. Sci.* 50, 23–36. doi:10.3109/10408363.2013.764490

Gan, Y., Chen, D., and Li, X. (2014). Proteomic analysis reveals novel proteins associated with progression and differentiation of colorectal carcinoma. *J. Cancer Res. Ther.* 10, 89–96. doi:10.4103/0973-1482.131396

He, R. R., Yue, G. L., Dong, M. L., Wang, J. Q., and Cheng, C. (2024). Sepsis biomarkers: advancements and clinical applications—a narrative review. *Int. J. Mol. Sci.* 25, 9010–9025. doi:10.3390/IJMS25169010

Helan, M., Malaska, J., Tomandl, J., Jarkovsky, J., Helanova, K., Benesova, K., et al. (2022). Kinetics of biomarkers of oxidative stress in septic shock: a pilot study. *Antioxidants* 11, 640. doi:10.3390/ANTIOX11040640

Magnusson, M. K., Strid, H., Isaksson, S., Bajor, A., Lasson, A., Ung, K. A., et al. (2015). Response to infliximab therapy in ulcerative colitis is associated with decreased monocyte activation, reduced CCL2 expression and downregulation of Tenascin C. *J. Crohns Colitis* 9, 56–65. doi:10.1093/ECCO-JCC/JJU008

Nimmo, E. R., Prendergast, J. G., Aldhous, M. C., Kennedy, N. A., Henderson, P., Drummond, H. E., et al. (2012). Genome-wide methylation profiling in Crohn's disease identifies altered epigenetic regulation of key host defense mechanisms including the Th17 pathway. *Inflamm. Bowel Dis.* 18, 889–899. doi:10.1002/IBD.21912

Richmond, R. C., and Smith, G. D. (2022). Mendelian randomization: concepts and scope. *Cold Spring Harb. Perspect Med.* 12, a040501. doi:10.1101/CSHPERSPECT.A040501

Sahoo, D. K., Allenspach, K., Mochel, J. P., Parker, V., Rudinsky, A. J., Winston, J. A., et al. (2022a). Synbiotic-IgY therapy modulates the mucosal microbiome and inflammatory indices in dogs with chronic inflammatory enteropathy: a randomized, double-blind, placebo-controlled study. *Veterinary Sci.* 10, 25. doi:10.3390/VETSCI10010025

Sahoo, D. K., Borcherding, D. C., Chandra, L., Jergens, A. E., Atherly, T., Bourgois-Mochel, A., et al. (2022b). Differential transcriptomic profiles following stimulation with lipopolysaccharide in intestinal organoids from dogs with inflammatory bowel disease and intestinal mast cell tumor. *Cancers (Basel)* 14, 3525. doi:10.3390/cancers14143525

Sahoo, D. K., Heilmann, R. M., Ackermann, M., Parker, V., Rudinsky, A., Winston, J. A., et al. (2024a). Tu1739 MicroRNAs as potential biomarkers for diagnosis and monitoring chronic inflammatory enteropathy in dogs. *Gastroenterology* 166, S-1401–1401. doi:10.1016/S0016-5085(24)03676-X

Sahoo, D. K., Wong, D., Patani, A., Paital, B., Yadav, V. K., Patel, A., et al. (2024b). Exploring the role of antioxidants in sepsis-associated oxidative stress: a comprehensive review. *Front. Cell Infect. Microbiol.* 14, 1348713. doi:10.3389/fcimb.2024.1348713

Schaefer, J. S., Attumi, T., Opekun, A. R., Abraham, B., Hou, J., Shelby, H., et al. (2015). MicroRNA signatures differentiate Crohn's disease from ulcerative colitis. *BMC Immunol.* 16, 5. doi:10.1186/S12865-015-0069-0

Shanahan, M. T., Kanke, M., Oyesola, O. O., Hung, Y. H., Koch-Laskowski, K., Singh, A. P., et al. (2021). Multiomic analysis defines the first microRNA atlas across all small intestinal epithelial lineages and reveals novel markers of almost all major cell types. *Am. J. Physiol. Gastrointest. Liver Physiol.* 321, G668–G681. doi:10.1152/ajpgi.00222.2021

Wang, X., Peng, J., Cai, P., Xia, Y., Yi, C., Shang, A., et al. (2024). The emerging role of the gut microbiota and its application in inflammatory bowel disease. *Biomed. and Pharmacother.* 179, 117302. doi:10.1016/J.BIOPHA.2024.117302

Wang, Y., Zhang, B., Zhu, Y., Zhang, Y., Li, L., Shen, T., et al. (2022). *hsa_circ_0000523/miR-let-7b/METTL3* axis regulates proliferation, apoptosis and metastasis in the HCT116 human colorectal cancer cell line. *Oncol. Lett.* 23, 186. doi:10.3892/OL.2022.13306

Wong, D., Sahoo, D. K., Faivre, C., Kopper, J., Dersh, K., Beachler, T., et al. (2025). Oxidative stress in critically ill neonatal foals. *J. Vet. Intern Med.* 39, e17297. doi:10.1111/JVIM.17297

Yamada, A., Horimatsu, T., Okugawa, Y., Nishida, N., Honjo, H., Ida, H., et al. (2015). Serum MIR-21, MIR-29a, and MIR-125b are promising biomarkers for the early detection of colorectal neoplasia. *Clin. Cancer Res.* 21, 4234–4242. doi:10.1158/1078-0432.CCR-14-2793

Zhang, Y., Wang, Y., Zhang, B., Li, P., and Zhao, Y. (2023). Methods and biomarkers for early detection, prediction, and diagnosis of colorectal cancer. *Biomed. and Pharmacother.* 163, 114786. doi:10.1016/J.BIOPHA.2023.114786

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 08 December 2023
DOI 10.3389/fmed.2023.1310391



OPEN ACCESS

EDITED BY
Dipak Kumar Sahoo,
Iowa State University, United States

REVIEWED BY
Andrea Glotta,
Ospedale Regionale di Lugano, Switzerland
Changsong Wang,
Hainan Medical University, China

*CORRESPONDENCE
Xiangrong Zuo
✉ zuoxiangrong@njmu.edu.cn

†These authors have contributed equally to this work and share first authorship

RECEIVED 11 October 2023
ACCEPTED 23 November 2023
PUBLISHED 08 December 2023

CITATION
Zhang Z, Yin Y, Chen T, You J, Zhang W, Zhao Y, Ren Y, Wang H, Chen X and Zuo X (2023) Investigating the impact of human blood metabolites on the Sepsis development and progression: a study utilizing two-sample Mendelian randomization.
*Front. Med.* 10:1310391.
doi: 10.3389/fmed.2023.1310391



# Investigating the impact of human blood metabolites on the Sepsis development and progression: a study utilizing two-sample Mendelian randomization

Zhongqi Zhang[1†], Yu Yin[2†], Tingzhen Chen[1†], Jinjin You[1], Wenhui Zhang[1], Yifan Zhao[1], Yankang Ren[1], Han Wang[1], Xiangding Chen[1] and Xiangrong Zuo[1*]

[1]Department of Critical Care Medicine, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China, [2]Department of Urology, The First Affiliated Hospital of Nanjing Medical University, Nanjing, China

**Background:** Existing data suggests a potential link between human blood metabolites and sepsis, yet the precise cause-and-effect relationship remains elusive. By using a two-sample Mendelian randomization (MR) analysis, this study aims to establish a causal link between human blood metabolites and sepsis.

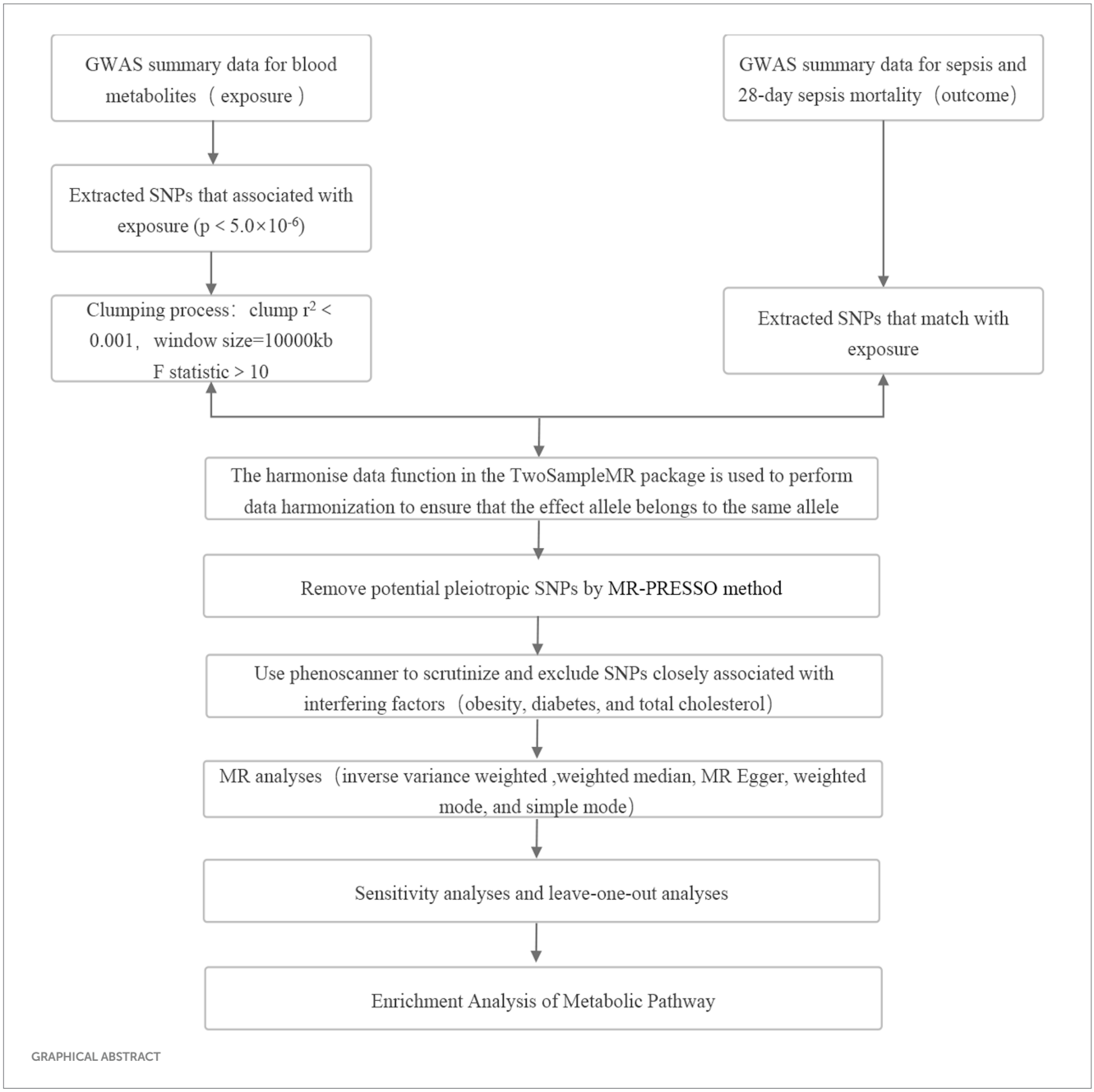
**Methods:** A two-sample MR analysis was employed to investigate the relationship between blood metabolites and sepsis. To assess the causal connection between sepsis and human blood metabolites, five different MR methods were employed, A variety of sensitivity analyses were conducted, including Cochrane's Q test, MR-Egger intercept test, MR-PRESSO and leave-one-out (LOO) analysis. In order to ensure the robustness of the causal association between exposure and outcome, the Bonferroni adjustment was employed. Additionally, we conducted analyses of the metabolic pathways of the identified metabolites using the Kyoto Encyclopedia of Genes and Genomes (KEGG) and the Small Molecule Pathway Database (SMPDB) database.

**Results:** The MR analysis revealed a total of 27 metabolites (16 known and 11 unknown) causally linked to the development and progression of sepsis. After applying the Bonferroni correction, 3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF) remained significant in relation to 28-day all-cause mortality in sepsis. By pathway enrichment analysis, we identified four significant metabolic pathways. Notably, the Alpha Linolenic Acid and Linoleic Acid metabolism pathway emerged as a pivotal contributor to the occurrence and progression of sepsis.

**Conclusion:** This study provides preliminary evidence of causal associations between human blood metabolites and sepsis, as ascertained by MR analysis. The findings offer valuable insights into the pathogenesis of sepsis and may provide insight into preventive and therapeutic approaches.

KEYWORDS

**sepsis, metabolites, two-sample mendelian randomization, causal inference, mortality**




## 1 Introduction

As a result of a dysregulated host response to infection, sepsis can lead to life-threatening organ dysfunction with inflammation and immune dysfunction. This condition exhibits a significant mortality rate (1, 2). Globally, there were approximately 48.9 million cases of sepsis reported in 2017, resulting in an estimated 11 million cases of sepsis-related deaths, which accounts for 19.7% of global mortality (3). Clinicians often encounter challenges in identifying individuals at risk of sepsis due to possible infections. Identifying sepsis as a major concern for global health and patient safety, the World Health Organization (WHO) stresses the importance of recognizing the contributing factors that either elevate or lower the risk of sepsis (4).

In recent years, the discipline of metabolomics has surfaced as a methodical strategy to explore small molecule metabolites in living beings, providing fresh prospects to improve our comprehension of the fundamental processes implicated in the initiation and advancement of illnesses (5). Inflammation can manifest in aseptic forms, such as those resulting from surgical procedures or trauma, or infectious forms, such as sepsis (6). Metabolomics holds promise in offering valuable insights to support clinical decision-making. Importantly, the profiles of metabolites have shown the capacity to successfully differentiate sterile inflammation from sepsis in both human and animal studies (7, 8). Simultaneously, metabolite profiles hold the potential to provide reasonably accurate predictions regarding the occurrence and progression of sepsis. In neonates, metabolite profiles can differentiate between healthy individuals and those with sepsis, as well as reveal distinct patterns between early-onset and late-onset sepsis (9). In addition, following traumatic injury, in adult patients

admitted to the intensive care unit (ICU), metabolite profiles can effectively distinguish between those who develop sepsis and those who do not (10). Metabolite profiles can also differentiate between the prognosis of septic patients, highlighting significant differences between survivors and non-survivors (11–13).

While several metabolites have been observed to be associated with sepsis in population-based cohorts (14), a systematic evaluation of the impact of blood metabolites on sepsis has not yet been conducted. Traditional studies face challenges in identifying and establishing potential causal relationships between blood metabolites and sepsis due to unavoidable confounding factors. MR analysis leverages genetic variants as instrumental variables (IVs) to mitigate confounding and assess the association between exposure and outcome, making it a widely used method for identifying reliable risk factors for various diseases (15). In contrast to observational studies, MR studies effectively minimize confounding variables and provide more robust causal evidence by utilizing natural random allocation (16, 17). As a result, the objective of this study is to utilize the strengths of the MR approach to investigate the correlation between blood metabolites and sepsis in a comprehensive way. The analysis will be based on extensive metabolomics data and clinical information. Statistical methods and genetic variant detection techniques will be employed to identify metabolites associated with sepsis and elucidate their potential contributions to the pathophysiological mechanisms underlying sepsis. Additional analyses will be performed to acquire more profound understanding of the function of these metabolites in sepsis.

# 2 Materials and methods

## 2.1 Data source

Blood metabolite data were obtained from the metabolomics Genome-wide association study (GWAS) server.[1] Summary data from a previously published GWAS study on human blood metabolites were utilized (18), a total of 7,824 European participants participated in this study, including 1768 participants from the KORA F4 study conducted in Germany, and 6,056 participants from the UK Twin Study. This GWAS dataset represents the most comprehensive genetic loci information for 2.1 million Single Nucleotide Polymorphisms (SNPs) associated with 486 blood metabolites. Among these metabolites, 177 metabolites remain unidentified due to their unknown chemical identity, while 309 metabolites have been classified into eight broad metabolic groups: amino acids, carbohydrates, cofactors and vitamins, energy, lipids, nucleotides, peptides, and xenobiotics, relying on data from KEGG (19).

Sepsis data were obtained from the IEU OpenGWAS project,[2] including sepsis and sepsis-related 28-day all-cause mortality. These data derive from a European population consisting of 486,484 participants drawn from the UK Biobank (20). Among them, 11,643 individuals had sepsis, and a total of 12,243,539 SNPs were considered. Within this group, 1,896 individuals succumbed to all causes within 28 days, while survivors were used as controls, involving a total of 12,243,487 SNPs.

## 2.2 Study design

MR was used to investigate the relationship between blood metabolites and sepsis, as well as sepsis-related mortality. The MR analysis conducted in this study adhered to three key assumptions (Figure 1): (1) Strong association between exposure factors and IVs; (2) Absence of confounding factors associated with the IVs; (3) IVs chosen did not have a direct impact on the outcome but influenced it solely through exposure factors (21).

## 2.3 Genetic instrument selection

The selection of IVs for analysis required a robust association with the exposure factor. To ensure precision and efficacy in establishing causal connections between blood metabolites and sepsis risk, SNPs with $p$-values below $5 \times 10^{-6}$, representing locus-wide significance, were chosen. Furthermore, selected instrumental variables needed to pass independence tests successfully, setting the linkage disequilibrium parameter ($R^2$) of SNPs to 0.001 and the genetic distance to 10,000 kb. Additionally, IVs with $F$ values <10 were excluded to ensure the strength of the association between IVs and exposure factors (22). IVs having $p$-values below $1.0 \times 10^{-5}$ concerning the outcome were also disregarded. The Phenoscanner software package was employed to identify covariates associated with potential confounding factors like obesity, diabetes, and total cholesterol to prevent these factors from confounding the impact of exposure on the outcome. Significant associations ($p < 1.0 \times 10^{-5}$) between SNPs and potential confounding factors led to the exclusion of those SNPs from the analysis. Subsequent MR analysis was conducted to validate the results' strength. MR analysis was specifically carried out on metabolites with at least 3 SNPs (23).

## 2.4 Statistical analysis

The inverse variance weighted (IVW) method was employed as the primary analysis method in this study to assess the significant causal relationship between metabolites and sepsis, as well as 28-day all-cause mortality in sepsis ($p < 0.05$). Furthermore, several other MR analysis methods, including weighted median, MR Egger, weighted mode, and simple mode, were used as complementary approaches. Analysis was conducted only on metabolites exhibiting consistent associations across all five methods. To assess heterogeneity and pleiotropy in the IVW method, both Cochran's Q test and MR-Egger intercept analysis were performed. Metabolites exhibiting pleiotropy in the IVW analysis ($p < 0.05$) were removed, and IVs were subjected to MR-PRESSO (version 1.0) to identify and eliminate outlier SNPs, addressing horizontal pleiotropy. LOO analysis was conducted to confirm that MR findings were not influenced by individual SNPs. The study focused on sepsis and 28-day mortality as outcomes, identifying core metabolites associated with the incidence and progression of sepsis. Given the numerous MR analyses during the metabolite screening process, Bonferroni adjustment was applied to rigorously

1 https://metabolomics.helmholtz-muenchen.de/gwas/

2 https://gwas.mrcieu.ac.uk/

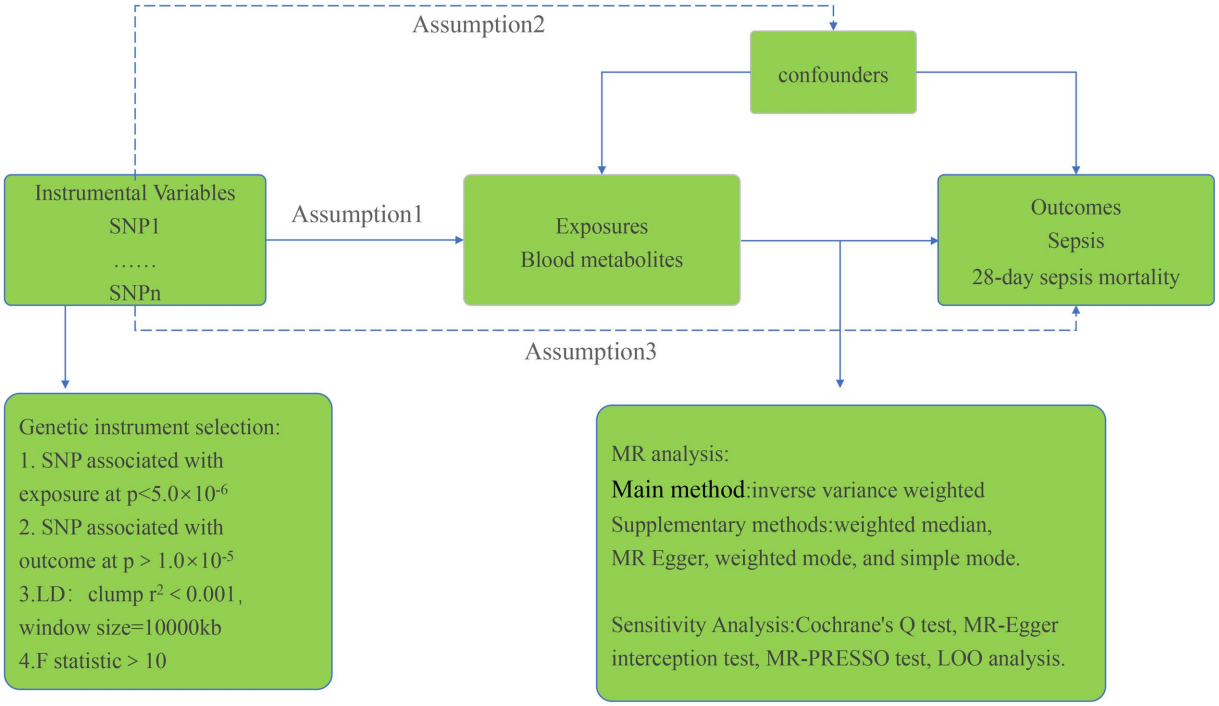



FIGURE 1
Study design overview and assumptions of the MR design. Notes: SNP: single nucleotide polymorphism, LD: linkage disequilibrium, IVW: inverse-variance weighted, MR-PRESSO: Mendelian randomization pleiotropy residual sum and outlier.

evaluate the causal relationships of identified metabolites ($p < 0.05/309 = 0.000167$). Using odds ratios and 95% confidence intervals, we estimated causal effects concerning the association between blood metabolites and sepsis risk and mortality. The statistical analyses were carried out using R (version 4.3.0).

## 2.5 Enrichment analysis of metabolic pathway

To gain further insights into metabolic pathways associated with sepsis occurrence and progression, MetaboAnalyst 5.0[3] was utilized for metabolic pathway analysis. First, we identified the corresponding ID of these metabolites in the MetaboAnalyst 5.0. Then, we used Pathway Analysis modules in the Annotated Features mode to perform the pathway analysis, relying on the KEGG and SMPDB databases. This analysis was performed following all the specified conditions for MR analysis.

# 3 Results

A thorough IV screening was performed for each of the 486 metabolites. Following the IV selection criteria, we included a comprehensive set of 5,538 SNPs related to sepsis in the analysis, ensuring that each metabolite was associated with a minimum of 3 SNPs. After scrutinizing Phenoscanner and excluding SNPs strongly correlated with confounding factors, a total of 17 SNPs were excluded from the analysis. Detailed information about the selected IVs in Tables S1, S2 in Supplementary material. Scatter plots and funnel plots illustrating the MR analysis results are presented in Figures S3–S6 in Supplementary material.

## 3.1 Two samples MR analysis

The IVW analysis was employed to screen metabolites, focusing on their associations with sepsis and 28-day all-cause mortality. Out of the initial pool, 34 blood metabolites were selected based on a significance threshold of $p < 0.05$. A visualization of these metabolites was displayed in a heatmap (Figure 2). Subsequently, these 34 metabolites underwent four other types of MR analysis. We selected metabolites that consistently showed significant associations across all five methods. Ultimately, 27 blood metabolites were identified that exhibited a causal relationship with the risk of sepsis and 28-day all-cause mortality in sepsis.

### 3.1.1 Sepsis

We found 13 causal relationships between blood metabolites and sepsis risk in this study (Figure 3 and Table 1). Among these, pantothenate, paraxanthine, propionylcarnitine, X-13477 were associated with a higher risk, whereas heptanoate, hypoxanthine, X-08988, X-11204, heme, Adrenate (22:4n6), X-12188, stearidonate, and 1-arachidonoylglycerophosphoethanolamine were protective factors. The Cochrane's Q test did not indicate any statistically significant heterogeneity ($p > 0.05$). Furthermore, both the MR-Egger

3 https://www.metaboanalyst.ca/

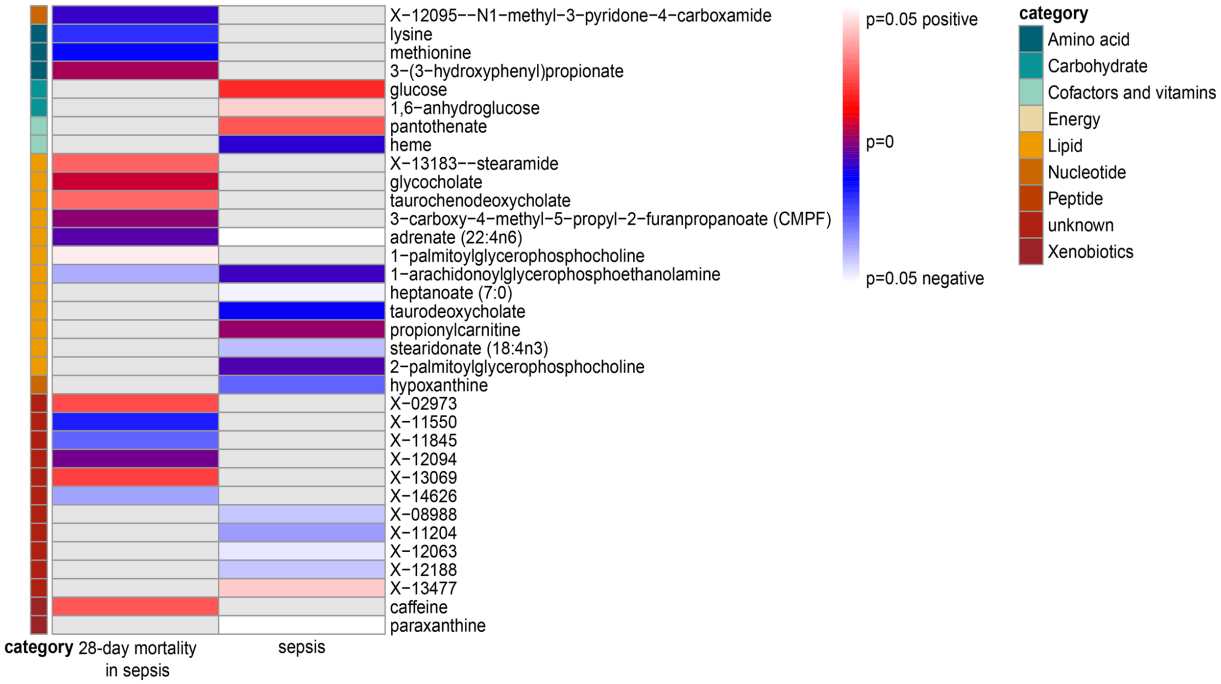



FIGURE 2
Heatmap of the IVW analysis results for sepsis and 28-day mortality in sepsis. Note: Heatmaps showing the *p*-values of the IVW analysis and the direction of the result.

interception test and MR-PRESSO test did not reveal any signs of pleiotropy ($p > 0.05$). Lastly, the LOO analysis (Figure S1 in Supplementary material) demonstrated that after systematically excluding each SNP, pantothenate, propionylcarnitine, and 1-arachidonoylglycerophosphoethanolamine consistently yielded stable results.

### 3.1.2 28-Day mortality in sepsis

We identified 16 causal relationships between blood metabolites and 28-day all-cause mortality in patients with sepsis identified (Figure 3 and Table 2). Among these, X-02973, glycocholate, taurochenodeoxycholate, 1-palmitoylglycerophosphocholine, X-13069, stearamide, 3-(3-hydroxyphenyl) propionate were risk factors, while methionine, 3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF), X-11550, Adrenate (22:4n6), X-11845, X-12094, N1-methyl-3-pyridone-4-carboxamide, 1-arachidonoylglycerophosphoethanolamine, X-14626 were protective factors. Similar to the sepsis analysis, the Cochrane's Q test did not indicate any statistically significant heterogeneity ($p > 0.05$). Moreover, both the MR-Egger interception test and MR-PRESSO test did not detect any signs of pleiotropy ($p > 0.05$). The LOO analysis (Figure S2 in Supplementary material) demonstrated that after systematically excluding each SNP, methionine, glycocholate, and CMPF consistently yielded stable results.

## 3.2 Bonferroni-corrected test and analysis of metabolic pathway enrichment

After applying the Bonferroni correction, we observed that CMPF maintained its significance in relation to 28-day all-cause mortality in sepsis ($p < 0.000167$). By incorporating data from the KEGG and SMPDB databases, the MR analysis identified four significant pathways (Table 3). Notably, the Alpha Linolenic Acid and Linoleic Acid Metabolism pathway played a crucial role in the occurrence and progression of sepsis.

# 4 Discussion

This study integrated two large GWAS datasets and used a rigorous MR design to investigate the causal associations between 486 blood metabolites and sepsis. Our analysis revealed that 13 blood metabolites exhibited a causal relationship with sepsis, while 16 blood metabolites demonstrated a causal relationship with 28-day mortality in sepsis. Specifically, we observed that 1-arachidonoylglycerophosphoethanolamine and Adrenate (22:4n6) were both significantly associated with the occurrence and progression of sepsis. After Bonferroni correction testing, CMPF showed a strong causal relationship with lower 28-day mortality in sepsis. These results have the potential to furnish valuable indications for identifying early diagnostic biomarkers and potential therapeutic targets for sepsis.

1-Arachidonoylglycerophosphoethanolamine is a novel metabolite related to the endogenous cannabinoid system, producing arachidonoylethanolamine (anandamide, AEA) through the cleavage of glycerophosphoethanolamine by phosphodiesterase. AEA, an endogenous cannabinoid, is synthesized by macrophages in response to pathological conditions like shock and is considered a pathogenic mediator in septic shock development (24). Studies on animal models have

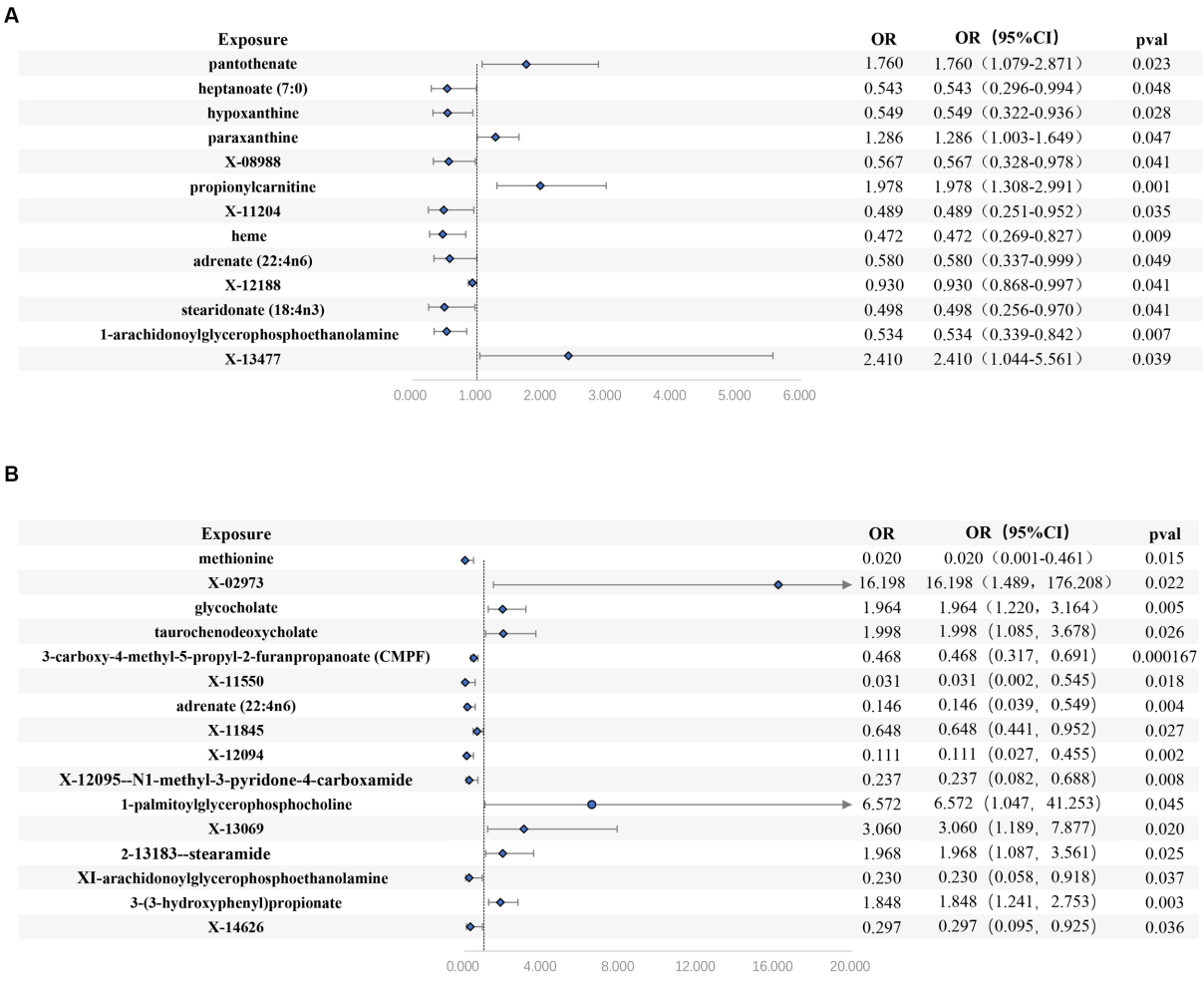



FIGURE 3
Forest plots of estimates identified with inverse variance weighted. **(A)** sepsis, **(B)** 28-day death in sepsis.

confirmed AEA's role in regulating the immune system, exhibiting various effects in sepsis, including anti-inflammatory, antioxidant, pro-apoptotic, and immunomodulatory effects (25–27). Activation of AEA receptors on immune cells reduces pro-inflammatory cytokine secretion and the recruitment of neutrophils and macrophages (28, 29). Recent research findings suggest that low baseline plasma AEA levels may serve as prognostic indicators for septic patients requiring prolonged mechanical ventilation. Furthermore, a lower concentration of AEA has been identified as a prognostic factor for hospital stays exceeding 10 days (30). In another study, AEA was found to attenuate acute respiratory distress syndrome induced by Staphylococcus Enterotoxin B by suppressing inflammation through the down-regulation of key miRNA that regulates immunosuppressive pathways (31). These findings imply that the endocannabinoid AEA may possess a protective effect against severe inflammation and could potentially be utilized in the management of sepsis cases with multiple complications. This study highlights that 1-arachidonoylglycerophosphoethanolamine exhibits a protective effect against sepsis, indicating its potential as a novel and promising therapeutic target for sepsis treatment.

Adrenate (22:4n6) has been identified as a protective factor against lacunar stroke in a previous MR study (32). Dihomo-isofurans (dihomo-IsoPs), which are peroxidation products derived from Adrenate (22:4n6), play a crucial role in the composition of white matter. These compounds hold potential as selective biomarkers for quantifying *in vivo* free radical damage to neuronal membranes Moreover, plasma biomarkers associated with Adrenate (22:4n6) and its derivatives hold promise for early and differential diagnosis of Alzheimer's disease (33). While current research on Adrenate (22:4n6) and its derivatives primarily focuses on neurodegenerative diseases (34), our study has revealed its potential significance in the development and progression of sepsis. However, the underlying mechanisms of this phenomenon remain unknown as our investigation primarily focused on correlation analysis. Therefore, further research is essential to explore and elucidate these mechanistic explanations.

The present study conducted a comprehensive MR analysis, revealing a robust causal association between CMPF and reduced mortality from sepsis within a 28-day period. In individuals who consume fish and fish oil as well as polyunsaturated fatty acids through high-temperature cooking, CMPF, a metabolite produced from furan

TABLE 1 MR results and sensitivity analysis of blood metabolites on sepsis.

| Level | Exposure | Outcome | Method | Nsnps | Se | P | OR (95% CI) | Cochran's Q *p* | MR-Egger intercept *p* | MR-PRESSO global test *p* |
|---|---|---|---|---|---|---|---|---|---|---|
| Cofactors and vitamins | Pantothenate | Sepsis | Inverse variance weighted | 15 | 0.250 | 0.023 | 1.760 (1.079 ~ 2.871) | 0.983 | 0.299 | 0.985 |
| | | | MR Egger | 15 | 0.470 | 0.778 | 1.145 (0.456 ~ 2.875) | | | |
| | | | Simple mode | 15 | 0.556 | 0.364 | 1.685 (0.567 ~ 5.010) | | | |
| | | | Weighted median | 15 | 0.351 | 0.098 | 1.787 (0.899 ~ 3.555) | | | |
| | | | Weighted mode | 15 | 0.497 | 0.260 | 1.794 (0.677 ~ 4.755) | | | |
| Lipid | Heptanoate (7:0) | Sepsis | Inverse variance weighted | 20 | 0.309 | 0.048 | 0.543 (0.296 ~ 0.994) | 0.698 | 0.951 | 0.718 |
| | | | MR Egger | 20 | 0.919 | 0.552 | 0.573 (0.094 ~ 3.472) | | | |
| | | | Simple mode | 20 | 0.708 | 0.488 | 0.606 (0.151 ~ 2.428) | | | |
| | | | Weighted median | 20 | 0.444 | 0.183 | 0.554 (0.232 ~ 1.322) | | | |
| | | | Weighted mode | 20 | 0.596 | 0.345 | 0.561 (0.174 ~ 1.805) | | | |
| Nucleotide | Hypoxanthine | Sepsis | Inverse variance weighted | 13 | 0.272 | 0.028 | 0.549 (0.322 ~ 0.936) | 0.412 | 0.752 | 0.460 |
| | | | MR Egger | 13 | 0.780 | 0.650 | 0.695 (0.151 ~ 3.209) | | | |
| | | | Simple mode | 13 | 0.608 | 0.211 | 0.448 (0.136 ~ 1.474) | | | |
| | | | Weighted median | 13 | 0.380 | 0.068 | 0.500 (0.238 ~ 1.053) | | | |
| | | | Weighted mode | 13 | 0.506 | 0.152 | 0.461 (0.171 ~ 1.242) | | | |
| Xenobiotics | Paraxanthine | Sepsis | Inverse variance weighted | 7 | 0.127 | 0.047 | 1.286 (1.003 ~ 1.649) | 0.603 | 0.973 | 0.567 |

*(Continued)*





10.3389/fmed.2023.1310391

TABLE 1 (Continued)

| Level | Exposure | Outcome | Method | Nsnps | Se | P | OR (95% CI) | Cochran's Q *p* | MR-Egger intercept *p* | MR-PRESSO global test *p* |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | MR Egger | 7 | 0.427 | 0.560 | 1.305 (0.565 ~ 3.011) | | | |
| | | | Simple mode | 7 | 0.296 | 0.680 | 1.137 (0.636 ~ 2.033) | | | |
| | | | Weighted median | 7 | 0.181 | 0.289 | 1.212 (0.850 ~ 1.728) | | | |
| | | | Weighted mode | 7 | 0.206 | 0.875 | 1.035 (0.690 ~ 1.550) | | | |
| unknown | X-08988 | Sepsis | Inverse variance weighted | 21 | 0.279 | 0.041 | 0.567 (0.328 ~ 0.978) | 0.309 | 0.517 | 0.367 |
| | | | MR Egger | 21 | 0.715 | 0.177 | 0.367 (0.090 ~ 1.491) | | | |
| | | | Simple mode | 21 | 0.697 | 0.699 | 0.760 (0.194 ~ 2.982) | | | |
| | | | Weighted median | 21 | 0.421 | 0.203 | 0.585 (0.257 ~ 1.334) | | | |
| | | | Weighted mode | 21 | 0.491 | 0.386 | 0.647 (0.247 ~ 1.694) | | | |
| Lipid | Propionylcarnitine | Sepsis | Inverse variance weighted | 24 | 0.211 | 0.001 | 1.978 (1.308 ~ 2.991) | 0.679 | 0.770 | 0.699 |
| | | | MR Egger | 24 | 0.462 | 0.096 | 2.233 (0.904 ~ 5.521) | | | |
| | | | Simple mode | 24 | 0.479 | 0.134 | 2.102 (0.823 ~ 5.372) | | | |
| | | | Weighted median | 24 | 0.293 | 0.006 | 2.236 (1.258 ~ 3.972) | | | |
| | | | Weighted mode | 24 | 0.374 | 0.055 | 2.134 (1.024 ~ 4.445) | | | |
| unknown | X-11204 | Sepsis | Inverse variance weighted | 24 | 0.340 | 0.035 | 0.489 (0.251 ~ 0.952) | 0.556 | 0.952 | 0.589 |

*(Continued)*





10.3389/fmed.2023.1310391

TABLE 1 (Continued)

| Level | Exposure | Outcome | Method | Nsnps | Se | P | OR (95% CI) | Cochran's Q *p* | MR-Egger intercept *p* | MR-PRESSO global test *p* |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | MR Egger | 24 | 1.275 | 0.620 | 0.527 (0.043 ~ 6.415) | | | |
| | | | Simple mode | 24 | 0.948 | 0.252 | 0.328 (0.051 ~ 2.105) | | | |
| | | | Weighted median | 24 | 0.485 | 0.030 | 0.349 (0.135 ~ 0.903) | | | |
| | | | Weighted mode | 24 | 0.898 | 0.219 | 0.321 (0.055 ~ 1.868) | | | |
| Cofactors and vitamins | heme* | Sepsis | Inverse variance weighted | 5 | 0.286 | 0.009 | 0.472 (0.269 ~ 0.827) | 0.344 | 0.254 | 0.414 |
| | | | MR Egger | 5 | 0.983 | 0.124 | 0.125 (0.018 ~ 0.857) | | | |
| | | | Simple mode | 5 | 0.550 | 0.163 | 0.391 (0.133 ~ 1.149) | | | |
| | | | Weighted median | 5 | 0.381 | 0.035 | 0.447 (0.212 ~ 0.943) | | | |
| | | | Weighted mode | 5 | 0.497 | 0.149 | 0.412 (0.156 ~ 1.092) | | | |
| Lipid | adrenate (22:4n6) | Sepsis | Inverse variance weighted | 5 | 0.277 | 0.049 | 0.580 (0.337 ~ 0.999) | 0.979 | 0.900 | 0.987 |
| | | | MR Egger | 5 | 1.018 | 0.553 | 0.508 (0.069 ~ 3.734) | | | |
| | | | Simple mode | 5 | 0.415 | 0.220 | 0.548 (0.243 ~ 1.234) | | | |
| | | | Weighted median | 5 | 0.338 | 0.077 | 0.55 (0.284 ~ 1.067) | | | |
| | | | Weighted mode | 5 | 0.364 | 0.170 | 0.544 (0.266 ~ 1.111) | | | |
| unknown | X-12188 | Sepsis | Inverse variance weighted | 11 | 0.035 | 0.041 | 0.930 (0.868 ~ 0.997) | 0.496 | 0.821 | 0.536 |

*(Continued)*





10.3389/fmed.2023.1310391

TABLE 1 (Continued)

| Level | Exposure | Outcome | Method | Nsnps | Se | P | OR (95% CI) | Cochran's Q *p* | MR-Egger intercept *p* | MR-PRESSO global test *p* |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | MR Egger | 11 | 0.075 | 0.271 | 0.916 (0.791 ~ 1.060) | | | |
| | | | Simple mode | 11 | 0.074 | 0.827 | 0.984 (0.851 ~ 1.137) | | | |
| | | | Weighted median | 11 | 0.048 | 0.448 | 0.964 (0.877 ~ 1.060) | | | |
| | | | Weighted mode | 11 | 0.045 | 0.376 | 0.959 (0.878 ~ 1.048) | | | |
| Lipid | stearidonate (18:4n3) | Sepsis | Inverse variance weighted | 5 | 0.340 | 0.041 | 0.498 (0.256 ~ 0.970) | 0.214 | 0.830 | 0.333 |
| | | | MR Egger | 5 | 1.361 | 0.793 | 0.676 (0.047 ~ 9.737) | | | |
| | | | Simple mode | 5 | 0.493 | 0.135 | 0.398 (0.151 ~ 1.045) | | | |
| | | | Weighted median | 5 | 0.352 | 0.013 | 0.417 (0.209 ~ 0.832) | | | |
| | | | Weighted mode | 5 | 0.389 | 0.081 | 0.406 (0.189 ~ 0.869) | | | |
| Lipid | 1-arachidonoylglycerophosphoethanolamine* | Sepsis | Inverse variance weighted | 14 | 0.232 | 0.007 | 0.534 (0.339 ~ 0.842) | 0.996 | 0.829 | 1.000 |
| | | | MR Egger | 14 | 0.602 | 0.237 | 0.473 (0.145 ~ 1.538) | | | |
| | | | Simple mode | 14 | 0.467 | 0.152 | 0.491 (0.196 ~ 1.228) | | | |
| | | | Weighted median | 14 | 0.318 | 0.027 | 0.496 (0.266 ~ 0.925) | | | |
| | | | Weighted mode | 14 | 0.366 | 0.074 | 0.491 (0.240 ~ 1.007) | | | |
| unknown | X-13477 | Sepsis | Inverse variance weighted | 4 | 0.427 | 0.039 | 2.410 (1.044 ~ 5.561) | 0.980 | 0.827 | 0.972 |

*(Continued)*





10.3389/fmed.2023.1310391

TABLE 1 (Continued)

| Level | Exposure | Outcome | Method | Nsnps | Se | P | OR (95% CI) | Cochran's Q *p* | MR-Egger intercept *p* | MR-PRESSO global test *p* |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | MR Egger | 4 | 0.993 | 0.382 | 3.013 (0.43 ~ 21.088) | | | |
| | | | Simple mode | 4 | 0.623 | 0.314 | 2.120 (0.625 ~ 7.189) | | | |
| | | | Weighted median | 4 | 0.494 | 0.083 | 2.356 (0.895 ~ 6.202) | | | |
| | | | Weighted mode | 4 | 0.554 | 0.269 | 2.114 (0.714 ~ 6.260) | | | |

CI, confidence interval; NSNPs, number of single nucleotide polymorphisms.

fatty acids, is found in higher concentrations (35). Numerous investigations have established a correlation between CMPF and the development of type 2 diabetes, while recent research has indicated that serum CMPF levels are inversely associated with the risk of type 2 diabetes (36–38). Given that diabetes mellitus is a recognized risk factor for the onset of sepsis, the inverse relationship between CMPF and diabetes mellitus could potentially serve as a significant protective factor. Additionally, a study on periodontitis has shown that elevated CMPF levels are linked to a reduced occurrence of gingival inflammation and a less severe form of periodontitis (39). Moreover, CMPF has exhibited potential anti-inflammatory properties. For instance, an extract derived from green-lipped mussels, known for their furan fatty acid content, has demonstrated promising outcomes in alleviating symptoms associated with rheumatoid arthritis in patients. This therapeutic effect is achieved through the reduction of interleukin (IL) 1β, prostaglandin (PGE2), and tumor necrosis factor α (TNF-α) levels (40). Furthermore, in experiments with mice, CMPF treatment resulted in improved fat removal from the liver and reduced fat storage, effectively preventing lipid buildup in the liver and the onset of hepatic insulin resistance induced by a high-fat diet (41). Sepsis patients, particularly those with obesity, often exhibit insulin resistance (42). Empirical research has confirmed that hyperglycemia in sepsis patients is associated with an unfavorable prognosis. Therefore, it is imperative to maintain sepsis patients' blood glucose levels within a reasonable range. Elevated levels of CMPF have the potential to enhance insulin sensitivity, mitigate lipid accumulation, and counteract insulin resistance induced by a high-fat diet. This, in turn, can effectively regulate blood glucose levels in sepsis patients and improve their prognosis (43).

The analysis of metabolic pathways in sepsis and the 28-day all-cause mortality has revealed a significant elevation in the value of Alpha Linolenic Acid (ALA) and Linoleic Acid (LA) metabolism. ALA and LA are crucial constituents of Omega-3 and Omega-6 polyunsaturated fatty acids. Additionally, they can undergo conversion in the body to form longer-chain Omega-3 and Omega-6 polyunsaturated fatty acids (44). The involvement of these fatty acid-derived metabolites holds significant importance in the development of sepsis. ALA, classified as an Omega-3 polyunsaturated fatty acid, has recently been identified as having the potential to mitigate sepsis-induced intestinal damage through several mechanisms, including the downregulation of miR-1-3p, increased expression of Notch3, and inhibition of the Smad pathway activation (45). Additionally, ALA possesses anti-inflammatory properties and can impede platelet aggregation and thrombus formation (46). LA, on the other hand, is an Omega-6 polyunsaturated fatty acid, and one of its significant metabolic byproducts is arachidonic acid (ARA) (47). ARA can undergo metabolism via the cyclooxygenase (COX) pathway, resulting in the production of prostaglandins (PGs) and thromboxanes (TXs). Additionally, ARA can be converted into leukotrienes (LTs) and lipoxins (LX) through the lipoxygenase (LOX) pathway. Consequently, ARA assumes a pro-inflammatory function within the inflammatory process, thereby facilitating the sequential progression of inflammation (47–49). In patients with sepsis, ARA levels are typically elevated, and this elevation is associated with increased inflammation (50). A recent MR study provided evidence supporting our hypothesis that omega-3 intake is associated with a lower risk of sepsis, while an elevation in the omega-6/omega-3 ratio is associated with a higher risk of sepsis-related mortality (51).

TABLE 2 MR results and sensitivity analysis of blood metabolites on 28-day death in sepsis.

| Level | Exposure | Outcome | Method | Nsnps | Se | *p* | OR (95% CI) | Cochran's Q *p* | MR-Egger intercept *p* | MR-PRESSO global test *p* |
|---|---|---|---|---|---|---|---|---|---|---|
| Amino acid | methionine | Sepsis (28 day death) | Inverse variance weighted | 12 | 1.605 | 0.015 | 0.020 (0.001 ~ 0.461) | 0.921 | 0.517 | 0.919 |
| | | | MR Egger | 12 | 4.945 | 0.878 | 0.459 (0.001 ~ 7434.500) | | | |
| | | | Simple mode | 12 | 3.251 | 0.562 | 0.143 (0.001 ~ 83.902) | | | |
| | | | Weighted median | 12 | 2.244 | 0.317 | 0.106 (0.001 ~ 8.624) | | | |
| | | | Weighted mode | 12 | 2.886 | 0.613 | 0.223 (0.001 ~ 63.648) | | | |
| unknown | X-02973 | Sepsis (28 day death) | Inverse variance weighted | 16 | 1.218 | 0.022 | 16.198 (1.489 ~ 176.208) | 0.792 | 0.719 | 0.797 |
| | | | MR Egger | 16 | 5.641 | 0.408 | 122.456 (0.002 ~ 7760268.183) | | | |
| | | | Simple mode | 16 | 2.726 | 0.194 | 40.819 (0.195 ~ 8538.371) | | | |
| | | | Weighted median | 16 | 1.588 | 0.029 | 31.860 (1.417 ~ 716.415) | | | |
| | | | Weighted mode | 16 | 2.674 | 0.212 | 32.611 (0.173 ~ 6155.931) | | | |
| Lipid | glycocholate | Sepsis (28 day death) | Inverse variance weighted | 6 | 0.243 | 0.005 | 1.964 (1.220 ~ 3.164) | 0.885 | 0.961 | 0.879 |
| | | | MR Egger | 6 | 0.333 | 0.108 | 1.988 (1.035 ~ 3.820) | | | |
| | | | Simple mode | 6 | 0.506 | 0.154 | 2.341 (0.868 ~ 6.315) | | | |
| | | | Weighted median | 6 | 0.336 | 0.048 | 1.942 (1.005 ~ 3.752) | | | |
| | | | Weighted mode | 6 | 0.325 | 0.115 | 1.857 (0.981 ~ 3.515) | | | |
| Lipid | taurochenodeoxycholate | Sepsis (28 day death) | Inverse variance weighted | 6 | 0.311 | 0.026 | 1.998 (1.085 ~ 3.678) | 0.489 | 0.863 | 0.662 |
| | | | MR Egger | 6 | 0.563 | 0.240 | 2.174 (0.722 ~ 6.549) | | | |
| | | | Simple mode | 6 | 0.598 | 0.413 | 1.706 (0.528 ~ 5.504) | | | |
| | | | Weighted median | 6 | 0.414 | 0.138 | 1.847 (0.821 ~ 4.154) | | | |

*(Continued)*





10.3389/fmed.2023.1310391

TABLE 2 (Continued)

| Level | Exposure | Outcome | Method | Nsnps | Se | *p* | OR (95% CI) | Cochran's Q *p* | MR-Egger intercept *p* | MR-PRESSO global test *p* |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Weighted mode | 6 | 0.393 | 0.162 | 1.906 (0.883 ~ 4.116) | | | |
| Lipid | 3-carboxy-4-methyl-5-propyl-2-furanpropanoate (CMPF) | Sepsis (28 day death) | Inverse variance weighted | 10 | 0.199 | 0.000 | 0.468 (0.317 ~ 0.691) | 0.471 | 0.780 | 0.602 |
| | | | MR Egger | 10 | 0.505 | 0.115 | 0.410 (0.152 ~ 1.102) | | | |
| | | | Simple mode | 10 | 0.388 | 0.042 | 0.400 (0.187 ~ 0.855) | | | |
| | | | Weighted median | 10 | 0.265 | 0.000 | 0.397 (0.236 ~ 0.667) | | | |
| | | | Weighted mode | 10 | 0.345 | 0.027 | 0.403 (0.205 ~ 0.793) | | | |
| unknown | X-11550 | Sepsis (28 day death) | Inverse variance weighted | 15 | 1.465 | 0.018 | 0.031 (0.002 ~ 0.545) | 0.262 | 0.770 | 0.304 |
| | | | MR Egger | 15 | 5.560 | 0.378 | 0.006 (0.001 ~ 337.742) | | | |
| | | | Simple mode | 15 | 3.457 | 0.156 | 0.006 (0.001 ~ 4.943) | | | |
| | | | Weighted median | 15 | 1.900 | 0.022 | 0.013 (0.001 ~ 0.540) | | | |
| | | | Weighted mode | 15 | 3.208 | 0.114 | 0.004 (0.001 ~ 2.393) | | | |
| Lipid | adrenate (22:4n6) | Sepsis (28 day death) | Inverse variance weighted | 5 | 0.676 | 0.004 | 0.146 (0.039 ~ 0.549) | 0.633 | 0.306 | 0.729 |
| | | | MR Egger | 5 | 2.482 | 0.145 | 0.008 (0.001 ~ 1.001) | | | |
| | | | Simple mode | 5 | 1.160 | 0.174 | 0.148 (0.015 ~ 1.433) | | | |
| | | | Weighted median | 5 | 0.831 | 0.012 | 0.123 (0.024 ~ 0.629) | | | |
| | | | Weighted mode | 5 | 0.949 | 0.087 | 0.118 (0.018 ~ 0.755) | | | |
| unknown | X-11845 | Sepsis (28 day death) | Inverse variance weighted | 11 | 0.196 | 0.027 | 0.648 (0.441 ~ 0.952) | 0.374 | 0.389 | 0.421 |
| | | | MR Egger | 11 | 0.471 | 0.922 | 0.954 (0.379 ~ 2.402) | | | |
| | | | Simple mode | 11 | 0.454 | 0.111 | 0.452 (0.185 ~ 1.101) | | | |

*(Continued)*





10.3389/fmed.2023.1310391

TABLE 2 (Continued)

| Level | Exposure | Outcome | Method | Nsnps | Se | *p* | OR (95% CI) | Cochran's Q *p* | MR-Egger intercept *p* | MR-PRESSO global test *p* |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Weighted median | 11 | 0.274 | 0.148 | 0.672 (0.393 ~ 1.151) | | | |
| | | | Weighted mode | 11 | 0.387 | 0.266 | 0.633 (0.297 ~ 1.353) | | | |
| unknown | X-12094 | Sepsis (28 day death) | Inverse variance weighted | 5 | 0.721 | 0.002 | 0.111 (0.027 ~ 0.455) | 0.554 | 0.211 | 0.569 |
| | | | MR Egger | 5 | 2.361 | 0.092 | 0.003 (0.001 ~ 0.320) | | | |
| | | | Simple mode | 5 | 1.359 | 0.146 | 0.087 (0.006 ~ 1.244) | | | |
| | | | Weighted median | 5 | 0.938 | 0.008 | 0.084 (0.013 ~ 0.530) | | | |
| | | | Weighted mode | 5 | 1.169 | 0.076 | 0.062 (0.006 ~ 0.616) | | | |
| Nucleotide | X-12095--N1-methyl-3-pyridone-4-carboxamide | Sepsis (28 day death) | Inverse variance weighted | 15 | 0.543 | 0.008 | 0.237 (0.082 ~ 0.688) | 0.620 | 0.671 | 0.604 |
| | | | MR Egger | 15 | 1.158 | 0.406 | 0.370 (0.038 ~ 3.582) | | | |
| | | | Simple mode | 15 | 1.499 | 0.263 | 0.174 (0.009 ~ 3.283) | | | |
| | | | Weighted median | 15 | 0.803 | 0.100 | 0.267 (0.055 ~ 1.290) | | | |
| | | | Weighted mode | 15 | 1.206 | 0.541 | 0.469 (0.044 ~ 4.991) | | | |
| Lipid | 1-palmitoylglycerophosphocholine | Sepsis (28 day death) | Inverse variance weighted | 20 | 0.937 | 0.045 | 6.572 (1.047 ~ 41.253) | 0.764 | 0.617 | 0.780 |
| | | | MR Egger | 20 | 4.305 | 0.363 | 55.687 (0.012 ~ 257066.749) | | | |
| | | | Simple mode | 20 | 2.523 | 0.133 | 52.689 (0.375 ~ 7401.713) | | | |
| | | | Weighted median | 20 | 1.322 | 0.013 | 26.630 (1.994 ~ 355.648) | | | |
| | | | Weighted mode | 20 | 2.473 | 0.133 | 48.505 (0.381 ~ 6172.563) | | | |

*(Continued)*





10.3389/fmed.2023.1310391

TABLE 2 (Continued)

| Level | Exposure | Outcome | Method | Nsnps | Se | *p* | OR (95% CI) | Cochran's Q *p* | MR-Egger intercept *p* | MR-PRESSO global test *p* |
|---|---|---|---|---|---|---|---|---|---|---|
| unknown | X-13069 | Sepsis (28 day death) | Inverse variance weighted | 11 | 0.482 | 0.020 | 3.060 (1.189 ~ 7.877) | 0.667 | 0.349 | 0.668 |
| | | | MR Egger | 11 | 1.497 | 0.127 | 12.416 (0.660 ~ 233.493) | | | |
| | | | Simple mode | 11 | 1.095 | 0.584 | 1.857 (0.217 ~ 15.872) | | | |
| | | | Weighted median | 11 | 0.645 | 0.222 | 2.200 (0.621 ~ 7.785) | | | |
| | | | Weighted mode | 11 | 1.116 | 0.550 | 1.994 (0.224 ~ 17.772) | | | |
| Lipid | X-13183--stearamide | Sepsis (28 day death) | Inverse variance weighted | 9 | 0.303 | 0.025 | 1.968 (1.087 ~ 3.561) | 0.602 | 0.520 | 0.628 |
| | | | MR Egger | 9 | 0.639 | 0.142 | 2.880 (0.823 ~ 10.078) | | | |
| | | | Simple mode | 9 | 0.688 | 0.409 | 1.820 (0.473 ~ 7.007) | | | |
| | | | Weighted median | 9 | 0.434 | 0.230 | 1.684 (0.719 ~ 3.944) | | | |
| | | | Weighted mode | 9 | 0.578 | 0.535 | 1.454 (0.469 ~ 4.513) | | | |
| Lipid | 1-arachidonoylglycerophosphoethanolamine* | Sepsis (28 day death) | Inverse variance weighted | 14 | 0.706 | 0.037 | 0.230 (0.058 ~ 0.918) | 0.091 | 0.418 | 0.127 |
| | | | MR Egger | 14 | 1.852 | 0.143 | 0.055 (0.001 ~ 2.070) | | | |
| | | | Simple mode | 14 | 1.707 | 0.093 | 0.045 (0.002 ~ 1.285) | | | |
| | | | Weighted median | 14 | 0.782 | 0.002 | 0.093 (0.020 ~ 0.430) | | | |
| | | | Weighted mode | 14 | 0.902 | 0.019 | 0.090 (0.015 ~ 0.528) | | | |
| Amino acid | 3-(3-hydroxyphenyl)propionate | Sepsis (28 day death) | Inverse variance weighted | 7 | 0.203 | 0.003 | 1.848 (1.241 ~ 2.753) | 0.411 | 0.871 | 0.424 |
| | | | MR Egger | 7 | 0.762 | 0.549 | 1.632 (0.366 ~ 7.268) | | | |
| | | | Simple mode | 7 | 0.495 | 0.114 | 2.499 (0.947 ~ 6.595) | | | |
| | | | Weighted median | 7 | 0.289 | 0.004 | 2.317 (1.316 ~ 4.080) | | | |

*(Continued)*





10.3389/fmed.2023.1310391

TABLE 2 (Continued)

| Level | Exposure | Outcome | Method | Nsnps | Se | *p* | OR (95% CI) | Cochran's Q *p* | MR-Egger intercept *p* | MR-PRESSO global test *p* |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | Weighted mode | 7 | 0.443 | 0.074 | 2.609 (1.095 ~ 6.217) | | | |
| unknown | X-14626 | Sepsis (28 day death) | Inverse variance weighted | 10 | 0.580 | 0.036 | 0.297 (0.095 ~ 0.925) | 0.749 | 0.948 | 0.799 |
| | | | MR Egger | 10 | 1.150 | 0.298 | 0.278 (0.029 ~ 2.646) | | | |
| | | | Simple mode | 10 | 1.163 | 0.320 | 0.294 (0.003 ~ 2.876) | | | |
| | | | Weighted median | 10 | 0.773 | 0.091 | 0.271 (0.060 ~ 1.232) | | | |
| | | | Weighted mode | 10 | 0.802 | 0.143 | 0.276 (0.057 ~ 1.330) | | | |

CI, confidence interval; NSNPs, number of single nucleotide polymorphisms.

TABLE 3 Results of pathway enrichment analysis.

| Metabolic pathway | Trait | Database | *p* |
|---|---|---|---|
| Caffeine metabolism | sepsis | KEGG | 0.031885 |
| alpha-Linolenic acid metabolism | sepsis | KEGG | 0.04129 |
| Alpha linolenic acid and linoleic acid metabolism | sepsis | SMPBD | 0.0047302 |
| | sepsis 28 | SMPBD | 0.046867 |
| Primary bile acid biosynthesis | sepsis 28 | KEGG | 0.0025375 |

KEGG, Kyoto Encyclopedia of Genes and Genomes; SMPDB, Small Molecule Pathway Database.

Therefore, considering dietary supplementation or adjustments in ALA and LA intake is crucial in mitigating the risk of sepsis and its associated mortality. Further investigation into the intricate mechanisms through which these fatty acids and their metabolites operate in sepsis is warranted to unveil innovative clinical treatment and prevention strategies.

Our study offers several advantages. Firstly, it is the first investigation to examine the causal association between blood metabolites and the occurrence and progression of sepsis using MR analysis. MR is a statistical technique grounded in whole-genome sequencing data, employed to reveal causal relationships. It effectively minimizes bias and yields more reliable outcomes compared to conventional observational studies, such as randomized controlled trials. The identified causal associations may provide potential blood metabolites for subsequent mechanistic investigations. Secondly, the SNPs linked to blood metabolites were obtained from the most extensive and comprehensive GWAS meta-analysis conducted to date. Thirdly, our selection criteria for IVs were more rigorous compared to other studies, ensuring the reliability of our research. Moreover, the large sample size enhances statistical power, and rigorous sensitivity analysis ensures the robustness of our findings. In spite of this, our study also has limitations. Firstly, all study participants are of European ancestry, so caution is necessary when extrapolating the results to other populations. Furthermore, there were 11 unknown blood metabolites in the preliminary analysis, necessitating further research to explore their specific associations with sepsis. Lastly, while MR analysis is effective in etiological research, the metabolites causally associated with sepsis identified in this study require further experimental validation and exploration of their specific mechanisms. Therefore, further refinement of our study in this area is needed.

## 5 Conclusion

In this MR study, we successfully identified 13 blood metabolites that exhibit a causal relationship with sepsis and 16 blood metabolites associated with a causal relationship to 28-day all-cause mortality in sepsis. The identification of these blood metabolites, whether beneficial or detrimental, holds significant promise for enhancing our understanding of sepsis etiology and informing the development of preventive and therapeutic approaches for managing sepsis.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary material, further inquiries can be directed to the corresponding author.

## Author contributions

ZZ: Writing – original draft, Writing – review & editing, Conceptualization, Data curation. YY: Data curation, Writing – original draft. TC: Data curation, Writing – original draft. JY: Writing – original draft. WZ: Writing – original draft. YZ: Writing – original draft. YR: Writing – original draft. HW: Writing – original draft. XC: Writing – original draft. XZ: Conceptualization, Resources, Supervision, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by the 333 High-Level Talent Training Project of Jiangsu Province (2022-3-25-045), and "Six one Project" of high-level health talents in Jiangsu Province (LGY2019067).

## Acknowledgments

We express our gratitude to the researchers for their generous sharing of the GWAS data utilized in our presented study.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fmed.2023.1310391/full#supplementary-material

## References

1. Shankar-Hari M, Phillips GS, Levy ML, Seymour CW, Liu VX, Deutschman CS, et al. Developing a new definition and assessing new clinical criteria for septic shock: for the third international consensus definitions for Sepsis and septic shock (Sepsis-3). *JAMA*. (2016) 315:775–87. doi: 10.1001/jama.2016.0289

2. Fleischmann C, Scherag A, Adhikari NK, Hartog CS, Tsaganos T, Schlattmann P, et al. Assessment of global incidence and mortality of hospital-treated Sepsis. Current estimates and limitations. *Am J Respir Crit Care Med*. (2016) 193:259–72. doi: 10.1164/rccm.201504-0781OC

3. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and National Sepsis Incidence and mortality, 1990-2017: analysis for the global burden of disease study. *Lancet*. (2020) 395:200–11. doi: 10.1016/s0140-6736(19)32989-7

4. Reinhart K, Daniels R, Kissoon N, Machado FR, Schachter RD, Finfer S. Recognizing Sepsis as a Global Health priority - a who resolution. *N Engl J Med*. (2017) 377:414–7. doi: 10.1056/NEJMp1707170

5. Holmes E, Wilson ID, Nicholson JK. Metabolic phenotyping in health and disease. *Cells*. (2008) 134:714–7. doi: 10.1016/j.cell.2008.08.026

6. Gong T, Liu L, Jiang W, Zhou R. Damp-sensing receptors in sterile inflammation and inflammatory diseases. *Nat Rev Immunol*. (2020) 20:95–12. doi: 10.1038/s41577-019-0215-7

7. Ferrario M, Cambiaghi A, Brunelli L, Giordano S, Caironi P, Guatteri L, et al. Mortality prediction in patients with severe septic shock: a pilot study using a target metabolomics approach. *Sci Rep*. (2016) 6:20391. doi: 10.1038/srep20391

8. Langley RJ, Tipper JL, Bruse S, Baron RM, Tsalik EL, Huntley J, et al. Integrative "Omic" analysis of experimental bacteremia identifies a metabolic signature that distinguishes human Sepsis from systemic inflammatory response syndromes. *Am J Respir Crit Care Med*. (2014) 190:445–55. doi: 10.1164/rccm.201404-0624OC

9. Fanos V, Caboni P, Corsello G, Stronati M, Gazzolo D, Noto A, et al. Urinary (1) H-Nmr and Gc-Ms metabolomics predicts early and late onset neonatal Sepsis. *Early Hum Dev*. (2014) 90:S78–83. doi: 10.1016/s0378-3782(14)70024-6

10. Blaise BJ, Gouel-Chéron A, Floccard B, Monneret G, Allaouchiche B. Metabolic phenotyping of traumatized patients reveals a susceptibility to Sepsis. *Anal Chem*. (2013) 85:10850–5. doi: 10.1021/ac402235q

11. Eckerle M, Ambroggio L, Puskarich MA, Winston B, Jones AE, et al. Metabolomics as a driver in advancing precision medicine in Sepsis. *Pharmacotherapy*. (2017) 37:1023–32. doi: 10.1002/phar.1974

12. Mickiewicz B, Duggan GE, Winston BW, Doig C, Kubes P, Vogel HJ. Metabolic profiling of serum samples by 1h nuclear magnetic resonance spectroscopy as a potential diagnostic approach for septic shock. *Crit Care Med*. (2014) 42:1140–9. doi: 10.1097/ccm.0000000000000142

13. Liu Z, Yin P, Amathieu R, Savarin P, Xu G. Application of Lc-Ms-based metabolomics method in differentiating septic survivors from non-survivors. *Anal Bioanal Chem*. (2016) 408:7641–9. doi: 10.1007/s00216-016-9845-9

14. Lee J, Banerjee D. Metabolomics and the microbiome as biomarkers in Sepsis. *Crit Care Clin*. (2020) 36:105–13. doi: 10.1016/j.ccc.2019.08.008

15. Thorkildsen MS, Gustad LT, Mohus RM, Burgess S, Nilsen TIL, Damås JK, et al. Association of Genetically Predicted Insomnia with risk of Sepsis: a Mendelian randomization study. *JAMA Psychiatry*. (2023) 80:1061–5. doi: 10.1001/jamapsychiatry.2023.2717

16. Smith GD, Ebrahim S. 'Mendelian Randomization': can genetic epidemiology contribute to understanding environmental determinants of disease? *Int J Epidemiol*. (2003) 32:1–22. doi: 10.1093/ije/dyg070

17. Richmond RC, Davey SG. Mendelian randomization: concepts and scope. *Cold Spring Harb. Perspect. Med*. (2022) 12. doi: 10.1101/cshperspect.a040501

18. Shin SY, Fauman EB, Petersen AK, Krumsiek J, Santos R, Huang J, et al. An atlas of genetic influences on human blood metabolites. *Nat Genet*. (2014) 46:543–50. doi: 10.1038/ng.2982

19. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. Kegg for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res*. (2012) 40:D109–14. doi: 10.1093/nar/gkr988

20. Ponsford MJ, Gkatzionis A, Walker VM, Grant AJ, Wootton RE, Moore LSP, et al. Cardiometabolic traits, Sepsis, and severe Covid-19: a Mendelian randomization investigation. *Circulation*. (2020) 142:1791–3. doi: 10.1161/circulationaha.120.050753

21. Boef AG, Dekkers OM, le Cessie S. Mendelian randomization studies: a review of the approaches used and the quality of reporting. *Int J Epidemiol*. (2015) 44:496–11. doi: 10.1093/ije/dyv071

22. Burgess S, Butterworth A, Thompson SG. Mendelian randomization analysis with multiple genetic variants using summarized data. *Genet Epidemiol*. (2013) 37:658–65. doi: 10.1002/gepi.21758

23. Gill D, Brewer CF, Monori G, Trégouët DA, Franceschini N, Giambartolomei C, et al. Effects of genetically determined Iron status on risk of venous thromboembolism

and carotid atherosclerotic disease: a Mendelian randomization study. *J Am Heart Assoc*. (2019) 8:e012994. doi: 10.1161/jaha.119.012994

24. Yamaji K, Sarker KP, Kawahara K, Iino S, Yamakuchi M, Abeyama K, et al. Anandamide induces apoptosis in human endothelial cells: its regulation system and clinical implications. *Thromb Haemost*. (2003) 89:875–84. doi: 10.1055/s-0037-1613475

25. Szafran B, Borazjani A, Lee JH, Ross MK, Kaplan BL. Lipopolysaccharide suppresses carboxylesterase 2g activity and 2-Arachidonoylglycerol hydrolysis: a possible mechanism to regulate inflammation. *Prostaglandins Other Lipid Mediat*. (2015) 121:199–06. doi: 10.1016/j.prostaglandins.2015.09.005

26. Dinu AR, Rogobete AF, Bratu T, Popovici SE, Bedreag OH, Papurica M, et al. *Cannabis Sativa* revisited-crosstalk between Microrna expression, inflammation, oxidative stress, and endocannabinoid response system in critically ill patients with Sepsis. *Cells*. (2020) 9:307. doi: 10.3390/cells9020307

27. Maccarrone M, Lorenzon T, Bari M, Melino G, Finazzi-Agro A. Anandamide induces apoptosis in human cells via Vanilloid receptors. Evidence for a protective role of cannabinoid receptors. *J Biol Chem*. (2000) 275:31938–45. doi: 10.1074/jbc.M005722200

28. Argenziano M, Tortora C, Bellini G, Di Paola A, Punzo F, Rossi F. The endocannabinoid system in pediatric inflammatory and immune diseases. *Int J Mol Sci*. (2019) 20:5875. doi: 10.3390/ijms20235875

29. Mariana Conceição S, Elaine CR. *Cannabinoid receptors as regulators of neutrophil activity in inflammatory diseases Neutrophils*. Rijeka: Intech Open (2018).

30. Šahinović I, Mandić S, Mihić D, Duvnjak M, Loinjak D, Sabadi D, et al. Endocannabinoids, anandamide and 2-Arachidonoylglycerol, as prognostic markers of Sepsis outcome and complications. *Cannabis Cannabinoid Res*. (2022) 8:802–11. doi: 10.1089/can.2022.0046

31. Sultan M, Alghetaa H, Mohammed A, Abdulla OA, Wisniewski PJ, Singh N, et al. The endocannabinoid anandamide attenuates acute respiratory distress syndrome by downregulating Mirna that target inflammatory pathways. *Front Pharmacol*. (2021) 12:644281. doi: 10.3389/fphar.2021.644281

32. Guo MN, Hao XY, Tian J, Wang YC, Li JD, Fan Y, et al. Human blood metabolites and lacunar stroke: a Mendelian randomization study. *Int J Stroke*. (2023) 18:109–16. doi: 10.1177/17474930221140792

33. Ferré-González L, Peña-Bautista C, Baquero M, Cháfer-Pericás C. Assessment of lipid peroxidation in Alzheimer's disease differential diagnosis and prognosis. *Antioxidants (Basel)*. (2022) 11:551. doi: 10.3390/antiox11030551

34. Galano JM, Lee JC, Gladine C, Comte B, Le Guennec JY, Oger C, et al. Non-enzymatic cyclic oxygenated metabolites of Adrenic, docosahexaenoic, Eicosapentaenoic and A-linolenic acids; bioactivities and potential use as biomarkers. *Biochim Biophys Acta*. (2015) 1851:446–55. doi: 10.1016/j.bbalip.2014.11.004

35. Yi J, Jin H, Zhang R, Zhang S, Chen P, Yu X, et al. Increased serum 3-Carboxy-4-Methyl-5-Propyl-2-Furanpropanoic acid (Cmpf) levels are associated with glucose metabolism in Chinese pregnant women. *J Endocrinol Investig*. (2018) 41:663–70. doi: 10.1007/s40618-017-0789-5

36. Zheng JS, Lin M, Imamura F, Cai W, Wang L, Feng JP, et al. Serum metabolomics profiles in response to N-3 fatty acids in Chinese patients with type 2 diabetes: a double-blind randomised controlled trial. *Sci Rep*. (2016) 6:29522. doi: 10.1038/srep29522

37. Miao Z, Zeng FF, Tian Y, Xiao C, Yan Y, Jiang Z, et al. Furan fatty acid metabolite Cmpf is associated with lower risk of type 2 diabetes, but not chronic kidney disease: a longitudinal population-based cohort study. *Am J Clin Nutr*. (2023) 118:637–45. doi: 10.1016/j.ajcnut.2023.07.016

38. Savolainen O, Lind MV, Bergström G, Fagerberg B, Sandberg AS, Ross A. Biomarkers of food intake and nutrient status are associated with glucose tolerance status and development of type 2 diabetes in older Swedish women. *Am J Clin Nutr*. (2017) 106:1302–10. doi: 10.3945/ajcn.117.152850

39. Ottosson F, Hultgren L, Fernandez C, Engström G, Orho-Melander M, Kennbäck C, et al. The inverse association between a fish consumption biomarker and gingival inflammation and periodontitis: a population-based study. *J Clin Periodontol*. (2022) 49:353–61. doi: 10.1111/jcpe.13602

40. Ruszała M, Pilszyk A, Niebrzydowska M, Kimber-Trojnar Ż, Trojnar M, Leszczyńska-Gorzelak B. Novel biomolecules in the pathogenesis of gestational diabetes mellitus 2.0. *Int J Mol Sci*. (2022) 23:4364. doi: 10.3390/ijms23084364

41. Mohan H, Brandt SL, Kim JH, Wong F, Lai M, Prentice KJ, et al. 3-Carboxy-4-Methyl-5-Propyl-2-Furanpropanoic acid (Cmpf) prevents high fat diet-induced insulin resistance via maintenance of hepatic lipid homeostasis. *Diabetes Obes Metab*. (2019) 21:61–72. doi: 10.1111/dom.13483

42. Rivas AM, Nugent K. Hyperglycemia, insulin, and insulin resistance in Sepsis. *Am J Med Sci*. (2021) 361:297–02. doi: 10.1016/j.amjms.2020.11.007

43. Prentice KJ, Wendell SG, Liu Y, Eversley JA, Salvatore SR, Mohan H, et al. Cmpf, a metabolite formed upon prescription Omega-3-acid ethyl Ester supplementation. *Prev Rev Steat EBioMedicine*. (2018) 27:200–13. doi: 10.1016/j.ebiom.2017.12.019

44. Russo GL. Dietary N-6 and N-3 polyunsaturated fatty acids: from biochemistry to clinical implications in cardiovascular prevention. *Biochem Pharmacol*. (2009) 77:937–46. doi: 10.1016/j.bcp.2008.10.020

45. Chen YL, Xie YJ, Liu ZM, Chen WB, Zhang R, Ye HX, et al. Omega-3 fatty acids impair Mir-1-3p-dependent notch 3 Down-regulation and alleviate Sepsis-induced intestinal injury. *Mol Med*. (2022) 28:9. doi: 10.1186/s10020-021-00425-w

46. Saini RK, Keum YS. Omega-3 and Omega-6 polyunsaturated fatty acids: dietary sources, metabolism, and significance - a review. *Life Sci*. (2018) 203:255–67. doi: 10.1016/j.lfs.2018.04.049

47. Yates CM, Calder PC, Ed RG. Pharmacology and therapeutics of Omega-3 polyunsaturated fatty acids in chronic inflammatory disease. *Pharmacol Ther*. (2014) 141:272–82. doi: 10.1016/j.pharmthera.2013.10.010

48. Calder PC. Marine Omega-3 fatty acids and inflammatory processes: effects, mechanisms and clinical relevance. *Biochim Biophys Acta*. (2015) 1851:469–84. doi: 10.1016/j.bbalip.2014.08.010

49. Wang T, Fu X, Chen Q, Patra JK, Wang D, Wang Z, et al. Arachidonic acid metabolism and kidney inflammation. *Int J Mol Sci*. (2019) 20:3683. doi: 10.3390/ijms20153683

50. Bruegel M, Ludwig U, Kleinhempel A, Petros S, Kortz L, Ceglarek U, et al. Sepsis-associated changes of the arachidonic acid metabolism and their diagnostic potential in septic patients. *Crit Care Med*. (2012) 40:1478–86. doi: 10.1097/CCM.0b013e3182416f05

51. Lei P, Xu W, Wang C, Lin G, Yu S, Guo Y. Mendelian randomization analysis reveals causal associations of polyunsaturated fatty acids with Sepsis and mortality risk. *Infect Dis Ther*. (2023) 12:1797–08. doi: 10.1007/s40121-023-00831-z

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 15 January 2024
DOI 10.3389/fimmu.2023.1327879



OPEN ACCESS

EDITED BY
Ashish Patel,
Hemchandracharya North Gujarat University, India

REVIEWED BY
Shaikhul Islam,
Bangladesh Agricultural Research Council, Bangladesh
Dipak Kumar Sahoo,
Iowa State University, United States
Jigna G. Tank,
Saurashtra University, India

*CORRESPONDENCE
Jianning Yao
Jianningyao@163.com

RECEIVED 25 October 2023
ACCEPTED 27 December 2023
PUBLISHED 15 January 2024

CITATION
Song Z, Li X, Xie J, Han F, Wang N, Hou Y and Yao J (2024) Associations of inflammatory cytokines with inflammatory bowel disease: a Mendelian randomization study. *Front. Immunol.* 14:1327879. doi: 10.3389/fimmu.2023.1327879



# Associations of inflammatory cytokines with inflammatory bowel disease: a Mendelian randomization study

Zhaoxiang Song, Xiangyu Li, Jinlin Xie, Fei Han, Nan Wang, Yuhan Hou and Jianning Yao*

Department of Gastroenterology, The First Affiliated Hospital of Zhengzhou University, ZhengZhou, China

**Objectives:** Previous studies have confirmed a link between specific inflammatory cytokines and inflammatory bowel disease (IBD), but the causal relationship between them is not completely clear. This Mendelian Randomization (MR) study aims to evaluate the causal relationship between 18 inflammatory cytokines and inflammatory bowel disease.

**Method:** Two-sample Mendelian randomization utilized genetic variances associated with IBD from two extensive publicly available genome-wide association studies (GWAS) (Crohn's Disease (CD): 12,194 cases and 28,072 controls; Ulcerative Colitis (UC): 12,336 cases and 33,609 controls). The data of inflammatory cytokines was acquired from a GWAS including 8,293 healthy participants. We used inverse variance weighted method, MR-Egger, weighted median, simple model and weighted model to evaluate the causal relationship between inflammatory cytokines and IBD. Sensitivity analysis includes heterogeneity and pleiotropy analysis to evaluate the robustness of the results.

**Results:** The findings indicated suggestive positive associations between Interleukin-13 (IL-13) and macrophage migration inhibitory factor (MIF) with CD (odds ratio, OR: 1.101, 95%CI: 1.021-1.188, p = 0.013; OR: 1.134, 95%CI: 1.024-1.255, p = 0.015). IL-13 also displayed a significant positive correlation with UC (OR: 1.099, 95%CI: 1.018-1.186, p = 0.016). Stem cell factor (SCF) was suggested to be associated with the development of both CD and UC (OR: 1.032, 95%CI: 0.973-1.058, p = 0.012; OR: 1.038, 95%CI: 1.005-1.072, p = 0.024).

**Conclusion:** This study proposes that IL-13 may be a factor correlated with the etiology of IBD (CD and UC), while MIF just be specifically associated with CD. Additionally, SCF appears more likely to be involved in the downstream development of IBD (CD and UC).

KEYWORDS

**inflammatory cytokines, biomarkers, inflammatory bowel disease, Mendelian randomization, GWAS**

# 1 Introduction

Inflammatory bowel disease (IBD) is a nonspecific immune-mediated, chronic recurrent gastrointestinal disease and can be subcategorized into Crohn's disease (CD), ulcerative colitis (UC), and idiopathic colitis (1). The global incidence of UC is on the rise, and it is projected that by 2030, the prevalence rate among Western populations will reach 1%. This poses a significant burden on both global health and the economy (2, 3). Until now, a comprehensive understanding of the etiology and pathogenesis of UC has eluded researchers. The factors implicated include genetic susceptibility, compromised gut mucosal barriers, environmental influences such as increased hygiene standards, urban living, dietary elements, and the dysregulation between gut microbiota and mucosal immunity, any of which may contribute to the onset of UC (4). The primary focus of numerous studies has been on investigating the involvement of immune responses in the pathogenesis of IBD (5, 6). With prolonged activation of the immune system within the intestinal mucosa, it promotes the release of various biomarkers, such as cytokines, including interleukin, chemokine and tumor necrosis factor (7). In canine IBD, there is an initiation of a pro-inflammatory pathway leading to Th cell differentiation, primarily driven by microbial dysbiosis. This imbalance in the microbial community stimulates the generation of mainly pro-inflammatory factors, notably IL-1b. Furthermore, mutations in pattern recognition receptors (PRRs), for example Toll-like receptor 5 (TLR5), heighten the response to flagellin. With dysbiosis characterized by increased Enterobacteriaceae, flagellin expression intensifies, enhancing the mucosa's pro-inflammatory reactions. Consequently, the inflammatory cytokines induce structural changes in epithelial cells, notably increasing permeability due to augmented leakage through tight junctions. This increased permeability establishes a vicious cycle, allowing more bacteria to breach the mucosal barrier, perpetuating the self-reinforcing cycle of inflammation (8). In CD, there's a higher expression of Th1 cell-related cytokines, particularly IFN-γ and IL-2, compared to both UC and individuals without the condition. Conversely, mucosal cells in UC exhibit a tendency to produce Th2-type cytokines like IL-5 and IL-13 (9–12). Multiple studies have indicated an elevated synthesis of Th17 cell-related cytokines, such as IL-17a and IL-17F, by mucosal T cells in both CD and UC (13). Meanwhile, a systematic review showed that six chemokines, including CCL2 (MCP-1), CCL11 (EOTAXIN), CCL26 (EOTAXIN-3), CXCL1 (GROa), CXCL8 (IL-8) and CXCL10 (IP-10), as biomarkers of CD activity are controversial (14). Though some observational studies have endeavored to clarify the connections between inflammatory cytokines and IBD. The conclusions derived from these investigations might be influenced by unforeseen confounding factors or reverse causation, complicating the establishment of definite causal correlations.

Mendelian randomization (MR) is recognized as the analytical method that infers the causal impact of an exposure on an outcome by utilizing genetic variations in non-experimental data (15). As alleles are randomly assigned during meiosis, Mendelian randomization (MR) has the capacity to minimize traditional confounding variables and reverse causation, thereby offering improved evidence for causal inference (16). Conducting a two-sample MR analysis permits researchers to assess the associations between the instrumental variables and both exposure and outcome across two distinct population samples, thereby improving the test's applicability and effectiveness (17). In this study, we first extracted valid genetic variants from the published genome-wide association study (GWAS) summary data of 18 inflammatory cytokines in order to investigate their associations with IBD, and then the direction of causation was further explored by reversing the exposures and outcomes.

# 2 Methods

## 2.1 Study design

The foundation of this study relies on a genetic association database derived from GWAS summary datasets (https://gwas.mrcieu.ac.uk/). Multiple single-nucleotide polymorphisms (SNPs) were chosen to represent genetic variability and used as instrumental variables in a two-sample MR analysis. Three primary hypotheses were established as outlined below (Figure 1): 1. The instrumental variables have a direct association with the exposure; 2. The instrumental variables are not influenced by any confounding variables; 3. Genetic variants solely impact outcomes through their effect on exposure (18). MR analysis was employed to evaluate the bidirectional causal connections between inflammatory cytokines and IBD, encompassing both UC and CD.

## 2.2 Data sources

Both datasets utilized in this MR analysis were sourced from publicly available summarized GWAS data. GWAS data of IBD, containing CD and UC, came from a meta-analysis study. 12194 CD cases (28072 healthy controls), and 12336 UC cases (33609 healthy controls) were available in this data set, with corresponding GWAS IDs of ebi-a-GCST004132 and ebi-a-GCST004133, respectively. For inflammatory cytokines, the data was from the study providing genome variant associations with cytokines and growth factors in 8,293 Finnish individuals. This study combined the results from The Cardiovascular Risk in Young Finns Study (YFS) and FINRISK surveys. The average participant ages are 37 years for YFS study and 60 years for FINRISK survey. There would be no overlap in population selection between the exposure group and the outcome group.

**Abbreviations:** MR, Mendelian randomization; IBD, Inflammatory bowel disease; CD, Crohn's disease; UC, Ulcerative colitis; GWAS, Genome-wide association study; SNPs, Single Nucleotide Polymorphisms; IVs, Instrumental variables; IVW, Inverse variance weighted; IL13, Interleukin-13; MIF, Macrophage migration inhibitory factor; SCF, Stem cell factor.

10.3389/fimmu.2023.1327879

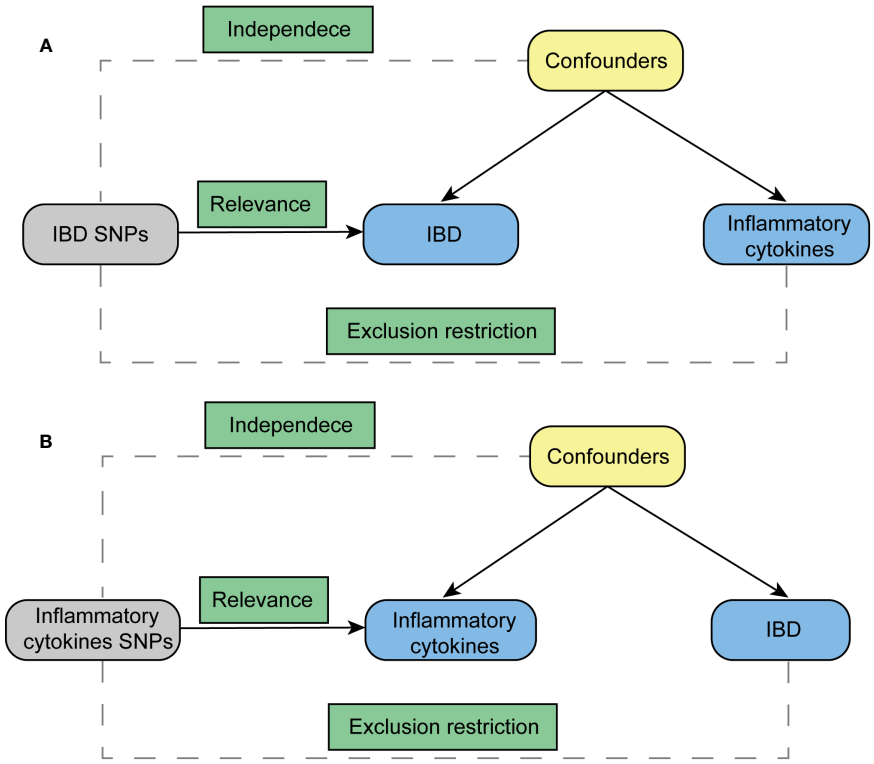



FIGURE 1
Diagram for key assumptions of MR analyses. **(A)** IBD SNPs were used as the genetic instruments to investigate the causal effect of IBD on inflammatory cytokines. **(B)** Genetic instruments in the form of inflammatory cytokine SNPs were utilized to explore the causal relationship between inflammatory cytokines and IBD. The lines with arrows signify the association of genetic instruments (SNPs) with the exposure, affecting the outcome solely through the exposure. Meanwhile, dashed lines represent the independence of the genetic instruments (SNPs) from any confounding variables concerning the outcomes. IBD stands for inflammatory bowel disease.

## 2.3 Instrumental variable selection

Initially, we established the genome-wide significance threshold as $p < 5 \times 10^{-8}$ to pinpoint highly associated SNPs linked with IBD and inflammatory cytokines. However, due to the limited number of identified SNPs for certain inflammatory cytokines when they were considered as the exposure, a higher cutoff ($p < 5 \times 10^{-6}$) was adopted. Next, for the purpose of evading linkage disequilibrium, we conducted SNP clumping (kb = 10,000, $r^2 = 0.001$). Palindromic SNPs were omitted due to uncertainty regarding their alignment in the same orientation for both exposure and outcome in the GWAS of systemic inflammatory regulators. Finally, we assessed the potency of each SNP utilizing the F-statistic, which integrates the extent and accuracy of the genetic impact on the trait: $F = R^2(N - 2)/(1 - R^2)$, where $R^2$ signifies the proportion of the trait's variance elucidated by the SNP, and N denotes the sample size of the GWAS encompassing SNPs associated with the trait (19). The $R^2$ values were estimated using the formula $R^2 = 2 \times EAF \times (1 - EAF) \times \beta^2$. The effect allele frequency (EAF) of the SNP is denoted as EAF, and β represents the estimated effect of the SNP on the trait. SNPs with an F-statistic less than 10 were excluded, as an F-statistic greater than 10 indicated ample strength, ensuring the credibility of the SNPs.

## 2.4 Statistical analysis

Main MR analysis was conducted using the inverse variance weighted (IVW) method. In the MR analysis, multiplicative random effects were applied when utilizing more than three SNPs or in cases of heterogeneity. Other MR methodologies employed to verify result consistency encompassed the weighted median, MR-Egger, simple mode, and weighted mode. Heterogeneity among SNPs was evaluated using the Cochran Q test analysis of IVW and MR-Egger. The MR-Egger intercept test served to identify potential horizontal pleiotropy (version 4.2.2).

# 3 Results

## 3.1 Influence of 18 inflammatory cytokines on IBD

The outcome of the MR analysis indicated a significant positive correlation between genetically predicted IL-13 (OR: 1.101; 95%CI: 1.021-1.188; p = 0.013) and macrophage migration inhibitory factor (MIF) (OR: 1.134; 95%CI: 1.024-1.255; p = 0.015) with CD (Figure 2A). IL13 (OR: 1.099; 95%CI: 1.018-1.186; p = 0.016)

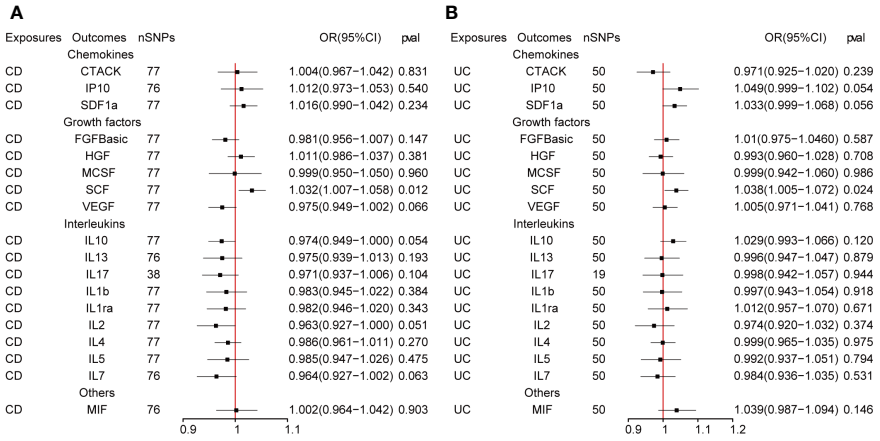



FIGURE 2
**(A, B)** The presented figures represent Mendelian randomization estimates illustrating the causal impacts of CD and UC on inflammatory cytokines. The estimates are displayed as OR and 95% CIs derived from bidirectional Mendelian randomization analyses. OR, odds ratio. 95% CI, 95% confidence interval.

exhibited a notable positive correlation with UC (Figure 2B). Figure 3 displays the scatter plots and funnel plots depicting the Mendelian randomization analyses for IL13 and MIF in IBD. The results from sensitivity analysis indicated that the MR-Egger regression analysis indicated no presence of horizontal pleiotropy, while the Cochran Q test demonstrated the absence of heterogeneity among IVs (Supplementary Table S1). Details of the SNPs are also presented in Supplementary Table S1. While the forest plots and leave-one-out sensitivity analyses of all suggestively significant regulators are presented in Supplementary Figure S1.

## 3.2 Influence of IBD on 18 inflammatory cytokines

Figure 4 presents the outcomes from the reverse MR analysis regarding the causality between IBD and inflammatory cytokines. The results obtained from the IVW method suggested a correlation between an increased level of Stem cell factor (SCF) and CD (OR: 1.032; 95% CI: 1.007-1.058; p = 0.012). Meanwhile, UC was also suggestively correlated with an elevated level of SCF (OR: 1.060; 95% CI: 1.006-1.118; p = 0.028). The scatter plots and funnel plots of SCF are displayed in Figure 5. The results from sensitivity analysis indicated that the MR-Egger regression analysis indicated no presence of horizontal pleiotropy, while the Cochran Q test demonstrated the absence of heterogeneity among IVs (Supplementary Table S1). Supplementary Figure S2 includes forest plots and leave-one-out sensitivity analyses for all regulators that showed suggestive significance.

# 4 Discussion

A recent Mendelian randomized study delved into the causal connections between five interleukins, six chemokines, and IBD. The findings indicated significant positive correlations of IL-16, IL-18, and CXCL10 with IBD, contrasting with IL-12p70 and CCL23, which showed significant negative correlations. Additionally, IL-16 and IL-18 suggested an increased risk of UC, while CXCL10 hinted at an increased risk of CD (20).

We expanded the number of inflammatory cytokines (which includes ILs, chemokines, growth factors and others) and explored the causal relationship between more inflammatory factors and IBD. To understand the causal relationship between IBD and inflammatory cytokines, we used publicly aggregated GWAS data for two-way MR analysis. In the forward MR analysis, elevated levels of IL-13 and MIF were associated with increased risk of CD, whereas IL-13 was also linked to an increased risk of UC. In our reverse MR analysis, CD and UC were suggestively associated with elevated levels of SCF.

Differentiating between CD and UC predominantly depends on the localization of inflammatory lesions and the specific cytokine involvement in their pathogenesis. CD manifests as a segmental, transmural disorder that can impact any segment of the gastrointestinal tract, while UC is identified by superficial, continuous mucosal ulcers restricted to the colon. Dysregulation between pro- and anti-inflammatory cytokines is widely acknowledged in both CD and UC (21). In CD, an association exists with a T helper type 1 (Th1) and T helper type 17 (Th17) immune response (22), leading to the secretion of diverse pro-inflammatory cytokines, such as IFNγ/IL12 and IL23/IL17, which encompass IL18, IL2, IL1, IL21, and IL22. Conversely, UC demonstrates a distinct Th17 and an altered Th2 cytokine profile, characterized by IL13 and IL5. Moreover, both Th1 and Th2 cells, alongside macrophages in both types of IBD, contribute to the production of IL6 and tumor necrosis alpha (TNFα) (23). Genetic polymorphisms in cytokine and cytokine receptor genes may significantly impact the progression of the inflammatory cascade, potentially elevating the susceptibility to developing IBD.

IL-13 is a typical Th2 cytokine produced from CD-1-reactive NKT cells, and its secretion mediates epithelial barrier dysfunction. An increase of IL-13 in lymphocytes of the lamina propria in the

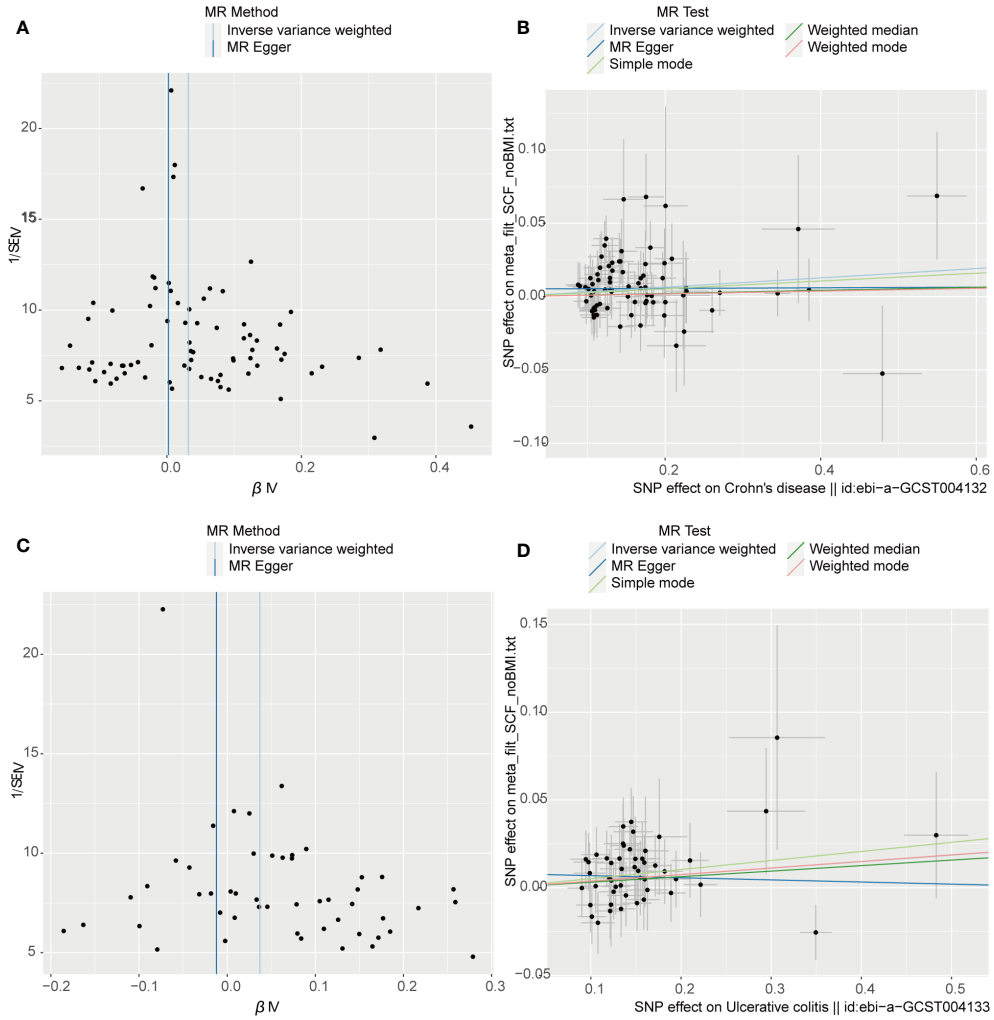



FIGURE 3
Visual aids like scatter plots and funnel plots were employed to illustrate the causal effects of CD and UC on inflammatory cytokines. **(A, C)** The funnel plots depict the inverse variance weighted MR estimates of single-nucleotide polymorphisms associated with CD and UC against cytokines versus 1/standard error (1/SEIV). **(B, D)** Black dots display individual inverse variance (IV) associations with the risk of CD and UC versus individual IV associations with cytokines. The 95%CI of odd ratio for each IV is shown by vertical and horizontal lines. The slope of the lines represents the estimated causal effect of the MR methods.

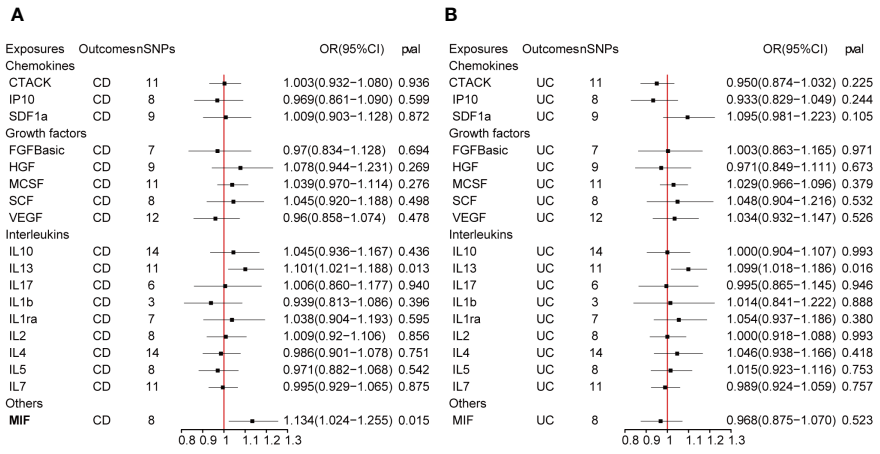



FIGURE 4
**(A, B)** represent mendelian randomized estimates of the causal effect of the ILs and chemokines on CD and UC. Estimates are presented as odds ratios (ORs) and 95% CIs from bidirectional mendelian randomization analyses.

10.3389/fimmu.2023.1327879

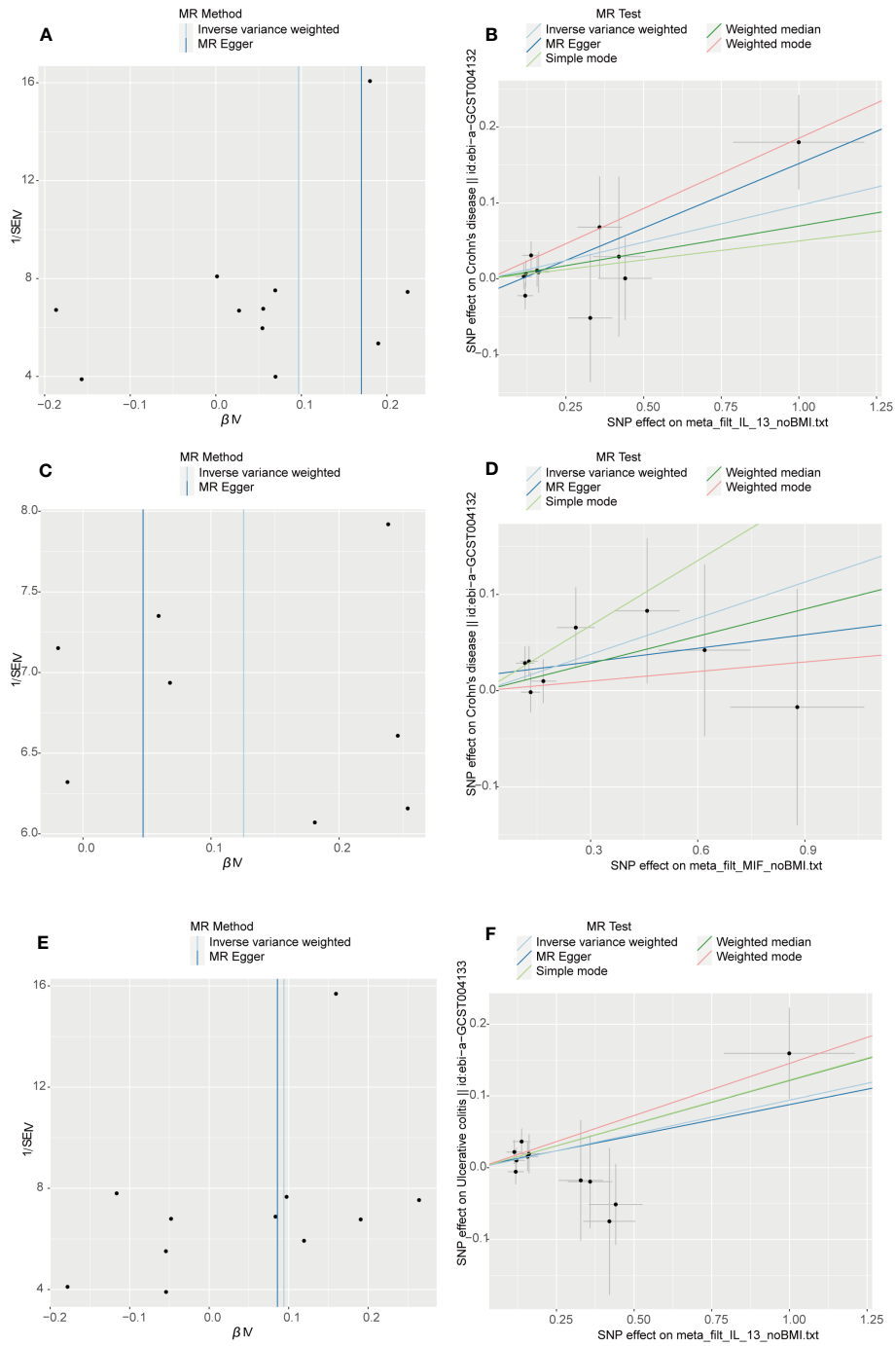



FIGURE 5
Visual aids like scatter plots and funnel plots were employed to illustrate the causal effects of inflammatory cytokines on CD and UC. **(A, C, E)** The funnel plots illustrate the inverse variance weighted MR estimates of each cytokine single-nucleotide polymorphism with CD and UC against 1/standard error (1/SEIV). **(B, D, F)** Black dots display individual inverse variance (IV) associations with the risk of cytokines versus individual IV associations with CD and UC. Vertical and horizontal lines depict the 95% confidence intervals (CI) of the odds ratio for each IV. The slope of these lines indicates the estimated causal effect determined by the MR methods.

affected area of UC represented a significant role of the Th2 immune response in UC pathogenesis. Furthermore, IL-13 was responsible for impairment of mucosal permeability that resulted in epithelial barrier damage, and There were alterations observed in the tight junctions of intestinal epithelial cells. In patients with UC, IL-13 showed a significant increase within apoptotic cells and the corresponding apoptotic area (The Th2 colitis model, Oxazolone colitis, resembling ulcerative colitis, is mediated through IL-13-producing NK-T cells). In the lamina propria's mononuclear cell culture, IL-13 heightened ion flux, leading to alterations in cellular tight junctions. IL-13 exerted an influence on mucosal repair, artificially reducing the rate of mucosal repair by 30% upon its

addition to the mucosal lesions (24). Two papers published in 2004 and 2005 reported that ex vivo stimulated lamina propria T cells, obtained from resected specimens of UC patients, exhibited heightened protein levels of IL-13 compared to individuals with CD and those who were healthy (24, 25). The inflammatory infiltrate of TNF and IL-13 triggers epithelial-to-mesenchymal transition and upregulation of matrix metalloproteinases, resulting in tissue remodeling and the formation of fistulas (26–28). In a Polish population, the presence of IL13 -1112 CT (rs1800925) genotypes indicated an increased likelihood of both IBD and UC occurrence (29).

Initially identified as a factor released by T cells, macrophage migration inhibitory factor (MIF) inhibits the random migration of macrophages (30). Later investigations disclosed that MIF acts as a pro-inflammatory factor., which has important roles in various chronic inflammatory diseases and immune disorders, including UC (31). Some studies showed the capacity of MIF to induce increased functional capacity of DC, and to produce IL-1b and IL-8 from monocytes and DC, indicate a role of MIF in the induction and/or perpetuation of the inflammatory environment in UC and CD (32).

Stem cell factor, also recognized as SCF, KIT-ligand, or steel factor, is a pleiotropic cytokine that governs regulatory impacts on inflammation, tissue remodeling, and fibrosis by binding to its receptor c-KIT (33, 34). SCF is extensively recognized for its role in governing the survival, proliferation, migration, and differentiation of hematopoietic progenitors, melanocytes, and germ cells. Recent investigations have indicated the expression of SCF in dermal and intestinal epithelial cells (35–37). Reports have indicated elevated SCF expression in the inflamed mucosa of individuals with IBD, including SCF248 (a 248 amino acid cleavable form) (38, 39).

Within the cascade of inflammatory events leading to the development of IBD, the involvement of inflammatory cytokines is intricate and potentially interactive. However, MR analysis can isolate their individual impacts and assess the relationship between IBD and these cytokines solely from a genetic standpoint. Nevertheless, our study faces limitations. Primarily, our findings stem from statistical analysis; further validation through extensive basic and clinical research is imperative. Additionally, while restricting the study to individuals of European descent diminishes population structure bias, it could constrain the generalizability of our findings to other populations.

## 5 Conclusion

The publicly available data information from the GWAS database was sourced and analyzed in this study to evaluate the causal relationship between IBD and ILs, and IBD and chemokines, by bidirectional MR analysis. Our results have shown that levels of IL-13 increase the risk of CD and UC, while MIF increase the risk of CD. CD and UC were suggestively correlated with an elevated level of SCF. The underlying mechanism behind these outcomes remains unclear, necessitating further investigation to substantiate our findings.

## Data availability statement

Publicly available datasets were analyzed in this study. This data can be found here: https://doi.org/10.5523/bris.3g3i5smgghp0s2uvm1doflkx9x.

## Author contributions

ZS: Writing – review & editing. XL: Writing – review & editing. JX: Writing – original draft. FH: Writing – original draft. NW: Writing – original draft. YH: Writing – review & editing. JY: Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work is supported by the key Scientific Research Project Plan of Henan Province Colleges and Universities in Department of Education of Henan Province (Grant Number 23A320020).

## Acknowledgments

The authors sincerely thank related investigators for sharing the GWAS summary statistics included in this study. We thank Yimeng Deng from the bottom of our heart for her contribution to manuscript checking and spelling revision.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2023.1327879/full#supplementary-material

# References


1. Nakase H, Uchino M, Shinzaki S, Matsuura M, Matsuoka K, Kobayashi T, et al. Evidence-based clinical practice guidelines for inflammatory bowel disease 2020. *J Gastroenterology.* (2021) 56(6):489–526. doi: 10.1007/s00535-021-01784-1

2. Ananthakrishnan AN, Kaplan GG, Ng SC. Changing global epidemiology of inflammatory bowel diseases: sustaining health care delivery into the 21st century. *Clin Gastroenterol Hepatology.* (2020) 18(6):1252–60. doi: 10.1016/j.cgh.2020.01.028

3. Alatab S, Sepanlou SG, Ikuta K, Vahedi H, Bisignano C, Safiri S, et al. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990–2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatology.* (2020) 5(1):17–30. doi: 10.1016/s2468-1253(19)30333-4

4. Armstrong H, Alipour M, Valcheva R, Bording-Jorgensen M, Jovel J, Zaidi D, et al. Host immunoglobulin G selectively identifies pathobionts in pediatric inflammatory bowel diseases. *Microbiome* (2019) 7(1):1. doi: 10.1186/s40168-018-0604-3

5. Zhang Y-Z. Inflammatory bowel disease: Pathogenesis. *World J Gastroenterology.* (2014) 20(1):91–9. doi: 10.3748/wjg.v20.i1.91

6. Sewell GW, Kaser A. Interleukin-23 in the pathogenesis of inflammatory bowel disease and implications for therapeutic intervention. *J Crohn's Colitis* (2022) 16 (Supplement_2):ii3–ii19. doi: 10.1093/ecco-jcc/jjac034

7. Velikova TV, Miteva L, Stanilov N, Spassova Z, Stanilova SA. Interleukin-6 compared to the other Th17/Treg related cytokines in inflammatory bowel disease and colorectal cancer. *World J Gastroenterology.* (2020) 26(16):1912–25. doi: 10.3748/wjg.v26.i16.1912

8. Kopper JJ, Iennarella-Servantez C, Jergens AE, Sahoo DK, Guillot E, Bourgois-Mochel A, et al. Harnessing the biology of canine intestinal organoids to heighten understanding of inflammatory bowel disease pathogenesis and accelerate drug discovery: A one health approach. *Front Toxicol* (2021) 3:773953. doi: 10.3389/ftox.2021.773953

9. Tan CXW, Brand HS, de Boer NKH, Forouzanfar T. Gastrointestinal diseases and their oro-dental manifestations: Part 2: Ulcerative colitis. *Br Dental J* (2017) 222(1):53–7. doi: 10.1038/sj.bdj.2017.37

10. de Vries SAG, Tan CXW, Bouma G, Forouzanfar T, Brand HS, de Boer NK. Salivary function and oral health problems in Crohn's disease patients. *Inflammatory Bowel Diseases.* (2018) 24(6):1361–7. doi: 10.1093/ibd/izy017

11. Hajishengallis G. Periodontitis: from microbial immune subversion to systemic inflammation. *Nat Rev Immunol* (2014) 15(1):30–44. doi: 10.1038/nri3785

12. Grössner-Schreiber B, Fetter T, Hedderich J, Kocher T, Schreiber S, Jepsen S. Prevalence of dental caries and periodontal disease in patients with inflammatory bowel disease: a case–control study. *J Clin Periodontology.* (2006) 33(7):478–84. doi: 10.1111/j.1600-051X.2006.00942.x

13. Brito F, FCd B, Zaltman C, Pugas Carvalho AT, De Vasconcellos Carneiro AJ, Fischer RG, et al. Prevalence of periodontitis and DMFT index in patients with Crohn's disease and ulcerative colitis. *J Clin Periodontology.* (2008) 35(6):555–60. doi: 10.1111/j.1600-051X.2008.01231.x

14. Mello JDC, MGomes LE, Silva JF, Siqueira NSN, Pascoal LB, Martinez CAR. The role of chemokines and adipokines as biomarkers of Crohn's disease activity: a systematic review of the literature. *Am J Transl Res* (2021) 13(8):8561–74.

15. Wootton RE, Richmond RC, Stuijfzand BG, Lawn RB, Sallis HM, Taylor GMJ, et al. Evidence for causal effects of lifetime smoking on risk for depression and schizophrenia: a Mendelian randomisation study. *psychol Med* (2019) 50(14):2435–43. doi: 10.1017/s0033291719002678

16. Burgess S, Scott RA, Timpson NJ, Davey Smith G, Thompson SG. Using published data in Mendelian randomization: a blueprint for efficient identification of causal risk factors. *Eur J Epidemiol* (2015) 30(7):543–52. doi: 10.1007/s10654-015-0011-z

17. Hartwig FP, Davies NM, Hemani G, Davey Smith G. Two-sample Mendelian randomization: avoiding the downsides of a powerful, widely applicable but potentially fallible technique. *Int J Epidemiol* (2016) 45(6):1717–26. doi: 10.1093/ije/dyx028

18. Davies NM, Holmes MV, Davey Smith G. Reading Mendelian randomisation studies: a guide, glossary, and checklist for clinicians. *Bmj* (2018) 362:k601. doi: 10.1136/bmj.k601

19. Palmer TM, Lawlor DA, Harbord RM, Sheehan NA, Tobias JH, Timpson NJ, et al. Using multiple genetic variants as instrumental variables for modifiable risk factors. *Stat Methods Med Res* (2011) 21(3):223–42. doi: 10.1177/0962280210394459

20. Fang GJ, Kong FZ, Zhang HQ, Huang B, Zhang JF, Zhang XL. Association between inflammatory bowel disease and interleukins, chemokines: a two-sample bidirectional mendelian randomization study. *Front Immunol* (2023) 14:1168188. doi: 10.3389/fimmu.2023.1168188

21. Fuss IJ, Neurath M, Boirivant M, Klein JS, de la Motte C, Strong SA, et al. Disparate CD4+ lamina propria (LP) lymphokine secretion profiles in inflammatory bowel disease. Crohn's disease LP cells manifest increased secretion of IFN-gamma, whereas ulcerative colitis LP cells manifest increased secretion of IL-5. *J Immunol* (1996) 157(3):1261–70.

22. Fujino S, Andoh A, Bamba S, Ogawa A, Hata K, Araki Y, et al. Increased expression of interleukin 17 in inflammatory bowel disease. *Gut* (2003) 52(1):65–70. doi: 10.1136/gut.52.1.65

23. Daniele F, Francesco P. What is the role of cytokines and chemokines in IBD? *Inflammatory Bowel Diseases.* (2008) 14:S117–8. doi: 10.1002/ibd.20677

24. Heller F, Florian P, Bojarski C, Richter J, Christ M, Hillenbrand B, et al. Interleukin-13 is the key effector Th2 cytokine in ulcerative colitis that affects epithelial tight junctions, apoptosis, and cell restitution. *Gastroenterology* (2005) 129 (2):550–64. doi: 10.1016/j.gastro.2005.05.002

25. Fuss IJ, Heller F, Boirivant M, Leon F, Yoshida M, Fichtner-Feigl S, et al. Nonclassical CD1d-restricted NK T cells that produce IL-13 characterize an atypical Th2 response in ulcerative colitis. *J Clin Invest* (2004) 113(10):1490–7. doi: 10.1172/jci19836

26. Scharl M, Frei S, Pesch T, Kellermeier S, Arikkat J, Frei P, et al. Interleukin-13 and transforming growth factor β synergise in the pathogenesis of human intestinal fistulae. *Gut* (2013) 62(1):63–72. doi: 10.1136/gutjnl-2011-300498

27. Frei SM, Pesch T, Lang S, Weber A, Jehle E, Vavricka SR, et al. A role for tumor necrosis factor and bacterial antigens in the pathogenesis of Crohn's disease–associated fistulae. *Inflammatory Bowel Diseases.* (2013) 19(13):2878–87. doi: 10.1097/01.Mib.0000435760.82705.23

28. Sun J, Frei SM, Hemsley C, Pesch T, Lang S, Weber A, et al. The role for dickkopf-homolog-1 in the pathogenesis of Crohn's disease-associated fistulae. *PloS One* (2013) 8(11):e78882. doi: 10.1371/journal.pone.0078882

29. Magyari L, Kovesdi E, Sarlos P, Javorhazy A, Sumegi K, Melegh B. Interleukin and interleukin receptor gene polymorphisms in inflammatory bowel diseases susceptibility. *World J Gastroenterology.* (2014) 20(12):3208–22. doi: 10.3748/wjg.v20.i12.3208

30. David JR. Delayed hypersensitivity *in vitro*: its mediation by cell-free substances formed by lymphoid cell-antigen interaction. *Proc Natl Acad Sci USA* (1966) 56(1):72–7. doi: 10.1073/pnas.56.1.72

31. Lue H, Kleemann R, Calandra T, Roger T, Bernhagen J. Macrophage migration inhibitory factor (MIF): mechanisms of action and role in disease. *Microbes Infect* (2002) 4(4):449–60. doi: 10.1016/s1286-4579(02)01560-5

32. Murakami H, Akbar SM, Matsui H, Horiike N, Onji M. Macrophage migration inhibitory factor activates antigen-presenting dendritic cells and induces inflammatory cytokines in ulcerative colitis. *Clin Exp Immunol* (2002) 128:504–10. doi: 10.1046/j.1365-2249.2002.01838.x

33. Flanagan J, Leder P. The kit ligand: a cell surface molecule altered in steel mutant fibroblasts. *Cell* (1990) 63:185–94. doi: 10.1016/0092-8674(90)90299-t

34. Zsebo K, Williams D, Geissler E, Broudy V, Martin F, Atkins H, et al. Stem cell factor is encoded at the Sl locus of the mouse and is the ligand for the c-kit tyrosine kinase receptor. *Cell* (1990) 63(1):213–24. doi: 10.1016/0092-8674(90)90302-u

35. Carter EL, O'Herrin S, Woolery C, Jack Longley B. Epidermal stem cell factor augments the inflammatory response in irritant and allergic contact dermatitis. *J Invest Dermatol* (2008) 128(7):1861–3. doi: 10.1038/sj.jid.5701247

36. Longley BJ, Morganroth GS, Tyrrell L, Ding TG, Anderson DM, Williams DE, et al. Altered metabolism of mast-cell growth factor (c-kit ligand) in cutaneous mastocytosis. *N Engl J Med* (1993) 328(18):1302–7. doi: 10.1056/NEJM199305063281803

37. Schmitt M, Schewe M, Sacchetti A, Feijtel D, van de Geer WS, Teeuwssen M, et al. Paneth cells respond to inflammation and contribute to tissue regeneration by acquiring stem-like features through SCF/c-kit signaling. *Cell Rep* (2018) 24(9):2312–2328.e2317. doi: 10.1016/j.celrep.2018.07.085

38. Comar M, Secchiero P, De Lorenzo E, Martelossi S, Tommasini A, Zauli G. JCV+ Patients with Inflammatory Bowel Disease show elevated plasma levels of MIG and SCF. *Inflammatory Bowel Diseases.* (2012) 18(6):1194–6. doi: 10.1002/ibd.22953

39. Garcia-Hernandez V, Raya-Sandino A, Azcutia V, Miranda J, Kelm M, Flemming S, et al. Inhibition of soluble stem cell factor promotes intestinal mucosal repair. *Inflammatory Bowel Diseases.* (2023) 29(7):1133–44. doi: 10.1093/ibd/izad003

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 20 February 2024
DOI 10.3389/fonc.2024.1320020



OPEN ACCESS

EDITED BY
Dipak Kumar Sahoo,
Iowa State University, United States

REVIEWED BY
Babak Pakbin,
Technical University of Munich, Germany
Mingsong Kang,
Canadian Food Inspection Agency (CFIA), Canada
Virendra Kumar Yadav,
Hemchandracharya North Gujarat University, India
Ashish Patel,
Hemchandracharya North Gujarat University, India

*CORRESPONDENCE
Xiao-Ling Zheng
shengli888220@163.com

†These authors have contributed equally to this work and share first authorship

RECEIVED 30 October 2023
ACCEPTED 02 February 2024
PUBLISHED 20 February 2024

CITATION
Pan H, Zhang Y-L, Fang C-Y, Chen Y-D, He L-P, Zheng X-L and Li X (2024) Retrospective cohort study investigating association between precancerous gastric lesions and colorectal neoplasm risk. *Front. Oncol.* 14:1320020. doi: 10.3389/fonc.2024.1320020



# Retrospective cohort study investigating association between precancerous gastric lesions and colorectal neoplasm risk

Hui Pan[1], Yu-Long Zhang[2†], Chao-Ying Fang[3†], Yu-Dai Chen[1], Li-Ping He[3], Xiao-Ling Zheng[1*] and Xiaowen Li[1]

[1]Gastrointestinal Endoscopy Center, Fujian Shengli Clinical Medical College, Fujian Medical University, Fuzhou, Fujian, China, [2]Department of Gynecology, Fujian Maternity and Child Health Hospital, Fuzhou, Fujian, China, [3]Gastrointestinal Endoscopy Center, Fujian Provincial Hospital South Branch, Fuzhou, Fujian, China

**Background:** Colorectal cancer (CRC) is considered the most prevalent synchronous malignancy in patients with gastric cancer. This large retrospective study aims to clarify correlations between gastric histopathology stages and risks of specific colorectal neoplasms, to optimize screening and reduce preventable CRC.

**Methods:** Clinical data of 36,708 patients undergoing gastroscopy and colonoscopy from 2005-2022 were retrospectively analyzed. Correlations between gastric and colorectal histopathology were assessed by multivariate analysis. Outcomes of interest included non-adenomatous polyps (NAP), conventional adenomas (CAs), serrated polyps (SPs), and CRC. Statistical analysis used R version 4.0.4.

**Results:** Older age (≥50 years) and *Helicobacter pylori* infection (HPI) were associated with increased risks of conventional adenomas (CAs), serrated polyps (SPs), non-adenomatous polyps (NAP), and colorectal cancer (CRC). Moderate to severe intestinal metaplasia specifically increased risks of NAP and CAs by 1.17-fold (95% CI 1.05-1.3) and 1.19-fold (95% CI 1.09-1.31), respectively. For CRC risk, low-grade intraepithelial neoplasia increased risk by 1.41-fold (95% CI 1.08-1.84), while high-grade intraepithelial neoplasia (OR 3.76, 95% CI 2.25-6.29) and gastric cancer (OR 4.81, 95% CI 3.25-7.09) showed strong associations. More advanced gastric pathology was correlated with progressively higher risks of CRC.

**Conclusion:** Precancerous gastric conditions are associated with increased colorectal neoplasm risk. Our findings can inform screening guidelines to target high-risk subgroups, advancing colorectal cancer prevention and reducing disease burden.

KEYWORDS

**helicobacter pylori, intestinal metaplasia, atrophic gastritis, colorectal adenoma, serrated lesions**

# Introduction

Colorectal cancer (CRC) stands as the third most diagnosed cancer globally and the second leading cause of cancer mortality (1). The incidence and mortality of CRC have been rising rapidly in China, with over 400,000 new cases and 195,600 deaths reported in 2016 (2), making it the second most common cancer diagnosis and fourth leading cause of cancer mortality nationwide. The surge in CRC incidence and mortality in China, underscores the urgent need for effective screening and preventive measures.

While population endoscopic screening is crucial for early detection and removal of premalignant polyps (3), the rates of colonoscopy screening in China lag behind those of upper endoscopy. This discrepancy becomes especially pertinent when clinicians, prompted by findings in gastrointestinal (GI) screening, must make decisions about the necessity of a follow-up colonoscopy, known for its superior detection of colorectal neoplasms.

Recent studies, primarily conducted in Western populations, suggest that certain upper GI pathologies identified during gastroscopy may indicate a higher concurrent or subsequent risk for colorectal neoplasms (4–6). Notably, gastric ulcers, Helicobacter pylori (*H. pylori*) infection, and chronic atrophic gastritis have been associated with an increased prevalence of colorectal adenomas and cancer, possibly due to downstream effects on the GI environment (7–9).

The connection between upper gastrointestinal diseases, such as gastric polyps (10, 11), *H. pylori* infection (12), and reflux esophagitis (13), and elevated colon neoplasm risk has been established in prior evidence. Although the exact mechanisms underlying gastrointestinal diseases remain unclear, lipopolysaccharide (LPS) may play a key role. As a component of Gram-negative bacteria, LPS promotes gastrointestinal diseases through multiple pathways. In the stomach, it activates TGF-beta and Wnt signaling, inducing epithelial-mesenchymal transition and immunotherapy resistance, thereby facilitating gastric cancer progression (14). In the intestines, LPS incites inflammation and tumorigenesis by modulating epithelial signaling cascades (15). Through these diverse mechanisms of stimulating cancer-associated signals in both the gastric and intestinal epithelia, LPS serves as an intermediary factor linking chronic inflammation to carcinoma in gastrointestinal diseases. Besides, this association is potentially linked to the impairment of the gastric acid barrier (16) and the use of proton pump inhibitors (17, 18). However, uncertainties persist regarding whether gastric histopathology, reflecting the pathological state and acid secretion, is directly related to colon neoplasms (19, 20).

A significant Shanghai study involving 5,986 patients shed light on the potential correlation between certain gastric histopathologies —such as atrophic gastritis, intestinal metaplasia, and gastric polyps —and the predisposition to advanced colorectal polyps, as opposed to non-advanced polyps or colorectal cancer (20). This observation underscores the potential value of gastric histopathology in predicting high-risk colon neoplasms specifically within the Chinese population.

However, to validate and generalize these findings, larger population studies across diverse regions are imperative. Such studies would confirm the links between gastric pathology, acid secretion, and susceptibility to colon neoplasms, thereby refining screening practices. Consequently, large-scale analyses leveraging national endoscopy data are essential to elucidate region-specific gastric-colorectal connections in China and optimize screening protocols. This comprehensive retrospective study aims to clarify correlations between gastric histopathology stages and risks of specific colorectal neoplasms, to optimize screening and reduce preventable CRC.

# Methods

## Study design and data selection

This retrospective cross-sectional study was approved by the Academic Ethics Committee of Fujian Provincial Hospital (K2022-01-019) prior to conducting research. Data were retrieved from a scientific research big data platform at Fujian Provincial Hospital, comprising patient demographics, pathological reports, and endoscopic reports (including *H. pylori* status and specimen requisitions). Patients who underwent endoscopy at the Digestive Endoscopy Center from Jan 01, 2005, to Jan 01, 2022 were included. Given the retrospective nature of the study, informed consent was not required.

The study enrolled patients who underwent simultaneous gastroscopy and colonoscopy with tissue biopsy within 2 months at Fujian Provincial Hospital campuses. Exclusion criteria were as follows: history of gastrectomy or colorectomy, history of colorectal polypectomy, inflammatory bowel disease, hereditary polyposis syndromes like Peutz–Jeghers or familial adenomatous polyposis, and incomplete cecal intubation. Data from the index visit were used, excluding repeat cases. Inclusion and exclusion criteria are detailed in the provided flowchart (Figure 1).

## Diagnosis of gastric and colorectal lesions

All gastric histopathological diagnoses were based on the guidelines from the updated Sydney System and Japanese Gastric Cancer Association classification (21). Pathological samples were collected into 10% buffered formalin, embedded in paraffin, sectioned at 2μm, and were sent to the Pathology Department for subsequent staining with hematoxylin and eosin (H&E) and evaluation. Gastric precursor lesions were categorized as atrophic gastritis (AG), intestinal metaplasia (IM - mild, moderate, severe), and dysplasia [low-grade intraepithelial neoplasia (LGIN), high-grade intraepithelial neoplasia (HGIN), gastric cancer (GC)] based on the pathology reports. *H. pylori* infection (HPI) status was assessed via rapid urease test of gastric antrum biopsies.

Colorectal lesions were classified into 5 subgroups per pathology reports which based on WHO classification of tumors of digestive system (22): (1) colorectal cancer (CRC); (2) serrated polyps (SPs) including traditional serrated adenomas and sessile serrated lesions; (3) conventional adenomas (CAs) with tubular/villous components, regardless of dysplasia grade; (4) non-

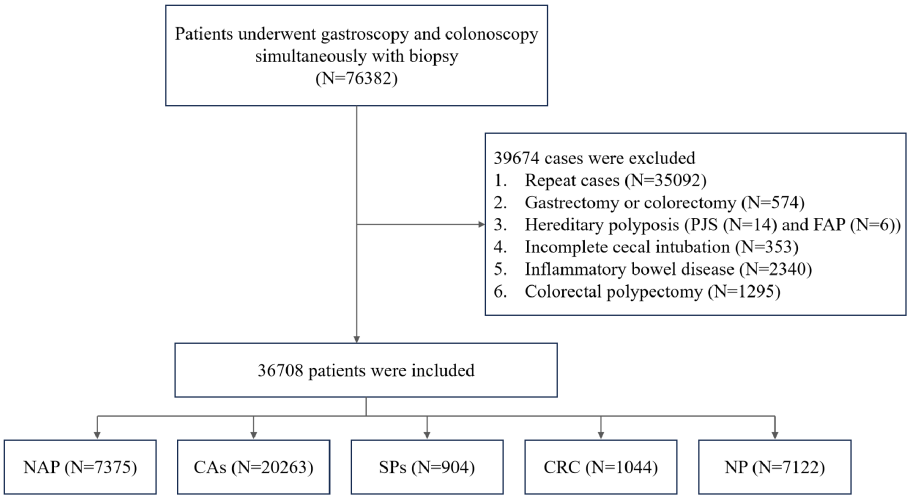



FIGURE 1
Flow chart of the study design. PJS, Peutz–Jeghers syndrome; FAP, familial adenomatous polyposis; NAP, non-adenomatous polyps; CAs, conventional adenomas; SPs, serrated polyps; CRC, colorectal cancer; NP, no polyps.

adenomatous polyps (NAP) like hyperplastic or inflammatory polyps; (5) no polyp (NP) control group.

## Statistical analysis

R language (version 4.0.4) software was used for data analysis and visualization. The histogram distribution or Q-Q plot was utilized to assess the normality of the variables. The mean ± standard deviation (SD) was used to express normally distributed continuous data, while skewed continuous variables were characterized using the median and interquartile range (IQR). The categorical variables were displayed as frequencies expressed as percentages. An analysis was conducted to compare continuous variables between groups using independent samples. The choice between Student's t-test or Mann-Whitney U-test was based on the normality of the distribution. Categorical data were compared using either the chi-square test or Fisher's exact test, depending on the appropriateness (Supplementary Figures 1–3). Univariate and multivariate logistic regression (adjusted for gender, age, H. pylori, and atrophy) was used to determine the relationship between gastric precursor lesions and colorectal conditions, including colorectal polyps and CRC. The results were reported as an adjusted odds ratio (OR) with a corresponding 95% confidence interval (CI). For all analyses, a P-value < 0.05 was statistically significant.

# Results

## Baseline characteristics of various colorectal conditions.

A total of 36,708 patients were included in this study and divided into 5 groups based on colonoscopy and biopsy findings: CRC (n=1,044); CAs (n=20,263); SPs (n=904); NAP (n=7,375); and NP (n=7,122). The proportion of males was significantly higher in the CRC (60.9%, 636/1,044), CAs (62.4%, 12,647/20,263), SPs (58.2%, 526/904), and NAP (61.2%, 4,512/7,375) groups compared to the NP control group (55%) (P < 0.001).

The mean age of patients with CRC (63.2 ± 11.6 years) was significantly higher than those with CAs (56.4 ± 10.5 years), SPs (56.1 ± 11.7 years), NAP (52.7 ± 10.9 years) and NP (53.4 ± 11.7 years) (P < 0.001).

The prevalence of gastric mucosal abnormalities was higher in the CRC, CAs, SPs and NAP groups compared to the NP control group. Specifically, the prevalence of gastric intestinal metaplasia (IM) was 41.3% in CRC (20.3% mild, 21% moderate/severe), 41.8% in CAs (21.6% mild, 20.2% moderate/severe), 38.5% in SPs (21.1% mild, 17.4% moderate/severe), 39.3% in NAP (21.2% mild, 18.1% moderate/severe), and 36.5% in NP (19.8% mild, 16.7% moderate/severe) (P<0.001). The prevalence of gastric dysplasia and cancer was 14.5% in CRC, 7.9% in CAs, 9% in SPs, 6.3% in NAP, and 6.1% in NP (P<0.001), with higher rates of low- and high-grade dysplasia and gastric cancer in the CRC group compared to the other groups (Table 1).

## Association between gastric conditions with NAP

Multivariate analysis revealed that older age (≥50 years) (OR= 1.29, 95% CI = 1.21–1.38) and HPI (OR = 1.09, 95% CI = 1.02–1.17) were independent risk factors for NAP. IM was also an independent risk factor for NAP, with moderate to severe IM having a higher odds ratio (OR = 1.17, 95% CI = 1.05–1.3) compared to mild IM (OR = 1.16, 95% CI = 1.05–1.29). In contrast, male sex (P = 0.334) and atrophic gastritis (AG) (P = 0.227) were not significantly associated with NAP. Additionally, there was no significant association between LGIN (P = 0.476) or HGIN (P = 0.419) and NAP (Figure 2).

TABLE 1 Baseline characteristics of various colorectal conditions.

| Variables | Total (n = 36708) | no polyps (NP) (n = 7122) | non adenomatous polyp (NAP) (n = 7375) | conventional adenomas (CAs) (n = 20263) | serrated polyps (SPs) (n = 904) | colorectal cancer (CRC) (n = 1044) | p |
|---|---|---|---|---|---|---|---|
| Gender, n (%) | | | | | | | < 0.001 |
| Male | 22236 (60.6) | 3915 (55) | 4512 (61.2) | 12647 (62.4) | 526 (58.2) | 636 (60.9) | |
| Female | 14472 (39.4) | 3207 (45) | 2863 (38.8) | 7616 (37.6) | 378 (41.8) | 408 (39.1) | |
| Age, (Mean ± SD) | 55.3 ± 11.1 | 53.4 ± 11.7 | 52.7 ± 10.9 | 56.4 ± 10.5 | 56.1 ± 11.7 | 63.2 ± 11.6 | < 0.001 |
| H. pylori (+), n (%) | 21142 (57.6) | 3923 (55.1) | 4232 (57.4) | 11766 (58.1) | 573 (63.4) | 648 (62.1) | < 0.001 |
| Atrophic gastritis, n (%) | 22040 (60.0) | 4081 (57.3) | 4348 (59) | 12446 (61.4) | 533 (59) | 632 (60.5) | < 0.001 |
| IM, n (%) | | | | | | | < 0.001 |
| Mild | 7755 (21.1) | 1411 (19.8) | 1564 (21.2) | 4377 (21.6) | 191 (21.1) | 212 (20.3) | |
| Mod-Severe | 7000 (19.1) | 1186 (16.7) | 1338 (18.1) | 4100 (20.2) | 157 (17.4) | 219 (21) | |
| Neoplasm, n (%) | | | | | | | < 0.001 |
| LGIN | 2153 (5.9) | 337 (4.7) | 399 (5.4) | 1274 (6.3) | 66 (7.3) | 77 (7.4) | |
| HGIN | 254 (0.7) | 36 (0.5) | 33 (0.4) | 154 (0.8) | 5 (0.6) | 26 (2.5) | |
| GC | 320 (0.9) | 63 (0.9) | 37 (0.5) | 162 (0.8) | 10 (1.1) | 48 (4.6) | |

Data are presented as number (%) or mean ± SD. A chi-square test was used for categorical variables, and a t-test was used for continuous variables. IM, intestinal metaplasia; Mod-Severe, Moderate-severe; LGIN, low-grade intraepithelial neoplasia; HGIN, high-grade intraepithelial neoplasia; GC, gastric cancer.

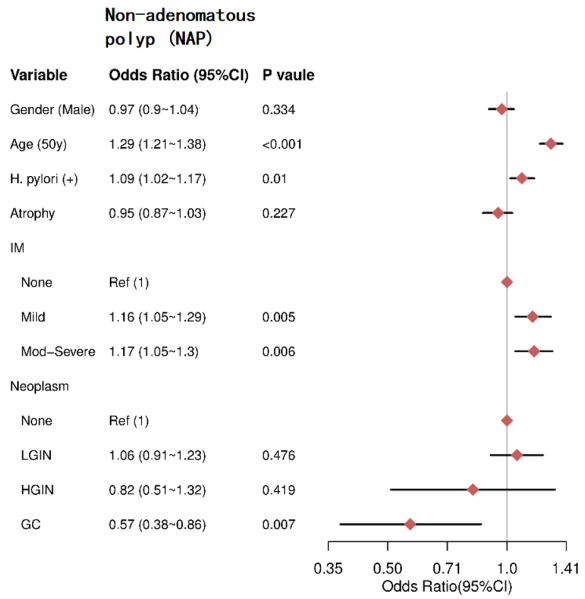



FIGURE 2
The result of multivariate logistic regression for the NAP group. IM, intestinal metaplasia; LGIN, low-grade intraepithelial neoplasia; HGIN, high-grade intraepithelial neoplasia; GC, gastric cancer.

## Association between gastric conditions with CAs

Multivariate analysis revealed that male sex (OR= 1.7, 95% CI = 1.6–1.8), older age (≥50 years) (OR= 1.4, 95% CI =1.33–1.48), and HPI (OR = 1.11, 95% CI = 1.05–1.17) were independent risk factors for CAs. IM was also a risk factor, with moderate to severe IM associated with higher risk (OR = 1.19, 95% CI = 1.09–1.31) compared to mild IM (OR = 1.14, 95% CI = 1.05–1.24). There was no significant association between AG and CAs (P = 0.473). For gastric neoplasms, HGIN (P = 0.33) and GC (P = 0.258) were not significantly associated with CAs (Figure 3).

## Association between gastric conditions with SPs

Multivariate analysis revealed that male sex (OR= 1.46, 95% CI = 1.25–1.71), HPI (OR = 1.41, 95% CI = 1.22–1.63), and older age (≥50 years) (OR = 1.16, 95% CI = 1–1.33) were independent risk factors for SPs. There was no significant association between AG and SPs (P = 0.477). Additionally, IM severity was not significantly associated with SPs, including mild IM (P = 0.537) and moderate-severe IM (P = 0.731). For gastric

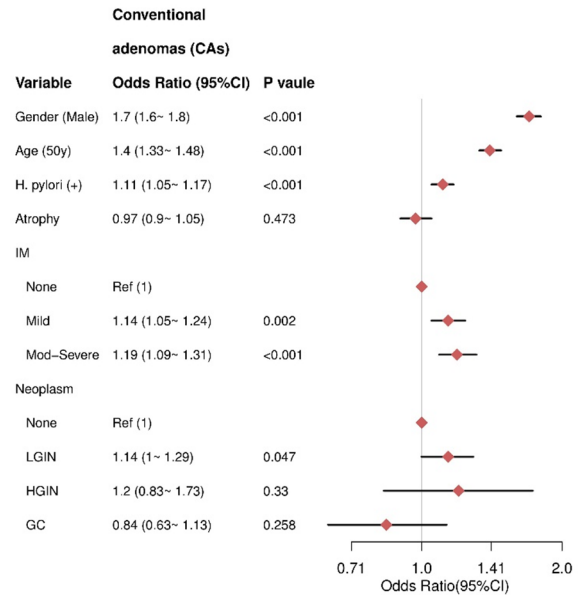



FIGURE 3
The result of multivariate logistic regression for the CAs group. IM, intestinal metaplasia; LGIN, low-grade intraepithelial neoplasia; HGIN, high-grade intraepithelial neoplasia; GC, gastric cancer.

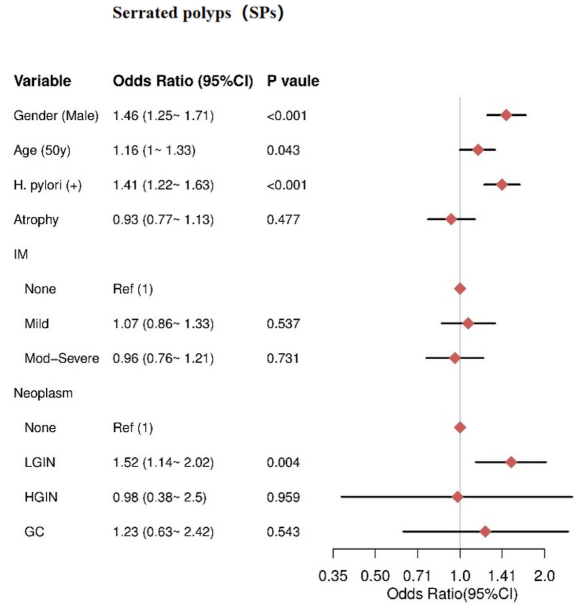



FIGURE 4
The result of multivariate logistic regression for the SPs group. IM, intestinal metaplasia; LGIN, low-grade intraepithelial neoplasia; HGIN, high-grade intraepithelial neoplasia; GC, gastric cancer.

neoplasms, only LGIN showed an increased risk for SPs (OR = 1.52, 95% CI = 1.14–2.02), while HGIN (P = 0.959) and GC (P = 0.543) were not significantly associated with SPs (Figure 4).

## Association between gastric conditions with CRC

Multivariate analysis revealed that male sex (OR= 4.03, 95% CI = 3.31–4.9), older age (≥50 years) (OR= 1.32, 95% CI = 1.15–1.51), and HPI (OR = 1.34, 95% CI = 1.17–1.53) were independent risk factors for CRC. Gastric neoplasms were also independent risk factors for CRC, including LGIN (OR = 1.41, 95% CI = 1.08–1.84), HGIN (OR = 3.76, 95% CI = 2.25–6.29), and GC (OR = 4.81, 95% CI = 3.25–7.09), with higher risks associated with more advanced neoplasms. There was no significant association between AG and CRC (P = 0.548) or IM severity and CRC, including mild IM (P = 0.73) and moderate-severe IM (P = 0.408) (Figure 5).

# Discussion

This comprehensive retrospective study, involving 36,708 patients who underwent gastroscopy and colonoscopy between 2005 and 2022, presents compelling evidence that precancerous gastric conditions may increase the risk of colorectal neoplasms. Multivariate analysis was employed to assess correlations between various gastric histopathological findings and different subtypes of colorectal neoplasms, including non-adenomatous polyps, conventional adenomas, serrated polyps, and colorectal cancer.

Our study's results demonstrate that advanced age (≥50 years) and *Helicobacter pylori* infection (HPI) significantly elevate the risk across all assessed colorectal neoplasm subtypes, including conventional adenomas, serrated polyps, non-adenomatous polyps, and colorectal cancer (CRC). These findings align with prior evidence suggesting that chronic inflammation induced by *H.*

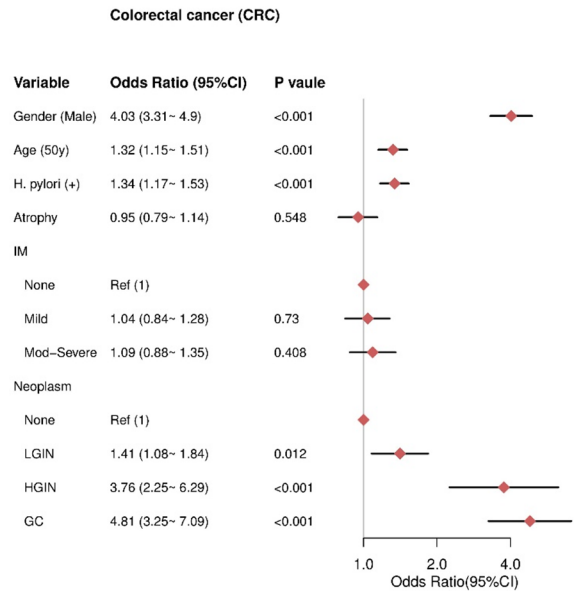



FIGURE 5
The result of multivariate logistic regression for the CRC group. IM, intestinal metaplasia; LGIN, low-grade intraepithelial neoplasia; HGIN, high-grade intraepithelial neoplasia; GC, gastric cancer.

*pylori* predisposes individuals to colorectal tumorigenesis. It is noteworthy that these risk factors share significant potential connections and similarities (23–25). HPI triggers chronic gastritis, which can progress to mucosal atrophy, intestinal metaplasia, and reduced gastric acid production (16), further heightening the risk of intestinal diseases.

Additionally, our study revealed an independent association between moderate to severe intestinal metaplasia (IM) and non-adenomatous polyps (NAP), implying that IM may serve as a risk factor for NAP. Unlike previous research that considered atrophy and IM together (20), our investigation differentiated between them. We found that atrophy alone was not linked to an increased risk of NAP, but moderate to severe IM was significantly associated with a higher risk of NAP development. This suggests that the severity of IM, rather than the mere presence of both atrophy and IM, may be a crucial factor affecting the association with NAP.

Similar results were observed for conventional adenomas (CAs). Gastric atrophy did not elevate the risk of CAs, whereas the presence of IM emerged as an independent risk factor for CAs development. Specifically, we found no significant association between gastric atrophy and the occurrence of CAs. In contrast, the presence of IM was linked to an increased risk of CAs.

However, our study did not establish a clear association between low-grade or high-grade gastric epithelial dysplasia (LGIN and HGIN) and NAP or CAs. In our study, only LGIN was identified as a significant risk factor for serrated polyps (SPs) (OR = 1.52, 95% CI = 1.14-2.02). No clear associations were observed between HGIN (P = 0.959) or gastric cancer (GC) (P = 0.543) and SPs. The heightened risk of SPs associated with LGIN but not advanced neoplasia suggests that early gastric lesions may play a particularly influential role in promoting serrated pathway colon carcinogenesis. These differential effects of gastric precancerous stages imply that specific interactions, rather than overall atrophic changes, may exert an influence on this subtype of colorectal neoplasms.

It is noteworthy that LGIN increases the risk of CRC, with an odds ratio (OR) of 1.41 (95% CI 1.08-1.84). HGIN exhibits an even stronger association with the risk of CRC, with an OR of 3.76 (95% CI 2.25-6.29). Gastric cancer itself demonstrates the highest correlation with CRC, with an OR of 4.81 (95% CI 3.25-7.09). This risk stratification based on the degree of gastric pathology aligns with the Correa cascade model of gastric carcinogenesis, supporting an underlying field effect that influences the risk of both gastric and colorectal neoplasia.

Previous research has documented bidirectional synchronous or metachronous occurrences of colorectal and gastric cancers, with 0.7-1.3% of gastric cancer patients subsequently developing colorectal cancer (26, 27), and approximately 5% of colorectal cancer patients experiencing gastric cancer development (28, 29). Prevalence studies have also highlighted an increased incidence of colorectal neoplasms among gastric cancer patients when compared to control groups. For instance, Lee et al. (7) reported a prevalence of colorectal neoplasms in 35.8% of gastric cancer patients, contrasting with 17.9% in the control group. Similarly, Park et al. (30) found a higher prevalence of colorectal adenoma (39.6% vs. 28.6%) and cancer (3.5% vs. 1.3%) in gastric cancer patients versus the control group. Collectively, these findings consistently demonstrate an epidemiological link between gastric and colorectal neoplasms, underscoring the potential for shared carcinogenic processes.

Several studies have identified specific pathways contributing to the development and metastasis of both gastric and colorectal cancers, providing deeper insights into the connections between gastrointestinal tumors (31, 32). The association between premalignant gastric lesions and colorectal neoplasms also suggests the necessity of updating screening recommendations for high-risk patient subgroups.

Additionally, while the significant Shanghai study involving 5,986 patients offered valuable insights into potential correlations between specific gastric histopathologies—such as atrophic gastritis, intestinal metaplasia, and gastric polyps—and the predisposition to advanced colorectal polyps (20), it did not differentiate between non-advanced polyps and colorectal cancer. Our research in Fujian, boasting a larger sample size, not only corroborates and expands upon the findings from the Shanghai study but also delves deeper into nuanced associations between gastric pathology and different stages of colorectal neoplasms.

The divergence in our findings may be attributed to diverse population characteristics in Fujian, shedding light on region-specific variations in the gastric-colorectal connection. These disparities underscore the importance of conducting large-scale studies in specific regions to tailor screening guidelines effectively. Our study provides nuanced insights that can significantly influence screening strategies, enabling the targeted identification and prevention of colorectal neoplasms in high-risk subgroups within the Fujian population.

It is important to acknowledge some limitations in our study. The cross-sectional design of our research prevented us from assessing temporality and establishing causation. Prospective cohort studies that track outcomes over time are needed to confirm predictive relationships between gastric abnormalities and subsequent colorectal neoplasm development. Furthermore, our analysis only adjusted for a limited set of confounding factors including gender, age, H. pylori, and atrophy, which may have introduced residual biases. Expanding adjustments for socioeconomic status, lifestyle factors, and comorbidities could improve isolation of the association with gastric lesions. Additionally, expanding the diversity of our study sample could enhance the generalizability of our results across different populations. However, despite these limitations, our findings strongly support the existence of associations between premalignant stomach lesions and the development of colorectal tumors, suggesting potential utility of this association in guiding screening approaches. Our results open new directions for future research with the potential to significantly impact the prevention and burden of CRC.

In conclusion, this study shows that certain precancerous gastric conditions are associated with a higher risk of colorectal neoplasms. Our results can be used to optimize secondary prevention initiatives targeting high-risk subgroups for enhanced screening. Further research that focuses on the connections between gastric and colorectal diseases may significantly contribute to curbing the global burden of preventable CRC.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

## Author contributions

HP: Writing – original draft, Writing – review & editing. YZ: Writing – original draft, Writing – review & editing. CF: Data curation, Writing – original draft. YC: Data curation, Writing – original draft. LH: Writing – original draft. XZ: Project administration, Writing – review & editing. XL: Visualization, Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This research study was supported by grants from Startup Fund for Scientific Research, Fujian Medical University, China (2019QH1181), and Science and Technology of Fujian Provincial Health Commission (2020QN01010151).

## Acknowledgments

The authors are grateful for the support provided by Chao Wang and Lanzai Liu.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1320020/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The age distribution histogram in the database.

SUPPLEMENTARY FIGURE 2
The distribution of age (box plot) across different tumor types. Median value (white text) was present in the plot. no polyps (NP) (n = 7122), non adenomatous polyp (NAP) (n = 7375), conventional adenomas (CAs) (n = 20263), serrated polyps (SPs) (n = 904, colorectal cancer (CRC) (n = 1044).

SUPPLEMENTARY FIGURE 3
Bar chart of categorical variables in the database.

## References

1. Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, et al. Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA Cancer J Clin* (2021) 71:209–49. doi: 10.3322/caac.21660.

2. Zheng R, Zhang S, Zeng H, Wang S, Sun K, Chen R, et al. Cancer incidence and mortality in China, 2016. *J Natl Cancer Center* (2022) 2:1–9. doi: 10.1016/j.jncc.2022.02.002.

3. Sánchez Gómez CA, Tejido Sandoval C, de Vicente Bielza N, Pin Vieito N, González A, Almazán R, et al. Surgical complications in a population-based colorectal cancer screening program: Incidence and associated factors. *Gastroenterol Hepatol* (2022) 45:660–7. doi: 10.1016/j.gastre.2022.03.001.

4. Lee KJ, Kim JH, Kim SI, Jang JH, Lee HH, Hong SN, et al. Clinical significance of colonoscopic examination in patients with early stage of gastric neoplasm undergoing endoscopic submucosal dissection. *Scand J Gastroenterol* (2011) 46:1349–54. doi: 10.3109/00365521.2011.613948.

5. Park W, Lee H, Kim EH, Yoon JY, Park JC, Shin SK, et al. Metabolic syndrome is an independent risk factor for synchronous colorectal neoplasm in patients with gastric neoplasm. *J Gastroenterol Hepatol* (2012) 27:1490–7. doi: 10.1111/j.1440-1746.2012.07128.x.

6. Kim SY, Jung SW, Hyun JJ, Koo JS, Choung RS, Yim HJ, et al. Is colonoscopic screening necessary for patients with gastric adenoma or cancer? *Dig Dis Sci* (2013) 58:3263–9. doi: 10.1007/s10620-013-2824-5.

7. Lee SS, Jung WT, Kim CY, Ha CY, Min HJ, Kim HJ, et al. The synchronous prevalence of colorectal neoplasms in patients with stomach cancer. *J Korean Soc Coloproctol* (2011) 27:246–51. doi: 10.3393/jksc.2011.27.5.246.

8. Yoo HM, Gweon TG, Seo HS, Shim JH, Oh SI, Choi MG, et al. Role of preoperative colonoscopy in patients with gastric cancer: a case control study of the prevalence of coexisting colorectal neoplasms. *Ann Surg Oncol* (2013) 20:1614–22. doi: 10.1245/s10434-012-2737-0.

9. Sonnenberg A. Review article: historic changes of Helicobacter pylori-associated diseases. *Aliment Pharmacol Ther* (2013) 38:329–42. doi: 10.1111/apt.2013.38.issue-4.

10. Cimmino DG, Mella JM, Luna P, González R, Pereyra L, Fischer C, et al. Risk of colorectal polyps in patients with sporadic gastric polyps: A case-control study. *World J Gastrointest Endosc* (2013) 5:240–5. doi: 10.4253/wjge.v5.i5.240.

11. Teichmann J, Weickert U, Riemann JF. Gastric fundic gland polyps and colonic polyps - is there a link, really? *Eur J Med Res* (2008) 13:192–5. doi: 10.1186/1750-1172-3-13

12. Yang F, Xu YL, Zhu RF. Helicobacter pylori infection and the risk of colorectal carcinoma: A systematic review and meta-analysis. *Minerva Med* (2019) 110:464–70. doi: 10.23736/S0026-4806.19.05942-1.

13. Sonnenberg A, Turner KO, Genta RM. Increased risk for colon polyps in patients with reflux disease. *Dig Dis Sci* (2018) 63:228–33. doi: 10.1007/s10620-017-4841-2.

14. Yuan T, Zhang S, He S, Ma Y, Chen J, Gu J. Bacterial lipopolysaccharide related genes signature as potential biomarker for prognosis and immune treatment in gastric cancer. *Sci Rep* (2023) 13:15916. doi: 10.1038/s41598-023-43223-6.

15. Sahoo DK, Borcherding DC, Chandra L, Jergens AE, Atherly T, Bourgois-Mochel A, et al. Differential transcriptomic profiles following stimulation with lipopolysaccharide in intestinal organoids from dogs with inflammatory bowel disease and intestinal mast cell tumor. *Cancers* (2022) 14:3525. doi: 10.3390/cancers14143525.

16. Yao X, Smolka AJ. Gastric parietal cell physiology and helicobacter pylori-induced disease. *Gastroenterology* (2019) 156:2158–73. doi: 10.1053/j.gastro.2019.02.036.

17. Thorburn CM, Friedman GD, Dickinson CJ, Vogelman JH, Orentreich N, Parsonnet J. Gastrin and colorectal cancer: A prospective study. *Gastroenterology* (1998) 115:275–80. doi: 10.1016/S0016-5085(98)70193-3.

18. Islam MM, Poly TN, Walther BA, Dubey NK, Anggraini Ningrum DN, Shabbir SA, et al. Adverse outcomes of long-term use of proton pump inhibitors: A systematic review and meta-analysis. *Eur J Gastroenterol Hepatol* (2018) 30:1395–405. doi: 10.1097/MEG.0000000000001198.

19. Hwang SM, Kim BW, Chae HS, Lee BI, Choi H, Ji JS, et al. Gastric fundic gland polyps and their relationship to colorectal neoplasia in Koreans: a 16-year retrospective study. *Korean J Gastroenterol* (2011) 58:20–4. doi: 10.4166/kjg.2011.58.1.20.

20. Li W, Zhang L, Jing Y, Yang Y, Wang Y. The potential value of gastric histopathology for predicting colorectal adenomatous polyps among the chinese population: A retrospective cross-sectional study. *Front Oncol* (2022) 12:889417. doi: 10.3389/fonc.2022.889417.

21. Dixon MF, Genta RM, Yardley JH, Correa P. Classification and grading of gastritis. The updated Sydney System. International Workshop on the Histopathology of Gastritis, Houston 1994. *Am J Surg Pathol* (1996) 20:1161–81. doi: 10.1097/00000478-199610000-00001.

22. Li ZS, Li Q. The latest 2010 WHO classification of tumors of digestive system. *Zhonghua Bing Li Xue Za Zhi* (2011) 40:351–4.

23. Ansari S, Yamaoka Y. Current understanding and management of Helicobacter pylori infection: an updated appraisal. *F1000Research* (2018) 7:F1000. doi: 10.12688/f1000research.

24. Peleteiro B, Bastos A, Ferro A, Lunet N. Prevalence of Helicobacter pylori infection worldwide: a systematic review of studies with national coverage. *Dig Dis Sci* (2014) 59:1698–709. doi: 10.1007/s10620-014-3063-0.

25. Nolen LD, Vindigni SM, Parsonnet J. Symposium leaders. Combating gastric cancer in alaska native people: an expert and community symposium. *Gastroenterology* (2020) 158:1197–201. doi: 10.1053/j.gastro.2019.11.299.

26. Eom BW, Lee HJ, Yoo MW, Cho JJ, Kim WH, Yang HK, et al. Synchronous and metachronous cancers in patients with gastric cancer. *J Surg Oncol* (2008) 98:106–10. doi: 10.1002/jso.21027.

27. Lee JH, Bae JS, Ryu KW, Lee JS, Park SR, Kim CG, et al. Gastric cancer patients at high-risk of having synchronous cancer. *World J Gastroenterol* (2006) 12:2588–92. doi: 10.3748/wjg.v12.i16.2588.

28. Kato T, Suzuki K, Muto Y, Sasaki J, Tsujinaka S, Kawamura YJ, et al. Multiple primary Malignancies involving primary sporadic colorectal cancer in Japan: incidence of gastric cancer with colorectal cancer patients may be higher than previously recognized. *World J Surg Oncol* (2015) 13:23. doi: 10.1186/s12957-014-0432-2.

29. Ochiai K, Kawai K, Nozawa H, Sasaki K, Kaneko M, Murono K, et al. Prognostic impact and clinicopathological features of multiple colorectal cancers and extracolorectal Malignancies: A nationwide retrospective study. *Digestion* (2021) 102:911–20. doi: 10.1159/000517271.

30. Park DI, Park SH, Yoo TW, Kim HS, Yang SK, Byeon JS, et al. The prevalence of colorectal neoplasia in patients with gastric cancer: a Korean Association for the Study of Intestinal Disease (KASID) Study. *J Clin Gastroenterol* (2010) 44:102–5. doi: 10.1097/MCG.0b013e3181a15849.

31. Jian F, Yanhong J, Limeng W, Guoping N, Yiqing T, Hao L, et al. TIMP2 is associated with prognosis and immune infiltrates of gastric and colon cancer. *Int Immunopharmacol* (2022) 110:109008. doi: 10.1016/j.intimp.2022.109008.

32. Pan JH, Zhou H, Cooper L, Huang JL, Zhu SB, Zhao XX, et al. LAYN is a prognostic biomarker and correlated with immune infiltrates in gastric and colon cancers. *Front Immunol* (2019) 10:6. doi: 10.3389/fimmu.2019.00006.

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 18 April 2024
DOI 10.3389/fimmu.2024.1352712



OPEN ACCESS

EDITED BY
Murugaiyan Gopal,
Harvard Medical School, United States

REVIEWED BY
Andrea Glotta,
Ospedale Regionale di Lugano, Switzerland
Marcos Edgar Herkenhoff,
University of São Paulo, Brazil
Zhipeng Zheng,
Zhejiang Chinese Medical University, China

*CORRESPONDENCE
Zhen Ding
13865650996@163.com

RECEIVED 08 December 2023
ACCEPTED 08 April 2024
PUBLISHED 18 April 2024

CITATION
Zhu S, Lin Y and Ding Z (2024) Exploring inflammatory bowel disease therapy targets through druggability genes: a Mendelian randomization study. *Front. Immunol.* 15:1352712. doi: 10.3389/fimmu.2024.1352712



# Exploring inflammatory bowel disease therapy targets through druggability genes: a Mendelian randomization study

Shuangjing Zhu, Yunzhi Lin and Zhen Ding*

Department of Hepatobiliary Surgery, Chaohu Hospital of Anhui Medical University, Hefei, China

**Background:** Inflammatory bowel disease is an incurable group of recurrent inflammatory diseases of the intestine. Mendelian randomization has been utilized in the development of drugs for disease treatment, including the therapeutic targets for IBD that are identified through drug-targeted MR.

**Methods:** Two-sample MR was employed to explore the cause-and-effect relationship between multiple genes and IBD and its subtypes ulcerative colitis and Crohn's disease, and replication MR was utilized to validate this causality. Summary data-based Mendelian randomization analysis was performed to enhance the robustness of the outcomes, while Bayesian co-localization provided strong evidential support. Finally, the value of potential therapeutic target applications was determined by using the estimation of druggability.

**Result:** With our investigation, we identified target genes associated with the risk of IBD and its subtypes UC and CD. These include the genes GPBAR1, IL1RL1, PRKCB, and PNMT, which are associated with IBD risk, IL1RL1, with a protective effect against CD risk, and GPX1, GPBAR1, and PNMT, which are involved in UC risk.

**Conclusion:** In a word, this study identified several potential therapeutic targets associated with the risk of IBD and its subtypes, offering new insights into the development of therapeutic agents for IBD.

KEYWORDS

**inflammatory bowel disease, ulcerative colitis, Crohn's disease, Mendelian randomization, therapeutic target**

## 1 Introduction

Inflammatory bowel disease (IBD) is a chronic inflammatory disease of the intestine represented by Crohn's disease (CD) and ulcerative colitis (UC), which seriously affect the quality of life and health of millions of people worldwide (1). As newly industrialized countries experience economic growth and lifestyle changes, factors such as diet,

environmental exposures, and genetic inheritance may be closely linked to the risk of developing IBD (2). By 2025, it is predicted that newly industrialized countries may have a higher number of people with IBD compared to the Western world. This trend has elevated IBD to an important global public health issue (3).

IBD symptoms encompass intestinal manifestations such as diarrhea, blood in the stool, and abdominal pain, greatly impacting patients' daily life and overall quality of life (4). Furthermore, IBD can lead to complications including malnutrition, cardiovascular diseases, liver and biliary tract disorders, and other intestinal manifestations, posing significant risks to human health (5). Currently, IBD is mostly treated through drug regimens, with commonly used therapeutic agents including anti-inflammatory drugs, immunosuppressants, and biologics (6). Yet, these drugs have possible side effects and are not consistently effective in a subset of patients (7). Firstly, for instance, the anti-inflammatory drug 5-aminosalicylic acid is a preferred treatment for mild-to-moderate UC disease, but an adverse gastrointestinal reaction such as nausea and vomiting can occur with them (7). Second, the biologic agent infliximab (IFLX) treats IBD by hastening the death of pro-inflammatory cells, while prolonged use of IFLX may lead to the possibility of infection, allergy, and even malignancy (8). Third, the immunomodulator azathioprine increases the potential for hepatotoxicity and pancreatitis (9). Hence, the quest for new therapeutic strategies and targets is becoming an ongoing important direction in IBD research.

Genome-wide association studies (GWAS) uncover disease-associated single nucleotide polymorphisms (SNPs) that enable scientists to pinpoint associations between genetic variants and specific diseases, which can be used to aid in the identification and validation of drug targets (10). Nevertheless, disease-causing genes cannot be fully characterized by GWAS analysis alone. Mendelian randomization (MR) is a genetic statistical method for assessing the causality among exposures and outcomes, which minimizes the interference of confounders and reverse cause and effect since genetic variants follow the principle of Mendelian random assignment at the time of conception and are independent of social background and lifestyle (11). Gene expression levels are influenced by genetic variation (eQTL), and cis-expression quantitative trait loci (cis-eQTL) serve as proxies for the gene expansion levels (12). Using drug target MR methods to analyze independent disease abstract GWAS summary datasets and gene cis-eQTL to identify relevant genes causing complex traits (13).

This research aims to contribute to the development of therapeutic targets for IBD by exploring causative genes associated with both the UC and CD subtypes of the disease.

# 2 Materials and methods

## 2.1 Research methods

The study methods were compliant with the STROBE-MR checklist (14), further details can be found in Additional File 1. The study design is depicted in Figure 1. In this study, our first step involved utilizing MR methods to evaluate the causal relationship between druggable genes in blood and IBD, encompassing both UC and CD subtypes (15, 16). Additionally, we employed replicated MR, Summary data-based MR analysis (SMR) approaches, along with the heterogeneity in dependent instruments (HEIDI) test, to enhance the robustness of our MR results (12). Subsequently, we conducted a Bayesian co-localization analysis to identify common causal SNPs shared between the genes and the risk of IBD (17). Lastly, we estimated the druggability of the identified genes to investigate their potential as effective therapeutic targets for IBD.

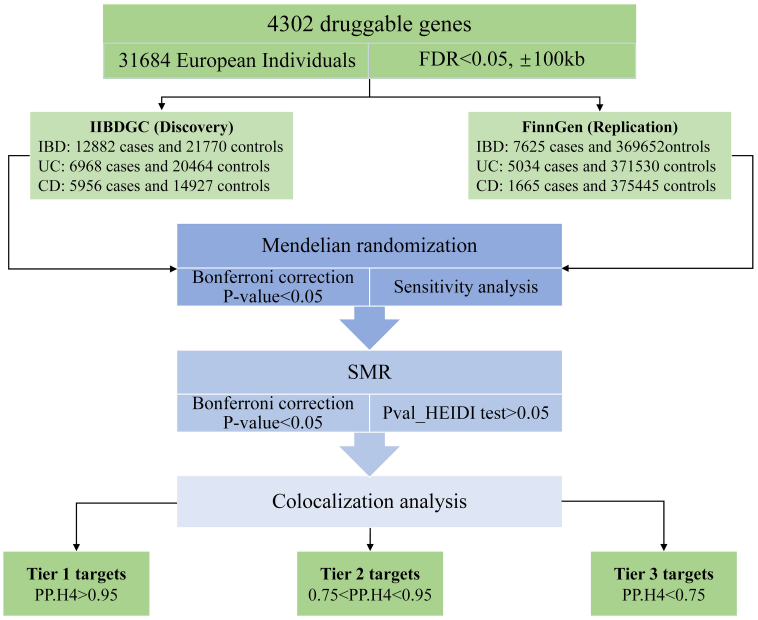



FIGURE 1
Research design.

## 2.2 Data sources

Exposure data were extracted from the eQTLGen consortium (https://eqtlgen.org/) and eQTL meta-analysis of peripheral blood samples from 31,684 individuals, with information about 16,987 genes (16). The druggable genes are from a previous study containing a total of 4,479 druggable genes. After removing the null values and genes not located on the autosomes, the number of actionable genes is 4302, We utilized the cis-eQTL for these 4302 genes as exposure (15). IBD GWAS abstract data from the International Inflammatory Bowel Disease Genetics Consortium (IIBDGC, https://www.ibdgenetics.org/), including 12,882 cases of IBD, 6,968 cases of UC, and 5,956 patients of European ancestry with CD (Table 1) (18). We also get GWAS summarized data for IBD, UC, and CD from FinnGen's version R9 database for replication MR analysis (https://r9.finngen.fi/), including 7625 IBD,5034 UC, and 1665 CD patients (Table 1). FinnGen Database is a project containing a large number of bio-samples and relevant diagnostic techniques to gather data from national health registries of the Finnish population since 1969. The program aims to delve deeper into the relationship between the genome and health as well as to provide valuable information to the general public health system to promote medical research into the etiology of diseases in the population (19).

## 2.3 Selection of cis-eQTLs associated with druggable genes

To obtain cis-eQTL data for drug target genes and allele frequency information, only statistically significant cis-eQTLs with an FDR <0.05 were included (13). To generate the genetic instrumental variables used to proxy the 4302 druggable targets, we performed a series of manipulations. First, we chose cis-eQTL within ±100kb from the genomic transcriptional start site, based on the 1000G Genome Europe reference panel setting r2 < 0. 1 to avoid the effect of chain imbalance (20). Second, we carried out a scan in the PhenoScanner database (https://www.phenoscanner.medschl.cam.ac.uk) to delete SNPs that were linked to confounders and IBD, to prevent the interference of confounders (21). Third, to guard against the biasing effects of weak instrumental variables, the F-value statistic was calculated by the formula $\beta^2/SE^2$, and when the F-value was less than 10 it would be excluded (22). Finally, the palindromic SNPs might not affect gene expression and protein functions, we will remove those palindromic SNPs with allelic frequencies.

## 2.4 MR analysis

For our MR study, we conducted two-sample MR analyses using cis-eQTL and outcomes. When screened exposures have only one SNP, Wald ratios were applied as the principal analysis method. In cases where there were more than two SNPs, inverse variance weighted (IVW) models were estimated to estimate the effect of each exposure on the outcome, with MR-Egger, MR-RAPS, Maximum likelihood, and Weighted median methods as additional methods (23–25). The genes were only incorporated into the next step of the analysis when four of the five methods were in alignment with each other in the same direction. We deployed Cochran's Q and MR-Egger intercept tests to examine possible heterogeneity and horizontal pleiotropy of the filtered instrumental variables (26, 27). MR-Steiger was enlisted to assess the potential reverse causality of exposure on outcome (28). When the gene was significant in both the primary MR and replication MR analyses, we proceeded to SMR analysis to further validate the MR results. SMR is the process of using GWAS-level summary data and eQTL to be used for investigating whether there is any causal relationship between one or more genes and specific phenotypes, using HEIDI to test the results (12). SMR software (https://yanglab.westlake.edu.cn/software/smr/) by using the SMR analysis and HEIDI assay. MR analysis using R software TwoSampleMR package (0.5.7) for analysis. We utilize the Bonferroni correction for multiple checks.

## 2.5 Co-localization analysis

We concluded with a Bayesian co-localization analysis of genes that were multiply corrected by MR and SMR. Co-localization analysis combines information from multiple SNPs or other genetic variants to determine whether genes and diseases exist at similar locations in the genome or interact with each other. We use default *a priori* probabilities p1 = 1E-4, p2 = 1E-4, and p12 = 1E-5, representing the likelihood that an SNP in a selected region is associated with significant gene expression, IBD risk, and both. The posterior probabilities were verified against five hypotheses: pp.H0, SNPs were not associated with any of the traits; PP.H1, SNPs were correlated with gene expression but

TABLE 1 IBD and its subtypes UC and CD data source details.

| Disease | Cases | Controls | Population | No SNPs |
|---|---|---|---|---|
| IBD(IIBDGC) | 12882 | 21770 | European | 12716084 |
| UC(IIBDGC) | 6968 | 20464 | European | 12255197 |
| CD(IIBDGC) | 5956 | 14927 | European | 12276506 |
| IBD(FinnGen) | 7625 | 369652 | European | 20170236 |
| UC(FinnGen) | 5034 | 371530 | European | 20170227 |
| CD(FinnGen) | 1665 | 375445 | European | 20170234 |

IBD, Inflammatory bowel disease; IIBDGC, International Inflammatory Bowel Disease Genetics Consortium; UC, Ulcerative colitis; CD, Crohn's disease.

not with IBD risk; PP.H2 were associated with the risk of developing IBD but not with gene expression; PP.H3 were related to both gene expression and IBD risk, but with different causal variants; and PP.H4, were related to IBD risk and gene expression, specifically the same genetic causal variant (17). We set the significance threshold for PP.H4 at 0.95 owing to the limited efficacy of the colocalization assessment. Bayesian co localization was analyzed using the software package coloc (version 5.0.1).

## 2.6 Druggability evaluation

DrugBank (https://go.drugbank.com/) brings together numerous data on the interactions between drugs and genes (29), integrating information from multiple public databases, including drug target prediction, mechanisms of action, and clinical applications to offer vital data and functionality. The potential of identified druggable target genes as therapeutic agents for IBD and its subtypes was further determined by using DrugBank to locate associations between characterized proteins and drugs.

# 3 Results

## 3.1 MR analysis reveals 27 genes associated with IBD, 21 genes associated with UC, and 17 genes associated with CD

In the current study cohort, we identified 49 genes with expressions associated with IBD (P<0.05/2641, Bonferroni corrected). Subsequently, in the replication MR analysis, 31 out of these 49 genes exhibited significance in the MR test (P<0.05, Figure 2). Sensitivity analysis revealed that the genes SLC22A5, RPS6KA2, and SENP7 did not pass the pleiotropy test (P<0.05), and the gene IMPDH2 did not pass the MR-Sterger test (P>0.05). Furthermore, the genes GPR25, JAK2, STAT3, SLC22A4, and NDFIP1 showed potential heterogeneity.

The expression of 36 genes was associated with UC (P<0.05/2641, Bonferroni corrected), and subsequent replication MR analysis demonstrated that 23 genes remained significant (P<0.05, Figure 3). Sensitivity analysis indicated that the gene SLC22A5 did not pass the pleiotropy test (P<0.05), and the gene IMPDH2 did not pass the MR-Sterger test (P>0.05). Additionally, the genes TNFRSF14 and GPBAR1 exhibited heterogeneity.

Thirty-six genes were causally connected to disease CD by expression (P<0.05/2641, Bonferroni correction). Replication MR showed that 19 of these genes passed the MR test again (P<0.05, Figure 4). There was reverse causality for genes DAG1 and SSR2 in the sensitivity analysis (P>0.05) and no pleiotropy (P<0.05). Genes NDFIP1, SLC22A4, THBS3, JAK2, and STAT3 had the presence of heterogeneity. Detailed information on significant gene MR results for IBD and its subtypes UC and CD are shown in Supplementary Tables 1-3.

## 3.2 SMR analysis validated 7 genes linked to IBD, 4 genes with UC, and 8 genes for CD

In the discovery cohort and replication cohort MR analyses, an SMR analysis was conducted for the examined genes. A total of 27 genes for IBD, 21 genes for UC, and 17 genes for CD were included in this analysis. Genes that did not pass the HEIDI test and were not consistently oriented (P<0.05, Figure 5) were removed from the analysis. As a result, 7 genes showed significant associations with IBD in the SMR analysis (P<0.05/27). For UC, 4 genes passed the SMR analysis (P<0.05/21, Figure 6A), and for CD, 8 genes demonstrated significant causal associations (p<0.05/17, Figure 6B). Supplementary Tables 4-6 provide detailed information on the SMR analysis.

## 3.3 Identification of genetic overlaps in IBD, UC, and CD using co-localization analysis

In our study, we performed the co-localization analysis of the screened genes associated with IBD, UC, and CD (Table 2,

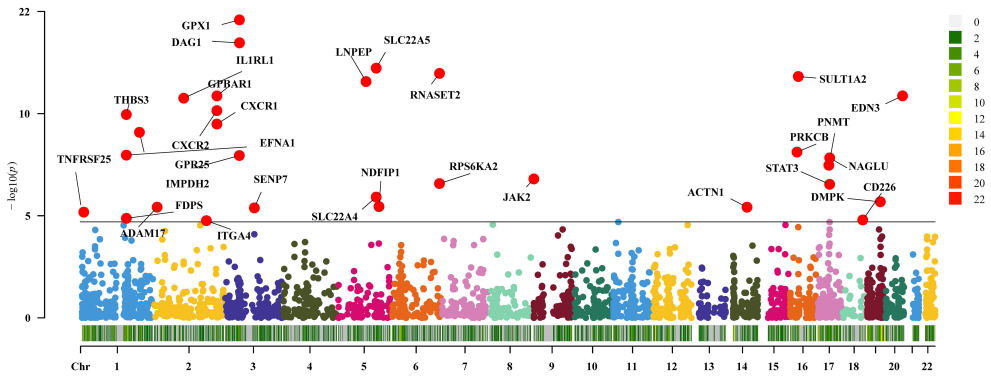



FIGURE 2
Two-sample Mendelian randomization results for druggable genes and inflammatory bowel disease Manhattan plot.

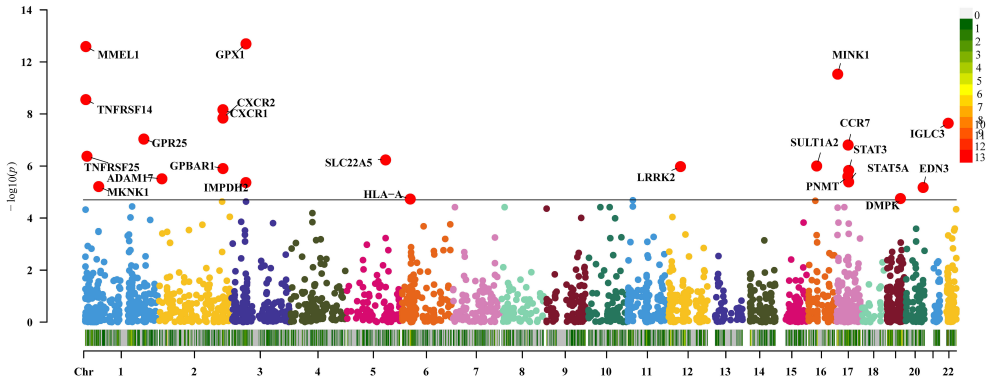



FIGURE 3
Two-sample Mendelian randomization results for druggable genes and ulcerative colitis Manhattan plot.

Figures 7, 8). Our findings revealed that for IBD, genes GPBAR1 (PP.H4 = 0.99), IL1RL1 (PP.H4 = 0.95), PRKCB (PP.H4 = 0.99), and PNMT (PP.H4 = 0.95) exhibited robust evidence of high co-localization support (Figure 7). Furthermore, genes GPX1 (PP.H4 = 0.98), GPBAR1 (PP.H4 = 0.99), and PNMT (PP.H4 = 0.99) demonstrated significant co-localization support with UC (Figure 8). Additionally, the gene IL1RL1 (PP.H4 = 0.98) showed strong co-localization support with CD (Figure 8). In our categorization, genes that passed all tests were considered primary targets (PP.H4 > 0.95), while genes that passed MR and SMR tests but had a PP.H4 less than 0.75 were categorized as tertiary targets. Genes that passed the MR and SMR tests and had a PP.H4 greater than 0.75 but less than 0.95 were classified as secondary targets.

## 3.4 Estimation of druggability

We conducted a comprehensive search of the drug database for several genes identified in this study as potential drug targets (Supplementary Table 7). Our investigation revealed that targeting GPBAR1 is approved for the treatment of primary biliary cirrhosis, bile acid synthesis disorders, and various other diseases. Various drugs targeting PRKCB are associated with antioxidant effects, therapeutic benefits for relapsed glioblastoma multiforme, and preventive measures for vitamin E deficiency. Drugs targeting GPX1 exhibit multiple effects, including antioxidant activity and pain relief. However, no relevant information was found for IL1RL1 and PNMT as potential drug targets. Additionally, we explored secondary and tertiary targets identified in this study and found that drugs targeting FDPS have been approved for the treatment of osteoporosis. Similarly, drugs targeting ITGA4 are utilized for the treatment of multiple sclerosis, UC, and CD.

# 4 Discussion

In this study, we conducted a comprehensive investigation of 4302 genes to determine their association with the risk of IBD, UC, and CD. Using various MR methods (including Wald ratio/IVW, MR-Egger, Weighted median, Maximum likelihood, and MR-RAPS) and multiple sensitivity analyses (such as Cochran's Q heterogeneity test, MR-Egger intercept pleiotropy test, and MR-Steiger directionality test), we explored target genes associated with the risk of IBD and its subtypes. To validate our results, we employed replicated MR, SMR, HEIDI tests, and multiple co-localization tests.

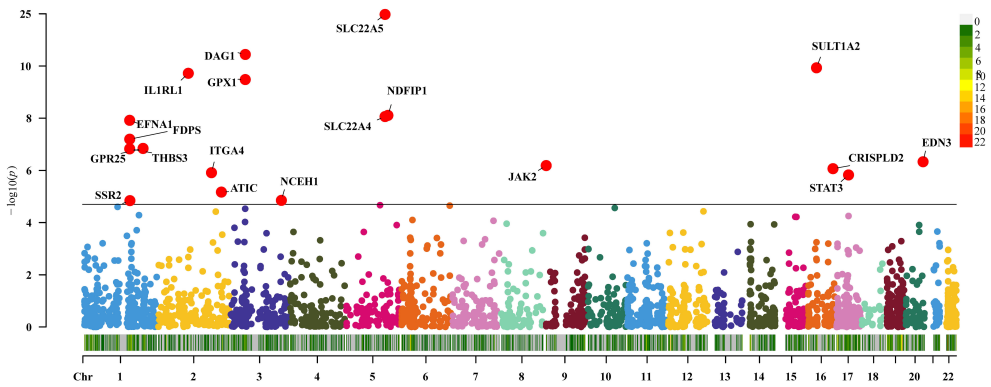



FIGURE 4
Two-sample Mendelian randomization results for druggable genes and Crohn's disease Manhattan plot.

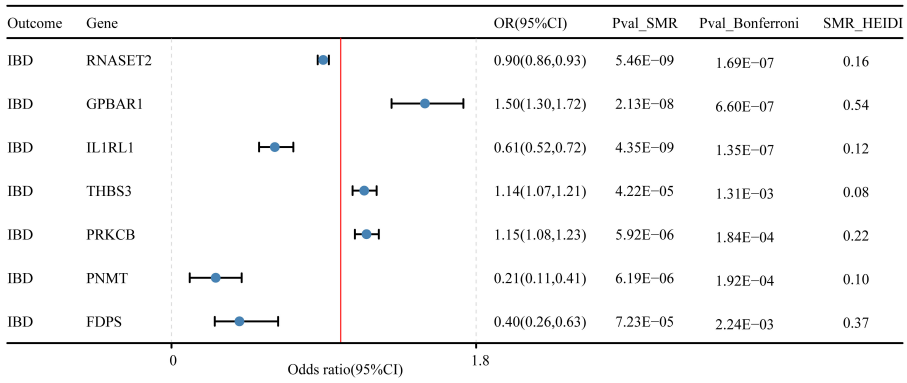



FIGURE 5
Druggable genes and inflammatory bowel disease SMR results.

Our comprehensive analyses revealed several significant findings. We found that the expression of the gene GPBAR1 was associated with an increased risk of IBD and UC. Conversely, the expression of PNMT was negatively associated with the risk of IBD and UC. Furthermore, higher levels of the genetically predicted gene IL1RL1 were linked to a reduced risk of IBD and CD. These findings provide valuable insights into the genetic factors influencing the risk of IBD and its subtypes, highlighting potential targets for further research and therapeutic interventions.

RNASET2 is the gene that encodes nuclease T2. It plays a key role in the intracellular context and its function involves processes such as RNA degradation and apoptosis. Prior work has identified RNASET2 as a susceptibility gene for IBD (30) and decreased RNASET2 expression has activating effects on pro-inflammatory cells, with an association with aggressive CD inflammation (31). The results of this study provide relatively strong evidence that RNASET2 levels may serve as an inflammatory biomarker for the prediction of progression in a novel disease.

The GPX1 gene is located at position 3p21.3 on the human genome and consists of five exons and four introns. The transcription product of this gene is a peptide containing 197 amino acid remnants that are translated into glutathione peroxidase 1 (GPX1) protein (32). GPX1 is mainly found in the cytoplasm of cells, where catalyzing the reaction between glutathione and substrates like hydrogen peroxide,

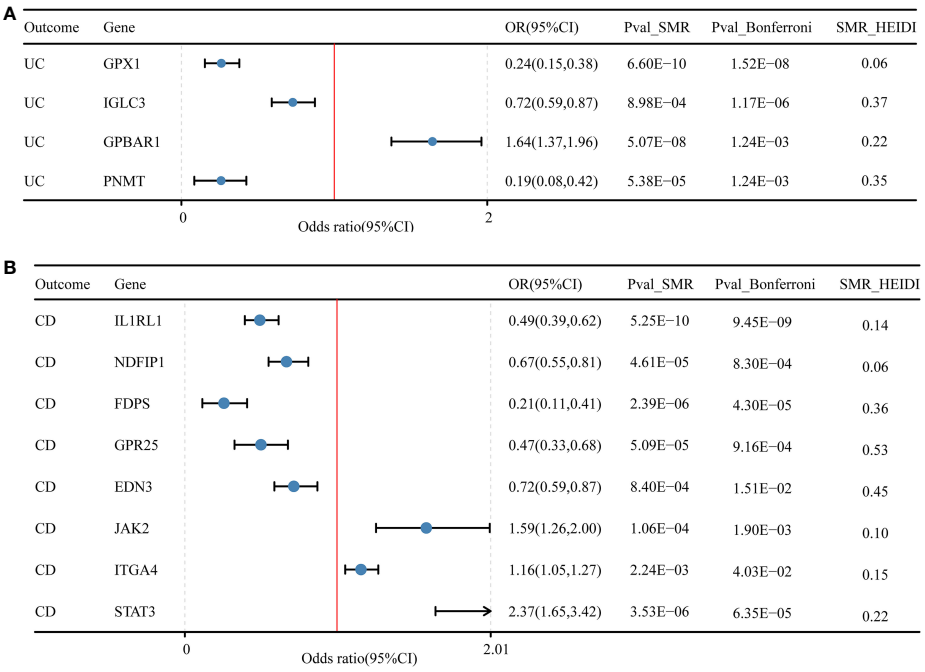



FIGURE 6
Druggable genes and SMR results for ulcerative colitis and Crohn's disease. **(A)** Druggable genes and ulcerative colitis SMR results. **(B)** Druggable genes and Crohn's disease SMR results.

10.3389/fimmu.2024.1352712

TABLE 2 Gene and outcome co-localization results.

| Disease | Gene | PP.H0 | PP.H1 | PP.H2 | PP.H3 | PP.H4 | Grade |
|---|---|---|---|---|---|---|---|
| IBD | GPBAR1 | 0 | 0 | 0 | 0.01 | 0.99 | Tier 1 target |
| | IL1RL1 | 0 | 0 | 0 | 0.05 | 0.95 | Tier 1 target |
| | PRKCB | 0 | 0 | 0 | 0.01 | 0.99 | Tier 1 target |
| | PNMT | 0 | 0 | 0 | 0.05 | 0.95 | Tier 1 target |
| | FDPS | 0 | 0.01 | 0 | 0.14 | 0.85 | Tier 2 target |
| | RNASET2 | 0 | 0 | 0 | 0.22 | 0.78 | Tier 2 target |
| | THBS3 | 0 | 0.14 | 0 | 0.47 | 0.52 | Tier 3 target |
| UC | GPX1 | 0 | 0 | 0 | 0.02 | 0.98 | Tier 1 target |
| | GPBAR1 | 0 | 0 | 0 | 0.01 | 0.99 | Tier 1 target |
| | PNMT | 0 | 0 | 0 | 0.01 | 0.99 | Tier 1 target |
| | IGLC3 | 0 | 0.23 | 0 | 0 | 0.77 | Tier 2 target |
| CD | IL1RL1 | 0 | 0 | 0 | 0.02 | 0.98 | Tier 1 target |
| | GPR25 | 0 | 0 | 0.01 | 0.05 | 0.94 | Tier 2 target |
| | FDPS | 0 | 0 | 0 | 0.08 | 0.92 | Tier 2 target |
| | STAT3 | 0 | 0 | 0 | 0.17 | 0.83 | Tier 2 target |
| | END3 | 0 | 0.22 | 0 | 0 | 0.78 | Tier 2 target |
| | ITGA4 | 0 | 0.45 | 0 | 0.03 | 0.52 | Tier 3 target |
| | NDFIP1 | 0 | 0 | 0 | 0.86 | 0.14 | Tier 3 target |
| | JAK2 | 0 | 0 | 0 | 0.99 | 0 | Tier 3 target |

IBD, Inflammatory bowel disease; UC, Ulcerative colitis; CD, Crohn's disease.
PP.H0–PP.H4 represents the posterior probabilities of different hypotheses.
PP.H4 > 0.95 represents a strong colocalization between gene expression and risk of IBD, UC, and CD.

decreases cellular damage by oxidative stress. Specifically, GPX1 uses reduced glutathione (GSH) to convert hydrogen peroxide to water and oxidizes GSH to oxidized glutathione (GSSG), which in turn generates GSH again via other reducing enzymes, maintaining the relative balance of GSSH and GSSG (33). Zhou et al. demonstrated the connection between endoplasmic reticulum stress-related genes and UC, and CD through a multi-omics approach and discovered that GPX1 expression lowered the risk of UC and CD (34). Oxidative stress leads to the inflammatory response exacerbated by oxidative damage to intracellular DNA, lipids, and proteins, which then triggers an inflammatory response in UC. GPX1 is known to be a toxicant through deleterious agents maintains redox balance, and can directly reverse the complex lipid peroxides in cells and tissues. It can also directly reduce complex lipid peroxides and minimize the damage of oxidative stress on cells and tissues. This may be the mechanism by which GPX1 reduces the risk of UC. Additionally, our study revealed that the gene GPX1 was relevant to UC, but did not find an association between GPX1 and CD. A study of 436 CD, 367 UC, and 434 controls showed that allele A in the gene GPX1 (rs1050450) was significantly observed to be associated with UC in a recessive model, and is a good candidate for a biological marker for the management of treatment of UC in the disease (35). An additional study using polymorphism-polymerase chain reaction in peripheral blood leukocytes from 1500 UC cases and 1500 healthy controls demonstrated that a genetic polymorphism in the GPX1 gene of 594TT is a danger factor for UC (36). They are consistent with the study we conducted, and future studies could explore the relationship of this gene with IBD as well as subtypes of CD to search for the more likely therapeutic targets.

STAT3 is an activator of signal transduction and transcription, playing a vital role within cells. Upon activation, STAT3 can enter the nucleus and regulate the transcription of several genes, thereby participating in the regulation of cell proliferation, apoptosis, inflammatory response, and other biological processes (37). According to a meta-analysis, the presence of the STAT3 rs744166 gene polymorphism may elevate the risk of developing CD, especially among Caucasians (38). A prior case-control study involving 232 CD patients and 272 controls indicated that the rs744166 and rs4796793 polymorphisms in the STAT3 gene may be linked to the onset of CD and are anticipated to serve as predictors of CD in the Chinese Han population (39). On one hand, STAT3 can regulate the activity of immune cells, such as macrophages and T cells, among others. It promotes the activation of immune cells and the release of inflammatory mediators while inhibiting the regulatory function of immune cells. This imbalance in the immune system leads to increased intestinal inflammatory response (40). On the other hand, IL-23 activates the STAT3 pathway, enhances the Th17 cell program, and contributes to the initiation and progression of

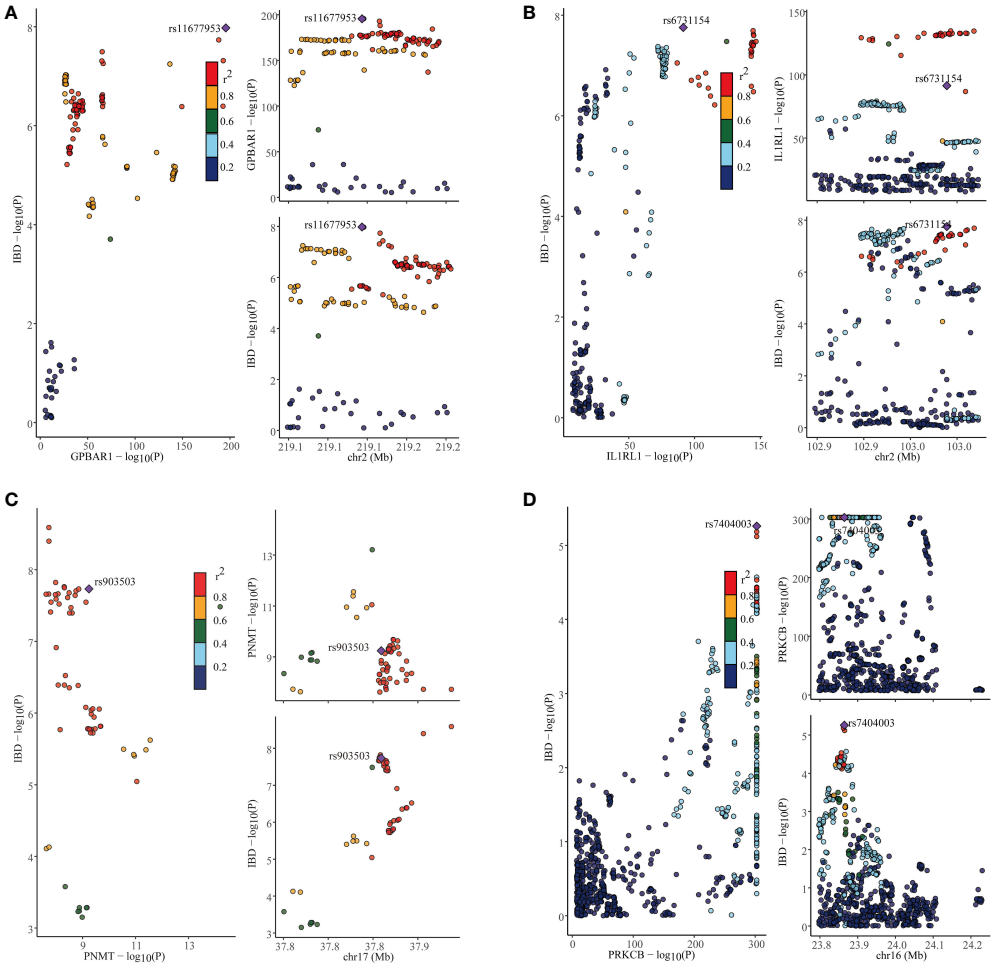



FIGURE 7
Co-localization results for inflammatory bowel disease. **(A)** Co-localization results of eQTL of the GPBAR1 gene and inflammatory bowel disease. **(B)** Co-localization results of eQTL of the IL1RL1 gene and inflammatory bowel disease. **(C)** Co-localization results of eQTL of the PNMT gene and inflammatory bowel disease. **(D)** Co-localization results of eQTL of the PRKCB gene and inflammatory bowel disease.

pathological reactions (41). These findings align with the outcomes of the current study, suggesting that the STAT3 gene could serve as a potential therapeutic target for CD, warranting further clinical trials.

JAK2 is a tyrosine kinase that is involved in a variety of cytosolic signaling pathways, including apoptosis, differentiation, survival, and immune response (42). Drugs targeting JAK2 are considered for the treatment of immunological diseases such as UC, rheumatoid arthritis, and myelofibrosis. The present study found that JAK2 is a potential target for CD therapy. JAK2 signaling pathway can regulate the proliferation, differentiation, and activation of immune cells (like T-cells, B-cells, macrophages, etc.), and thus affects the normal function of the Immune system (43). Besides, alteration of intestinal barrier function would probably be one of the mechanisms by which JAK2 is involved in the pathogenesis of CD. One report of 464 CD patients, 292 UC patients, and 508 healthy controls in Germany revealed that patients carrying the C risk allele of the JAK2 rs10758669 gene polymorphism were at a higher frequency of increased risk of intestinal permeability (43). Targeted inhibitors of JAK2 have been studied and developed, and these drugs can interfere with the JAK2 modeling pathway and inhibit its aberrant activation, resulting in a reduction in the production of inflammatory mediators, alleviation of the inflammatory response, and amelioration of symptoms and disease progression in CD.

IBD is a complex group of diseases that includes multiple subtypes such as UC and CD. Despite sharing certain pathophysiological features, they differ dramatically in their clinical manifestations, histologic and immunologic features, and gene expression levels (44). Such differences point to the possibility that they may have different genetic mechanisms. Additional risk genes associated with IBD were identified in our survey, however, they are not currently validated by larger numbers of experimental studies. More studies may be needed to probe these genes in the future to prioritize IBD drug development. The strength of this study lies in its comprehensive screening of genes associated with IBD risk using the two-sample MR method, which effectively mitigates confounding bias. Moreover, the utilization of replicated MR, SMR, and co-localization to corroborate the experimental findings significantly bolsters the study's conclusions, enhancing the robustness of the results and minimizing the potential for false

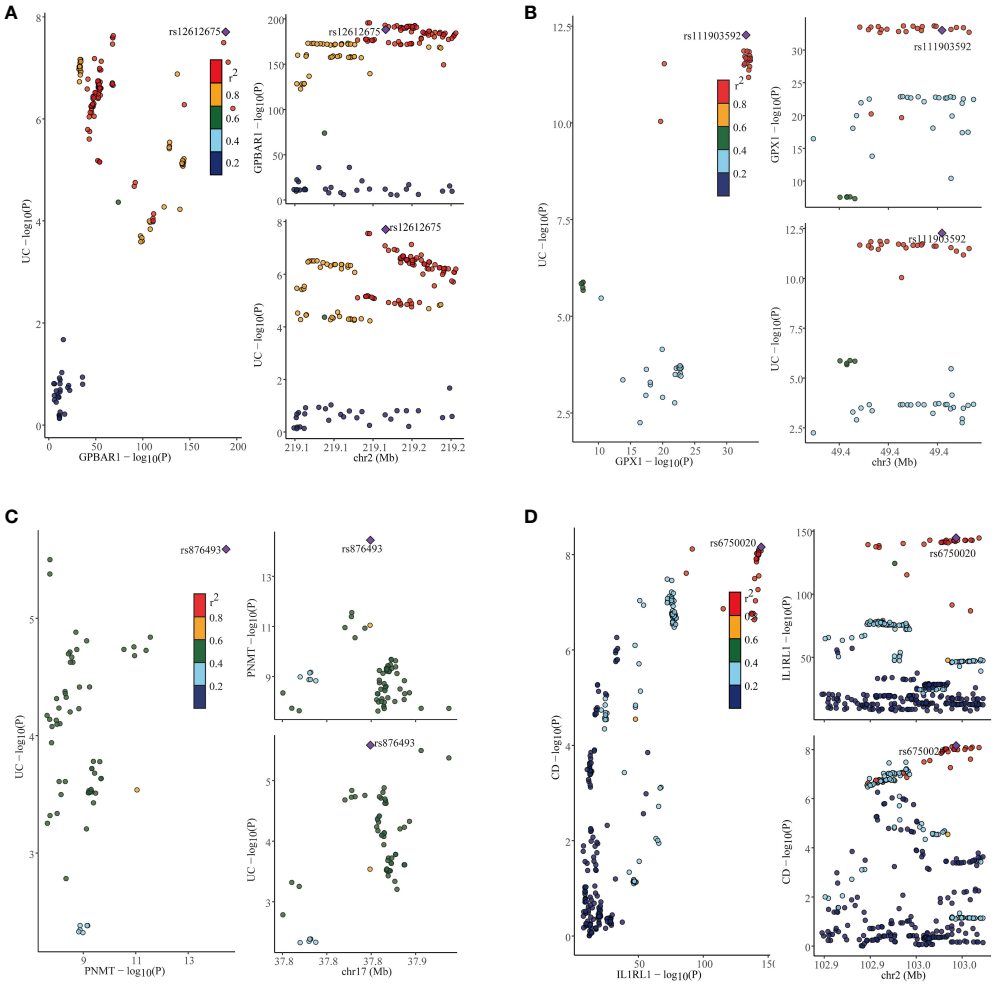



FIGURE 8
Co-localization results for ulcerative colitis and Crohn's disease. **(A)** Co-localization results of eQTL of the GPBAR1 gene and ulcerative colitis. **(B)** Co-localization results of eQTL of the GPX1 gene and ulcerative colitis. **(C)** Co-localization results of eQTL of the PNMT gene and ulcerative colitis. **(D)** Co-localization results of eQTL of the IL1RL1 gene and Crohn's disease.

positives. The evaluation of druggability offers promise for IBD treatment. However, there are several limitations to consider. Firstly, the genetic data were derived from a European population and necessitate further validation for extrapolation to other ethnic groups. Secondly, genetic regulatory mechanisms may exhibit tissue-specific variability, and focusing solely on blood eQTL may not afford a comprehensive understanding of the disease and its therapeutic avenues. Therefore, it is imperative to account for genetic regulatory diversity across multiple tissues and organs to gain a more nuanced comprehension of disease pathogenesis and identify effective treatments. Thirdly, while some genes associated with IBD risk have been experimentally validated, exploration of certain genes and their correlation with IBD risk remains deficient.

## 5 Conclusion

To summarize, our study employed sophisticated methods such as MR, SMR, and co293 localization to identify key genes intricately associated with the risk of IBD and its subtypes, UC and CD. Specifically, our analysis revealed the crucial roles of GPBAR1, IL1RL1, PRKCB, and PNMT genes in IBD pathogenesis while implicating GPX1, GPBAR1, and PNMT genes in UC susceptibility. Additionally, we found that IL1RL1 exhibits a protective effect against CD risk. These groundbreaking findings not only offer promising targets for the development of more effective biomarkers and therapeutic interventions but also deepen our understanding of the underlying molecular mechanisms driving IBD etiology. Nevertheless, further rigorous experimental and clinical investigations are required to validate and substantiate these findings before their translation into clinical practice.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

## Author contributions

SZ: Data curation, Formal analysis, Investigation, Methodology, Resources, Software, Validation, Visualization, Writing – original draft. YL: Conceptualization, Data curation, Formal analysis, Validation, Writing – review & editing. ZD: Conceptualization, Data curation, Investigation, Methodology, Visualization, Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## Acknowledgments

Thanks to the researchers who have contributed to GWAS data.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1352712/full#supplementary-material

## References

1. Collaborators GBDIBD. The global, regional, and national burden of inflammatory bowel disease in 195 countries and territories, 1990-2017: a systematic analysis for the Global Burden of Disease Study 2017. *Lancet Gastroenterol Hepatol.* (2020) 5:17–30. doi: 10.1016/S2468-1253(19)30333-4

2. Kaplan GG. The global burden of IBD: from 2015 to 2025. *Nat Rev Gastroenterol Hepatol.* (2015) 12:720–7. doi: 10.1038/nrgastro.2015.150

3. Buie MJ, Quan J, Windsor JW, Coward S, Hansen TM, King JA, et al. Global hospitalization trends for crohn's disease and ulcerative colitis in the 21st century: A systematic review with temporal analyses. *Clin Gastroenterol Hepatol.* (2023) 21:2211–21. doi: 10.1016/j.cgh.2022.06.030

4. Ungaro R, Mehandru S, Allen PB, Peyrin-Biroulet L, Colombel JF. Ulcerative colitis. *Lancet.* (2017) 389:1756–70. doi: 10.1016/S0140-6736(16)32126-2

5. Rogler G, Singh A, Kavanaugh A, Rubin DT. Extraintestinal manifestations of inflammatory bowel disease: current concepts, treatment, and implications for disease management. *Gastroenterology.* (2021) 161:1118–32. doi: 10.1053/j.gastro.2021.07.042

6. Seyedian SS, Nokhostin F, Malamir MD. A review of the diagnosis, prevention, and treatment methods of inflammatory bowel disease. *J Med Life.* (2019) 12:113–22. doi: 10.25122/jml-2018-0075

7. Murray A, Nguyen TM, Parker CE, Feagan BG, MacDonald JK. Oral 5-aminosalicylic acid for maintenance of remission in ulcerative colitis. *Cochrane Database Syst Rev.* (2020) 8:CD000544. doi: 10.1002/14651858.CD000544.pub5

8. Ferretti F, Cannatelli R, Monico MC, Maconi G, Ardizzone S. An update on current pharmacotherapeutic options for the treatment of ulcerative colitis. *J Clin Med.* (2022) 11(9):2302. doi: 10.3390/jcm11092302

9. Chaparro M, Ordas I, Cabre E, Garcia-Sanchez V, Bastida G, Penalva M, et al. Safety of thiopurine therapy in inflammatory bowel disease: long-term follow-up study of 3931 patients. *Inflammation Bowel Dis.* (2013) 19:1404–10. doi: 10.1097/MIB.0b013e318281f28f

10. Tam V, Patel N, Turcotte M, Bosse Y, Pare G, Meyre D. Benefits and limitations of genome-wide association studies. *Nat Rev Genet.* (2019) 20:467–84. doi: 10.1038/s41576-019-0127-1

11. Burgess S, Small DS, Thompson SG. A review of instrumental variable estimators for Mendelian randomization. *Stat Methods Med Res.* (2017) 26:2333–55. doi: 10.1177/0962280215597579

12. Zhu Z, Zhang F, Hu H, Bakshi A, Robinson MR, Powell JE, et al. Integration of summary data from GWAS and eQTL studies predicts complex trait gene targets. *Nat Genet.* (2016) 48:481–7. doi: 10.1038/ng.3538

13. Chen Y, Xu X, Wang L, Li K, Sun Y, Xiao L, et al. Genetic insights into therapeutic targets for aortic aneurysms: A Mendelian randomization study. *EBioMedicine.* (2022) 83:104199. doi: 10.1016/j.ebiom.2022.104199

14. Skrivankova VW, Richmond RC, Woolf BAR, Yarmolinsky J, Davies NM, Swanson SA, et al. Strengthening the reporting of observational studies in epidemiology using mendelian randomization: The strobe-Mr statement. *JAMA* (2021) 326(16):1614–21. doi: 10.1001/jama.2021.18236

15. Finan C, Gaulton A, Kruger FA, Lumbers RT, Shah T, Engmann J, et al. The druggable genome and support for target identification and validation in drug development. *Sci Transl Med.* (2017) 9(383):eaag1166. doi: 10.1126/scitranslmed.aag1166

16. Vosa U, Claringbould A, Westra HJ, Bonder MJ, Deelen P, Zeng B, 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) 53:1300–10. doi: 10.1038/s41588-021-00913-z

17. Giambartolomei C, Vukcevic D, SChadt EE, Franke L, Hingorani AD, Wallace C, et al. Bayesian test for colocalisation between pairs of genetic association studies using summary statistics. *PloS Genet.* (2014) 10:e1004383. doi: 10.1371/journal.pgen.1004383

18. Liu JZ, van Sommeren S, Huang H, Ng SC, Alberts R, Takahashi A, et al. Association analyses identify 38 susceptibility loci for inflammatory bowel disease and highlight shared genetic risk across populations. *Nat Genet.* (2015) 47:979–86. doi: 10.1038/ng.3359

19. Peltonen L, Jalanko A, Varilo T. Molecular genetics of the Finnish disease heritage. *Hum Mol Genet.* (1999) 8:1913–23. doi: 10.1093/hmg/8.10.1913

20. Genomes Project C, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. *Nature.* (2015) 526:68–74. doi: 10.1038/nature15393

21. Kamat MA, Blackshaw JA, Young R, Surendran P, Burgess S, Danesh J, et al. PhenoScanner V2: an expanded tool for searching human genotype-phenotype associations. *Bioinformatics.* (2019) 35:4851–3. doi: 10.1093/bioinformatics/btz469

22. Burgess S, Thompson SG, Collaboration CCG. Avoiding bias from weak instruments in Mendelian randomization studies. *Int J Epidemiol.* (2011) 40:755–64. doi: 10.1093/ije/dyr036

23. Bowden J, Davey Smith G, Haycock PC, Burgess S. Consistent estimation in mendelian randomization with some invalid instruments using a weighted median estimator. *Genet Epidemiol.* (2016) 40:304–14. doi: 10.1002/gepi.21965

24. Burgess S, Davey Smith G, Davies NM, Dudbridge F, Gill D, Glymour MM, et al. Guidelines for performing Mendelian randomization investigations: update for summer 2023. *Wellcome Open Res.* (2019) 4:186. doi: 10.12688/wellcomeopenres

25. Zhao Q, Wang J, Hemani G, Bowden J, Small DS. Statistical inference in two-sample summary-data Mendelian randomization using robust adjusted profile score. *Ann Stat.* (2020) 48:1742–69. doi: 10.1214/19-AOS1866

26. Greco MF, Minelli C, Sheehan NA, Thompson JR. Detecting pleiotropy in Mendelian randomisation studies with summary data and a continuous outcome. *Stat Med.* (2015) 34:2926–40. doi: 10.1002/sim.6522

27. Bowden J, Davey Smith G, Burgess S. Mendelian randomization with invalid instruments: effect estimation and bias detection through Egger regression. *Int J Epidemiol*. (2015) 44:512–25. doi: 10.1093/ije/dyv080

28. Hemani G, Tilling K, Davey Smith G. Orienting the causal relationship between imprecisely measured traits using GWAS summary data. *PloS Genet*. (2017) 13: e1007081. doi: 10.1371/journal.pgen.1007081

29. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, et al. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res*. (2018) 46: D1074–82. doi: 10.1093/nar/gkx1037

30. Brant SR, Okou DT, Simpson CL, Cutler DJ, Haritunians T, Bradfield JP, et al. Genome-Wide association study identifies african-Specific susceptibility loci in african americans with inflammatory bowel disease. *Gastroenterology*. (2017) 152:206–217 e202. doi: 10.1053/j.gastro.2017.02.041

31. Gonsky R, Fleshner P, Deem RL, Biener-Ramanujan E, Li D, Potdar AA, et al. Association of ribonuclease T2 gene polymorphisms with decreased expression and clinical characteristics of severity in crohn's disease. *Gastroenterology*. (2017) 153:219–32. doi: 10.1053/j.gastro.2017.04.002

32. Li S, Yan T, Yang JQ, Oberley TD, Oberley LW. The role of cellular glutathione peroxidase redox regulation in the suppression of tumor cell growth by manganese superoxide dismutase. *Cancer Res*. (2000) 60:3927–39.

33. Trenz TS, Delaix CL, Turchetto-Zolet AC, Zamocky M, Lazzarotto F, Margis-Pinheiro M. Going forward and back: the complex evolutionary history of the GPx. *Biol (Basel)*. (2021) 10:1165. doi: 10.3390/biology10111165

34. Zou M, Liang Q, Zhang W, Zhu Y, Xu Y. Endoplasmic reticulum stress related genome-wide Mendelian randomization identifies therapeutic genes for ulcerative colitis and Crohn's disease. *Front Genet*. (2023) 14:1270085. doi: 10.3389/fgene.2023.1270085

35. Costa Pereira C, Duraes C, Coelho R, Gracio D, Silva M, Peixoto A, et al. Association between polymorphisms in antioxidant genes and inflammatory bowel disease. *PloS One*. (2017) 12:e0169102. doi: 10.1371/journal.pone.0169102

36. Zhang C, Guo L, Qin Y. [Interaction of MIF gene -173G/C polymorphism and GPX1 gene 594C/T polymorphism with high-fat diet in ulcerative colitis]. *Zhonghua Yi Xue Yi Chuan Xue Za Zhi*. (2016) 33:85–90. doi: 10.3760/cma.j.issn.1003-9406.2016.01.021

37. Darnell JE Jr., Kerr IM, Stark GR. Jak-STAT pathways and transcriptional activation in response to IFNs and other extracellular signaling proteins. *Science*. (1994) 264:1415–21. doi: 10.1126/science.8197455

38. Zhang J, Wu J, Peng X, Song J, Wang J, Dong W. Associations between STAT3 rs744166 polymorphisms and susceptibility to ulcerative colitis and Crohn's disease: a meta-analysis. *PloS One*. (2014) 9:e109625. doi: 10.1371/journal.pone.0109625

39. Wang Z, Xu B, Zhang H, Fan R, Zhou J, Zhong J. Association between STAT3 gene polymorphisms and Crohn's disease susceptibility: a case-control study in a Chinese Han population. *Diagn Pathol*. (2014) 9:104. doi: 10.1186/1746-1596-9-104

40. Robinson P, Magness E, Montoya K, Engineer N, Eckols TK, Rodriguez E, et al. Genetic and small-molecule modulation of stat3 in a mouse model of crohn's disease. *J Clin Med*. (2022) 11:7020. doi: 10.3390/jcm11237020

41. Bai A, Moss A, Kokkotou E, Usheva A, Sun X, Cheifetz A, et al. CD39 and CD161 modulate Th17 responses in Crohn's disease. *J Immunol*. (2014) 193:3366–77. doi: 10.4049/jimmunol.1400346

42. Villarino AV, Kanno Y, O'Shea JJ. Mechanisms and consequences of Jak-STAT signaling in the immune system. *Nat Immunol*. (2017) 18:374–84. doi: 10.1038/ni.3691

43. Prager M, Buttner J, Haas V, Baumgart DC, Sturm A, Zeitz M, et al. The JAK2 variant rs10758669 in Crohn's disease: altering the intestinal barrier as one mechanism of action. *Int J Colorectal Dis*. (2012) 27:565–73. doi: 10.1007/s00384-011-1345-y

44. Hodson R. Inflammatory bowel disease. *Nature*. (2016) 540:S97. doi: 10.1038/540S97a

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 30 May 2024
DOI 10.3389/fonc.2024.1368120



OPEN ACCESS

EDITED BY
Dipak Kumar Sahoo,
Iowa State University, United States

REVIEWED BY
Songyun Deng,
Central South University, China
Paul Willemsen,
Hospital Network Antwerp (ZNA), Belgium
Jung Kyong Shin,
Sungkyunkwan University, Republic of Korea

*CORRESPONDENCE
Nastaran Mohammadian Rad
nastaran.mrad@maastrichtuniversity.nl

†These authors have contributed equally to this work and share last authorship

RECEIVED 16 January 2024
ACCEPTED 06 May 2024
PUBLISHED 30 May 2024

CITATION
Mohammadian Rad N, Sosef O, Seegers J, Koolen LJER, Hoofwijk JJWA, Woodruff HC, Hoofwijk TAGM, Sosef M and Lambin P (2024) Prognostic models for colorectal cancer recurrence using carcinoembryonic antigen measurements.
*Front. Oncol.* 14:1368120.
doi: 10.3389/fonc.2024.1368120



# Prognostic models for colorectal cancer recurrence using carcinoembryonic antigen measurements

Nastaran Mohammadian Rad[1*], Odin Sosef[2], Jord Seegers[2], Laura J. E. R. Koolen[3], Julie J. W. A. Hoofwijk[2], Henry C. Woodruff[1,4], Ton A. G. M. Hoofwijk[2], Meindert Sosef[2†] and Philippe Lambin[1†]

[1]The D-Lab, Department of Precision Medicine, GROW – Research Institute for Oncology & Reproduction, Maastricht University, Maastricht, Netherlands, [2]Department of Surgery, Zuyderland Medisch Centrum, Sittard-Geleen, Netherlands, [3]St.Elisabeth Krankenhaus Geilenkirchen, Geilenkirchen, Germany, [4]Department of Radiology and Nuclear Medicine, GROW – Research Institute for Oncology & Reproduction, Maastricht University Medical Center, Maastricht, Netherlands

**Objective:** Colorectal cancer (CRC) is one of the most prevalent cancers worldwide. A considerable percentage of patients who undergo surgery with curative intent will experience cancer recurrence. Early identification of individuals with a higher risk of recurrence is crucial for healthcare professionals to intervene promptly and devise appropriate treatment strategies. In this study, we developed prognostic models for CRC recurrence using machine learning models on a limited number of CEA measurements.

**Method:** A dataset of 1927 patients diagnosed with Stage I-III CRC and referred to Zuyderland Hospital for surgery between 2008 and 2016 was utilized. Machine learning models were trained using this comprehensive dataset, which included demographic details, clinicopathological factors, and serial measurements of Carcinoembryonic Antigen (CEA). In this study, the predictive performance of these models was assessed, and the key prognostic factors influencing colorectal cancer (CRC) recurrence were pinpointed

**Result:** Among the evaluated models, the gradient boosting classifier demonstrated superior performance, achieving an Area Under the Curve (AUC) score of 0.81 and a balanced accuracy rate of 0.73. Recurrence prediction was shown to be feasible with an AUC of 0.71 when using only five post-operative CEA measurements. Furthermore, key factors influencing recurrence were identified and elucidated.

**Conclusion:** This study shows the transformative role of machine learning in recurrence prediction for CRC, particularly by investigating the minimum number of CEA measurements required for effective recurrence prediction. This approach not only contributes to the optimization of clinical workflows but also facilitates the development of more effective, individualized treatment plans, thereby laying the groundwork for future advancements in this area. Future directions involve validating these models in larger and more diverse cohorts. Building on these efforts, our ultimate goal is to develop a risk-based follow-up strategy that can improve patient outcomes and enhance healthcare efficiency.

KEYWORDS

colorectal cancer, machine learning, carcinoembryonic antigen, cancer recurrence, prognosis

# 1 Introduction

Colorectal cancer (CRC) ranks as the third most prevalent cancer worldwide (1). Advances in chemotherapy and increased use of hepatic resection surgery have contributed to significant improvements in the survival rate for patients with this type of cancer (2). Despite these improvements, cancer recurrence remains a prolonged challenge, and delays in detection can compromise the effectiveness of surgical intervention (3). Studies have revealed that approximately 85% of recurrences occur within 30 months after surgery, with nearly all cases appearing within 5 years (4). Thus, it is essential to maintain continuous monitoring of patients even after successful therapeutic intervention to detect potential cancer recurrence at the earliest possible stage.

The current guidelines for identifying recurrence involve regular testing of CEA levels in post-operative patients (5, 6). CEA level is a widely used clinical marker, demonstrating associations with the occurrence and severity of CRC (7, 8). However, studies have revealed the fact that single CEA measurements lack strong prognostic potential for monitoring CRC, exhibiting a balanced accuracy of 0.65 (9, 10).

In recent years, machine learning (ML) techniques have gained significant traction in oncology (11, 12). These techniques are applied for both diagnosis and prognosis, aiming to enhance patient outcomes and optimize treatment strategies (13, 14). While ML models have been employed for recurrence prediction in CRC (see Section 2), there is a need for CRC prognostication models that simultaneously achieve high accuracy and offer clear explainability. This study aims to bridge this gap by employing ML techniques to accurately prognosticate CRC recurrence and also to identify the underlying factors contributing to it. Our contributions are three-fold:

- In this study, we apply and evaluate four various machine learning models, integrating demographic information, clinicopathological factors, and CEA measurements. Through the progressive integration of CEAs, we also investigate the minimal number of CEA measurements necessary to effectively predict recurrence.
- We use permutation importance method to identify the key clinical factors influencing our model's predictions, providing valuable information about the variables most impactful in CRC recurrence.
- We investigate the impact of data imputation on the predictive performance of CRC recurrence models.

# 2 Related work

In recent years, CRC prognosis and diagnosis have gained attention in clinical and research areas. Commercial tools such as Oncotype DX Colon (15) and ColoPrint assay (16), which incorporate gene expression profiling, have emerged as resources for assessing the risk of recurrence. However, these tools showed a relatively modest performance (area under the receiver operating characteristic curve (AUC) of 0.63 for ColoPrint and 0.55 for OncoDefender-CRC) (14).

Previous studies on CRC prognosis applied ML through retrospective analyses on diverse data types, mainly as a single modality, including clinical, epidemiological, and genetic data (11, 12). Through the analysis of genetic data, Grudner et al. (17) predicted diverse clinical outcomes, including cancer recurrence. Their model demonstrated a stratification between recurrence and non-recurrence patients, surpassing the effectiveness of sub-categorization based on prior literature, reporting an accuracy of 0.71 for their predictions. In (11), the authors explored the feasibility of using ML models, mainly decision-tree-based learning algorithms, to predict recurrence in Stage IV CRC patients. The reported AUC score for their top-performing model was 0.76. Elsewhere, Castellanos et al. (12), employed an ensemble model to predict recurrence in Stage II-III CRC patients. Their dataset included gene expression data, protein-protein interaction details, and tumor suppressor and driver mutation information. Their experimental results showed superior predictive capabilities on molecular data compared to clinical data alone. Their most effective model achieved an AUC of 0.79. In (13), the authors applied a range of ML models on a relatively small dataset with 904 CRC patients to predict recurrence. Their best-performing model, a support vector machine (SVM), applied to the structured data yielded an AUC of 0.83.

# 3 Methods

## 3.1 Data collection

This study used a dataset of patients diagnosed with Stage I-III CRC who were referred to Zuyderland Hospital for surgery with curative intent and follow-up of the primary tumor between 2008 to 2016. This study was approved by the Medical Ethics Committee, and informed consent was not obligatory. The dataset is composed of static and time-series (dynamic) features. The static features consist of 24 predictor variables that are associated with recurrence, such as demographic information, comorbidities, tumor characteristics, and treatment parameters. As shown in Figure 1, the dynamic features contain 40 CEA measurements. Patients were followed up post-operatively according to the Dutch National Guidelines, every 3 to 6 months on average (for a detailed description of all included predictor variables, see Table 1).

## 3.2 Data preprocessing

Comprehensive data preprocessing steps were performed to ensure the integrity of the dataset. Initially, 13 patients who presented inconsistencies in their data with the time of data collection appearing in descending order contrary to the expected chronological progression post-surgery were excluded. CEA measurements obtained before surgery were disregarded for the remaining patients. Missing data were then imputed using data imputation techniques based on each feature type. In line with

10.3389/fonc.2024.1368120

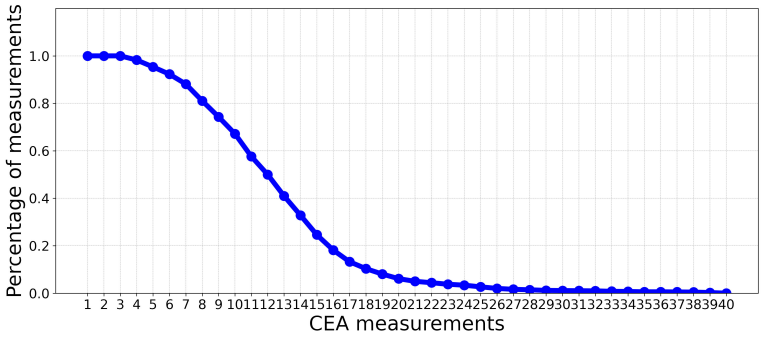



FIGURE 1
The percentage of CEA measurements (non-missing values) in each time point over all patients.

TABLE 1 An overview of the different variables used for the colorectal cancer recurrence prediction model.

| Category | Variable (Type) | Description |
|---|---|---|
| **Demographics** | Age (con)<br>Smoking status (cat) | Age at time of surgery<br>Divided into three classes: never, past, or current smokers |
| | Sex (bin) | Male or female |
| **Comorbidity** | Irritable Bowel Syndrome (bin)<br>Inflammatory Bowel Disease (bin) | Gastrointestinal tract functional disorder characterized by chronic abdominal pain and altered bowel habits<br>Chronic auto-inflammatory condition of the gastrointestinal tract |
| | Diabetes (bin) | Absolute or relative insulin deficiency |
| | Familial Adenomatous Polyposis (bin) | Presence of familial adenomatous polyposis; rare inherited disease causing extensive polyp formation |
| | Lynch (bin) | Presence of hereditary nonpolyposis colorectal cancer (HNPCC, or Lynch syndrome) |
| | Cardiac disease (bin) | Cardiac disease present (e.g., congestive heart failure, ischemic heart disease) |
| **Tumor Characteristics** | Organ (bin)<br>Synchronic metastasis (bin)<br>Location (bin) cTNM stage (cat) ycTNM (cat) | Location of tumor (colon/rectum)<br>Presence of metastasis detected at or before diagnosis of the primary tumor<br>Location of metastasis (liver/other location)<br>Clinical TNM (5th edition)<br>Clinical TNM after neoadjuvant therapy |
| | p(y)TNM (cat) | Pathological TNM (5th edition) |
| | Tumor type (cat) | Tumor type (adenocarcinoma, mucinous carcinoma, or other) |
| | Cancer staging (cat) | Cancer stage according to pTNM |
| | lymph invasion (bin) | Presence of lymph invasion |
| | Angioinvasion (bin) | Presence of angio invasion |
| **Treatment parameters** | Neoadjuvant therapy (cat) | Radiotherapy (5x5 Gray), chemotherapy, or radiochemotherapy |
| | Adjuvant chemotherapy (bin) | Use of any form of adjuvant chemotherapy |
| | Adjuvant radiotherapy (bin) | Use of any form of adjuvant radiotherapy |
| **Treatment outcome** | Resection marge free (bin) | Surgical outcome in achieving complete tumor removal |
| **Tumor marker** | CEA measurements (con) | Tumor marker used for detection recurrence |

The variables are ordered based on their category. The variable types and descriptions are provided. Con, continuous; Cat, categorical; Bin, binary; TNM, TNM-classification; CEA, Carcinoembryonic antigen; HNPCC, Hereditary Non-Polyposis Colorectal Cancer; 5x5 Gray, 5x5 rectal cancer radiation protocol.

previous studies (13), missing entries within each binary and categorical feature (e.g., smoking status) were imputed using the most frequent value present in that particular feature (See Supplementary Table S1 in Supplementary Material for the number of missing values in each static feature). For continuous values (e.g., CEA measurements), median value-based imputation was employed, effectively maintaining the overall distribution characteristics. Then, categorical features were encoded using the one-hot encoding scheme, resulting in 67 features for subsequent modeling. Quantile transform was applied to features to mitigate

the impact of outliers and non-normality in the original data. The final preprocessed dataset consisted of 1927 patients (See Supplementary Table S2 in Supplementary Material for the distribution of patients by cancer stage). The dataset was imbalanced, with the positive class (recurrence) constituting approximately 15% of the total dataset which equates to 285 patients.

## 3.3 Experimental setup

### 3.3.1 Experiment 1 (prognostic models using static features)

This experiment aims to investigate the influence of static clinical data on the accuracy of recurrence prediction in patients with CRC. We evaluated four diverse classifiers: 1) logistic regression (LR), a linear classifier; 2) support vector machine (SVM) with a radial basis function kernel, a non-linear classifier; 3) random forest (RF), a decision-tree-based classifier; and 4) gradient boosting (GB), an ensemble model of decision-tree based classifiers. Furthermore, to identify the key clinical factors contributing to the recurrence prediction, we applied the permutation importance technique (18), a model-agnostic method for assessing feature importance, on the static features using our top-performing classifier.

### 3.3.2 Experiment 2 (prognostic models using static features and step-wise incorporation of CEA measurements)

This experiment aims to assess the impact of incorporating CEA measurements alongside static features for recurrence prediction. We evaluated the performance of the classifiers introduced in Experiment 3.3.1 using a limited number of CEA measurements after surgery. In this context, we progressively incorporated CEA measurements with static features. This iterative process involved gradually adding individual CEA measurements at a time to the existing set of static features, incrementally building a comprehensive set of combined features. Subsequent to each inclusion of a new CEA measurement, we trained ML models, outlined in Experiment 3.3, with the updated input for the prediction. As depicted in Figure 1, by moving beyond the first 10 CEA measurements, the percentage of measurements (non-missing values) in each time point over all the patients significantly decreases. Consequently, we have restricted our analysis to these initial 10 CEA measurements. This selection ensures a more reliable and complete dataset, with less than 50% missing values.

### 3.3.3 Experiment 3 (the impact of data imputation)

Considering the presence of missing values in our dataset and the use of imputation as a preprocessing step, this experiment examines the impact of data imputation on recurrence prediction. This is achieved by comparing the performance of the best-performing classifier, which was applied to the imputed data, with that of the Histogram-based Gradient Boosting (HGB) classifier. Unlike all classifiers used in this study, the HGB classifier can handle missing values without data imputation. By using the HGB classifier, we aim to evaluate the impact of its missing value-handling capabilities on the predictive accuracy of our recurrence prediction task. Through a comparative analysis, we can assess the benefits of incorporating the HGB classifier's missing value-handling mechanism in our prediction model.

All classifiers were implemented using the Sklearn library (19). To tackle the challenge of data imbalance, a weighted training approach was adopted, wherein class weights were set to be inversely proportional to their frequencies in the dataset. Hyperparameters for these classifiers were optimized using a grid search algorithm, which was applied to a validation set to ensure optimal model performance and generalizability.

## 3.4 Evaluation

In this study, the samples were randomly divided into training and testing sets at a ratio of 8:2. All the experiments were repeated 10 times to evaluate the variability in performances and ensure reliable estimates of the model's performance. For each repetition, the following evaluation metrics were calculated to measure the classification performance:

- Area Under the Curve (AUC): This metric measures the model's discriminative power, reflecting its ability to differentiate between the positive and negative classes accurately.
- Balanced Accuracy (BAC): This metric assesses the overall accuracy of a classification model, considering both sensitivity and specificity (20). Unlike traditional accuracy, which may be misleading in the presence of imbalanced datasets, BAC is useful when the dataset is imbalanced. BAC inherently encompasses both specificity and sensitivity, crucial metrics often employed in evaluating clinical assay performance. Therefore, in line with prevalent ML practices (13), while we prioritize models with superior AUC scores, we also value models with high BAC scores.

# 4 Results

Table 2 shows that the LR classifier achieved the best performance of all the ML models trained on the static data, with an AUC of 0.65 and a BAC of 0.60. Furthermore, the results indicated a boost in classifiers' performance upon adding CEA measurements. Up until the inclusion of the first 7 CEA measurements, performance enhancements were observed for GB, RF, and SVM models (AUC and BAC increased by approximately 12–17% and 7–11%, respectively). Conversely, the performance improvement of the LR model was comparatively more gradual within the same range of measurements (both AUC and BAC increased 5%). The improvement rate diminished after the inclusion of the first 7 CEA measurements.

TABLE 2 AUC and BAC measures of four ML models when trained on static data, and combination of static data and 10 CEA measurements.

| Models | Experiment | BAC | AUC |
|---|---|---|---|
| LR | Static | **0.60 ± 0.00** | **0.65 ± 0.00** |
| | Static+10 CEA | 0.64 ± 0.00 | 0.70 ± 0.00 |
| SVM | Static | 0.54 ± 0.02 | 0.58 ± 0.02 |
| | Static+10 CEA | 0.68 ± 0.01 | 0.74 ± 0.01 |
| RF | Static | 0.58 ± 0.01 | 0.62 ± 0.00 |
| | Static+10 CEA | 0.71 ± 0.02 | 0.77 ± 0.01 |
| GB | Static | 0.59 ± 0.02 | 0.60 ± 0.01 |
| | Static+10 CEA | **0.73 ± 0.01** | **0.81 ± 0.02** |

The values indicating higher performance are highlighted in bold.

As illustrated in Figure 2, for the first 3 post-operative post-operative CEA measurements, the LR model showed the highest performance in terms of AUC. Among the employed models, GB and RF classifiers outperformed other ML classifiers when applied to the combination of static and CEA measurements. GB trained on the combined static data and 10 CEA measurements performed the best, achieving the highest performance with an AUC score of 0.81 and BAC of 0.73. Furthermore, our results showed that using only the first 5 post-operative CEA measurements in combination with static data, the GB model was able to predict recurrence with an AUC of 0.71. Although this marked a reduction of around 10% from the final model's performance, which used the entire dynamic data, the performance demonstrated an incremental enhancement with the inclusion of more time points.

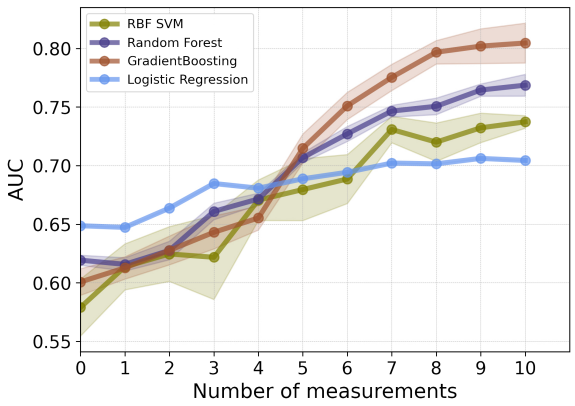



FIGURE 2
The mean AUC scores of four ML models on combination of static features and CEA measurements taken after the date of surgery.

Furthermore, the results of permutation importance method depicted in Figure 3 identified tumor characteristics and demographic information as key determinants. As expected, p(y)TNM was the most important feature, demonstrating a substantial effect on the prediction of recurrence. While p(y)TNM and cancer stadium are measurements for advanced tumor stages, p(y)TNM provides a more detailed account of tumor size and metastasis. In contrast, cancer stadium is a more compressed or simplified version of the p(y)TNM classification. This simplification is primarily evident in stage III cancer, where we did not differentiate between sub-stages A, B, and C but rather considered it as a single stage. Thus, our analyses suggested that p(y)TNM remains the most detailed and informative variable for inclusion in the model, mainly because of its comprehensive detailing of the extensiveness of tumor growth and spread. Among other features, age also showed a significant influence on recurrence prediction, reinforcing its importance as a prognostic factor (21–24).

As an alternative solution to data imputation, one can use HGB, which offers a mechanism to handle missing values directly without the need for data imputation. By comparing the results of HGB with the best-performing model, GB, which requires data imputation, we observed that HGB achieved comparable performance without the additional step of data imputation (see Figure 4). Using HGB can streamline the modeling pipeline and simplify the data preprocessing, ultimately leading to more efficient and reliable predictions.

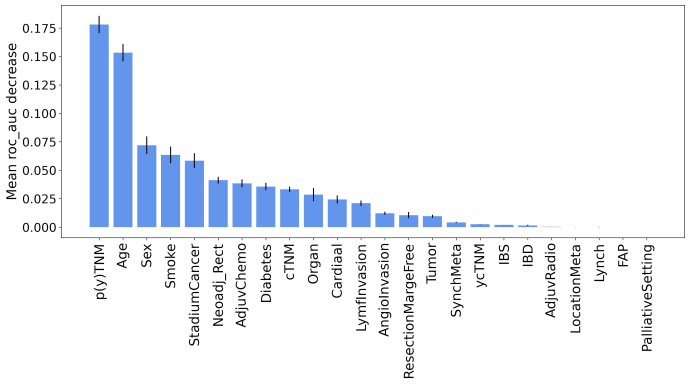



FIGURE 3
GB feature importance via permutation importance method.

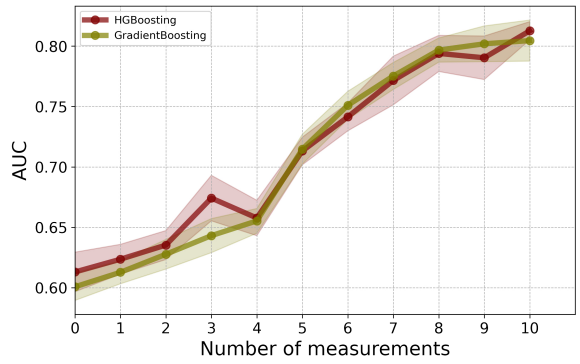



FIGURE 4
Performance comparison between GB with data imputation and HGB without data imputation.

## 5 Discussion

In this study, we proposed the application of ML for recurrence prediction in patients with CRC using a combination of longitudinal CEA measurements with clinical information, including demographic data, tumor staging, and treatment parameters. Our best-performing classifier, GB, achieved remarkable AUC and BAC scores. In summary, the model's predictive ability for recurrence, based on limited and early post-surgical CEA measurements, suggests the potential for devising personalized monitoring schedules. In addition, our analysis underscored the significance of demographic information, including age and sex, as well as tumor attributes such as p(y) TNM in predicting the risk of recurrence. These findings are consistent with earlier studies, highlighting a high risk of recurrence in older patients (21–24) and align with evidence of an association between advanced tumor stages and an increased risk of recurrence (21, 25). Furthermore, the analysis suggested that the impact of comorbidities on recurrence prediction was less pronounced when compared to these other factors. Furthermore, we showed that using the HGB model can remove the need for data imputation while preserving the model's performance.

One of the major strengths of this study is the large sample size and the availability of data on a wide range of variables. Our dataset comprises 1927 patients representing a significant increase in size compared to datasets used in prior studies (13, 14). The ample sample size in our dataset supports the application of deep learning. Considering the presence of temporal information in the longitudinal CEA measurements, recurrent neural networks are considered suitable candidates for recurrence prediction in future studies.

Despite the promising results, this study has several limitations. We evaluated our models using data collected from a single hospital while the heterogeneity of patient demographics, disease presentations, and treatment protocols across different hospitals and geographic locations can significantly influence model performance. To address this critical aspect of our research and ensure the robustness and generalizability of our models, we need to further validate the developed models on a broader range of patient data, reflecting diverse demographic and clinical characteristics. While the outcomes highlighted in this study are promising, it is worth noting that these achievements have been obtained by directly applying ML models to the raw data without involving any feature extraction processes. This shows the potential inherent in the original data to contribute to the predictive capabilities of the models. In future work, we will investigate the advantages of incorporating feature extraction methods from clinical data. This could encompass manually curating features that align with domain knowledge or deploying advanced techniques that enable the models to learn informative features automatically. Furthermore, all the analyses were conducted retrospectively. Consequently, the performance of the model in predicting cancer recurrence on new, yet-to-be-observed data could not be directly assessed or validated in real-time. The ability of the model to accurately predict cancer recurrence in future patients remains to be tested through prospective studies. By developing an application (26) to frequently capture patient symptoms in short intervals after surgery, we can bridge the gap between real-time patient experiences and medical interventions. Such a system facilitates timely prediction and management of recurrence and promotes a patient-centric approach by allowing individuals to participate in their care actively. The adoption of such platforms has the potential to revolutionize recurrence prediction and overall patient management.

Additionally, in future work, we aim to explore integrating multi-modal healthcare data, recognizing its potential to enhance the prediction of CRC recurrence. This approach will involve diverse data types, such as molecular prognostic factors (27) and incorporation of radiomic analysis (28), each contributing unique information about the disease's progression and the prognosis of the patient. The inclusion of molecular prognostic markers, offers a deeper understanding of the tumor's biological behavior. These markers can provide information about the aggressiveness of the cancer, its likelihood of recurrence, and potential response to therapy. Incorporating radiomic analysis into our model can enhance our understanding of the tumor's characteristics and its interaction with surrounding tissues, further refining our predictions of recurrence risk.

## 6 Conclusion

CRC remains a significant global health challenge, with a notable percentage of patients experiencing recurrence after curative surgery. This study showed the value of CEA as a non-invasive and efficient marker for recurrence prediction. Through the application of ML, specifically GB classifier, we demonstrated an accurate recurrence prediction using comprehensive clinical data combined with CEA measurements, even when limited to early CEA measurements. We further showed that age and tumor characteristics are the most important factors influencing the risk of recurrence. Finally, we showed that HGB yields performance comparable to GB for this particular dataset while eliminating the need for data imputation. As healthcare moves towards more patient-centric models, the integration of web-based

platforms and real-time symptom monitoring will be crucial. The findings of this study highlight the need for further prospective studies and show the transformative potential of ML in revolutionizing patient centered care in CRC management.

## Data availability statement

The data analyzed in this study is subject to the following licenses/restrictions: patient confidentiality and participant privacy. Requests to access these datasets should be directed to OS, odin.sosef@me.com, and JS, jordseegers@gmail.com.

## Ethics statement

The studies involving humans were approved by Medical Ethical Review Committee Zuyderland and Zuyd University of Applied Sciences (METC Z). The studies were conducted in accordance with the local legislation and institutional requirements. The human samples used in this study were acquired from primarily isolated as part of your previous study for which ethical approval was obtained. Written informed consent for participation was not required from the participants or the participants' legal guardians/next of kin in accordance with the national legislation and institutional requirements.

## Author contributions

NM: Writing – original draft, Visualization, Methodology, Formal analysis, Data curation, Conceptualization. OS: Writing – review & editing, Resources, Data curation. JS: Writing – review & editing, Resources, Data curation. LK: Data curation, Writing – review & editing, Resources. JH: Writing – review & editing, Resources. HW: Writing – review & editing. TH: Writing – review & editing, Resources. MS: Writing – review & editing, Resources. PL: Conceptualization, Writing – review & editing, Supervision.

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1368120/full#supplementary-material

## References

1. Benson AB, Venook AP, Cederquist L, Chan E, Chen YJ, Cooper HS, et al. Colon cancer, version 1.2017, nccn clinical practice guidelines in oncology. *J Natl Compr Cancer Network*. (2017) 15:370–98. doi: 10.6004/jnccn.2017.0036

2. Kopetz S, Chang GJ, Overman MJ, Eng C, Sargent DJ, Larson DW, et al. Improved survival in metastatic colorectal cancer is associated with adoption of hepatic resection and improved chemotherapy. *J Clin Oncol*. (2009) 27:3677. doi: 10.1200/JCO.2008.20.5278

3. Hellinger MD, Santiago CA. Reoperation for recurrent colorectal cancer. *Clinics colon rectal Surg*. (2006) 19:228–36. doi: 10.1055/s-2006-956445

4. Guthrie J. Colorectal cancer: follow-up and detection of recurrence. *Abdominal Radiol*. (2002) 27:570. doi: 10.1007/s00261-001-0081-0

5. Locker GY, Hamilton S, Harris J, Jessup JM, Kemeny N, Macdonald JS, et al. Asco 2006 update of recommendations for the use of tumor markers in gastrointestinal cancer. *J Clin Oncol*. (2006) 24:5313–27. doi: 10.1200/JCO.2006.08.2644

6. Graham RA, Wang S, Catalano PJ, Haller DG. Postsurgical surveillance of colon cancer: preliminary cost analysis of physician examination, carcinoembryonic antigen testing, chest x-ray, and colonoscopy. *Ann Surg*. (1998) 228:59. doi: 10.1097/00000658-199807000-00009

7. Lee JH, Lee SW. The roles of carcinoembryonic antigen in liver metastasis and therapeutic approaches. *Gastroenterol Res Pract*. (2017) 2017. doi: 10.1155/2017/7521987

8. Tong G, Xu W, Zhang G, Liu J, Zheng Z, Chen Y, et al. The role of tissue and serum carcinoembryonic antigen in stages i to iii of colorectal cancer—a retrospective cohort study. *Cancer Med*. (2018) 7:5327–38. doi: 10.1002/cam4.1814

9. Shinkins B, Nicholson BD, Primrose J, Perera R, James T, Pugh S, et al. The diagnostic accuracy of a single cea blood test in detecting colorectal cancer recurrence: Results from the facs trial. *PloS One*. (2017) 12:e0171810. doi: 10.1371/journal.pone.0171810

10. Sørensen CG, Karlsson WK, Pommergaard HC, Burcharth J, Rosenberg J. The diagnostic accuracy of carcinoembryonic antigen to detect colorectal cancer recurrence–a systematic review. *Int J Surg*. (2016) 25:134–44. doi: 10.1016/j.ijsu.2015.11.065

11. Xu Y, Ju L, Tong J, Zhou CM, Yang JJ. Machine learning algorithms for predicting the recurrence of stage iv colorectal cancer after tumor resection. *Sci Rep*. (2020) 10:2519. doi: 10.1038/s41598-020-59115-y

12. Castellanos JA. *Predicting colorectal cancer recurrence by utilizing multiple-view multiple-learner supervised learning*. (2016). doi: 10.1200/JCO.2017.35.4_suppl.635

13. Ho D, Chong DQ, Tay B, Tan IB, Motani M. (2021). Prognosticating colorectal cancer recurrence using machine learning techniques, in: *2020 IEEE International Conference on E-health Networking, Application & Services (HEALTHCOM)*, . pp. 1–6. IEEE.

14. Ho D, Tan IBH, Motani M. (2021). Predictive models for colorectal cancer recurrence using multi-modal healthcare data, in: *Proceedings of the Conference on Health, Inference, and Learning*, Proceedings of Machine Learning Research (PMLR). pp. 204–13.

15. Lenehan PF, Boardman LA, Riegert-Johnson D, De Petris G, Fry DW, Ohrnberger J, et al. Generation and external validation of a tumor-derived 5-gene prognostic signature for recurrence of lymph node-negative, invasive colorectal carcinoma. *Cancer*. (2012) 118:5234–44. doi: 10.1002/cncr.27628

16. Kopetz S, Tabernero J, Rosenberg R, Jiang ZQ, Moreno V, Bachleitner-Hofmann T, et al. Genomic classifier coloprint predicts recurrence in stage ii colorectal cancer patients more accurately than clinical factors. *oncologist*. (2015) 20:127–33. doi: 10.1634/theoncologist.2014-0325

17. Gründner J, Prokosch HU, Stürzl M, Croner R, Christoph J, Toddenroth D. Predicting clinical outcomes in colorectal cancer using machine learning. *MIE*. (2018) 247:101–5. doi: 10.3233/978-1-61499-852-5-101

18. Breiman L. Random forests. *Mach Learn*. (2001) 45:5–32. doi: 10.1023/A:1010933404324

19. Pedregosa F, Varoquaux G, Gramfort A, Michel V, Thirion B, Grisel O, et al. Scikit-learn: Machine learning in python. *J Mach Learn Res*. (2011) 12:2825–30. doi: 10.5555/1953048.2078195

20. Brodersen KH, Ong CS, Stephan KE, Buhmann JM. (2010). The balanced accuracy and its posterior distribution, in: *2010 20th international conference on pattern recognition*, United States. pp. 3121–4. IEEE.

21. Zare-Bandamiri M, Fararouei M, Zohourinia S, Daneshi N, Dianatinasab M. Risk factors predicting colorectal cancer recurrence following initial treatment: a 5-year cohort study. *Asian Pacific J Cancer prevention: APJCP*. (2017) 18:2465. doi: 10.22034/APJCP.2017.18.9.2465

22. Aghili M, Izadi S, Madani H, Mortazavi H. Clinical and pathological evaluation of patients with early and late recurrence of colorectal cancer. *Asia-pacific J Clin Oncol*. (2010) 6:35–41. doi: 10.1111/j.1743-7563.2010.01275.x

23. Westberg K, Palmer G, Johansson H, Holm T, Martling A. Time to local recurrence as a prognostic factor in patients with rectal cancer. *Eur J Surg Oncol (EJSO)*. (2015) 41:659–66. doi: 10.1016/j.ejso.2015.01.035

24. Badic B, Oguer M, Cariou M, Kermarrec T, Bouzeloc S, Nousbaum JB, et al. Prognostic factors for stage iii colon cancer in patients 80 years of age and older. *Int J Colorectal Dis*. (2021) 36:811–9. doi: 10.1007/s00384-021-03861-6

25. Sloothaak D, Sahami S, van der Zaag-Loonen HJ, van der Zaag E, Tanis P, Bemelman W, et al. The prognostic value of micrometastases and isolated tumour cells in histologically negative lymph nodes of patients with colorectal cancer: a systematic review and meta-analysis. *Eur J Surg Oncol (EJSO)*. (2014) 40:263–9. doi: 10.1016/j.ejso.2013.12.002

26. Denis F, Basch E, Septans AL, Bennouna J, Urban T, Dueck AC, et al. Two-year survival comparing web-based symptom monitoring vs routine surveillance following treatment for lung cancer. *Jama*. (2019) 321:306–7. doi: 10.1001/jama.2018.18085

27. Formica V, Sera F, Cremolini C, Riondino S, Morelli C, Arkenau HT, et al. Kras and braf mutations in stage ii and iii colon cancer: a systematic review and meta-analysis. *JNCI: J Natl Cancer Institute*. (2022) 114:517–27. doi: 10.1093/jnci/djab190

28. Lv L, Xin B, Hao Y, Yang Z, Xu J, Wang L, et al. Radiomic analysis for predicting prognosis of colorectal cancer from preoperative 18f-fdg pet/ct. *J Trans Med*. (2022) 20:66. doi: 10.1186/s12967-022-03262-5

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 23 July 2024
DOI 10.3389/fimmu.2024.1398719



OPEN ACCESS

EDITED BY
Ashish Patel,
Hemchandracharya North Gujarat University, India

REVIEWED BY
Anil Patani,
Sankalchand Patel University, India
Snehal Bagatharia,
Gujarat State Biotechnology Mission, India

*CORRESPONDENCE
Wendong Sun
swd2102@163.com

†These authors have contributed equally to this work

RECEIVED 10 March 2024
ACCEPTED 05 July 2024
PUBLISHED 23 July 2024

CITATION
Jin S, Zhang H, Lin Q, Yang J, Zeng R, Xu Z and Sun W (2024) Deciphering the immune-metabolic nexus in sepsis: a single-cell sequencing analysis of neutrophil heterogeneity and risk stratification. *Front. Immunol.* 15:1398719. doi: 10.3389/fimmu.2024.1398719



# Deciphering the immune-metabolic nexus in sepsis: a single-cell sequencing analysis of neutrophil heterogeneity and risk stratification

Shaoxiong Jin[1†], Huazhi Zhang[2†], Qingjiang Lin[1], Jinfeng Yang[1], Rongyao Zeng[1], Zebo Xu[1] and Wendong Sun[1*]

[1]Department of Emergency Surgery, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, Fujian, China, [2]Department of Radiology, The Second Affiliated Hospital of Fujian Medical University, Quanzhou, Fujian, China

**Background:** Metabolic dysregulation following sepsis can significantly compromise patient prognosis by altering immune-inflammatory responses. Despite its clinical relevance, the exact mechanisms of this perturbation are not yet fully understood.

**Methods:** Single-cell RNA sequencing (scRNA-seq) was utilized to map the immune cell landscape and its association with metabolic pathways during sepsis. This study employed cell-cell interaction and phenotype profiling from scRNA-seq data, along with pseudotime trajectory analysis, to investigate neutrophil differentiation and heterogeneity. By integrating scRNA-seq with Weighted Gene Co-expression Network Analysis (WGCNA) and machine learning techniques, key genes were identified. These genes were used to develop and validate a risk score model and nomogram, with their efficacy confirmed through Receiver Operating Characteristic (ROC) curve analysis. The model's practicality was further reinforced through enrichment and immune characteristic studies based on the risk score and *in vivo* validation of a critical gene associated with sepsis.

**Results:** The complex immune landscape and neutrophil roles in metabolic disturbances during sepsis were elucidated by our in-depth scRNA-seq analysis. Pronounced neutrophil interactions with diverse cell types were revealed in the analysis of intercellular communication, highlighting pathways that differentiate between proximal and core regions within atherosclerotic plaques. Insight into the evolution of neutrophil subpopulations and their differentiation within the plaque milieu was provided by pseudotime trajectory mappings. Diagnostic markers were identified with the assistance of machine learning, resulting in the discovery of PIM1, HIST1H1C, and IGSF6. The identification of these markers culminated in the development of the risk score model, which demonstrated remarkable precision in sepsis prognosis. The model's capability to categorize patient profiles based on immune characteristics was confirmed, particularly in identifying individuals at high risk with suppressed immune cell activity and inflammatory responses. The role of PIM1 in modulating the immune-inflammatory response during sepsis was

further confirmed through experimental validation, suggesting its potential as a therapeutic target.

**Conclusion:** The understanding of sepsis immunopathology is improved by this research, and new avenues are opened for novel prognostic and therapeutic approaches.

KEYWORDS

sepsis, single-cell sequencing, neutrophils, metabolic dysregulation, risk score model

## Introduction

Sepsis is a life-threatening condition characterized by a dysregulated host response to infection, which can lead to organ dysfunction. Globally, sepsis is estimated to affect over 30 million people annually, potentially resulting in 6 million deaths (1). Currently, the management of sepsis relies on prompt antibiotic therapy, removal of the infection source, and supportive measures to maintain hemodynamic stability and preserve organ function (2). However, patient-specific response variability complicates management, reflecting the limited understanding of sepsis pathogenesis and signaling the need for more effective, individualized treatment approaches. Tailored therapies, founded on patient-specific biomarkers and stratification based on immunological or genetic profiles, can enhance effectiveness and reduce the likelihood of adverse effects.

Sepsis initiates a dynamic immune response that evolves over time, marked by concurrent pro-inflammatory and anti-inflammatory processes. Consequently, most sepsis patients rapidly exhibit signs of profound immune suppression, resulting in detrimental outcomes (3). Recent research has emphasized the significance of metabolic dysfunction, epigenetic changes, myeloid-derived suppressor cells, immature neutrophils with suppressive properties, and immune variations in main lymphoid organs during sepsis (4–6).

Metabolic dysfunction plays a crucial role in the initiation and progression of sepsis. During sepsis, the body undergoes an advanced level of metabolic emergency which can potentially lead to organ dysfunction (7). Metabolic shifts are prominently observed in various cell types during sepsis, particularly as immune cells undergo transformation. Cellular metabolism, which exhibits variable metabolic profiles across different cell types and stages of the disease, plays a key role in the immune dysregulation and organ failure associated with sepsis (3, 8). Metabolic reprogramming, wherein glycolysis supersedes oxidative phosphorylation (OXPHOS) as the primary source of energy production, is crucial for immune cell activation while simultaneously contributing to immunosuppression (9). Additionally, metabolites from OXPHOS and glycolysis may serve as signaling molecules, modulating the immune response throughout sepsis. The "energy crisis" induced by sepsis leads to impaired cellular functions and potentially severe organ dysfunction (10). Although metabolic reprogramming can partially mitigate this energy deficit, fostering host tolerance and enhancing cell survival, reversion to OXPHOS is imperative for cellular function restoration (11). In the intricate landscape of molecular and cellular biology, significant rewiring of metabolic pathways and epigenetic modifications has been identified as a pivotal factor in triggering and perpetuating immune system changes linked to sepsis. These alterations precipitate profound changes in gene expression patterns which lie at the heart of sepsis-induced immunological transformations (12). From a broader perspective, immune cells require metabolic profile alterations to achieve effective functionality. These metabolic changes are tentatively linked to the progression of immune responses during sepsis (13). This metabolic deceleration is akin to the cellular hibernation noted in organ dysfunction related to sepsis (14). Therefore, exploring the interplay between metabolism and immunity in the context of sepsis is a critical area of research, pivotal for identifying novel therapeutic targets to restore immune homeostasis following sepsis.

In this study, scRNA-seq was employed to investigate the immune cell composition of sepsis patients, revealing a specific enrichment of immune cell types. The application of the MuSiC algorithm and intercellular communication assessments uncovered notable interactions among immune cells, highlighting the crucial role of neutrophils in sepsis and their connection to metabolic activity. The analysis of neutrophil heterogeneity has led to the identification of four distinct subtypes, each characterized by unique functional attributes. Furthermore, the developmental trajectories of neutrophils were traced, leading to the identification of essential genes and the characterization of subpopulation lineage differentiation. By utilizing WGCNA, gene co-expression networks were constructed to identify significant genes for further investigation. Gene enrichment assays were then performed to elucidate the biological functions of these genes. Machine learning algorithms were employed to identify potential biomarkers for atherosclerosis, leading to the development of a diagnostic model with enhanced predictive capabilities. Using a customized riskScore model, we stratified patients based on their risk profiles and investigated the molecular and immune characteristics associated

with different risk levels. The validation of characteristic genes *in vivo* sepsis models underscored the significance of these targeted genes in the disease's pathology.

# Methods

## Acquisition of raw data

scRNA-seq data of peripheral blood from a cohort of five healthy individuals and four patients with advanced-stage sepsis were retrieved from the Gene Expression Omnibus (GEO) repository, specifically under accession number GSE175453. Concurrently, aggregated transcriptomic datasets associated with sepsis were acquired from GEO (accession numbers: GSE65682, GSE95233, GSE63042) and the ArrayExpress database (accession number: E-MTAB-5273). After the acquisition, these datasets underwent a logarithmic transformation to base 2 and normalization utilizing the Robust Multi-array Average (RMA) algorithm available within the "affy" package in the R statistical environment.

## scRNA-seq data processing and cell annotation

Utilizing the R package "Seurat," single-cell RNA sequencing (scRNA-seq) data was analyzed with meticulous attention to precision. Initially, the dataset underwent a rigorous gene filtering process where only genes present in no fewer than three individual cells were considered for further examination. This initial step ensures that the focus remains on genes with sufficient representation across the cell population, thus enhancing the robustness of downstream analyses. Concomitantly, cells were filtered based on their gene expression profiles, retaining cells that exhibited an expression range of 200 to 3000 genes. This specific criterion was set to exclude cells with abnormally low or high gene counts, which could otherwise introduce noise into the dataset.

Additionally, cells were subjected to further filtering based on two additional parameters: the total RNA count (nCount_RNA) and mitochondrial gene expression. Specifically, a threshold was established to retain cells with an nCount_RNA below 20,000 to exclude potential doublets or multiplets that could distort the results. Mitochondrial gene expression was also monitored and kept under 10%, as an elevated mitochondrial gene percentage is often indicative of low-quality or dying cells. These stringent filtering criteria ensured that only high-quality cells were retained, culminating in a dataset comprising 40,584 cells deemed suitable for advanced analyses.

To prepare the selected cell population for subsequent steps, the data was normalized and scaled using Seurat's "NormalizeData" and "ScaleData" functions. Normalization adjusted the gene expression measurements for each cell to account for differences in sequencing depth, resulting in the expression levels on a comparable scale across all cells. Scaling further refined these measurements by centering the data and scaling each gene to unit variance, thereby mitigating the effects of any highly variable or abundant genes.

Following this preliminary processing, the most variable genes were identified to capture the underlying biological heterogeneity within the cell population. Using the "FindVariableFeatures" function in Seurat, the top 3,000 genes exhibiting the highest variability across the dataset were pinpointed. Given the dataset's multi-sample origin, it was crucial to address potential batch effects that could confound the analyses. This was achieved using the "RunHarmony" function, which harmonizes the data across different samples, thereby reducing batch-induced biases.

Subsequent dimensionality reduction was performed using principal component analysis (PCA), a technique that enables the condensation of the data's complexity by projecting it into a set of orthogonal components. We focused on the top 20 principal components, which encapsulated the most significant variance in the dataset. To further elucidate cell population structures, these components were subjected to t-distributed stochastic neighbor embedding (t-SNE) analysis, which projected the high-dimensional data into a two-dimensional space. This visualization technique facilitated the discernment of significant cellular conglomerates and patterns.

In the clustering phase, Seurat's "FindNeighbors" and "FindClusters" functions were executed, with the latter set to a resolution parameter of 0.3. This clustering approach partitioned the cells into 13 distinct clusters. The resolution parameter was tuned to balance the granularity of the clusters, ensuring a meaningful yet interpretable clustering outcome. The resulting clusters were visually represented in a t-SNE plot, providing an intuitive overview of the cellular landscape.

Subsequent cluster annotation involved a thorough manual examination, wherein each cluster was classified into major cell types based on established marker gene profiles. Marker genes serve as distinctive identifiers for various cell types, allowing for accurate classification. To characterize the markers within each cellular contingent, the "COSG" package in R was employed. The parameters were specifically configured with a mean expression threshold of 10 and a user-defined gene count of 100, facilitating a comprehensive and precise marker characterization essential for downstream biological interpretations.

## Evaluation of metabolic activity at single-cell resolution

In each cell population, the metabolic processes of singular cells were mapped and measured utilizing the 'scMetabolism' package in R, a cutting-edge tool designed for single-cell metabolic activity quantification (15). This tool processes a matrix of single-cell data, employing the VISION algorithm to assess individual cell metabolic pathway scores. Embedded within the 'scMetabolism' tool are the comprehensive KEGG and Reactome pathway databases. Before the metabolic examination, the dataset underwent a uniform transformation. The VISION algorithm played a pivotal role in computing the metabolic scores. Comparative analysis of metabolic activities across different cellular groupings pinpointed pathways

with notable variances. For this investigation, the analysis harnessed KEGG metabolic gene sets coupled with the "VISION" technique. Subsequently, for graphical representation, we utilized the "DotPlot.metabolism" and "BoxPlot.metabolism" functions.

## Annotating cell types in bulk RNA –seq dataset

Single Cell Multi-Subject (SCMS) serves as an effective approach to determining the prevalence of distinct cell populations. This methodology applies gene expression profiles particular to each cell type obtained from scRNA-seq to ascertain the comparative frequency of a range of cell subsets within a composite RNA-seq dataset. To appraise the respective contributions of cell types within aggregate peripheral blood samples, a uniform procedure was employed. Subsequently, the variations across diverse cell categories among different cohorts were graphically represented.

## Trajectory analysis

The differentiation pathways within the identified cell clusters were examined using the Monocle2 algorithm (16). To isolate the cell clusters of interest, we employed the "subset" function from the Seurat package, followed by the construction of a CellDataSet object with the "newCellDataSet" method in Monocle2, setting the "lowerDetectionLimit" to 0.5. To enhance the quality of the dataset, low-quality cells and genes were filtered out by employing the "detectGenes" and "subset" methods with the "min_expr" threshold set at 0.1. This step occurred after size factor computation and dispersion estimation. Differential gene expression along the determined trajectories was identified using the "differentialGeneTest". Dimensionality reduction was accomplished through the "reduceDimension" function, leveraging the "DDRTree" approach. For visualization, functions such as "plot cell trajectory", "plot genes in pseudotime", and "plot genes branched heatmap" were implemented following cell ordering. Additionally, a CytoTRACE analysis, which is a method for the unsupervised prediction of cells' relative differentiation states from their single-cell transcriptomes, was conducted (17). This analysis was carried out using the default parameters specified in recommended protocols to augment our understanding of cell trajectory. Visualization of the results was achieved through the "plotCytoGenes" and "plotCytoTRACE" functions.

## Cell communication analysis

The "CellChat" R package (available at https://www.github.com/sqjin/CellChat) (4) facilitated the construction of CellChat objects, with the UMI count matrices pertinent to each subset (Normal and AD) serving as the foundation. The "CellChatDB.human" database was prioritized for ligand-receptor pairings during the analysis. The examination of cellular communication was executed with the preconfigured default settings. Subsequently, to discern the cumulative interaction count and the comparative intensity of these interactions, CellChat objects respectively to each subset were amalgamated via the "mergeCellChat" command. To display the variances in interaction numbers or strengths across different cellular types between Normal and AD groups, both "compareInteractions" and "netVisual_circle" functions were employed. Lastly, the "netVisual_bubble" function allowed for the illustration of the signaling gene expression distribution across the groups.

## WGCNA analysis

WGCNA, a method used for the construction of gene co-expression networks in GSE65682, was facilitated by the WGCNA package in R. The steps for processing were as outlined: initially, genes with missing values were filtered out using the 'goodSamplesGenes' function. An optimal soft-thresholding power was then visually selected to ensure a robust network construction. Subsequently, the gene expression data were transformed into an adjacency matrix, and this was further converted into a topological overlap matrix (TOM) to map out genetic interconnections. By examining TOM dissimilarities, genes were clustered using average linkage hierarchical clustering. The clustering dendrogram was dynamically cut to delineate highly correlated modules. The module eigengenes (MEs) served as the representative core of each gene cluster, capturing the module's overall gene expression profile. The association between MEs and clinical traits was assessed using Pearson correlation to establish their relevance. In conclusion, the focus was on genes within modules that exhibited the strongest correlation to the sphingolipid score for downstream investigation.

## Construction and validation of the risk scoring

To conduct a univariate examination of the intersecting genes to uncover those statistically linked to the patient's overall survival rate, a significance threshold of $P<0.05$ was adopted. This analysis was implemented using R, initiating with data preparation which involved importing the gene expression and survival data into R. After ensuring proper data structuring with the 'survival' package, univariate Cox proportional hazards regression was employed. This facilitated the identification of genes with significant prognostic value based on their P-values being less than 0.05 through the coxph function.

Subsequent refinement involved leveraging the LASSO (Least Absolute Shrinkage and Selection Operator) Cox regression analysis, carried out using the 'glmnet' package. Here, a matrix was constructed from the expression data of the significant genes identified from the univariate analysis, and a corresponding response vector containing survival times and event status was prepared. With the LASSO method being sensitive to the values' scales, standardization was ensured before model fitting. The model

fitting was performed using the cv.glmnet function to identify the optimal parameters via cross-validation, focusing on the lambda value that minimized the cross-validated error (17, 18). Through coef, the best set of genes and their associated risk coefficients, having significant associations with patient outcomes, were selected.

For survival analyses, the log-rank (Mantel-Cox) methodology was operationalized to find the gene group with the most significant prognostic value. This process was facilitated through the survdiff function, which compared survival curves across different gene expression groups, and the gene group achieving the lowest P-value was noted.

Risk scores for each sepsis patient were subsequently calculated from the coefficients derived from the log-rank test. These scores allowed for stratification of patients into high-risk and low-risk groups based on the median value of the risk scores, ensuring clear demarcation between the two cohorts.

Kaplan-Meier plots, generated using the survfit function from the 'survival' package, visually represented the survival probabilities over time for both risk groups, providing a clear prognosis evaluation through survival curves. To further scrutinize the predictive model's performance, ROC (Receiver Operating Characteristic) curves were constructed using the 'pROC' package, focusing on the measurement of sensitivity and specificity across varying thresholds.

Finally, the robustness and generalizability of the derived prognostic signature were assessed across four independent datasets. The model's Area Under the Curve (AUC) values were calculated using roc function from the 'pROC' package, serving as a critical validation measure to confirm the model's consistent performance across different patient cohorts.

## Assessment of the prognostic model

To estimate the 28-day overall survival probabilities, a predictive nomogram was constructed, which includes age, gender, and a composite risk score as separate prognostic determinants. To assess the predictive precision of the nomogram, calibration plots were generated. Additionally, the clinical utility and added value of the nomogram were evaluated through decision curve analysis (DCA), comparing its net benefit to the use of clinical characteristics in isolation.

## Enrichment analysis

Utilizing the "clusterProfiler" R package, as previously specified in the literature (8), we executed enrichment analyses for the KEGG and GO. The scope of the GO biological function covered three domains: BP, MF, and CC. To determine statistical relevance, p-values less than 0.05 were identified as significant.

Furthermore, the Gene Set Variation Analysis (GSVA) was conducted using the 'GSVA' R package to elucidate the heterogeneity of biological processes and the activity of various pathways (19). For GSVA, hallmark gene sets from the MSigDB database were selected as the targeted gene sets. The "limma" R package was instrumental in identifying significant differences between biological functions and signaling pathways, with the threshold for statistical significance set to GSVA scores exceeding an absolute t-value of 2.

Additionally, gene set enrichment analysis (GSEA) was conducted to probe the differences in pathway activities, using "clusterProfiler" (20). Pathway activities were ranked according to the Normalized Enrichment Score (NES), with a p-value below 0.05 maintained as the criterion for statistical significance.

Lastly, the activity scores of key disease-related signaling pathways across different cohorts were assessed using the progeny R package, with p-values under 0.05 considered to ascertain statistical significance.

## Assessing the scores of different phenotypes

To discern the distinct phenotypic signatures—namely, those relevant to cholesterol efflux, lysosomal activity, endoplasmic reticulum (ER) stress, angiogenesis, phagocytic function, hypoxic response, acute inflammation, autophagy, and ferroptosis—pertinent gene markers were retrieved from the Molecular Signatures Database (MSigDB). Subsequently, we employed the AUCell algorithm, applying its standard parameters, to calculate phenotype-associated scores across various groups. This process was facilitated by utilizing the irGSEA package.

## Sepsis immunity

The levels of immune cell infiltration were analyzed utilizing the ssGSEA method incorporated within the GSVA softwere (9). In essence, the relative proportions of diverse immune cells were quantified across all samples by leveraging universally recognized gene markers. Subsequently, these algorithms were implemented to ascertain the degree of enrichment or relative quantities for each category of the immune cell. Assessment of the variations in immune cell infiltration across different groups was performed using the Wilcoxon rank-sum test. To depict the extent of immune cell penetration within each AD specimen, divided by algorithm, heatmaps served as a visual aid. Furthermore, the "ESTIMATE" R script played a role in deducing the levels of immune infiltration in patients afflicted with sepsis. Moreover, immune checkpoints consist of an array of molecules such as those involved in antigen presentation, cellular adhesion, co-inhibition, co-stimulation, ligand engagement, and receptor activity—found on immune cells—which modulate the intensity of the immune response. As critical regulators in averting overactive immune responses, we scrutinized and contrasted the expression rates of renowned immune checkpoint genes between the cohorts.

## Establishment and verification of a sepsis rat model with altered PIM1 expression

To investigate the role of PIM1 in sepsis, two cohorts of Sprague-Dawley male rats weighing 250-300g were developed. These animals were raised in a controlled environment with

regulated temperature, humidity, and a 12-hour light/dark cycle, and were given unrestricted access to food and water. The animal procedures were approved by the animal ethics committee of Fujian Medical University. The sepsis condition was induced via the well-established cecal ligation and puncture (CLP) technique, which was performed under aseptic conditions and after administering anesthesia (50 mg/kg sodium pentobarbital intraperitoneally). The cecum was ligated, punctured while preserving intestinal continuity, and then returned to the abdomen. Sham-operated rats received all surgical interventions except the CLP procedure. Postoperative care included rehydration through subcutaneous administration of saline. After 24 hours post-operatively, whether CLP or sham, the rats were sedated, and peripheral blood was drawn from the heart into EDTA tubes, subsequently centrifuged, and the samples were preserved for future examination.

For a detailed study on the role of PIM1 in sepsis, a model with diminished PIM1 expression was additionally created via *in vivo* silencing. Adenoviral vectors containing shRNA sequences that specifically target the PIM1 gene in rats (shPIM1) were employed, in comparison with a non-targeting control shRNA sequence (shNC), both of which were procured from RiboBio, located in Guangzhou, China. The experimental group rats were injected via the tail vein with about 30 billion PFU of shPIM1 in 200 μL saline, whereas the control group received an equivalent dosage of shNC. The potency of gene suppression was evaluated on the 14th day following injection through qRT-PCR. Blood RNA isolated with Trizol reagent was subjected to qRT-PCR with PIM1-specific primers for quantitative expression analysis. On the day of analysis, sepsis was induced in the genetically altered subjects, and blood samples were taken using the same collection and preservation method as before for further analysis.

## RT-qPCR

Peripheral blood samples were used to isolate total RNA employing Trizol reagent (Life Technologies, USA). The isolated RNA was subsequently reverse-transcribed to cDNA using the RevertAid First Strand cDNA Synthesis Kit, following the manufacturer's protocol. Quantitative RT-PCR analyses were performed with the ABI PRISM 7500 system (Applied Biosystems, USA) using the SYBR Premix EX Taq (Takara, Japan) kit. The relative quantification of mRNA expression levels was achieved by normalizing the CT values of the target gene against those of β-actin, with results presented as relative fold changes calculated by the comparative 2-ΔΔCT method.

## Enzyme-linked immunosorbent assay

Peripheral rat blood was collected and the concentrations of cytokines IL-17A, IL-6, TNF-α, and IL-10 were measured using ELISA following protocols supplied by R&D Systems, USA. The blood samples were centrifuged at 2000g for 10 minutes and the supernatants were subsequently harvested for analysis. A reagent diluent was dispensed into each microplate well before the addition of either a blood sample or a standard control. The microplates were sealed and incubated for 2 hours at ambient temperature. After incubation, the contents of the wells were discarded, and the wells were washed thrice. A conjugate reagent (100 μL) was then added to each well, followed by a secondary incubation at room temperature for 2 hours. A subsequent aspiration and washing step was performed before the addition of 100 μL of substrate solution to each well. After a 20-minute incubation, the enzymatic reaction was halted with 50 μL of stop solution. Optical densities at 450 nm were immediately recorded using a spectrophotometer. Cytokine concentrations were quantified against established standard curves.

## Statistical analysis

The R platform was utilized for the management and calculation of our dataset and statistical figures. The survival comparison across the two cohorts was conducted by analyzing Kaplan-Meier plots in conjunction with a log-rank assessment. The 'ggsurvplot' package in R facilitated the construction of all survival plots. Prognostic determinants were assessed through univariate Cox regression. The Lasso technique within Cox regression was applied to pinpoint factors with a more substantial impact on the outcomes. Visualization of data points was conducted using ggplot2 in R, while overall survival computations were performed with the survival package. To deduce the association between a pair of continuous variables, Spearman's rank correlation was executed. The disparities in continuous data between the cohorts were probed via either the Wilcoxon rank-sum test or the two-tailed t-test. Chi-square assessments were put to use for the analysis of categorical variable differences between groups. All statistical evaluations were conducted within the R environment. A P-value below 0.05 was regarded as a threshold for statistical significance.

# Result

## scRNA-seq analysis of GSE175453

The methodology of this study was delineated in a flowchart (Figure 1). scRNA-seq analysis was employed to extensively characterize the immune cell landscape within the dataset GSE175453. After quality control, a total of 40,584 high-quality cells were obtained, with 22,196 cells derived from healthy controls and 18,388 from sepsis samples, all deemed appropriate for further analysis. Figure 2A illustrates the distribution of cell clusters in GSE175453, revealing 15 clusters and 11 immune cell types, categorized as follows: Neutrophils (CD3FR-marked), CD4^+ T cells (CD4-marked), CD8^+ T cells (CD8B-marked), Natural Killer (NK) cells (GNLY-marked), megakaryocytes (TUBB1-marked), macrophages (C1QA-marked), B cells (MS4A1-marked), dendritic cells (DCs; FCER1A-marked), mast cells (CPA3-marked), plasma cells (DERL3-marked), and monocytes (VCAN-marked) with respective cell counts shown in Supplementary Figure S1. The distribution of these cell clusters within each sample, control group, and sepsis group is depicted in Figures 2B-D,

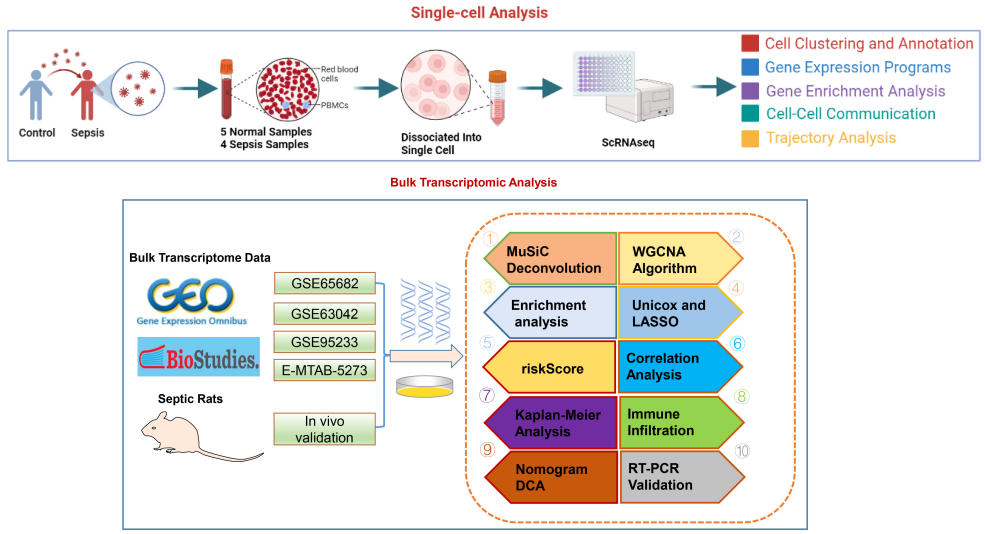



FIGURE 1
The study flow chart.

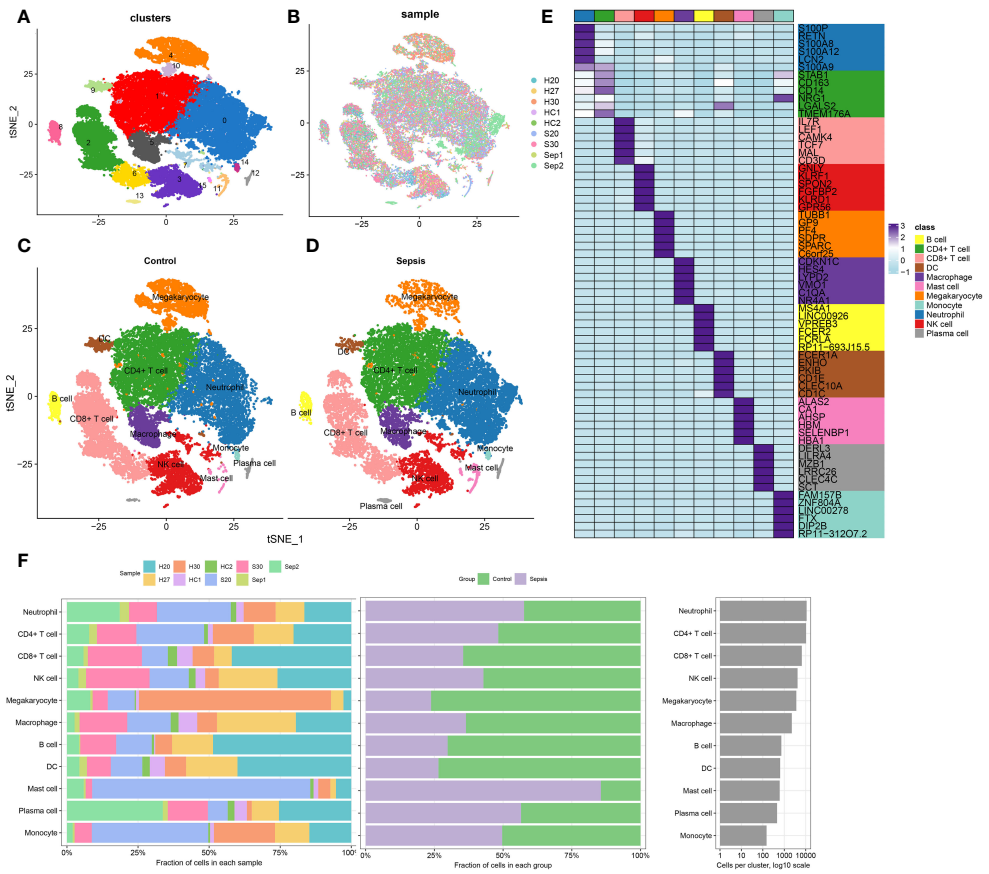



FIGURE 2
scRNA-seq cell annotation. **(A)** The UMAP plot display distribution of the cell clusters of GSE175453. **(B)**The UMAP plot display distribution of the cell clusters of 5 Healthy control and 4 late septic patients. **(C)** The UMAP plot display distribution of the cell types of Healthy control. **(D)** The UMAP plot display distribution of the cell types of late sepsis patients. **(E)** A heatmap displayed the distribution of the top 6 differentially expressed genes specific to different cell subtypes. **(F)** Cell type fractions of 5 Healthy control and 4 late septic patients.

respectively. Furthermore, Figure 2E illustrates the six most characteristic genes for each cell type, and Figure 2F depicts the proportional representation of each cell type across all samples in dataset GSE175453.

## Evaluation of metabolic activity at single-cell resolution

In this section, the metabolic activities of individual cells in transcriptomic dataset GSE175453 are analyzed. Diverse cell types exhibited enrichment in distinct metabolic pathways, reflecting their unique metabolic roles in the context of sepsis. To summarize, B cells are associated with the one-carbon pool by folate metabolism, CD4+ T cells with drug metabolism involving other enzymes and the pentose phosphate pathway, and CD8+ T cells with propanoate metabolism, as well as cysteine and methionine metabolism. Dendritic cells (DC) were linked to oxidative phosphorylation, glycolysis/gluconeogenesis, drug metabolism involving other enzymes, and cysteine and methionine metabolism. Macrophages were noted for oxidative phosphorylation, while mast cells were involved in riboflavin metabolism, porphyrin, and chlorophyll metabolism, phosphonate and phosphinate metabolism, nitrogen metabolism, and fatty acid biosynthesis. Megakaryocytes were related to glutathione metabolism and arachidonic acid metabolism, monocytes to pantothenate and CoA biosynthesis, neutrophils to the pentose phosphate pathway, natural killer (NK) cells to fatty acid elongation, and plasma cells to propanoate metabolism, phenylalanine metabolism, oxidative phosphorylation, N-Glycan biosynthesis, and cysteine and methionine metabolism (Figure 3A). The metabolic pathway activity for each cell type is presented in Figure 3B. Neutrophils, CD8+ T cells, B cells, and monocytes demonstrated relatively low metabolic activity, whereas NK cells, plasma cells, and mast cells exhibited higher activity. Notably, compared to the control group, the metabolic activity in the immune cells from the sepsis group was significantly reduced (Figure 3C). Moreover, the differentially enriched pathways among the global cell subtypes are depicted in Figure 3D. Additionally, variations in classical phenotypes between the

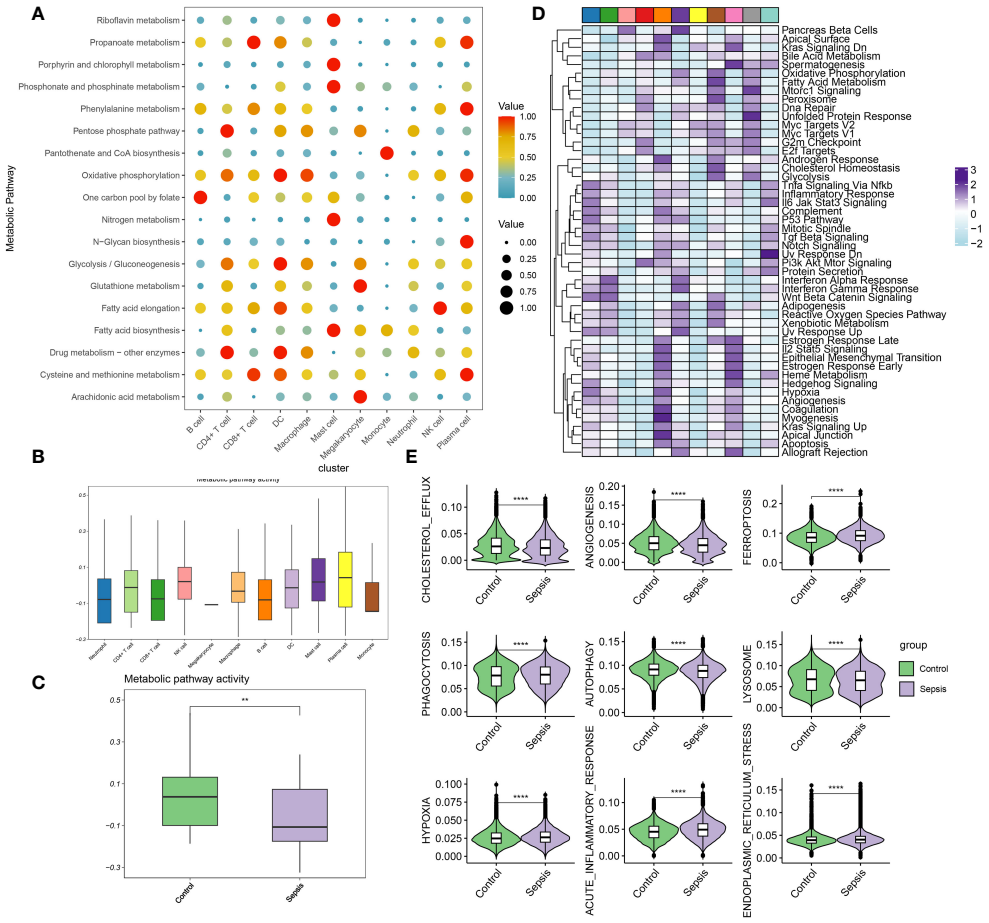



FIGURE 3
Evaluation of metabolic activity at single-cell resolution **(A)** Dot plots showing differentially metabolic pathways among the global cell subtypes. **(B)** Boxplot showing the metabolic pathway activity among the global cell subtypes. **(C)** Boxplot showing the metabolic pathway activity between control and sepsis group. **(D)** Heatmap showing the differentially enriched pathways among the global cell subtypes. **(E)** Boxplots showing phenotypic scores (cholesterol efflux, lysosome, endoplasmic reticulum stress, angiogenesis, phagocytosis, hypoxia, acute inflammatory response, autophagy, and ferroptosis) between control and sepsis groups. **p < 0.01; ****, $P < 0.0001$.

control and sepsis groups were analyzed, revealing that phenotypes such as cholesterol efflux, angiogenesis, phagocytosis, autophagy, and lysosome activity were more pronounced in the control group, whereas hypoxia, acute inflammatory response, and endoplasmic reticulum stress were predominantly observed in the sepsis group (Figure 3E). In conclusion, the findings indicated that metabolic activity is suppressed during sepsis. Among the cell types studied, neutrophils exhibit the lowest metabolic activity, suggesting that neutrophil function may critically regulate metabolic processes in the context of sepsis.

## Intercellular communication analysis of neutrophils in sepsis

The distribution of cell subgroups in the bulk transcriptome dataset GSE65682 was estimated within the single-cell set using the MuSiC algorithm. Notably, it was observed that neutrophils were most prominently enriched in the sepsis group, correlating significantly with metabolic activity (Figure 4A). Therefore, in subsequent analyses, Neutrophil was separately extracted for further analysis. In the subsequent analysis, cellular interactions between neutrophils and other cell types were investigated in both control and sepsis groups. As illustrated in Figure 4B, a greater number of inferred interactions between neutrophils and other cells were observed in the control group, whereas the interaction strength, depicted in Figure 4C, was found to be weaker. In the control group, Neutrophils showed intensive interaction strength and large interaction number with CD4+ T cell, CD8+ T cell, B cell, plasma, macrophage, and Neutrophil (Figure 4D). In the sepsis group, Neutrophil displayed strong interaction strength and large interaction number with CD4+ T cell, CD8+ T cell, NK cell, B cell, macrophage, megakaryocyte, DC, mast cell, monocyte, plasma, and Neutrophil (Figure 4E). The significant ligand-receptor pairs between neutrophils and other cell types were subsequently further explored. Functions as a ligand, Neutrophil strongly increased the activity of RETN-CAP1 to interact with the majority of receptor cells (CD4+ T cell, CD8+ T cell, NK cell, macrophage, megakaryocyte, B cell, DC, and plasma cell) in the sepsis group and lightly decreased the activity of ANXA1-FPR1 interact with CD4+ T cell (strong), GRN-SORT1 interact with macrophage (light), TNFSF13B-TNFRSF17 interact with plasma cell (light). While Neutrophil only up-regulated TNFSF13B-TNFRSF17 as ligands to interact with plasma cell in control group and decreased the activity of ANXA1-FPR1 interact with CD4+ T cell (strong), MIF-(CD74+CXCR4) interact with CD8+ T cell and plasma cell (light), GRN-SORT1 interact with macrophage (light), MIF-(CD74+CD44) interact with B cell (strong) (Figure 4F). Nevertheless, while acting as a receptor, Neutrophils connected with plasma cells by up-regulating MIF-(CD74+CD44) (strong) in

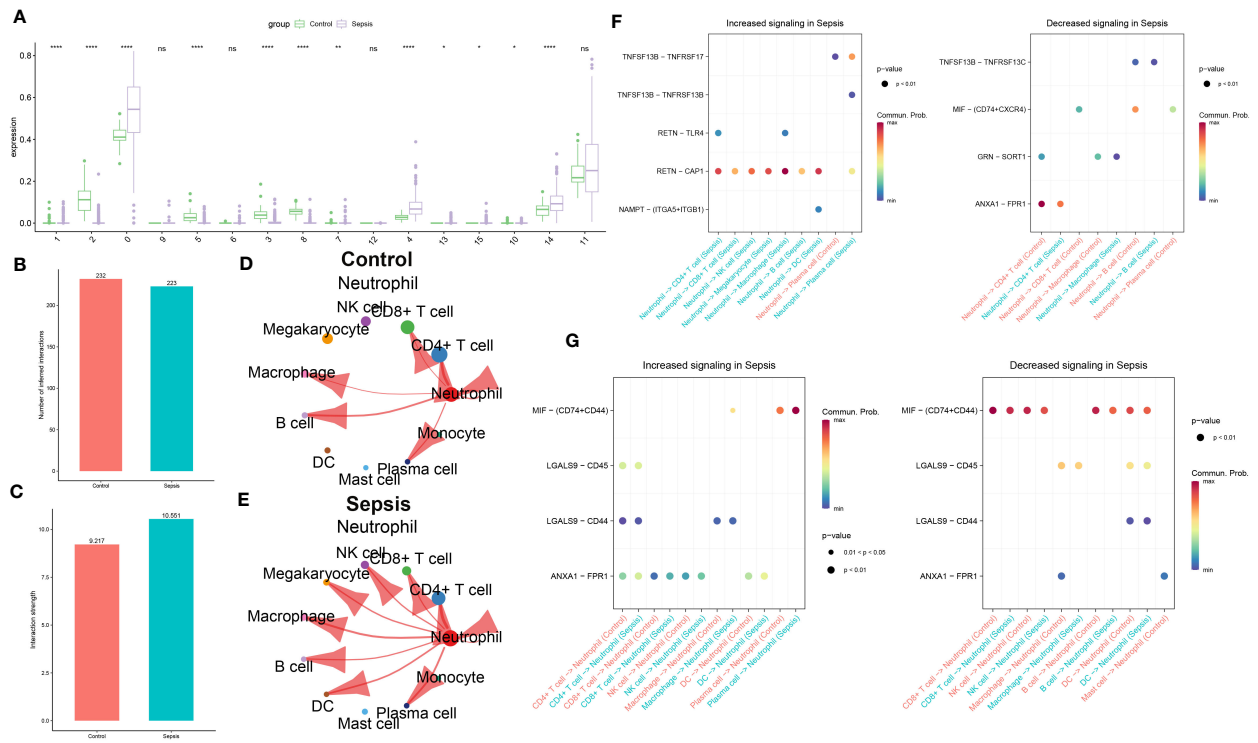



FIGURE 4
Intercellular communication analysis of Neutrophil in sepsis. **(A)** The distribution of cellular subpopulations within the single-cell cluster in the bulk transcriptome dataset GSE65682. The bar graph illustrates the numbers of inferred interactions **(B)** and interaction strength **(C)** between neutrophils and other cells in the control and sepsis groups in GSE175453. **(D)** The strength of interaction between Neutrophil and other cells in the control group in GSE175453. **(E)** The strength of interaction between Neutrophil and other cells in the sepsis group in GSE175453. **(F)** The neutrophil functions as a ligand that mediates intercellular communication in sepsis. **(G)** Neutrophils functions as a receptor in the pathogenic intercellular communication in sepsis.

both sepsis and control group, but connected with CD8+ T cell, NK cell, B cell, and DC by down-regulating MIF-(CD74+CD44) in both sepsis and control group (Figure 4G). In this section, intensive communication between neutrophils and other cell types, particularly within the sepsis group, was observed.

## The development trajectory of neutrophils from control and sepsis samples

To further elucidate the dynamics of the immune response, a pseudoprime developmental trajectory analysis was conducted on neutrophils, with the objective of fitting the most optimal trajectory curve of cellular development or differentiation in sepsis. This analysis inferred the lineage structure of neutrophils within the atherosclerotic plaque milieu based on the developmental trajectory. As time advanced, the pseudotime analysis delineated the principal evolutionary trajectory of neutrophils, which bifurcated into two unique cellular fates (Figure 5A). Subsequently, the developmental trajectories of neutrophils were segregated into control and sepsis groups. Predominantly, neutrophils from the control group were clustered within cellular fate 1, whereas those from the sepsis group were sparsely distributed between both cellular fate 1 and 2 (Figure 5B). Neutrophils from the sepsis group were classified into three distinct differentiation states (Figure 5C). Furthermore, the differential expression of specific genes (S100A9, VCAN, and IFITM2) was validated within the sepsis trajectory. Of these, S100A9 showed high expression in state 3, with VCAN being chiefly expressed in states 1 and 3, and IFITM2 uniformly present in all three states (Figure 5D). The trajectories of lineage-dependent gene expression patterns, accompanying cellular transformations, were further visualized in Figure 5E. CytoTRACE predictions suggested neutrophils in states 2 and 3 possess a higher potential for differentiation in sepsis, contrasting with those in state 1 who showed minimal potential (Figure 5F). Finally, the phenotypes present within the three cellular states were illustrated, with state 3 incorporating the widest spectrum of phenotypes and state 2 encompassing the narrowest (Figure 5F).

## Identification of characteristic genes

The WGCNA algorithm was employed to construct a gene co-expression network for GSE65682. By using an optimal soft-thresholding power (β) of 9, a hierarchical clustering algorithm was implemented on the sample data, leading to the identification of nine unique gene co-expression modules, each differentiated by

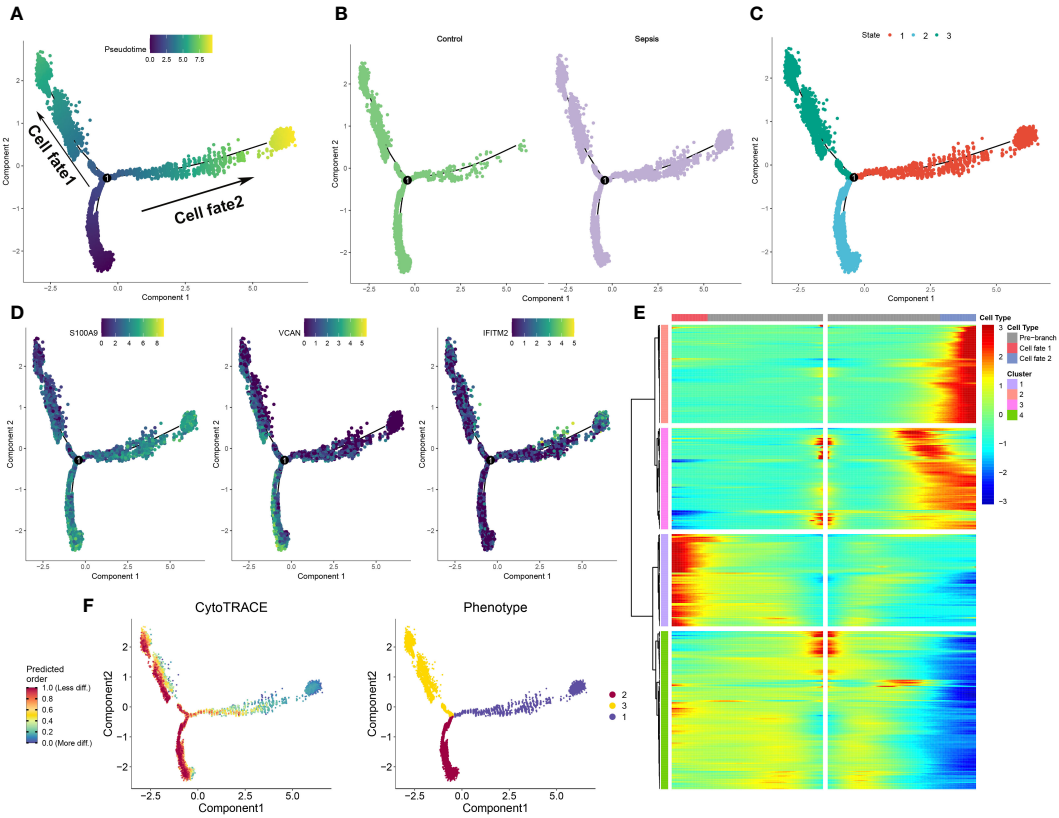



FIGURE 5
Development trajectory of neutrophils from control and sepsis samples. **(A-C)** The developmental trajectory of neutrophils, colored-coded by the pseudotime **(A)**, group **(B)**, and states **(C)**. **(D)** Representative gene expression in neutrophils during sepsis initiation and progression. Intensity of color indicates normalized gene expression. **(E)**Heatmap showing different blocks of DEGs in each cell fate along the pseudotime of sepsis initiation and progression, colored by cell fates. **(F)** Development trajectory of neutrophils colored by the CytoTRACE scores and Phenotype.

color, in the clustering dendrogram (Figures 6A, B). Interestingly, the black module showed the most significant correlation (R=-0.77) with neutrophils, yielding a total of 1,089 genes for further scrutiny (Figure 6C). The development trajectory analysis of neutrophils provided 444 neutrophil Differentiation-Related Genes (NDRGs). Following this, an intersection of data from WGCNA and the trajectory analysis resulted in the recognition of 29 characteristic genes, as portrayed in Figure 6D.

In the next step, enrichment analyses were carried out to illuminate the potential biological functions of these 29 genes. GO analysis revealed that these genes have a wide-ranging involvement in BP, such as the metabolic process of porphyrin-containing compounds and heme. In terms of CC, the genes could be found in ubiquitin ligase complex, cullin-RING ubiquitin ligase complex, and basal plasma membrane, among others. With regards to MF, these genes were involved in activities such as ubiquitin-protein transferase and ubiquitin-like protein transferase (Supplementary Figure S2A). Additionally, the KEGG analysis uncovered substantial enrichment in areas such as bacterial and viral infections, and metabolisms of substances inside and outside cells (Supplementary Figure S2B).

## Construction of neutrophils related riskScore system in sepsis

The univariate Cox proportional hazards analysis was performed on 29 NDRGs, revealing 12 genes that demonstrated a statistically significant association with the overall survival of patients in the bulk sepsis transcriptome data GSE65682 (represented as a univariate analysis hazard ratio [HR]) (Figure 7A). This was followed by the least absolute shrinkage and selection operator (LASSO) Cox regression analysis and the log-rank (Mantel-Cox) tests to refine the identification of survival-associated genes (Figures 7B–E). The analysis culminated in the identification of three hallmark genes (IGSF6, HIST1H1C, and PIM1), based on which the neutrophil-related riskScore model was created. The riskScore calculation is (-0.3196490 × IGSF6) + (0.1483832 × HIST1H1C) + (0.3325431 × PIM1). Patients from the bulk sepsis transcriptome data were divided into high- and low-risk categories using the median riskScore.

## The evaluation of the riskScore system

The effectiveness of the riskScore-based prognosis predictive model was evaluated using survival analysis, exhibiting consistency across all assessments. The prediction accuracy of riskScore reflected robustly in the four datasets: GSE65682, GSE63042, GSE95233, and E-MTAB-5273) with AUC values for 28-day mortality exceeding 0.65 (Figures 8A–D). This high accuracy continued to prevail in combined dataset evaluations for 7, 14, 21, and 28-day mortality, wherein AUC values all surpassed 0.65 (Figure 8E). Further division of sepsis samples into two risk categories, high-risk and low-risk, revealed trends of reduced mean survival periods in high-risk patients, often succumbing in the early illness phase. Low-risk patients revealed a consistent increase of IGSF6 expression as opposed to their high-risk

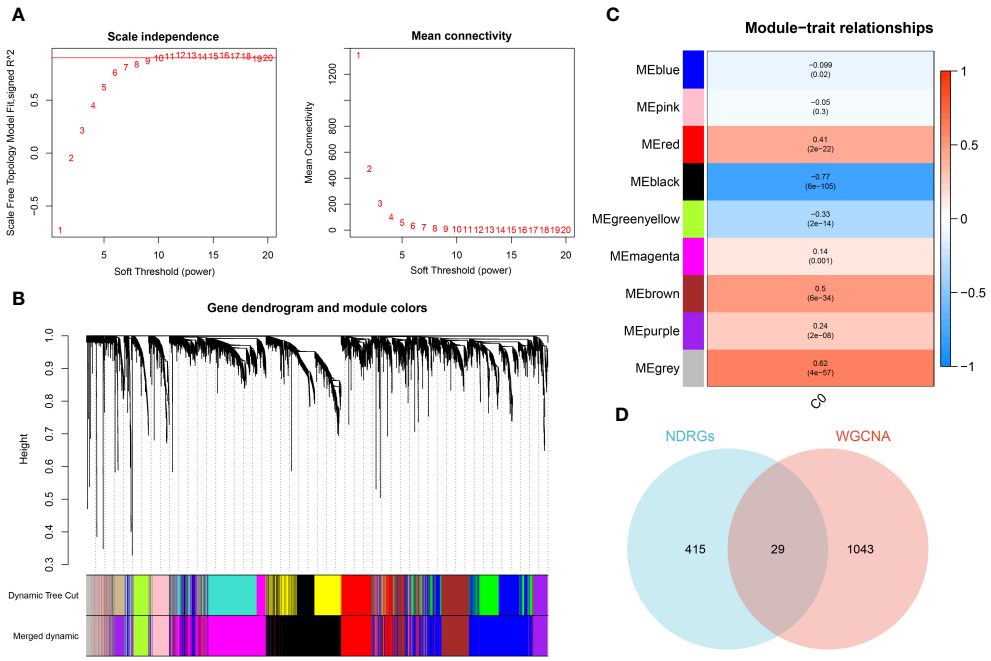



FIGURE 6
Identification of characteristic genes. **(A)** Ideal soft threshold for adjacency computation of WGCNA. **(B)** Dendrogram of co-expression module clustering. **(C)** The WGCNA analysis investigated the modules of with the most remarkable correlation to neutrophils. **(D)** Interaction of characteristic genes screened from WGCNA and development trajectory analysis of neutrophils.

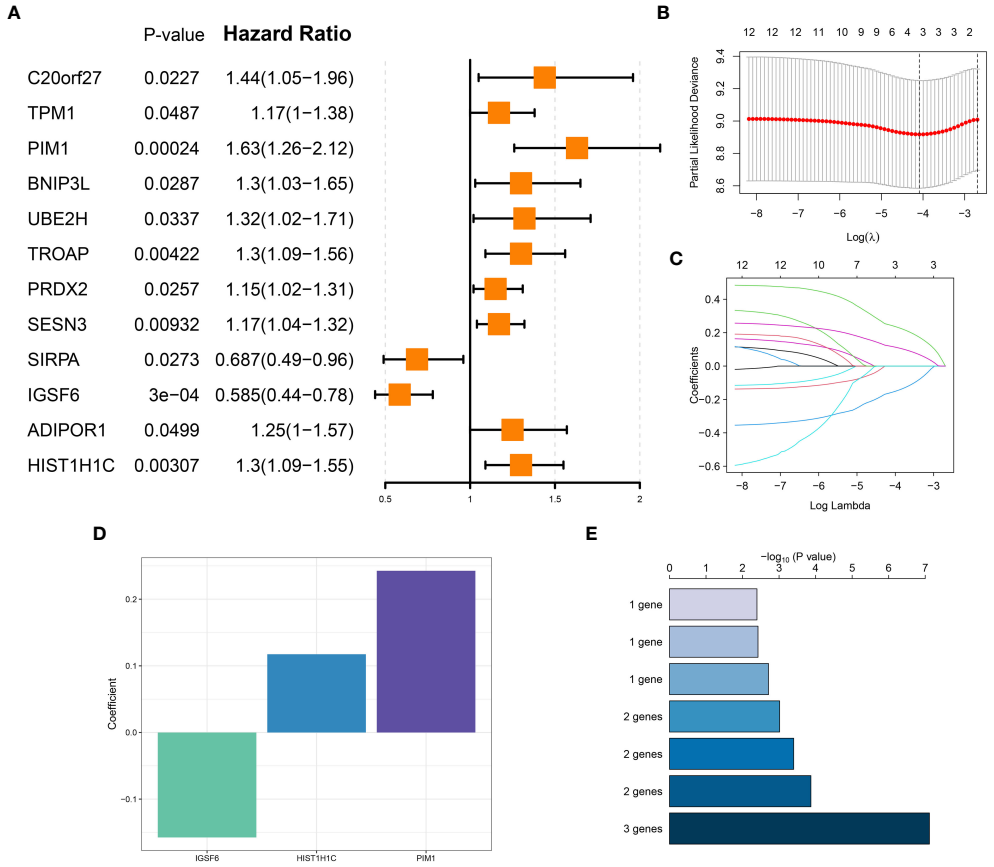



FIGURE 7
Construction of neutrophils related riskScore system in sepsis. **(A)** Univariate cox analysis on the intersection genes. **(B)** Tuning feature selection in the LASSO model. **(C)** LASSO coefficient profiles of the DDR-related characteristic genes. **(D)** The specific coefficient value of the 3 Genes associated with GM identified by the optimal lambda value. **(E)** Kaplan-Meier analysis of gene combinations, the top 7 signatures were ranked and the signature comprising four genes was selected due to its relatively large negative logarithm (-log10) of the p-value combined with a minimal gene count.

counterparts who showed increased expressions of PIM1 and HIST1H1C (Figure 8F). Interestingly, survival analysis reiterated the enhanced survival probabilities for low-risk patients compared to high-risk patients (Figure 8G).

In addition, a prognostic nomogram for sepsis was developed, integrating demographic variables such as age and sex, based on the neutrophils-related risk score. Each predictor in the nomogram warranted a particular score, with the total score across all predictors designating a cumulative score reflecting the likelihood of a negative outcome in sepsis. This cumulative score was visibly represented in Figure 9A. The calibration plot verifies the predictive accuracy of the nomogram as shown in Figure 9B. The clinical applicability of our nomogram, standing on the calculated risk score, was further substantiated by DCA (Figure 9C). Moreover, a schematic representation of the demographic distribution by age, sex, and survival statuses, categorized into two risk groups, has been provided. No significant variation in the age and gender distribution across cohorts was brought to light by this analysis (Figure 9D).

To further elucidate the neutrophil-associated mechanisms in sepsis, characteristic genes of both high- and low-risk groups in GSE65682 were examined. Upon identification, these said genes were put through enrichment analysis using both GSVA and GSEA methods. Divergent pathway enrichment patterns were observed between the two groups. In the high-risk group, notable enrichment was seen in pathways relating to Metabolic Processes, Cellular Stress Responses, and Cell Cycle. This encompassed pathways such as Heme Metabolism, Hypoxia, Oxidative Phosphorylation, Estrogen Response Early, Pi3k Akt Mtor Signaling, Mtorc1 Signaling, E2f Targets, Unfolded Protein Response, Xenobiotic Metabolism, Notch Signaling, Reactive Oxygen Species Pathway, Mitotic Spindle, and P53 Pathway. Conversely, the low-risk group demonstrated significant involvement in several biological functions critical to immuno-inflammatory responses, namely the Interferon Alpha Response, Androgen Response, Apoptosis, Complement, Protein Secretion, Interferon Gamma Response, Allograft Rejection, Jak-Stat3 Signaling, Bile Acid Metabolism, Tnf-a Signaling Via Nf-kb, and Wnt Beta Catenin Signaling pathways (Figure 10A). Proceeding with the investigation, the top 5 up-regulated (Porphyrin And Chlorophyll Metabolism, Nitrogen Metabolism, Nitrogen Metabolism, Purine Metabolism, Ubiquitin Mediated Proteolysis) and top 5 down-regulated pathways (Natural Killer Cell-Mediated Cytotoxicity, B Cell Receptor Signaling Pathway, Nod Like Receptor Signaling Pathway, Cytokine Cytokine Receptor Interaction, Toll-Like Receptor Signaling Pathway) within the high-risk group were further discerned through

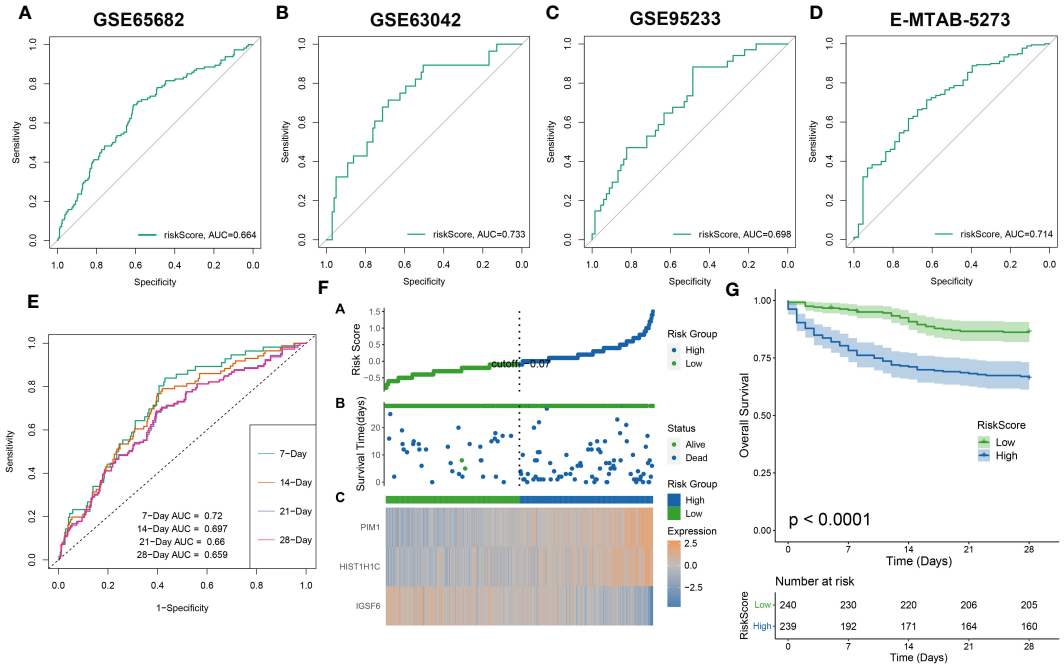



FIGURE 8
The evaluation of the RiskScore system. The ROC curve was used to evaluate the performance of the riskScore model in the GSE65682 **(A)**, GSE63602 **(B)**, GSE95233 **(C)**, E-MTAB-5273 **(D)**, and the combination dateset **(E)**. **(F)** The distribution of the riskscore, patients' survival status as well as gene expression signature in the combination dateset. **(G)** Overall survival situation between the low- and high-risk group.

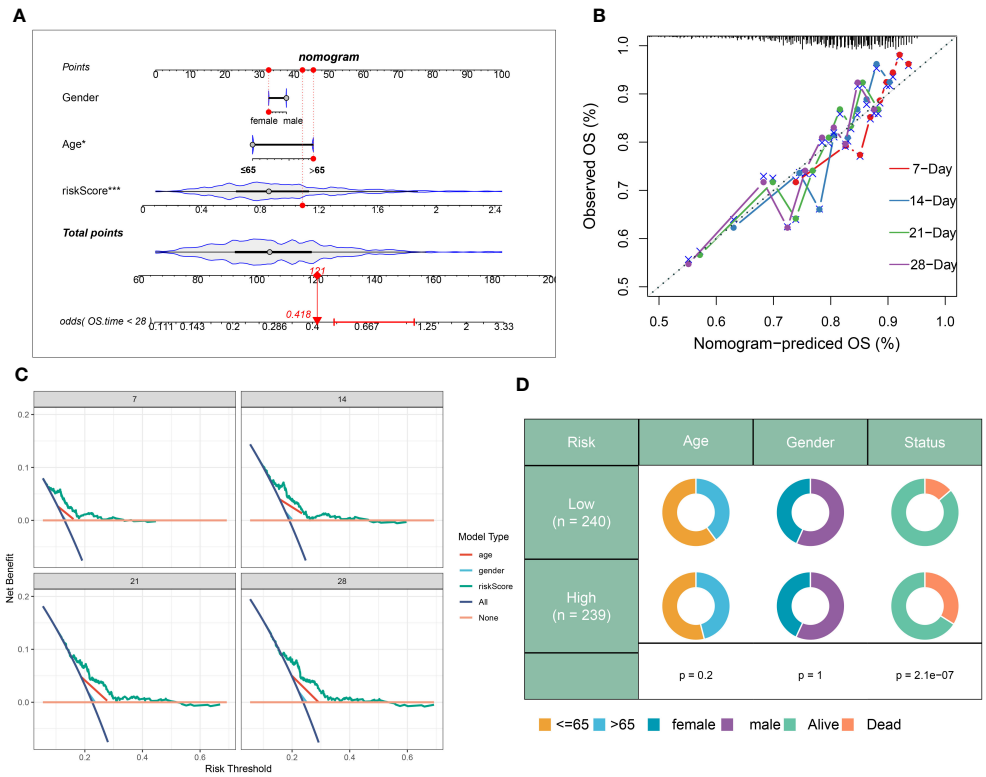



FIGURE 9
Construction and validation of a prognostic prediction model based on the riskScore. **(A)** Construction of a nomogram based on riskScore and clinical characteristics in the combination dateset. **(B)** Correction of the characteristic curve based on riskscore and pathological characteristic. **(C)** DCA indicating the clinical benefit of the nomogram. **(D)** The distribution of clinical features and survival status in the low- and high-risk groups.

GSEA (Figures 10B, C). Pathogenetic pathway variability was also evident among different-risk sepsis patients. Particularly, high-risk patients demonstrated significant activity in the EGFR, Estrogen, and Trail pathways. On the other hand, low-risk patients showed hyperactivity in the WNT, TNF-α, NF-KB, PI3K, and VEGF pathways compared to their high-risk counterparts (Figure 10D).

## Immunological features of sepsis patients at low and high risk

To elucidate the infiltration of immune cells in patients with sepsis categorized into high- and low-risk groups, each further classified by stable or unstable clinical statuses, a comparative analysis of 26 immune cell subtypes was initially performed. This was executed through the calculation of the 26 immune cell scores using the ssGSEA algorithm (Figure 11A). Generally, the low-risk group displayed higher levels of immune cell infiltration compared to the high-risk group, aligning with previous results that exhibited higher immune cell scores in the majority of immune cell types within the low-risk group (Figure 11B). Moreover, the variations in immune modulators between the high- and low-risk groups were evaluated, based on different statuses, genders, and ages, as an attempt to further clarify the immune characteristics of sepsis patients (Figure 11C). In summary, immune genes associated with antigen presentation (HLA-DQA), cell adhesion (SELP), co-inhibitor (CD276 and PDCD1LG2), co-stimulator (ICOSLG), ligand (CCL5, CD40LG, CD70, CX3CL1, and VEGFB), receptor (CD27, EDNRB, IL2RA, LAG3, and PDCD1), among others (PRF1), were visibly elevated in the high-risk group. However, low-risk samples illustrated substantial expression of antigen presentation (HLA-A, HLA-B, HLA-C, MICA, and MICB), cell adhesion (ICAM1 and ITGB2), co-inhibitor (CD274 and SLAMF7), ligand IL1B, TGFB, and TNF), receptor (CD40, HAVCR2, TIGIT, TLR4, and TNFRSF14), among others (ENTPD1) (Supplementary Figures S3A–G). Furthermore, a comparative analysis of immune scores from each risk group was carried out, yielding a comprehensive review of immunological attributes. Patients in the low-risk group achieved higher immune scores compared to those in the high-risk group (Figure 11D). Additionally, a correlation analysis indicated that elevated risk scores negatively affected the entirety of immune cell types and demonstrated higher immune

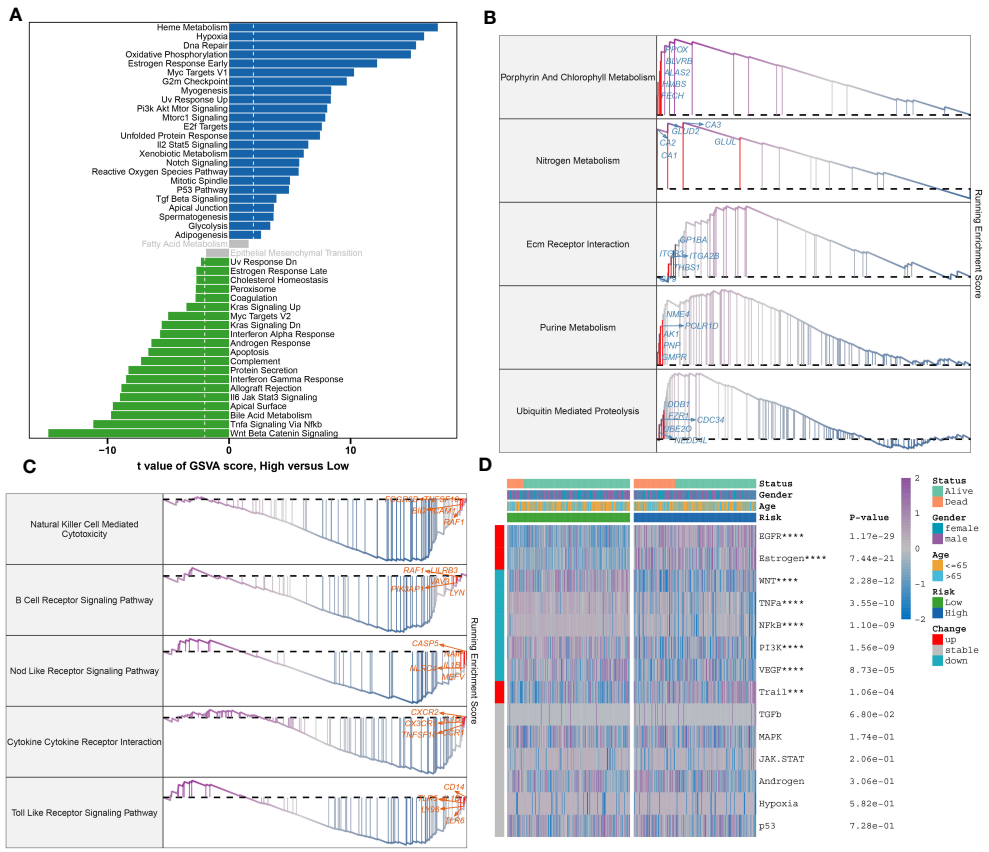



FIGURE 10
Molecular characteristic and functional annotation of the neutrophils-related riskScore model in sepsis. **(A)** The GSVA identified significant differences in biological functions between the high- and low-risk groups. Positive values indicate that the biological function is enriched in the high-risk group, while negative values indicate that the biological function is enriched in the low-risk group. **(B)**Top five up-regulated pathways in the high-risk group. **(C)**Top five pathways down-regulated in the high-risk group. **(D)** Heatmap displaying the difference of pathogenic pathways in sepsis patients at low and high risk. Age, gender, and survival status are displayed as patient annotations. ***$p < 0.001$, ****$p < 0.0001$.

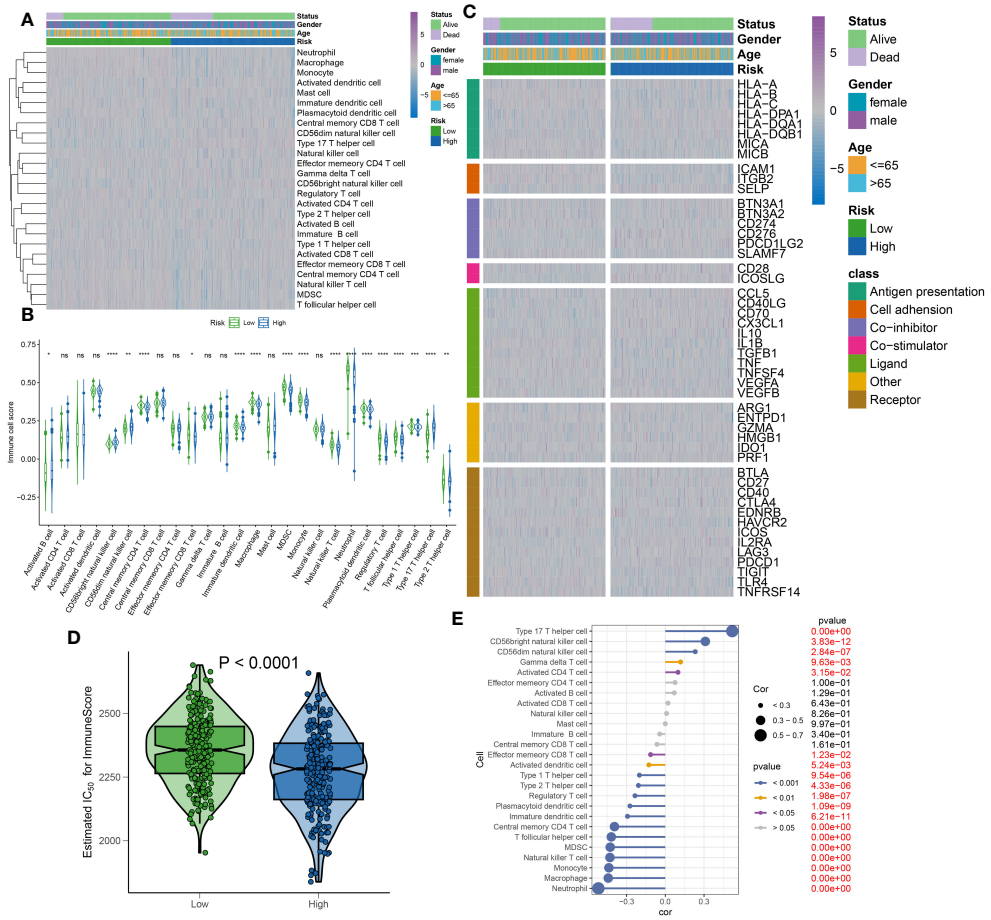



FIGURE 11
Immunological features of sepsis patients at low and high risk. **(A)** The heatmap showing the degree of infiltration of 26 immune cell subtypes in high- and low-risk groups. **(B)** Differences in immune cell scores between high- and low-risk groups. **(C)** Heatmap depicting the differences in immune-modulators and patients' survival status between high- and low-risk groups. **(D)** A comparison of the immuneScore between high- and low-risk groups. **(E)** The interaction between riskScore and 26 immune cell subtypes.

infiltration levels (Figure 11E). Based on the findings of our study, a relative decrease in the activity of immune cells, immune responses, and immune-related pathways was observed in high-risk sepsis patients, indicating a symptomatic immune suppression during sepsis.

## Validation of hallmark genes in a rat model of sepsis

To substantiate the involvement of signature genes in the development of a neutrophil-related risk score model for sepsis, *in vivo* validation experiments were conducted. Initially, a sepsis model was established in rats, followed by an analysis of gene expression in their peripheral blood via reverse transcription-quantitative polymerase chain reaction (RT-qPCR). Among the genes studied, PIM1, HIST1H1C, and IGSF6 demonstrated a marked upregulation in the septic rats (see Figure 12A). PIM1 was selected for in-depth validation due to its significant contribution to the risk score model. Subsequent RT-qPCR evaluations revealed that PIM1 expression in the Sepsis+shPIMI group was reduced to nearly one-third compared with the Sepsis +shNC group, confirming effective gene silencing within our framework (refer to Figure 12B). Survival analyses further showed that rats subjected to PIM1 knockdown presented enhanced survival rates relative to the Sepsis+shNC cohorts (as indicated in Figure 12C). Additionally, the levels of pro-inflammatory cytokines, such as IL-17A, IL-6, and TNF-α, were notably elevated in the Sepsis+shPIMI group, whereas the anti-inflammatory cytokine IL-10 was reduced (depicted in Figure 12D). These findings identify PIM1 as a potential pivotal modulator of immune and inflammatory responses during sepsis. The section that follows will provide additional evidence of the critical role played by PIM1 in sepsis, suggesting its potential involvement in the immunosuppressive mechanisms of the disease.

# Discussion

During sepsis, metabolic changes in the patient's body not only contribute to early inflammation and organ damage but also play significant roles in immune tolerance and immune exhaustion (11).

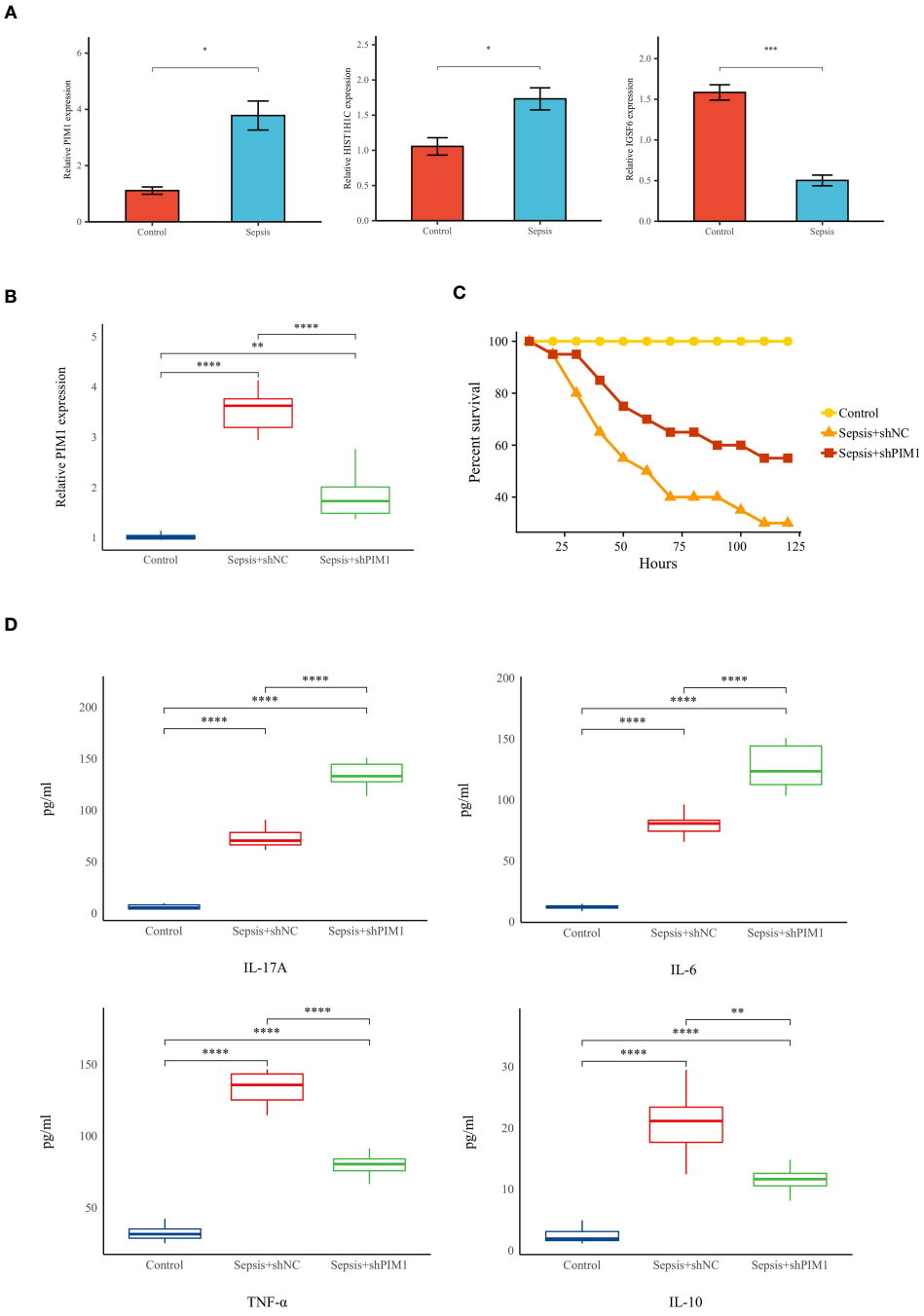



FIGURE 12
Validation of hallmark genes in a rat model of sepsis **(A)** Relative expression levels of hallmark genes in Control and Sepsis groups (n=5 in each group). **(B)**Relative expression levels of PIMI in Control, Sepsis+shNC, and sepsis+shPIMI group (n=8 in each group). **(C)** Survival status of rats in each group (n=10 in each group). **(D)**The level of pro-inflammatory cytokines (IL-17A, TNF-a, and IL-6) and anti-inflammatory cytokines (IL-10) in the peripheral blood of rats in each group (n=8 in each group). *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$, ****$p < 0.001$.

Sepsis is marked by significant metabolic dysregulation across various pathways, including carbohydrate, amino acid, and fat metabolism (21). Leukocytes from patients with severe sepsis exhibited profound defects in cellular energy metabolism, which were correlated with a diminished capacity to respond to secondary stimulation (11). In the pathogenesis and progression of sepsis, further research is needed to elucidate the intricate mechanisms and heterogeneity of various immune cells influencing metabolism. Such research is crucial for establishing a robust theoretical foundation to advance personalized clinical interventions for sepsis.

In this study, scRNA-seq was utilized to delineate the immune landscape in both healthy controls and patients with late-stage sepsis. We identified distinct distributions of immune cells and metabolic activity profiles between the groups. Remarkably,

immune cells from septic patients exhibited a broadly reduced metabolic activity compared to those from healthy controls, likely due to metabolic exhaustion related to the severe inflammatory response during sepsis. The observed hypometabolic state in sepsis may serve as a protective mechanism against excessive inflammation or energy depletion, or it may indicate a dysfunctional immune response. Among all immune cell types, neutrophils played a pivotal role in the immune response and metabolic activity during sepsis. Neutrophils were the most abundantly expressed and demonstrated the lowest metabolic pathway activity in the septic group, while also interacting significantly with other immune cells. During sepsis, neutrophils exhibited enhanced longevity and reduced migratory capabilities, leading to their retention within the vascular system. Consequently, this promotes excessive vascular inflammation through the secretion of cytokines, reactive oxygen species, and neutrophil extracellular traps (22). As the first responders of the innate immune defense against infection, neutrophils utilize traditional mechanisms such as phagocytosis alongside the release of inflammatory cytokines and ROS. In addition to these mechanisms, activated neutrophils release web-like structures comprising decondensed DNA, histones, myeloperoxidase, and other granular contents, known as neutrophil extracellular traps (NETs), which effectively ensnare bacteria within the bloodstream (23). Although the prevailing response of immune cells in sepsis is to undergo apoptosis, thus promoting an immunosuppressive environment, neutrophils uniquely exhibit delayed apoptosis, further perpetuating the inflammatory response (24). Polymorphonuclear neutrophils possess limited mitochondria and predominantly rely on the comparatively inefficient process of glycolysis for their energy metabolism, which is responsible for generating the bulk of ATP needed for neutrophil functionality (25, 26). During phagocytosis, there is an elevated consumption rate of ATP, and in the context of sepsis, systemic ATP levels can impede neutrophil activation and chemotaxis by disrupting intrinsic purinergic signaling pathways (27). Nevertheless, the specific metabolic traits and immunomodulatory routes of neutrophils during sepsis remain inadequately explored.

This research revealed that intercellular communication demonstrates a complex interaction network between neutrophils and other cell types, potentially underlying the septic process. Ligand-receptor analyses indicated active crosstalk between neutrophils and other cell types during sepsis, highlighting elevated levels of specific proinflammatory mediators in the septic milieu. Moreover, the exploration of developmental trajectories suggested neutrophil plasticity in sepsis, with distinct phenotypes correlating to varying sepsis severity. This plasticity likely represented an adaptive response to the multifaceted stimuli encountered during sepsis. Moreover, 29 neutrophil differentiation-related genes during sepsis were obtained by intersecting feature genes from WGCNA and trajectory analysis. Then we acquired 3 hallmark genes (IGSF6, HIST1H1C, and PIM1) by machine learning approaches, and the neutrophils-related riskScore model consisting of 3 genes was constructed. The immunoglobulin superfamily member IGSF6 was involved in immune regulation and has been linked to the immunological landscape of tumors (28). IGSF6 expression was associated with the infiltration of CD8+ T cells and CD4+ T cells in tumors, indicating an active immune response within the tumor microenvironment (29). Some studies reported the involvement of IGSF6 in the immunoregulation of atherosclerosis and inflammatory bowel disease (30, 31). A recent study identified that IGSF6 regulates ER stress and the inflammatory response in intestinal macrophages. IGSF6 expression is sustained by microbiota and significantly upregulated upon bacterial infection (32). HIST1H1 proteins bind to nucleosomes and facilitate chromatin compaction 1, although their biological functions are poorly understood. According to a recent authoritative study, HIST1H1 was identified as a bona fide tumor suppressor and show that mutations in H1 drive malignant transformation primarily through three-dimensional genome reorganization, which leads to epigenetic reprogramming and derepression of developmentally silenced genes (33). Moloney murine leukemia virus-1 (PIM1) functions as a kinase influenced by cytokine signaling, and its role is particularly pivotal in the context of IFN-γ signaling pathways during infections (34). It appears to act as a sensor detecting a wide array of pathogens that disrupt IFN-γ signaling. PIM1 has a short lifespan within infected cells. PIM1 appears to play a regulatory role in the immune response by controlling the parasiticidal function of GBP1. The regulation of GBP1's antimicrobial function by PIM1 suggests that this interaction is a part of an IFNγ-induced pathway which provides post-translational control of innate immune defense (35). Additionally, PIM1 also promotes the survival and immunosuppressive function of neutrophils during chronic viral infection, influencing CD8 T cell function and viral control (36).

The neutrophil-related riskScore system reflects vital prognostic information and predicts patient outcomes informatively in sepsis. This aligns with studies advocating for personalized medicine approaches based on immune profiling. Based on the risk scoring, patients were stratified into high-risk and low-risk groups. GSVA and GSEA highlighted marked functional disparities between high- and low-risk sepsis patients. The high-risk group was associated with enrichment of metabolic pathways and stress responses, potentially indicative of the metabolic demands of a sustained inflammatory response. In contrast, the low-risk group demonstrated enrichment in immune functions, suggesting less compromised immune responses. Assessment of immune cell infiltration and immune-modulators unveiled a robust immune phenotype in low-risk patients, likely contributing to the effective response against infection. In contrast, high-risk patients exhibited a subdued immunological profile, which may predispose them to adverse outcomes. In all, in this study, we identify the high-risk group as "immune suppression phenotype", while the low-risk group is "Immunoactive type".

Research has established that sepsis-induced immunosuppression stems from dysfunctions in both innate and adaptive immunity. This condition is marked by elevated levels of anti-inflammatory cytokines, the apoptosis of immune cells, T-cell dysfunction, and a heightened presence of immuno-regulatory cells such as regulatory T

cells and myeloid-derived suppressor cells (37, 38). Immunological suppression associated with inflammation is a critical determinant in the onset of secondary infections and multiple organ dysfunction syndrome (MODS), which are chief contributors to the adverse prognoses observed in septic patients (39). The results are consistent with the conclusions drawn in this study. It was demonstrated that within the neutrophil-based risk model, patients classified in the high-risk group exhibited significant immunosuppression and metabolic dysregulation. This finding indicates the potential utility of using neutrophil-based metrics as immunological prognostic markers to aid in risk assessment and to identify potential therapeutic targets.

The three genes constituting the risk score were further validated, revealing that their expression levels were significantly higher in the peripheral blood of sepsis-induced rats compared to the control group. Additional experimental validation was subsequently performed on PIM1, the gene with the highest risk coefficient. The inhibition of PIM1 resulted in a significant increase in the inflammatory levels in the peripheral blood of the septic rats. Moreover, the survival rate of the septic rats in the PIM1 knockdown group was higher than that of the septic rats in the control group. These experimental findings were consistent with the conclusions of our previous bioinformatics analysis, confirming that PIM1 may be one of the critical genes involved in the immune suppression observed following the onset of sepsis. These findings highlight the heterogeneity in immune responses among sepsis patients and suggest that a personalized medicine approach, informed by detailed immunophenotyping, could lead to more tailored and effective treatment strategies. Understanding the role of specific immune cells and their metabolic pathways in sepsis may open avenues for the development of immunomodulatory therapies aimed at restoring immune balance rather than just controlling the infection.

Throughout the duration of the research, a series of challenges were encountered, and unexpected discoveries were made: (1) Inter-individual Variability: considerable variability in immune responses and metabolic profiles among sepsis patients was observed. This variability underscores the complexity of sepsis as a syndrome and indicates the potential necessity for personalized therapeutic approaches. (2) Unanticipated dynamic changes in neutrophil subpopulations were revealed by pseudotime analysis. Certain neutrophil states demonstrated unexpected gene expression patterns, suggesting novel roles in the immune response to sepsis. (3) Unexpected interactions between different metabolic pathways, typically studied in isolation, were identified. This cross-talk implies more intricate metabolic reprogramming in immune cells during sepsis than previously recognized.

Several limitations need to be acknowledged in the present study. First, the sample size of sepsis patients included in the scRNA-seq was relatively small, potentially impacting the generalizability of our findings. Future research with larger cohorts is essential to validate the robustness of the identified biomarkers and the risk score model across diverse populations. Second, the current analysis primarily focuses on neutrophils, which, although critical, represent only a fraction of the complex immune response in sepsis. Expanding the scope to include a broader range of immune cells and their interactions will provide a more comprehensive understanding of sepsis immunopathology.

## Conclusion

A comprehensive analysis has provided insights into the complex immune cell interactions and functional pathways associated with metabolic dysregulation in sepsis, with a particular emphasis on neutrophils. Distinct neutrophil subpopulations and their dynamic differentiation patterns have been discovered, contributing to the understanding of immune response variability in sepsis. Key diagnostic biomarkers, including PIM1, HIST1H1C, and IGSF6, have been identified and incorporated into an accurate riskScore model for the prognosis of sepsis. This model stratifies patients into risk categories and provides insights into immune dysfunction associated with poor outcomes. Furthermore, PIM1 has been experimentally validated as a negative regulator of immune-inflammatory response, indicating its therapeutic potential. These findings collectively enhance the understanding of sepsis immunopathology and offer promising directions for prognosis and treatment interventions.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

## Ethics statement

This study was approved by the Institutional Animal Care and Use Committee of The Second Affiliated Hospital of Fujian Medical University and conducted in accordance with the Guidelines for the Care and Use of Laboratory Animals." (No humans were accessed in this study).

## Author contributions

SJ: Writing – original draft. HZ: Writing – review & editing. QL: Writing – review & editing. JY: Writing – review & editing. RZ: Writing – review & editing. ZX: Writing – review & editing. WS: Writing – review & editing, Writing – original draft.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. Natural Science Foundation of Fujian Province (2021N003S).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1398719/full#supplementary-material

## References

1. Bauer M, Gerlach H, Vogelmann T, Preissing F, Stiefel J, Adam D. Mortality in sepsis and septic shock in Europe, North America and Australia between 2009 and 2019- results from a systematic review and meta-analysis. *Crit Care (London England)*. (2020) 24:239. doi: 10.1186/s13054-020-02950-2

2. Vincent J-L. Current sepsis therapeutics. *EBioMedicine*. (2022) 86:104318. doi: 10.1016/j.ebiom.2022.104318

3. Lelubre C, Vincent J-L. Mechanisms and treatment of organ failure in sepsis. *Nat Rev Nephrol*. (2018) 14:417–27. doi: 10.1038/s41581-018-0005-7

4. Venet F, Monneret G. Advances in the understanding and treatment of sepsis-induced immunosuppression. *Nat Rev Nephrol*. (2018) 14:121–37. doi: 10.1038/nrneph.2017.165

5. Venet F, Demaret J, Gossez M, Monneret G. Myeloid cells in sepsis-acquired immunodeficiency. *Ann New York Acad Sci*. (2021) 1499(1):3-17. doi: 10.1111/nyas.14333

6. Zou L, Chen HH, Li D, Xu G, Feng Y, Chen C, et al. Imaging lymphoid cell death *in vivo* during polymicrobial sepsis. *Crit Care Med*. (2015) 43:2303–12. doi: 10.1097/CCM.0000000000001254

7. Wasyluk W, Zwolak A. Metabolic alterations in sepsis. *J Clin Med*. (2021) 10 (11):2412. doi: 10.3390/jcm10112412

8. Hotchkiss RS, Monneret G, Payen D. Sepsis-induced immunosuppression: from cellular dysfunctions to immunotherapy. *Nat Rev Immunol*. (2013) 13:862–74. doi: 10.1038/nri3552

9. Irahara T, Sato N, Otake K, Matsumura S, Inoue K, Ishihara K, et al. Alterations in energy substrate metabolism in mice with different degrees of sepsis. *J Surg Res*. (2018) 227:44–51. doi: 10.1016/j.jss.2018.01.021

10. Stienstra R, Netea-Maier RT, Riksen NP, Joosten LAB, Netea MG. Specific and complex reprogramming of cellular metabolism in myeloid cells during innate immune responses. *Cell Metab*. (2017) 26:142–56. doi: 10.1016/j.cmet.2017.06.001

11. Liu J, Zhou G, Wang X, Liu D. Metabolic reprogramming consequences of sepsis: adaptations and contradictions. *Cell Mol Life Sci CMLS*. (2022) 79:456. doi: 10.1007/s00018-022-04490-0

12. Arts RJW, Gresnigt MS, Joosten LAB, Netea MG. Cellular metabolism of myeloid cells in sepsis. *J Leukocyte Biol*. (2017) 101:151–64. doi: 10.1189/jlb.4MR0216-066R

13. Chapman NM, Boothby MR, Chi H. Metabolic coordination of T cell quiescence and activation. *Nat Rev Immunol*. (2020) 20:55–70. doi: 10.1038/s41577-019-0203-y

14. Singer M, De Santis V, Vitale D, Jeffcoate W. Multiorgan failure is an adaptive, endocrine-mediated, metabolic response to overwhelming systemic inflammation. *Lancet (London England)*. (2004) 364:545–8. doi: 10.1016/S0140-6736(04)16815-3

15. Wu Y, Yang S, Ma J, Chen Z, Song G, Rao D, et al. Spatiotemporal immune landscape of colorectal cancer liver metastasis at single-cell level. *Cancer Discovery*. (2022) 12:134–53. doi: 10.1158/2159-8290.CD-21-0316

16. Qiu X, Mao Q, Tang Y, Wang L, Chawla R, Pliner HA, et al. Reversed graph embedding resolves complex single-cell trajectories. *Nat Methods*. (2017) 14:979–82. doi: 10.1038/nmeth.4402

17. Gulati GS, Sikandar SS, Wesche DJ, Manjunath A, Bharadwaj A, Berger MJ, et al. Single-cell transcriptional diversity is a hallmark of developmental potential. *Sci (New York N.Y.)*. (2020) 367:405–11. doi: 10.1126/science.aax0249

18. Zeng D, Ye Z, Wu J, Zhou R, Fan X, Wang G, et al. Macrophage correlates with immunophenotype and predicts anti-PD-L1 response of urothelial cancer. *Theranostics*. (2020) 10:7002–14. doi: 10.7150/thno.46176

19. Hänzelmann S, Castelo R, Guinney J. GSVA: gene set variation analysis for microarray and RNA-seq data. *BMC Bioinf*. (2013) 14:7. doi: 10.1186/1471-2105-14-7

20. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci United States America*. (2005) 102:15545–50. doi: 10.1073/pnas.0506580102

21. De Waele E, Malbrain MLNG, Spapen H. Nutrition in sepsis: A bench-to-bedside review. *Nutrients*. (2020) 12(2):395. doi: 10.3390/nu12020395

22. Sônego F, Castanheira FVES, Ferreira RG, Kanashiro A, Leite CAVG, Nascimento DC, et al. Paradoxical roles of the neutrophil in sepsis: Protective and deleterious. *Front In Immunol*. (2016) 7:155. doi: 10.3389/fimmu.2016.00155

23. Zhang H, Wang Y, Qu M, Li W, Wu D, Cata JP, et al. Neutrophil, neutrophil extracellular traps and endothelial cell dysfunction in sepsis. *Clin Trans Med*. (2023) 13: e1170. doi: 10.1002/ctm2.1170

24. Zhu C-L, Wang Y, Liu Q, Li H-R, Yu C-M, Li P, et al. Dysregulation of neutrophil death in sepsis. *Front In Immunol*. (2022) 13:963955. doi: 10.3389/fimmu.2022.963955

25. Richer BC, Salei N, Laskay T, Seeger K. Changes in neutrophil metabolism upon activation and aging. *Inflammation*. (2018) 41:710–21. doi: 10.1007/s10753-017-0725-z

26. Döring Y, Libby P, Soehnlein O. Neutrophil extracellular traps participate in cardiovascular diseases: Recent experimental and clinical insights. *Circ Res*. (2020) 126:1228–41. doi: 10.1161/CIRCRESAHA.120.315931

27. Burn GL, Foti A, Marsman G, Patel DF, Zychlinsky A. The neutrophil. *Immunity*. (2021) 54:1377–91. doi: 10.1016/j.immuni.2021.06.006

28. Verschueren E, Husain B, Yuen K, Sun Y, Paduchuri S, Senbabaoglu Y, et al. The immunoglobulin superfamily receptome defines cancer-relevant networks associated with clinical outcome. *Cell*. (2020) 182(2):329–344.e19. doi: 10.1016/j.cell.2020.06.007

29. Rong Y-M, Xu Y-C, Chen X-C, Zhong M-E, Tan Y-X, Liang Y-F, et al. IGSF6 is a novel biomarker to evaluate immune infiltration in mismatch repair-proficient colorectal cancer. *Sci Rep*. (2023) 13:20368. doi: 10.1038/s41598-023-47739-9

30. Shen Y, Xu L-R, Tang X, Lin C-P, Yan D, Xue S, et al. Identification of potential therapeutic targets for atherosclerosis by analysing the gene signature related to different immune cells and immune regulators in atheromatous plaques. *BMC Med Genomics*. (2021) 14:145. doi: 10.1186/s12920-021-00991-2

31. Bates EE, Kissenpfennig A, Péronne C, Mattei MG, Fossiez F, Malissen B, et al. The mouse and human IGSF6 (DORA) genes map to the inflammatory bowel disease 1 locus and are embedded in an intron of a gene of unknown function. *Immunogenetics*. (2000) 52:112–20. doi: 10.1007/s002510000259

32. Wu Y, Zhang P, Shi T, Cao D, Pan W. Deficiency of immunoglobulin IgSF6 enhances antibacterial effects by promoting ER stress and the inflammatory response in intestinal macrophages. *Mucosal Immunol*. (2024) 17(2): 288–302. doi: 10.1016/j.mucimm.2024.02.006

33. Yusufova N, Kloetgen A, Teater M, Osunsade A, Camarillo JM, Chin CR, et al. Histone H1 loss drives lymphoma by disrupting 3D chromatin architecture. *Nature*. (2021) 589:299–305. doi: 10.1038/s41586-020-3017-y

34. Ma J, Arnold HK, Lilly MB, Sears RC, Kraft AS. Negative regulation of Pim-1 protein kinase levels by the B56beta subunit of PP2A. *Oncogene*. (2007) 26:5145–53. doi: 10.1038/sj.onc.1210323

35. Fisch D, Pfleiderer MM, Anastasakou E, Mackie GM, Wendt F, Liu X, et al. PIM1 controls GBP1 activity to limit self-damage and to guard against pathogen infection. *Sci (New York NY)*. (2023) 382:eadg2253. doi: 10.1126/science.adg2253

36. Volberding PJ, Xin G, Kasmani MY, Khatun A, Brown AK, Nguyen C, et al. Suppressive neutrophils require PIM1 for metabolic fitness and survival during chronic viral infection. *Cell Rep*. (2021) 35:109160. doi: 10.1016/j.celrep.2021.109160

37. Wiersinga WJ, van der Poll T. Immunopathophysiology of human sepsis. *EBioMedicine*. (2022) 86:104363. doi: 10.1016/j.ebiom.2022.104363

38. Liu D, Huang S-Y, Sun J-H, Zhang H-C, Cai Q-L, Gao C, et al. Sepsis-induced immunosuppression: mechanisms, diagnosis and current treatment options. *Military Med Res*. (2022) 9:56. doi: 10.1186/s40779-022-00422-y

39. van der Poll T, Shankar-Hari M, Wiersinga WJ. The immunology of sepsis. *Immunity*. (2021) 54:2450–64. doi: 10.1016/j.immuni.2021.10.012

frontiers | Frontiers in Immunology

TYPE Original Research
PUBLISHED 13 August 2024
DOI 10.3389/fimmu.2024.1420208



OPEN ACCESS

EDITED BY
Dipak Kumar Sahoo,
Iowa State University, United States

REVIEWED BY
Ashish Patel,
Hemchandracharya North Gujarat University, India
Carlo Maria Rossi,
San Matteo Hospital Foundation (IRCCS), Italy

*CORRESPONDENCE
Craig Smail
csmail@cmh.edu

†These authors have contributed equally to this work and share senior authorship

RECEIVED 19 April 2024
ACCEPTED 26 July 2024
PUBLISHED 13 August 2024

CITATION
Keever-Keigher MR, Harvey L, Williams V, Vyhlidal CA, Ahmed AA, Johnston JJ, Louiselle DA, Grundberg E, Pastinen T, Friesen CA, Chevalier R, Smail C and Shakhnovich V (2024) Genomic insights into pediatric intestinal inflammatory and eosinophilic disorders using single-cell RNA-sequencing. *Front. Immunol.* 15:1420208. doi: 10.3389/fimmu.2024.1420208



# Genomic insights into pediatric intestinal inflammatory and eosinophilic disorders using single-cell RNA-sequencing

Marissa R. Keever-Keigher[1], Lisa Harvey[1], Veronica Williams[2], Carrie A. Vyhlidal[3], Atif A. Ahmed[4], Jeffery J. Johnston[1], Daniel A. Louiselle[1], Elin Grundberg[1,5], Tomi Pastinen[1,5], Craig A. Friesen[1,5], Rachel Chevalier[1,5], Craig Smail[1,5*†] and Valentina Shakhnovich[1,5,6†]

[1]Children's Mercy Kansas City, Kansas, MO, United States, [2]Nemours Children's Health, Jacksonville, FL, United States, [3]KCAS Bioanalytical & Biomarker Services, Shawnee, KS, United States, [4]Seattle Children's Hospitals, University of Washington, Seattle, WA, United States, [5]School of Medicine, University of Missouri-Kansas City, Kansas, MO, United States, [6]Ironwood Pharmaceuticals, Boston, MA, United States

**Introduction:** Chronic inflammation of the gastrointestinal tissues underlies gastrointestinal inflammatory disorders, leading to tissue damage and a constellation of painful and debilitating symptoms. These disorders include inflammatory bowel diseases (Crohn's disease and ulcerative colitis), and eosinophilic disorders (eosinophilic esophagitis and eosinophilic duodenitis). Gastrointestinal inflammatory disorders can often present with overlapping symptoms necessitating the use of invasive procedures to give an accurate diagnosis.

**Methods:** This study used peripheral blood mononuclear cells from individuals with Crohn's disease, ulcerative colitis, eosinophilic esophagitis, and eosinophilic duodenitis to better understand the alterations to the transcriptome of individuals with these diseases and identify potential markers of active inflammation within the peripheral blood of patients that may be useful in diagnosis. Single-cell RNA-sequencing was performed on peripheral blood mononuclear cells isolated from the blood samples of pediatric patients diagnosed with gastrointestinal disorders, including Crohn's disease, ulcerative colitis, eosinophilic esophagitis, eosinophilic duodenitis, and controls with histologically healthy gastrointestinal tracts.

**Results:** We identified 730 (FDR < 0.05) differentially expressed genes between individuals with gastrointestinal disorders and controls across eight immune cell types.

**Discussion:** There were common patterns among GI disorders, such as the widespread upregulation of *MTRNR2L8* across cell types, and many differentially expressed genes showed distinct patterns of dysregulation among the different gastrointestinal diseases compared to controls, including upregulation of *XIST* across cell types among individuals with ulcerative colitis and upregulation of Th2-associated genes in eosinophilic disorders. These findings indicate both overlapping and distinct alterations to the transcriptome of individuals with gastrointestinal disorders compared to controls, which provide insight as to which genes may be useful as markers for disease in the peripheral blood of patients.

KEYWORDS

**single-cell, RNA-seq, gastrointestinal disorder, pediatrics, genomics**

# 1 Introduction

In genetically predisposed individuals, chronic overactivation of the inflammatory response damages tissues along the gastrointestinal (GI) tract frequently resulting in painful and debilitating symptoms (1, 2). GI inflammatory disorders include inflammatory bowel disease (IBD) and eosinophilic gastrointestinal diseases such as eosinophilic esophagitis (EoE) and eosinophilic duodenitis (EoD). IBD is characterized by chronic relapsing neutrophilic inflammation of the intestine and can be divided into two main subtypes based on the site and characteristics of inflammation, with Crohn's disease (CD) occurring within any portion of the gut and ulcerative colitis (UC) being confined to the colon. IBD affects patients of all ages, with approximately a quarter of patients diagnosed before adulthood and incidence of pediatric IBD increasing (3, 4). EoE and EoD are Th-2 mediated inflammatory disorders which can cause dysphagia, vomiting, abdominal pain, and structuring. Histologically, EoE and EoD are characterized by mucosal eosinophilia (5) and may exist independently or as a comorbid condition with either form of IBD. However, eosinophilic infiltration of the mucosa may also precede histologic evidence of IBD (crypt distortion, cryptitis with crypt abscesses, mucus depletion from goblet cells, granulomas), sometimes by years (6, 7), further complicating the ability to differentiate concomitant eosinophilic disease from early harbingers of IBD.

Differentiating between UC and CD or between early IBD and EoE/EoD can help direct therapy. For example, CD patients benefit from early biologic therapy (8), and colectomy is only curative in UC (9). Patients with two concomitant diseases (i.e. EoE and CD) may have symptoms affecting the same area but exhibit different histology and symptoms and require different treatments (10). Additionally, early knowledge of whether mucosal eosinophilia is burgeoning IBD can allow early and appropriate intervention. The pediatric population with IBD and EoE/EoD are particularly vulnerable to growth failure (11, 12) and interruptions in social-emotional development (13) and will require decades of healthcare for their condition. Since disease diagnosis and follow up evaluation currently require invasive endoscopic testing, non-invasive diagnostics and targeted therapies are particularly valuable to this subset of patients.

Characterizing the role of specific immune cell populations in GI diseases has aided in recognizing aberrant processes that underlie these conditions, initiating an important shift in the treatment paradigm away from systemic, non-targeted immunosuppression (fraught with many unwanted side effects) to targeted modulation at the site of disease activity (14). Continued identification of novel therapeutic targets and molecular signatures of disease is pivotal for advancing and optimizing treatment options for chronic immune-mediated inflammatory disorders. In the IBD-affected GI tract, dendritic cells (DCs)—antigen presenting cells belonging to the innate immune system—exhibit up-regulation of microbial recognition receptors and increased cytokine production (15) that appears to induce inflammation through activation of T cells (16). T cells play a crucial role in immune homeostasis (17, 18), and dysregulation of cytokine signaling in $CD4^+$ T cells of the GI tract has been shown to lead to pathogenic inflammation (18, 19). T cells also play a key role in eosinophilic disorders of the GI tract, as overexpression of interleukin 5 (IL-5) in $CD2^+$ T cells is sufficient to produce eosinophilia in the esophagus and small intestine of transgenic mice (20).

In addition to contributing to inflammation and tissue damage at lesion sites in the GI tract, evidence of altered gene expression and signaling among immune cells in peripheral blood may be reflective of luminal inflammation (21–23). Information gathered from peripheral blood has the potential to identify minimally-invasive, diagnosis-specific and/or disease location-specific genetic markers for GI diseases. Discovery of such biomarkers could potentially decrease the need for repeat endoscopy, which is invasive, associated with risks, and costly. Furthermore, identification of altered gene expression within these disorders at the cellular level could yield a more complete understanding of impacted pathways within specific cell types, and aid in characterizing genetic signatures for future use in disease sub-typing and drug response applications. However, due to the complex and multifactorial nature of GI diseases and differences in immune cell response across GI disease sub-types, reliable indicators of active inflammation have been difficult to characterize within the peripheral blood of patients to date.

In this study, we identified cell-type specific differential gene expression and enrichment of functional gene ontology terms and pathways in individuals diagnosed with CD, UC, EoE, and EoD using single-cell RNA-sequencing of peripheral blood mononuclear cells (PBMCs) in a pediatric patient cohort. Results from this study assist in uncovering the genomic landscape of these phenotypes, which often present with overlapping symptoms in patients, and aid in identifying robust markers of disease types within the peripheral blood mononuclear cells of patients.

# 2 Materials and methods

## 2.1 Patient information

Potential study participants were identified via review of the clinical endoscopy schedule and the electronic medical record (EMR) at Children's Mercy Hospital (CMH), a tertiary regional pediatric hospital in the Midwestern United States. To be considered for study inclusion, patients had to be between 1 month and 21 years of age (inclusive), undergoing both upper and lower endoscopy with biopsies for clinical purposes, having a reasonable clinical suspicion for a new diagnosis of immune-mediated inflammatory disease or another clinical indication for undergoing endoscopy (e.g., abdominal pain), and not receiving systemic immunomodulating, immunosuppressive, or biological drugs. Subjects were recruited on the day of procedure, prior to endoscopy. All subjects were fasting at least 8 hours for procedural purposes as part of routine medical care. Only those subjects who provided informed consent (if 18 years of age), or informed assent with parental/legal guardian consent (if under 18 years of age) were included. All research activities were approved by the CMH Institutional Review Board. A total of 35 patients seen in the

CMH operating room for routine endoscopy (Kansas City, MO, USA) were included in the study. Diagnosis of CD was determined by a pediatric gastroenterologist after evaluation of clinical symptoms (e.g. abdominal pain, weight loss, etc.), laboratory data (e.g. anemia, hypoalbuminemia, elevated fecal calprotectin, etc.), and histopathology (e.g. findings of cryptitis, granulomas, etc.). Diagnosis of EoD was determined by a pediatric gastroenterologist after evaluation of clinical symptoms (e.g. abdominal pain, diarrhea, etc.) and histopathology (e.g. duodenal eosinophils >20 eos/hpf) (24). Diagnosis of EoE was determined by a pediatric gastroenterologist after evaluation of clinical symptoms (e.g. dysphagia, vomiting, abdominal pain, etc.) and histopathology (e.g. eosinophils >15/hpf in the esophagus) (24). This cohort consists of 16 males and 19 females ranging in age from 6.17 to 19.25 years with a mean age of 13.3 years. Seven patients were subsequently diagnosed with CD, nine with EoD, ten with EoE, and three with UC. Six patients were identified as controls who had no relevant GI pathology on visual or histologic examination of tissue. Review of individuals' medical charts indicated no bias toward a single drug therapy in any sub-cohort.

## 2.2 PBMC isolation

Up to 4 mL of whole blood was collected from patients in a sodium heparin tube and stored on ice until PBMCs were isolated. Automated PBMC isolation was performed using a STEMCELL Technologies RoboSep-S using the EasySep Direct Human PBMC Isolation Kit (STEMCELL Technologies Cat No. 19654RF) and following the manufacturer's protocol. After PBMC isolation, the resulting cell suspension was centrifuged at 300 x g for 8 min, and the supernatant was carefully aspirated. The cell pellet was resuspended in 1 mL of ACK Lysing Buffer (Thermo Fisher Cat No. A1049201) and incubated at room temperature for 5 min to remove any remaining RBCs. The cell suspension was centrifuged at 300 x g for 8 min, and the supernatant was carefully aspirated. Cells were washed twice with PBS (Thermo Fisher Cat No. 14190144) supplemented with 2% heat-inactivated FBS (GE Healthcare Cat No. SH30088.03HI), and cell count and viability were assessed using a Countess II automated cell counter. An aliquot of 300,000 cells was diluted in a total volume of 200 µL of PBS + 2% FBS and frozen at -80°C for downstream DNA isolation and genotyping. The remaining cells were cryopreserved in aliquots of at least one million cells by centrifuging at 300 x g for 8 min, aspirating the supernatant, and resuspending the cell pellets in Recovery Cell Culture Freezing Medium (Thermo Fisher Cat No. 12648010). The cell suspensions were transferred to cryogenic storage vials and were slow-frozen overnight to a temperature of -80°C in a Corning CoolCell FTS30.

## 2.3 DNA isolation and genotyping

Aliquots of 300,000 PBMCs frozen in PBS + 2% FBS were thawed at room temperature, and DNA was isolated using the Qiagen DNeasy Blood & Tissue Kit (Qiagen Cat No. 69506) according to the manufacturer's protocol. Eluate was concentrated to approximately 50 µL using an Eppendorf Vacufuge Plus, and DNA was quantified using a Qubit dsDNA BR Assay Kit following the manufacturer's protocol. All DNA samples were selected for high-density genotyping using the Illumina Global Screening Array (GSAMD-24v1-0) according to protocols recommended by Illumina.

## 2.4 Cell pooling

Two pools of PBMCs were made. Thawing Medium for PBMC samples consisted of IMDM (ATCC Cat No. 30-2005) supplemented with 10% heat-inactivated FBS, 100 units/mL of penicillin, and 100 µg/mL of streptomycin. For each sample to be thawed, 10 mL of Thawing Medium was prewarmed in a 37°C bead bath. Cells were thawed in groups of up to five samples at a time. The cryovials were placed in a 37°C bead bath. When thawed, the cryovials and 15-mL conical tubes containing Thawing Medium were aseptically transferred to the biosafety cabinet. For each sample, 1 mL of Thawing Medium was added, dropwise, to the cell suspension. The cell suspension was pipette-mixed and then diluted in the remaining 9 mL of Thawing Medium. The thawed and diluted cells were left at room temperature while the remaining cells to be pooled were similarly thawed. When all samples were thawed, the samples were centrifuged at 300 x g for 8 min. The supernatant was carefully aspirated, and the cell pellets were resuspended in 0.5 mL of room-temperature Thawing Medium. All samples were placed on ice and then pooled together. The pool was passed through a 40-µm nylon mesh cell strainer to remove cell aggregates. The pool was centrifuged at 300 x g for 8 min at 4°C, and the supernatant was carefully aspirated. The cell pellet was resuspended in 1 mL of cold Thawing Medium, and cell count and viability were assessed using a Countess II automated cell counter. No fewer than three aliquots per pool were cryopreserved by centrifuging at 300 x g for 8 min at 4°C and resuspending the cell pellets in Recovery Cell Culture Freezing Medium. The cell suspensions were transferred to cryogenic storage vials and were slow-frozen overnight to a temperature of -80°C in a Corning CoolCell FTS30.

## 2.5 Single-cell sequencing

Aliquots from each pool were thawed for scRNA-seq. The 10x Genomics Chromium Single Cell 3′ Reagent Kit v3 was used according to the manufacturer's protocol to target approximately 15,000 cells per scRNA-seq capture. Libraries were sequenced on an Illumina NovaSeq 6000 platform using 2x94 cycle paired-end sequencing.

## 2.6 scRNA-seq alignment and quality control

CellRanger v 4.0.0 (10x Genomics) was used for read alignment to the GRCh38 (2020) reference genome, gene counting, and cell

calling. Demuxlet (25) was used to demultiplex single-cell data, assigning reads back to the patient of origin using VCF files associated with each patient. Additionally, deumuxlet was used to remove data in instances where barcodes were assigned to more than a single cell.

Quality control of cells was performed with the Seurat v 4.4.0 (26) package in R v 4.2.3. Cells with greater than 20% MT-RNA, fewer than 500 UMIs, and fewer than 0.8 $\log_{10}$ genes per UMI were removed. Annotation of the cell types of remaining cells was performed with the Azimuth v 0.4.6 package in R using a previously published PBMC reference (27) to annotate cells to eight broad level one cell types and 30 more specific level two cell types (Supplementary File 1 Table A).

## 2.7 Pseudobulk differential expression analysis and functional analysis

Gene expression data was aggregated by genotype within cell type using AggregateExpression() in the Seurat R package v 4.4.0. Pseudobulk differential expression analysis of the single-cell data was performed on the aggregated count data for each defined cell type with edgeR v 3.40.2 (28) in R v 4.2.3. Patient sex and sequencing pool were added to the statistical model to account for biological variation and batch effects, and a generalized linear model was used to identify differentially expressed genes (DEGs) between each GI disorder group (CD, UC, EoE, and EoD) and controls across eight level one cell types and 29 level two cell types. P-values were adjusted for multiple testing using the Benjamini-Hochberg false discovery rate (FDR), and genes with an FDR < 0.05 were considered significantly differentially expressed.

Functional and pathway analysis for DEGs was carried out using the ToppFun (Transcriptome, ontology, phenotype, proteome, and pharmacome annotations based gene list functional enrichment analysis) tool with default settings from the web-based software ToppGene Suite (http://toppgene.cchmc.org) to identify enriched gene ontology (GO) terms and pathways from databases including KEGG, Reactome, and BioCarta (29). Terms and pathways with a Benjamini-Hochberg FDR < 0.05 and a minimum number of three hits in query list were considered to be significantly enriched.

## 2.8 Protein-protein interaction networks

Networks of interactions of DEGs were generated in STRING v 12.0 (30) to visualize gene relationships and trends within and between cell types and GI disorders. Networks generated with STRING were imported into Cyotscape v 3.10.0 (31) where node color was used to designate the direction of fold change observed in GI disorders compared to controls, with red corresponding to upregulation and blue corresponding to downregulation. Furthermore, functional terms of interest identified to be enriched in STRING were added to networks.

# 3 Results

## 3.1 Cell clustering

A total of 39,622 cells from the 35 individuals in this study passed quality control measures and were mapped to eight level one cell types and 29 level two cell types (Figure 1; Supplementary File 1 Tables A-C).

## 3.2 Characterizing immune cell transcriptomes across gastrointestinal disorders

Comparison of gene expression between individuals diagnosed with GI disorders (CD, UC, EoD, and EoE) and controls yielded 730 (FDR < 0.05) DEGs (Figure 2) across eight level one cell types and 807 (FDR < 0.05) across 25 level two cell types. A list of significant DEGs found across GI disorders within each cell type can be found in Supplementary File 1 Tables D, E, and volcano plots depicting the results of differential expression analysis for between GI disorders and controls within all level one cell types can be found in Supplementary File 2 Figures A-H.

Relatively few genes showed common patterns of dysregulation across all GI disorders within cell types: four genes in B cells and DCs, three genes in $CD4^+$ and $CD8^+$ T cells, and one gene in both NK cells and other T cells (Supplementary File 1 Table D). Among these genes with shared patterns of dysregulation was upregulated *MTRNR2L8* in six of the eight level one cell types (B, $CD4^+$ T, $CD8^+$ T, DC, NK, and other T) (Table 1; Supplementary File 1 Table D).

Additionally, relatively few genes showed similar patterns of differential expression specific to eosinophilic disorders (EoD and EoE) or IBD (CD and UC) compared to controls. Notable among genes that shared expression patterns among eosinophilic disorders compared to controls was the upregulation of *MTRNR2L1* in four level one cell types (B, $CD4^+$ T, $CD8^+$ T, and DC) and upregulation of several cell cycle associated genes in $CD4^+$ T cells (Table 1; Figure 3; Supplementary File 1 Table D). In IBD subtypes, *OR11G2* was upregulated in $CD4^+$ T and $CD8^+$ T cells relative to controls (Table 1; Supplementary File 1 Table D).

The majority of DEGs did not present with dysregulation in common between GI disorder subtypes; thus, many genes had patterns of dysregulation specific to each GI disorder. Notably, in CD, we observed upregulation of *CC8T* in $CD4^+$ T cells, downregulation of *BTN3A2* in $CD4^+$ TCM cells, and upregulation of *ETS2* in DCs. In EoD, we observed upregulation of *FOXM1* and *CCR9* in $CD4^+$ T cells and downregulation of *GADD45B* and *GADD45G* in $CD4^+$ T cells. In EoE, we observed upregulation of *IFNG* in $CD4^+$ T cells; upregulation of *IER2* in B and $CD8^+$ naïve T cells; upregulation of *EGR1* in three level one cell types (B, $CD4^+$ T, and $CD8^+$ T cells) and several level two cell types, including naïve $CD8^+$ T cells and $CD4^+$ TCM cells, and upregulation of both *EGR3* and *EGR4* in B cells. Within UC, we observed downregulation of several genes in $CD8^+$ T cells associated with cytotoxicity and

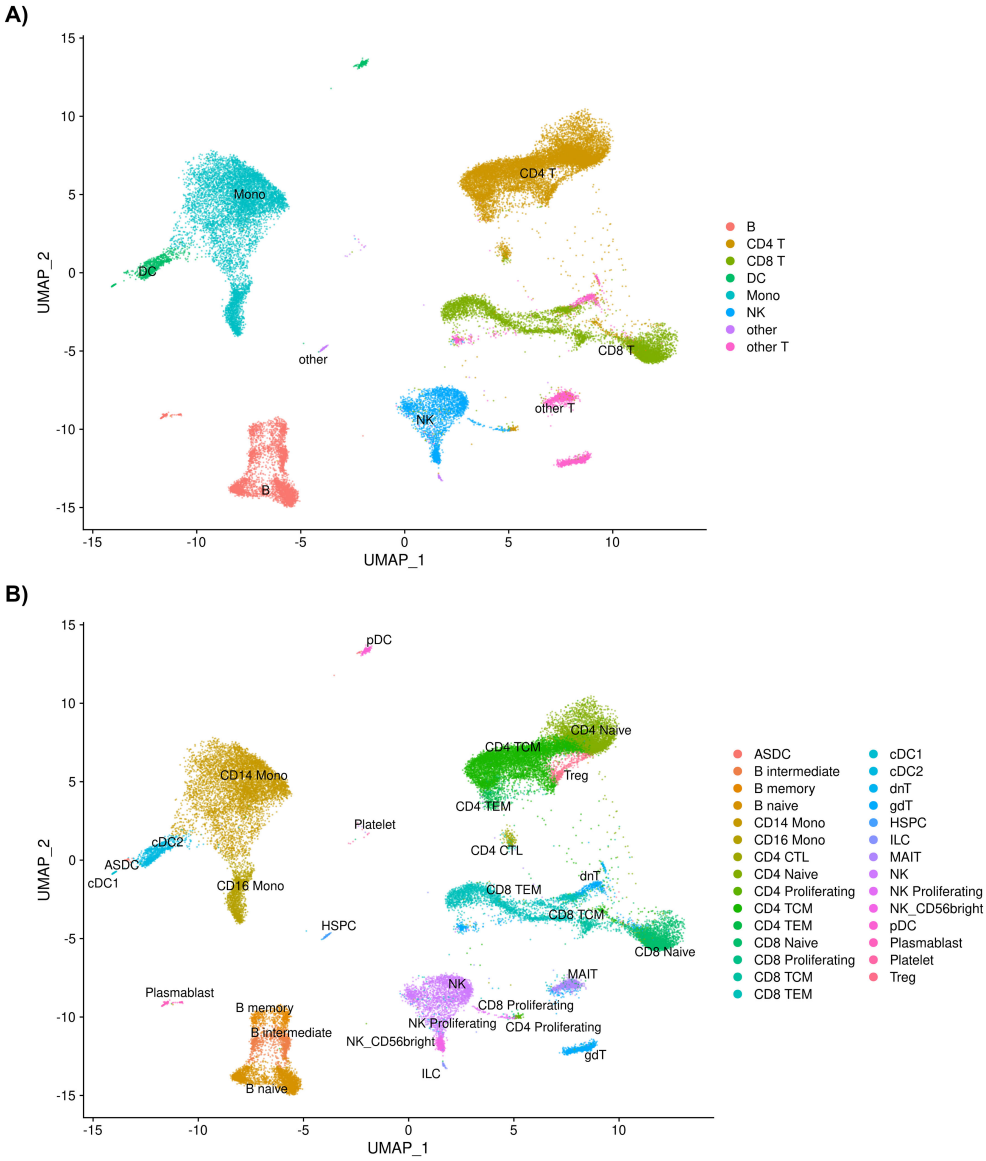



FIGURE 1
**(A)** UMAP of 39,622 cells passing quality control measures sorted into eight level one cell types and into **(B)** 29 level two cell types after annotation using Azimuth.

upregulation *XIST* in five level one cell types (B, CD4$^+$ T, DC, monocytes, and NK) (Table 1; Supplementary File 1 Tables D, E).

A subset of DEGs along with citations of literature supporting their role in GI disorders or in the pathogenesis of inflammation can be found in Supplementary File 1 Table F.

## 3.3 Functional annotation of DEGs associated with gastrointestinal disorders

To better understand the context of gene dysregulation observed within cell types, and uncover which cellular processes may be affected, functional enrichment analysis of gene ontology terms and pathways was performed. Functional analysis of DEGs to identify enriched gene ontology terms and pathways yielded 1037 terms in B cells, 892 terms in CD4$^+$ T cells, 760 terms in CD8$^+$ T cells, 421 terms in DCs, and 62 terms in NK. An abbreviated list of enriched terms for each set of DEGs among GI disorders (CD, UC, EoD, and EoE) within each cell type is shown in Figure 3. A full list of significantly enriched GO terms and pathways can be found in Supplementary File 1 Tables G-K.

GO terms and pathways associated with cell cycle activity in CD4$^+$ T cells were enriched among genes differentially expressed in EoD and EoE compared to controls, including Reactome cell cycle mitotic (M5336), Reactome cell cycle (M543) and mitotic cell cycle process (GO:1903047). Other commonly enriched terms among DEGs found in EoD and EoE included transition metal ion binding (GO:0046914) within NK cells and PID AP1 pathway (M167) within CD4$^+$ T cells (Figure 3; Supplementary File 1 Tables H, K).

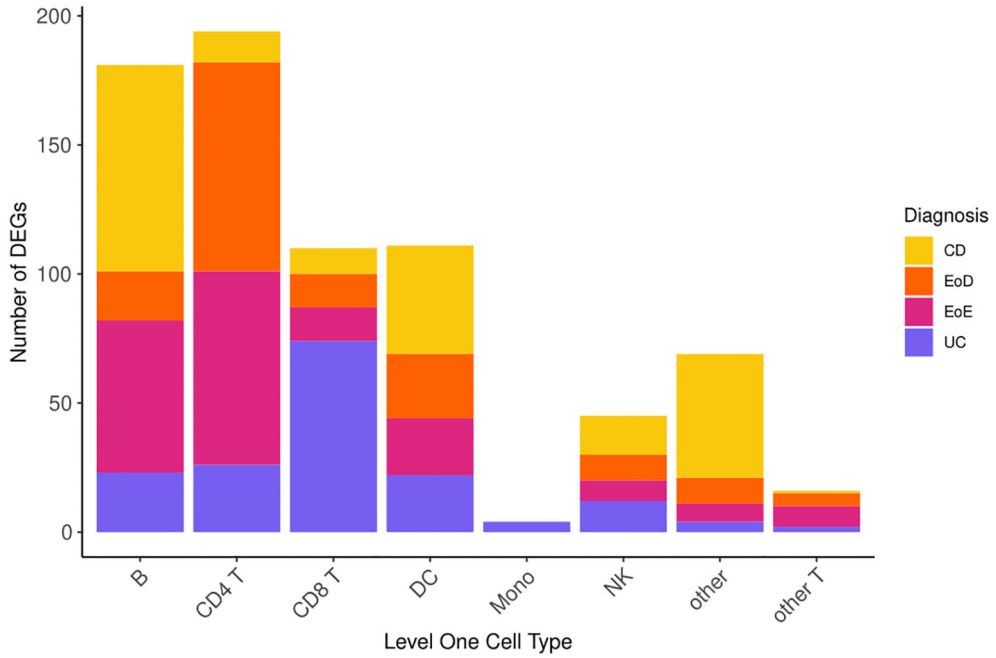



FIGURE 2
Number of differentially expressed genes (DEGs) across level one cell types and GI disorder subtypes.

**TABLE 1** Abbreviated list of differentially expressed genes detected in individuals with Crohn's disease (CD), eosinophilic duodenitis (EoD), eosinophilic esophagitis (EoE), and ulcerative colitis (UC) compared to controls in level one immune cell types: B cells (B), CD4+T cells (CD4 T), CD8+T cells (CD8 T), dendritic cells (DC), monocytes (Mono), natural killer cells (NK), other T cells (other T), and all other cells (other).

| Cell Type | Gene | [a]CD $log_2$FC | [b]CD FDR | EoD $log_2$FC | EoD FDR | EoE $log_2$FC | EoE FDR | UC $log_2$FC | UC FDR |
|---|---|---|---|---|---|---|---|---|---|
| B | MTRNR2L8 | 2.76 | 6.18E-28 | 2.52 | 7.58E-26 | 2.28 | 4.16E-16 | 2.98 | 4.44E-26 |
| | MTRNR2L1 | – | – | 2.02 | 3.31E-02 | 2.63 | 3.35E-08 | – | – |
| | EGR1 | – | – | – | – | 1.46 | 5.42E-09 | – | – |
| | IER2 | – | – | – | – | 0.76 | 8.29E-03 | – | – |
| | XIST | – | – | – | – | – | – | 2.16 | 2.47E-03 |
| CD4 T | MTRNR2L8 | 2.69 | 1.47E-05 | 2.39 | 1.36E-05 | 2.18 | 1.73E-03 | 3.16 | 6.79E-07 |
| | CST3 | -1.27 | 1.47E-05 | -1.38 | 2.15E-10 | -1.12 | 6.57E-06 | -1.33 | 5.11E-04 |
| | CPA5 | -2.14 | 8.39E-05 | -1.78 | 8.82E-05 | -1.85 | 1.22E-05 | -2.99 | 3.77E-04 |
| | OR11G2 | 4.07 | 1.71E-04 | – | – | – | – | 4.15 | 1.63E-03 |
| | RRM2 | – | – | 2.40 | 1.27E-06 | 2.21 | 1.09E-05 | – | – |
| | MTRNR2L1 | – | – | 3.02 | 4.03E-06 | 3.44 | 6.63E-09 | – | – |
| | MKI67 | – | – | 1.69 | 1.80E-04 | 1.65 | 1.58E-04 | – | – |
| | TOP2A | – | – | 1.36 | 4.01E-02 | 1.43 | 1.67E-02 | – | – |
| | CCT8 | 0.67 | 3.24E-03 | – | – | – | – | – | – |
| | CCR9 | – | – | 2.88 | 1.44E-03 | – | – | – | – |
| | GADD45B | – | – | -0.59 | 2.17E-03 | – | – | – | – |
| | GADD45G | – | – | -0.90 | 8.41E-03 | – | – | – | – |
| | EGR1 | – | – | – | – | 1.32 | 5.81E-06 | – | – |

*(Continued)*

TABLE 1 Continued

| Cell Type | Gene | [a]CD $log_2$FC | [b]CD FDR | EoD $log_2$FC | EoD FDR | EoE $log_2$FC | EoE FDR | UC $log_2$FC | UC FDR |
|---|---|---|---|---|---|---|---|---|---|
| | IFNG | – | – | – | – | 1.55 | 6.06E-05 | – | – |
| | XIST | – | – | – | – | – | – | 2.57 | 1.67E-21 |
| CD8 T | MTRNR2L8 | 2.96 | 1.52E-38 | 2.85 | 6.60E-42 | 2.47 | 4.32E-25 | 3.35 | 4.23E-39 |
| | CPA5 | -1.43 | 1.87E-03 | -2.53 | 4.40E-15 | -1.60 | 8.98E-07 | -3.42 | 1.42E-06 |
| | OR11G2 | 4.63 | 2.67E-05 | – | – | – | – | 4.21 | 3.88E-03 |
| | MTRNR2L1 | – | – | 2.93 | 6.71E-06 | 3.42 | 1.95E-09 | – | – |
| | EGR1 | – | – | – | – | 1.05 | 2.54E-04 | – | – |
| | IFNG | – | – | – | – | – | – | -2.86 | 5.26E-11 |
| | NKG7 | – | – | – | – | – | – | -1.35 | 6.70E-06 |
| | FGFBP2 | – | – | – | – | – | – | -1.33 | 1.03E-04 |
| | GZMA | – | – | – | – | – | – | -1.22 | 2.59E-04 |
| | GZMB | – | – | – | – | – | – | -1.34 | 2.59E-04 |
| | GZMH | – | – | – | – | – | – | -1.21 | 3.43E-04 |
| | NOG | – | – | – | – | – | – | 1.71 | 2.66E-03 |
| | CCL5 | – | – | – | – | – | – | -1.01 | 3.26E-03 |
| | REG4 | – | – | – | – | – | – | 1.66 | 1.79E-02 |
| | AIF1 | – | – | – | – | – | – | 0.84 | 2.12E-02 |
| | PRF1 | – | – | – | – | – | – | -1.15 | 2.67E-02 |
| DC | MTRNR2L8 | 3.56 | 2.76E-47 | 3.30 | 1.44E-46 | 3.08 | 3.98E-38 | 3.61 | 4.45E-39 |
| | MTRNR2L1 | – | – | 5.26 | 1.11E-08 | 4.80 | 4.96E-07 | – | – |
| | ETS2 | 1.11 | 4.58E-02 | – | – | – | – | – | – |
| | XIST | – | – | – | – | – | – | 2.36 | 2.92E-04 |
| Mono | XIST | – | – | – | – | – | – | 3.65 | 2.10E-05 |
| NK | MTRNR2L8 | 3.50 | 5.25E-35 | 3.02 | 2.18E-30 | 2.71 | 9.14E-23 | 3.93 | 4.30E-39 |
| | XIST | – | – | – | – | – | – | 2.38 | 8.48E-13 |
| other T | MTRNR2L8 | 3.97 | 2.55E-31 | 3.53 | 7.12E-29 | 3.06 | 1.13E-19 | 4.19 | 8.08E-28 |
| other | MTRNR2L8 | 5.45 | 2.18E-11 | – | – | – | – | 3.87 | 7.50E-04 |

[a]$Log_2$(Fold Change) between Crohn's disease (CD), eosinophilic duodenitis (EoD), eosinophilic esophagitis (EoE), and ulcerative colitis (UC) versus controls.
[b]False Discovery Rate adjusted P-values for genes differentially expressed between Crohn's disease (CD), eosinophilic duodenitis (EoD), eosinophilic esophagitis (EoE), and ulcerative colitis (UC) versus controls.
The symbol "–" indicates that no significant difference in expression was found for this gene in the provided comparison.

Among DEGs in CD, there was an enrichment of GO terms and pathways associated with endoplasmic reticulum function and protein folding in B cells, such as endoplasmic reticulum protein-containing complex (GO:0140534), endoplasmic reticulum chaperone complex (GO:0034663), and Reactome pathways ATF6 alpha activates chaperone genes (M801). Notably, among DEGs in EoD, there was enrichment of terms associated with the p38 MAPK pathway in $CD4^+$ T cells, including p38MAPK cascade (GO:0038066) and regulation of p38 MAPK cascade (GO:1900744). Analysis of DEGs within EoE yielded enrichment of zinc ion binding (GO:0008270) within B cells (Figure 3; Supplementary File 1 Tables G, H).

## 3.4 Protein-protein interaction networks of DEGs

We generated networks for protein interaction among significant DEGs to identify relationships from curated databases and mined from high-throughput studies and primary literature. A selection of the major components of PPI networks generated in STRING and enhanced with Cytoscape illustrating the relationships of DEGs detected within cell types are shown in Figure 4. Additionally, genes associated with enriched gene ontology terms and pathways of interest have been highlighted.

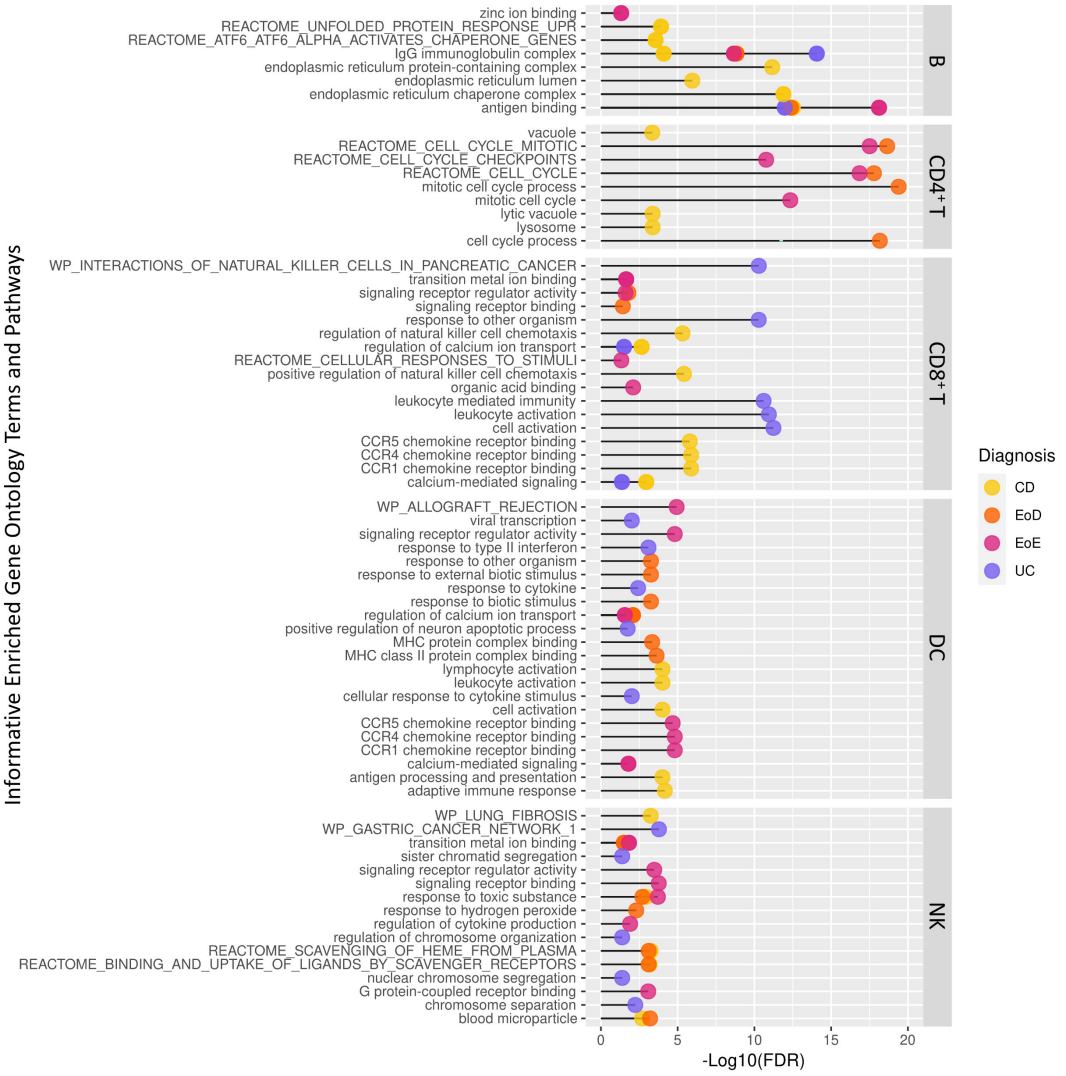



FIGURE 3
Abbreviated list of significantly enriched Gene Ontology terms and pathways among differentially expressed genes across level one cell types and GI disorder subtypes identified through ToppGene.

In B cells of individuals with CD, most of the genes associated with endoplasmic reticulum terms are downregulated. Additionally, many of those genes are associated with chaperone functions of the endoplasmic reticulum (Figure 4A). We also observed widespread downregulation of DEGs within CD8$^+$ T cells among individuals with UC, including markers of cytotoxicity (*CCL5*, *FGFBP2*, *GZMA*, *GZMB*, *GZMH*, *IFNG*, *NKG7*, and *PRF1*) along with the upregulation *NOG*, *REG4*, and *AIF1* (Figure 4B). We further illustrate through PPIs, the similarity of gene dysregulation and dominance of DEGs associated with the cell cycle for both EoE and EoD within CD4$^+$ T cells (Figures 4C, D), highlighting the similarity of these disorders. However, within these networks we can also see distinct profiles of genes dysregulated between EoE and EoD, including the downregulation of genes associated with p38 MAPK (*GADD45B*, *GADD45G*, and *DUSP1*) in EoD (Figure 4C).

# 4 Discussion

Aberrant immune signaling due to genetic and environmental factors contributes to the development of GI disorders (32). In this study we focused on the characterization of transcriptomic patterns in PBMCs of pediatric patients with active CD, UC, EoE, and EoD to identify genes and pathways associated with active inflammation and the pathogenesis of each disease. Insights into these transcriptomic phenotypes provide potential indicators of active inflammation and identify genetic markers for improved diagnosis, as well as possible therapeutic targets for treatment.

Genes showing similar patterns of differential expression among all GI disorders (CD, UC, EoE, and EoD) include the upregulation of *MTRNR2L8* within six level one cell types (B, CD4$^+$ T, CD8$^+$ T, DC, NK, and other T), may be helpful in identifying active inflammation in GI disorders. *MTRNR2L8* is

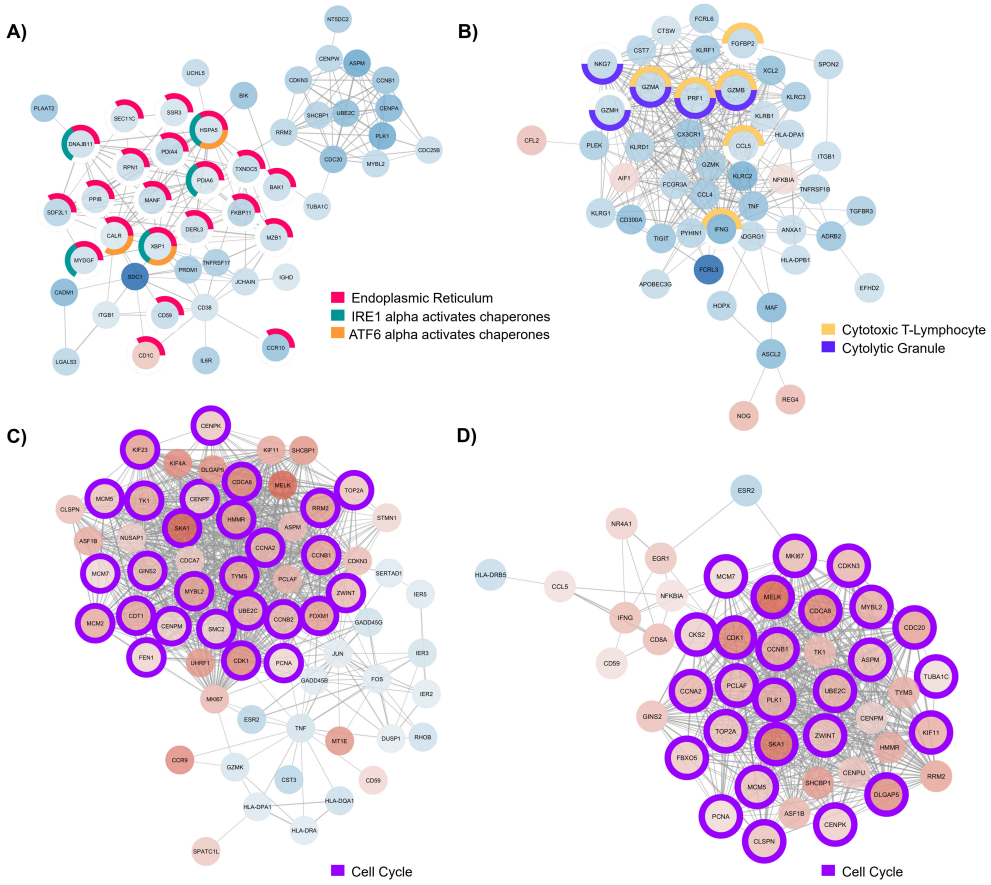



FIGURE 4
Major component of protein-protein interaction (PPI) network generated with **(A)** differentially expressed genes (DEGs) detected in B cells of individuals with CD, **(B)** DEGs detected in $CD8^+$ T cells from individuals with UC, **(C)** DEGs detected in $CD4^+$ T cells from individuals with EoD, and **(D)** DEGs detected in $CD4^+$ T cells from individuals with EoE. Node color indicates the direction of fold change of expression of each gene, with red indicating upregulation of gene expression in patients with GI disorders compared to control and blue indicating downregulation of gene expression in patients with GI disorders compared to control. Node outline color indicates enriched functional annotations detected in STRING.

believed to be a marker of cellular stress (33), and upregulation of *MTRNR2L8* has been observed among PBMCs of patients with primary mitochondrial disease (34) and among immune cell types in individuals with aspirin-exacerbated respiratory disease (35).

The common upregulation of *MTRNR2L1* for EoD and EoE was identified in four level one cell types (B, $CD4^+$ T, $CD8^+$ T, and DC). Similar to *MTRNR2L8*, *MTRNR2L1* also appears to be upregulated in response to cellular stress (33) and has been found to be upregulated in myeloid cells of patients with autoimmune disorders (36). There was also enrichment of functional terms associated with cell cycle activity in $CD4^+$ T cells among DEGs in EoE and EoD compared to controls, prominent among which were Reactome cell cycle mitotic (M5336), Reactome cell cycle (M543) and mitotic cell cycle process (GO:1903047), which may be indicative of dysregulation associated with Th2-mediated inflammation; genes associated with cell cycle progression and proliferation, specifically *MKI76*, *RRM2*, and *TOP2A*, which were found among upregulated cell cycle genes in EoD and EoE, have also been observed to be upregulated in the epithelium of asthmatics with high levels of Th2 (37).

Functional analysis of DEGs in CD detected in B cells yielded enrichment of terms associated with endoplasmic reticulum function and protein folding. Endoplasmic reticulum stress and the dysregulation of the unfolded protein response have been implicated in the development of IBD, which are essential for the maintenance of homeostasis in both the intestine and immune cells (38, 39). Furthermore, the specific enrichment of Reactome pathway ATF6 alpha activates chaperone genes in CD is supported by evidence that diminished *ATF6* activity can contribute to intestinal barrier dysfunction in mouse models of IBD (40).

In individuals with EoD, upregulation of *CCR9* and downregulation of *GADD45B* and *GADD45G* was found in $CD4^+$ T cells of patients compared to controls. *CCR9* has been demonstrated to play a key role in Th2-mediated inflammation, with $Ccr9^{-/-}$ mice treated with ovalbumin to induce allergic inflammation exhibiting an impaired immune response and diminished recruitment of eosinophils to inflamed tissue (41). Deficiency in *GADD45* expression has been associated with autoimmune disease (42), and the increased expression of *GADD45G* through administration of IL-27, has been shown to

attenuate Th2-mediated allergic response possibly through the activation of the p38 MAPK pathway (43), a pathway that was found to be enriched among DEGs in $CD4^+$ T cells in EoD.

Among identified DEGs in EoE relative to controls was *EGR1* in three level one cell types (B, $CD4^+$ T, and $CD8^+$ T cells) and several level two cell types, including naïve $CD8^+$ T cells and $CD4^+$ TCM cells. Among $CD4^+$ T cells, *EGR1* is preferentially expressed by Th2 and plays a role in the production of IL-4 cytokines (44), which mediate the allergic inflammatory response (45). Additionally, increased intracellular zinc levels have been shown to upregulate *EGR1* (46). Relevantly, the term zinc ion binding was enriched among DEGs in B cells in EoE. Zinc exposure has been demonstrated to elicit cellular damage (47), induce eosinophilia in mice, and evoke Th2 cytokine production (48). Furthermore, dysregulation of zinc signaling resulting from depletion of zinc within mucosal tissues and the release of zinc from airway discharge has been associated with eosinophilia (49). Together, these data support evidence that zinc homeostasis is critical in regulating inflammatory responses (50) and suggests that zinc homeostasis may be involved the development of eosinophilic disorders.

Several DEGs were found to be dysregulated between UC and controls within $CD8^+$ T cells, such as the downregulation of several genes involved in T cell cytotoxicity (*CCL5*, *FGFBP2*, *GZMA*, *GZMB*, *GZMH*, *IFNG*, *NKG7*, and *PRF1*). Previously, it has been shown that while elevated levels of *GZMB* were present in mucosal biopsies of treatment-naïve individuals with CD, there was not upregulation of *GZMB* in treatment-naïve individuals with UC compared to controls, suggesting enhanced $CD8^+$ T cytotoxicity in CD but not UC (51). Reduced T-cell cytotoxicity has also been detected in sub-populations of individuals with UC, including those who develop persistent low-grade dysplasia (52). Additionally, widespread upregulation of *XIST* was detected within five level one cell types (B, $CD4^+$ T, DC, monocytes, and NK) in UC compared to controls. In mouse models of IBD, *Xist* expression has been demonstrated to be upregulated after inducing colitis, and silencing of *Xist* in these models has shown to reduce colitis-associated symptoms (53). Moreover, previous transcriptome analysis of intestinal mucosa biopsies from individuals with UC have identified *XIST* as a key mediator of inflammation within this disease, as well as a possible therapeutic target (54).

The results of this study may indicate future directions of study for researchers and clinicians. DEGs from peripheral blood show promise of indicating degree of inflammation allowing clinicians to stratify severity of active disease and adjust treatment plans accordingly. Zinc homeostasis dysregulation in eosinophilic disorders provides a potential treatment target in a condition where presently there are few pharmacologic treatment options (55). Patients with EoE are known to have zinc deficiencies associated with elimination diets (56), but further investigation using these DEGs as targets may help elucidate pathways that affect overall zinc regulation.

While this study is limited by a small sample size, the ability to confirm trends of dysregulation in relevant affected GI tissues, and contrast these findings with pathological controls, overall, it demonstrates preliminary evidence for the utility of single-cell RNA-sequencing of patient blood cells to characterize the genomic landscape of pediatric IBD subtypes and eosinophilic disorders. These data indicate both overlapping and distinct DEGs, enriched Gene Ontology terms, and enriched pathways associated the pathogenesis of inflammation among individuals with CD, UC, EoD, and EoE, offering further insight into which genes and pathways may serve as useful markers of disease in the peripheral blood mononuclear cells of patients. Future studies should focus on expanded cell type profiling from disease-specific and peripheral tissues in larger patient cohorts and include pathological controls, such as patients with reflux. Additionally, future studies evaluating changes in DEG expression within an individual based on disease activity could determine utilization for monitoring remission without the need for more invasive endoscopy.

## Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found below: https://zenodo.org/records/10998303, 10.5281/zenodo.10998302.

## Ethics statement

The studies involving humans were approved by CMH Institutional Review Board. The studies were conducted in accordance with the local legislation and institutional requirements. Written informed consent for participation in this study was provided by the participants' legal guardians/next of kin.

## Author contributions

MK-K: Writing – original draft. LH: Writing – review & editing. VW: Writing – review & editing. CV: Writing – review & editing. AA: Writing – review & editing. JJ: Writing – review & editing. DL: Writing – review & editing. EG: Writing – review & editing. TP: Writing – review & editing. CF: Writing – review & editing. RC: Writing – review & editing. CS: Writing – original draft. VS: Writing – review & editing.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work was supported by Children's Mercy Kansas City and a philanthropic donation to CMKC by Emily and Todd Novikoff.

## Conflict of interest

The work reported herein was conducted while VS was affiliated with CMKC. VS is employed by Ironwood Pharmaceuticals. CV is employed by KCAS Bioanalytical & Biomarker Services.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The author(s) declared that they were an editorial board member of Frontiers, at the time of submission. This had no impact on the peer review process and the final decision.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fimmu.2024.1420208/full#supplementary-material

## References


1. Abraham C, Cho JH. MECHANISMS OF DISEASE inflammatory bowel disease. *New Engl J Med*. (2009) 361:2066–78. doi: 10.1056/NEJMra0804647

2. Silva FAR, Rodrigues BL, Ayrizono MDS, Leal RF. The immunological basis of inflammatory bowel disease. *Gastroent Res Pract*. (2016) 2016(1):2097274. doi: 10.1155/2016/2097274

3. Molodecky NA, Soon IS, Rabi DM, Ghali WA, Ferris M, Chernoff G, et al. Increasing incidence and prevalence of the inflammatory bowel diseases with time, based on systematic review. *Gastroenterology*. (2012) 142:46–54. doi: 10.1053/j.gastro.2011.10.001

4. Kaplan GG. The global burden of IBD: from 2015 to 2025. *Nat Rev Gastroenterol Hepatol*. (2015) 12:720–7. doi: 10.1038/nrgastro.2015.150

5. Furuta GT, Katzka DA. Eosinophilic esophagitis. *New Engl J Med*. (2015) 373:1640–8. doi: 10.1056/NEJMra1502863

6. Bass JA, Friesen CA, Deacy AD, Neilan NA, Bracken JM, Shakhnovich V, et al. Investigation of potential early Histologic markers of pediatric inflammatory bowel disease. *BMC Gastroenterol*. (2015) 15:129. doi: 10.1186/s12876-015-0359-2

7. Jacobs I, Ceulemans M, Wauters L, Breynaert C, Vermeire S, Verstockt B, et al. Role of eosinophils in intestinal inflammation and fibrosis in inflammatory bowel disease: an overlooked villain? *Front Immunol*. (2021) 12:754413. doi: 10.3389/fimmu.2021.754413

8. Berg DR, Colombel JF, Ungaro R. The role of early biologic therapy in inflammatory bowel disease. *Inflammation Bowel Dis*. (2019) 25:1896–905. doi: 10.1093/ibd/izz059

9. Kelay A, Tullie L, Stanton M. Surgery and paediatric inflammatory bowel disease. *Transl Pediatr*. (2019) 8:436–48. doi: 10.21037/tp

10. Urquhart SA, Quinn KP, Ravi K, Loftus EV Jr. The clinical characteristics and treatment outcomes of concomitant eosinophilic esophagitis and inflammatory bowel disease. *Crohns Colitis 360*. (2021) 3:otab018. doi: 10.1093/crocol/otab018

11. Sawczenko A, Ballinger AB, Savage MO, Sanderson IR. Clinical features affecting final adult height in patients with pediatric-onset Crohn's disease. *Pediatrics*. (2006) 118:124–9. doi: 10.1542/peds.2005-2931

12. Heuschkel R, Salvestrini C, Beattie RM, Hildebrand H, Walters T, Griffiths A. Guidelines for the management of growth failure in childhood inflammatory bowel disease. *Inflammation Bowel Dis*. (2008) 14:839–49. doi: 10.1002/ibd.20378

13. Greenley RN, Hommel KA, Nebel J, Raboin T, Li SH, Simpson P, et al. A meta-analytic review of the psychosocial adjustment of youth with inflammatory bowel disease. *J Pediatr Psychol*. (2010) 35:857–69. doi: 10.1093/jpepsy/jsp120

14. Cohen NA, Rubin DT. New targets in inflammatory bowel disease therapy: 2021. *Curr Opin Gastroenterol*. (2021) 37:357–63. doi: 10.1097/MOG.0000000000000740

15. Hart AL, Al-Hassi HO, Rigby RJ, Bell SJ, Emmanuel AV, Knight SC, et al. Characteristics of intestinal dendritic cells in inflammatory bowel diseases. *Gastroenterology*. (2005) 129:50–65. doi: 10.1053/j.gastro.2005.05.013

16. Coombes JL, Powrie F. Dendritic cells in intestinal immune regulation. *Nat Rev Immunol*. (2008) 8:435–46. doi: 10.1038/nri2335

17. Chaudhry A, Rudensky AY. Control of inflammation by integration of environmental cues by regulatory T cells. *J Clin Invest*. (2013) 123:939–44. doi: 10.1172/JCI57175

18. Weaver CT, Elson CO, Fouser LA, Kolls JK. The Th17 pathway and inflammatory diseases of the intestines, lungs, and skin. *Annu Rev Pathol*. (2013) 8:477–512. doi: 10.1146/annurev-pathol-011110-130318

19. Ohnmacht C, Marquee R, Presley L, Sawa S, Lochner M, Eberl G. Intestinal microbiota, evolution of the immune systemand the bad reputation of pro-inflammatory immunity. *Cell Microbiol*. (2011) 13:653–9. doi: 10.1111/cmi.2011.13.issue-5

20. Mishra A, Hogan SP, Lee JJ, Foster PS, Rothenberg ME. Fundamental signals that regulate eosinophil homing to the gastrointestinal tract. *J Clin Invest*. (1999) 103:1719–27. doi: 10.1172/JCI6560

21. Burczynski ME, Peterson RL, Twine NC, Zuberek KA, Brodeur BJ, Casciotti L, et al. Molecular classification of Crohn's disease and ulcerative colitis patients using transcriptional profiles in peripheral blood mononuclear cells. *J Mol Diagn*. (2006) 8:51–61. doi: 10.2353/jmoldx.2006.050079

22. Martin JC, Chang C, Boschetti G, Ungaro R, Giri M, Grout JA, et al. Single-cell analysis of Crohn's disease lesions identifies a pathogenic cellular module associated with resistance to anti-TNF therapy. *Cell*. (2019) 178:1493–+. doi: 10.1016/j.cell.2019.08.008

23. Chen P, Zhou GS, Lin JX, Li L, Zeng ZR, Chen MH, et al. Serum biomarkers for inflammatory bowel disease. *Front Med-Lausanne*. (2020) 7. doi: 10.3389/fmed.2020.00123

24. Yang HR. Update on eosinophilic gastrointestinal disease beyond eosinophilic esophagitis in children. *Clin Exp Pediatr*. (2023) 66:233–9. doi: 10.3345/cep.2022.01046

25. Kang HM, Subramaniam M, Targ S, Nguyen M, Maliskova L, McCarthy E, et al. Multiplexed droplet single-cell RNA-sequencing using natural genetic variation. *Nat Biotechnol*. (2018) 36:89–94. doi: 10.1038/nbt.4042

26. Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat Biotechnol*. (2024) 42:293–304. doi: 10.1038/s41587-023-01767-y

27. Hao YH, Hao S, Andersen-Nissen E, Mauck WM, Zheng SW, Butler A, et al. Integrated analysis of multimodal single-cell data. *Cell*. (2021) 184:3573–87. doi: 10.1016/j.cell.2021.04.048

28. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*. (2010) 26:139–40. doi: 10.1093/bioinformatics/btp616

29. Chen J, Bardes EE, Aronow BJ, Jegga AG. ToppGene Suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Res*. (2009) 37:W305–W11. doi: 10.1093/nar/gkp427

30. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. *Nucleic Acids Res*. (2023) 51:D638–D46. doi: 10.1093/nar/gkac1000

31. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Res*. (2003) 13:2498–504. doi: 10.1101/gr.1239303

32. Turpin W, Goethel A, Bedrani L, Croitoru K. Determinants of IBD heritability: genes, bugs, and more. *Inflammation Bowel Dis*. (2018) 24:1133–48. doi: 10.1093/ibd/izy085

33. Yen K, Lee C, Mehta H, Cohen P. The emerging role of the mitochondrial-derived peptide humanin in stress resistance. *J Mol Endocrinol*. (2013) 50:R11–R9. doi: 10.1530/JME-12-0203

34. Gordon-Lipkin EM, Banerjee P, Franco JLM, Tarasenko T, Kruk S, Thompson E, et al. Primary oxidative phosphorylation defects lead to perturbations in the human B cell repertoire. *Front Immunol*. (2023) 14:1142634. doi: 10.3389/fimmu.2023.1142634

35. Bangert C, Villazala-Merino S, Fahrenberger M, Krausgruber T, Bauer WM, Stanek V, et al. Comprehensive analysis of nasal polyps reveals a more pronounced type 2 transcriptomic profile of epithelial cells and mast cells in aspirin-exacerbated respiratory disease. *Front Immunol*. (2022) 13. doi: 10.3389/fimmu.2022.850494

36. Taft J, Markson M, Legarda D, Patel R, Chan M, Malle L, et al. Human TBK1 deficiency leads to autoinflammation driven by TNF-induced cell death. *Cell*. (2021) 184:4447–63. doi: 10.1016/j.cell.2021.07.026

37. Hachim MY, Elemam NM, Ramakrishnan RK, Salameh L, Olivenstein R, Hachim IY, et al. Derangement of cell cycle markers in peripheral blood mononuclear cells of asthmatic patients as a reliable biomarker for asthma control. *Sci Rep-Uk*. (2021) 11:11873. doi: 10.1038/s41598-021-91087-5

38. Cao SS. Endoplasmic reticulum stress and unfolded protein response in inflammatory bowel disease. *Inflammation Bowel Dis*. (2015) 21:636–44. doi: 10.1097/MIB.0000000000000238

39. Li C, Grider JR, Murthy KS, Bohl J, Rivet E, Wieghard N, et al. Endoplasmic reticulum stress in subepithelial myofibroblasts increases the TGF-β1 activity that regulates fibrosis in Crohn's disease. *Inflammation Bowel Dis*. (2020) 26:809–19. doi: 10.1093/ibd/izaa015

40. Huang SS, Xie Z, Han J, Wang HL, Yang G, Li MY, et al. Protocadherin 20 maintains intestinal barrier function to protect against Crohn's disease by targeting ATF6. *Genome Biol*. (2023) 24(1):159. doi: 10.1186/s13059-023-02991-0

41. López-Pacheco C, Soldevila G, Du Pont G, Hernández-Pando R, García-Zepeda EA. CCR9 is a key regulator of early phases of allergic airway inflammation. *Mediat Inflamm*. (2016) 2016(1):3635809. doi: 10.1155/2016/3635809

42. Schmitz I. Gadd45 proteins in immunity. *Gadd45 Stress Sensor Genes*. (2013) 793:51–68.

43. Su XQ, Pan J, Bai FX, Yuan HL, Dong N, Li DD, et al. IL-27 attenuates airway inflammation in a mouse asthma model via the STAT1 and GADD45γ/p38 MAPK pathways. *J Transl Med*. (2016) 14:1–12. doi: 10.1186/s12967-016-1039-x

44. Lohoff M, Giaisi M, Köhler R, Casper B, Krammer PH, Li-Weber M. Early growth response protein-1 (Egr-1) is preferentially expressed in T helper type 2 (Th2) cells and is involved in acute transcription of the th2 cytokine interleukin-4. *J Biol Chem*. (2010) 285:1643–52. doi: 10.1074/jbc.M109.011585

45. Ricci M, Matucci A, Rossi O. IL-4 as a key factor influencing the development of allergen-specific Th2-like cells in atopic individuals. *J Invest Allerg Clin*. (1997) 7:144–50.

46. Barbato JC, Catanescu O, Murray K, DiBello PM, Jacobsen DW. Targeting of metallothionein by L-homocysteine - A novel mechanism for disruption of zinc and redox homeostasis. *Arterioscl Throm Vas*. (2007) 27:49–54. doi: 10.1161/01.ATV.0000251536.49581.8a

47. Fukui H, Horie M, Endoh S, Kato H, Fujita K, Nishio K, et al. Association of zinc ion release and oxidative stress induced by intratracheal instillation of ZnO nanoparticles to rat lung. *Chem Biol Interact*. (2012) 198:29–37. doi: 10.1016/j.cbi.2012.04.007

48. Huang KL, Lee YH, Chen HI, Liao HS, Chiang BL, Cheng TJ. Zinc oxide nanoparticles induce eosinophilic airway inflammation in mice. *J Hazard Mater*. (2015) 297:304–12. doi: 10.1016/j.jhazmat.2015.05.023

49. Suzuki M, Ramezanpour M, Cooksley C, Lee TJ, Jeong B, Kao S, et al. Zinc-depletion associates with tissue eosinophilia and collagen depletion in chronic rhinosinusitis. *Rhinology*. (2020) 58:451–9. doi: 10.4193/Rhin

50. Devirgiliis C, Zalewski PD, Perozzi G, Murgia C. Zinc fluxes and zinc transporter genes in chronic diseases. *Mutat Res-Fund Mol M*. (2007) 622:84–93. doi: 10.1016/j.mrfmmm.2007.01.013

51. Jenkins D, Seth R, Kummer JA, Scott BB, Hawkey CJ, Robins RA. Differential levels of granzyme B, regulatory cytokines, and apoptosis in Crohn's disease and ulcerative colitis at first presentation. *J Pathol*. (2000) 190:184–9. doi: 10.1002/(ISSN)1096-9896

52. Kotsafti A, D'Incà R, Scarpa M, Fassan M, Angriman I, Mescoli C, et al. Weak cytotoxic T cells activation predicts low-grade dysplasia persistence in ulcerative colitis. *Clin Transl Gastroen*. (2019) 10(7):e00061. doi: 10.14309/ctg.0000000000000061

53. Gu D, Cao T, Yi S, Li X, Liu Y. Transcription suppression of GABARAP mediated by lncRNA XIST-EZH2 interaction triggers caspase-11-dependent inflammatory injury in ulcerative colitis. *Immunobiology*. (2024) 229:152796. doi: 10.1016/j.imbio.2024.152796

54. Xu M, Kong Y, Chen N, Peng W, Zi R, Jiang M, et al. Identification of immune-related gene signature and prediction of ceRNA network in active ulcerative colitis. *Front Immunol*. (2022) 13:855645. doi: 10.3389/fimmu.2022.855645

55. Tamarit-Sebastian S, Ferrer-Soler FM, Lucendo AJ. Current options and investigational drugs for the treatment of eosinophilic esophagitis. *Expert Opin Investig Drugs*. (2022) 31:193–210. doi: 10.1080/13543784.2022.2033207

56. Votto M, De Filippo M, Lenti MV, Rossi CM, Di Sabatino A, Marseglia GL, et al. Diet therapy in eosinophilic esophagitis. Focus on a personalized approach. *Front Pediatr*. (2021) 9:820192. doi: 10.3389/fped.2021.820192

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 11 October 2024
DOI 10.3389/fonc.2024.1402847



OPEN ACCESS

EDITED BY
Alessandro Passardi,
Scientific Institute of Romagna for the Study and Treatment of Tumors (IRCCS), Italy

REVIEWED BY
Andrea Glotta,
Ospedale Regionale di Lugano, Switzerland
Ashish Patel,
Hemchandracharya North Gujarat University, India

*CORRESPONDENCE
Shiyong Huang
huangshiyong@xinhuamed.com.cn
Yun Liu
liuyun@xinhuamed.com.cn

†These authors have contributed equally to this work

RECEIVED 18 March 2024
ACCEPTED 24 September 2024
PUBLISHED 11 October 2024

CITATION
Dai X, Li Y, Wang H, Dai Z, Chen Y, Liu Y and Huang S (2024) Development and validation of nomograms based on pre-/post-operative CEA and CA19-9 for survival predicting in stage I-III colorectal cancer patients after radical resection.
*Front. Oncol.* 14:1402847.
doi: 10.3389/fonc.2024.1402847



# Development and validation of nomograms based on pre-/post-operative CEA and CA19-9 for survival predicting in stage I-III colorectal cancer patients after radical resection

Xuan Dai[1†], Yifan Li[2†], Haoran Wang[3], Zhujiang Dai[1], Yuanyuan Chen[1], Yun Liu[1*] and Shiyong Huang[1*]

[1]Department of Colorectal and Anal Surgery, Xinhua Hospital, Shanghai Jiao Tong University School of Medicine, Shanghai, China, [2]Department of Gastrointestinal Surgery, Huaihe Hospital of Henan University, Kaifeng, Henan, China, [3]The First Clinical School, Xinxiang Medical University, Xinxiang, Henan, China

**Background:** Carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9) are the predominant serum tumour markers (STMs) for predicting the prognosis of colorectal cancer (CRC). The objective of this research is to develop clinical prediction models based on preoperative and postoperative CEA and CA19-9 levels.

**Methods:** 1,452 consecutive participants with stage I-III colorectal cancer were included. Kaplan-Meier method, log-rank test, and multivariate COX regression were used to evaluate the significance of preoperative and postoperative STMs. Patients were grouped into a discovery cohort (70%) and a validation cohort (30%). Variables for the nomograms were selected according to the Akaike information criterion (AIC). Subsequently, two clinical predictive models were constructed, evaluated, validated, and then compared with the AJCC 8th TNM stage.

**Results:** The overall survival (OS) rate and disease-free survival(DFS) rate declined progressively as the number of positive tumour markers(NPTMs) before and after surgery increased. For both OS and DFS, age, sex, pN stage, and NPTMs before and after surgery were independent prognostic factors, and then clinical prediction models were developed. The Concordance index (C-index), Receiver operating characteristic (ROC) curve, calibration curve, Decision curve analysis (DCA), and risk score stratification all indicated that the models possessed robust predictive efficacy and clinical applicability. The Net reclassification index (NRI) and Integrated discrimination improvement (IDI) indicated that the performance of models was significantly superior to the TNM stage.

**Conclusion:** Nomograms based on pre-and postoperative CEA and CA19-9 can accurately predict survival and recurrence for stage I-III CRC patients after radical surgery, and were significantly better than the AJCC 8th TNM stage.

KEYWORDS

**CEA, CA19-9, nomogram, colorectal cancer, overall survival, disease-free survival**

# 1 Introduction

Colorectal cancer (CRC) ranks as the world's second most deadly malignancy (1). Despite advancements in surgical techniques and integrated therapies, the clinical outcomes for CRC patients remain unsatisfactory. Approximately fifteen percent of stage II patients and thirty percent of stage III patients experience recurrence even after radical resection (2, 3). The high recurrence and mortality rates have increasingly drawn attention to the need for individualized treatment and prognosis of this disease. Clinicians currently rely on the TNM staging system to predict and assess the prognosis of patients with colorectal cancer (4). While the current staging system provided essential insights into tumour behavioral characteristics, it doesn't fully encompass vital determinants of patient prognosis, such as age, serum tumour markers(STMs) and so on. Consequently, there's an imperative demand to unearth novel markers for individualized prognostic assessment, empowering clinicians to offer more precise counsel on survival forecasts and therapeutic approaches for CRC patients.

Owning to its simplicity and cost-effectiveness, tumour marker detection is extensively performed in medical institutions. Carcinoembryonic antigen (CEA) and carbohydrate antigen 19-9 (CA19-9) represent the primary STMs for preoperative evaluation and postoperative follow-up examination of CRC patients. CEA is an acidic glycoprotein associated with oncogenic advancement (5). Some clinical guidelines recommend CEA as a prognostic biomarker for CRC and endorse its routine measurement after radical resection in CRC patients (6, 7). CA19-9 is closely linked to recurrence and survival in colorectal cancer (8). Significantly, combined tumour marker testing has significantly improved predictive accuracy compared with single marker testing (9). Concurrently, the number of positive tumour markers(NPTMs) is gaining attention (10, 11). Previous research has demonstrated its feasibility as a prognostic factor for stage II-III CRC (12). However, while this study has underscored the impact of NPTMs before surgery on prognosis, the significance of postoperative STMs remains underexplored. Recently, some researches have paid attention to the role of postoperative STMs and found that they are also promising indicators (9, 13, 14). It has also been shown that the number of positive tumour markers before and after treatment is important for the prognosis of rectal cancer (15, 16). Therefore, we believe that combining both preoperative and postoperative CEA and CA19-9 measures might enhance predictive accuracy.

While clinical predictive models are endorsed for estimating the recurrence and survival of diverse malignancies due to their utility and comprehensiveness (17, 18), no research has incorporated NPTMs before and after surgery into these models for stage I-III CRC. Recognizing the vital prognostic implications of NPTMs, we evaluated the association of preoperative and postoperative CEA and CA199 with OS and DFS in patients with stage I-III CRC who underwent radical resection. Age, sex, pN stage, NPTMs before and after surgery were chosen to construct the clinical prediction models of overall survival(OS) and disease-free survival(DFS). Additionally, we further compared the clinical value of these models with that of the AJCC 8th TNM stage.

# 2 Patients and methods

## 2.1 Study population

This study included consecutive CRC patients who underwent radical resection at the Department of Colorectal and Anal Surgery, Xinhua Hospital Affiliated to Shanghai Jiao Tong University School of Medicine from January 2010 to August 2017. Exclusion criteria were as follows (Figure 1): (1) patients with distant metastasis; (2) patients without radical resection; (3) patients with pathological non-adenocarcinoma or undetailed pathological data; (4) patients with preoperative neoadjuvant therapy; (5) patients with incomplete data of preoperative or postoperative CEA or CA19-9. Finally, 1,452 patients were involved in the study. The entire population was randomly grouped into a discovery cohort of 70% (n = 966) and a validation cohort of 30% (n = 486). All patients were staged according to the latest NCCN guidelines. All patients included in the study underwent radical (R0) resection of the primary tumour. Chemotherapy was administered according to NCCN guidelines to patients who met the criteria for postoperative chemotherapy.

## 2.2 Detection of CEA and CA19-9

Preoperative STMs (CEA, CA19-9) were tested within 7 days before radical surgery for colorectal cancer. Postoperative STMs

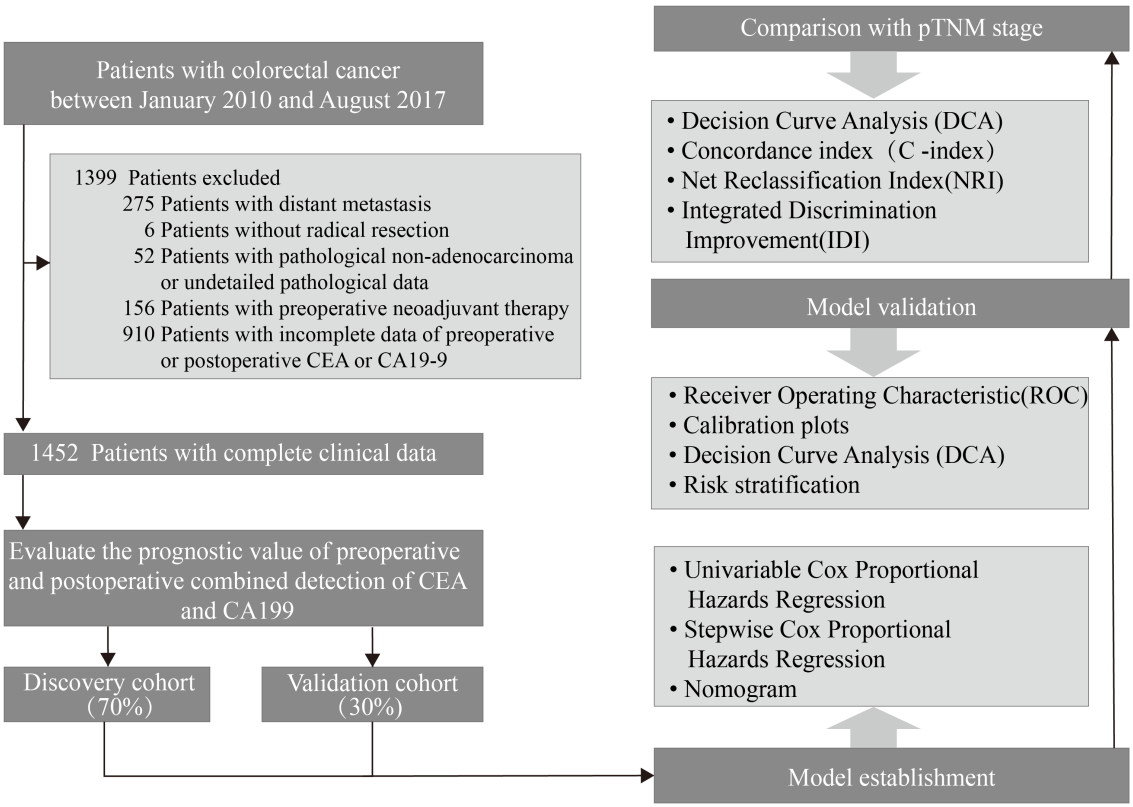



FIGURE 1
Study flow chart.

(CEA, CA19-9) were tested in serum samples obtained at the patient's first visit during the postoperative 2.5 – 3.5 months. A cutoff of 10 ng/ml was utilized to determine CEA positivity, while CA19-9 positivity was ascertained using a threshold of 39 U/ml (19–21). Patients were stratified based on the NPTMs before and after surgery as follows: (1) NPTMs was zero (both CEA and CA19-9 negative); (2) NPTMs was one (either CEA or CA19-9 positive); and (3) NPTMs was two (both CEA and CA19-9 positive). Patients were categorized based on NPTMs, followed by an analysis of their clinical characteristics and survival outcomes.

## 2.3 Follow-up study

Follow-up evaluations were conducted quarterly for the first two years after surgery. Subsequent assessments occurred biannually from the third to the fifth year, and then annually thereafter. In both cohorts, the follow-up protocol included physical examination, chest CT scan, measurement of CEA and CA19-9, abdominal and pelvic MRI or CT, etc. Colonoscopy was carried out once a year. OS is the time from radical resection to either death from any cause or the last follow-up, while DFS spans from radical resection to the first recurrence, any cause of death, or the last follow-up. The follow-up evaluation of this study concluded on August 2022.

## 2.4 Data analysis

The $\chi 2$ test or Fisher's exact test was utilized to compare categorical variables. The Kaplan-Meier method and the log-rank test were employed, so as to assess the survival curves across groups. In the discovery cohort, traditional clinicopathological variables underwent the univariate analysis. Factors with $P < 0.2$ were incorporated as independent variables into the COX regression for a multivariate assessment. Variables were selected for inclusion in the nomograms based on the Akaike information criterion (AIC). Until the optimal model was obtained, AIC (Akaike information criterion, a standard for measuring statistical model fitting) was gradually reduced. The model with the lowest AIC value is usually chosen as the best model. The nomograms were used to predict the probability of survival and recurrence. The discrimination ability was evaluated by the concordance index (C-index) and receiver operating characteristic curve (ROC). The calibration curve was used to evaluate the calibration power. The net reclassification index (NRI) and integrated discrimination improvement (IDI) are designed to evaluate enhancements in risk forecasting and gauge the efficacy of the novel nomogram. They were used to compare the clinical value between nomograms and TNM stage. Decision curve analysis (DCA) is a method to evaluate the clinical applicability, quantifying its net benefit across various threshold probabilities.

Curves representing all patients treated (indicating the highest clinical cost) and no treatment (indicating no clinical benefit) were used as references. All tests were conducted on both sides, with a significance level established at P<0.05. All data were analyzed using SPSS(26) and R software (4.2.1).

# 3 Result

## 3.1 Clinicopathological features

The study had 1,452 participants. The 5-year OS and DFS rates were 80.7% and 76.7%, respectively, with a median age of 63 years (IQR: 57 – 72 years). The discovery cohort included 966 cases, while the validation cohort had 486 cases (Table 1). The 5-year OS rates for the discovery and validation cohorts were 81.7% and 78.8%, respectively, while the 5-year DFS rates were 78.0% and 74.0%. In the discovery cohort, there were 531 men and 435 women. According to the TNM staging system, stages I, II, and III included 169 (17.5%), 395 (40.9%), and 402 (41.6%) cases, respectively. The age of participants in the validation cohort was obviously younger than that in the discovery cohort. ($P < 0.05$). Except for age, other variables showed no significant difference. ($P > 0.05$).

## 3.2 Clinicopathological features based on preoperative and postoperative tumour markers

Table 2 summarizes the association between NPTMs and the characteristics of patients. Preoperatively, 1,062 patients (73.1%) were negative for both markers, 300 patients (20.7%) were positive for one marker, and 90 patients (6.2%) were positive for both markers. Postoperatively,1,346 patients (92.7%) were negative for both markers, 86 (5.9%) patients were positive for one marker, and 20 (1.4%) patients were positive for both markers. There was a significant correlation between NPTMs before surgery and tumour location, histological type, pT stage, pN stage, pTNM stage, and nerve/vascular invasion (all $P < 0.05$; Table 2). The NPTMs after surgery was also significantly associated with age, histological type, pN stage, pTNM stage, and nerve/vascular invasion (all $P < 0.05$; Table 2).

## 3.3 OS and DFS based on preoperative and postoperative tumour markers

Kaplan-Meier survival curve results displayed obvious decreases in 5-year survival with increasing NPTMs before surgery (5-year OS rate: n = 0: 85.1%; n = 1: 69.9%; n = 2: 64.4%, $P < 0.0001$, Figure 2A; 5-year DFS rate: n = 0: 81.3%; n = 1: 65.6%; n = 2: 59.0%, $P < 0.001$, Figure 2B); similarly, there was also a obvious correlation between the NPTMs after surgery and patients' OS and DFS (5-year OS rate: n = 0: 83.3%; n = 1: 54.9%; n = 2: 16.9%, $P < 0.0001$, Figure 2C; 5-year DFS rate: n = 0: 79.2%; n = 1: 50.3%; n = 2: 116.4%, $P < 0.001$, Figure 2D).

## 3.4 Nomogram variable screening

Tables 3 and 4 showed the consequences of the variables analyses concerning survival in CRC patients. Multivariate COX regression analysis showed that sex, age, pN stage, NPTMs before and after surgery were independent prognostic factors for OS (Table 3); sex, age, pN stage, NPTMs before and after surgery were also independent prognostic factors for DFS (Table 4).

## 3.5 Construction and validation of nomograms for CRC

Age, sex, pN stage, NPTMs before and after surgery were selected to construct nomograms for OS and DFS, respectively (Figure 3). For OS, the C-index was 0.760 for the discovery cohort and 0.772 for the validation cohort. For the discovery cohort, the model's AUC values stood at 0.793 for 3 years and 0.773 for 5 years(Figure 4A). Concurrently, for the validation cohort, they were 0.785 and 0.769 (Figure 4C). For DFS, the C-index was 0.724 for the discovery cohort and 0.748 for the validation cohort. For the discovery cohort, the model's AUC values stood at 0.755 for 3 years and 0.743 for 5 years (Figure 4B). Concurrently, for the validation cohort, they were 0.745 and 0.760 (Figure 4D). In addition, the calibration curves exhibited strong concordance between the models' predictions and actual observations in both cohorts (Supplementary Figure 1).

## 3.6 Clinical value of nomograms compared with TNM stage

The DCA showed that the nomograms offered superior net clinical benefits for both OS and DFS compared with TNM stage (Supplementary Figure 2). To further compare the accuracy of the models with the conventional TNM stage, we also analyzed the C-index change, NRI, and IDI (Table 5). Within the discovery cohort, the C-index change for OS was 0.105, The NRI for OS at 3 and 5 years registered at 0.519 and 0.515, respectively. IDI was 0.131 and 0.117; The C-index change for DFS was 0.090, and the NRI for DFS at 3 and 5 years registered at 0.481 and 0.444, respectively. IDI was 0.101 and 0.098, respectively. This result was further verified in the validation cohort. In addition, Participants were grouped into two different risk groups based on the median of the risk group scores in the discovery cohort. Results from the Kaplan-Meier survival curves revealed notable distinctions between two different risk cohorts ($P < 0.01$, Figure 5). Overall, our nomograms demonstrated superior predictive performance and clinical applicability compared with the traditional TNM stage, offering a more precise prognosis and survival prediction for patients.

# 4 Discussion

Researches on biomarkers of gastrointestinal cancer have been widely concerned. A study on dogs has explored the important role

10.3389/fonc.2024.1402847

TABLE 1 Comparison of baseline clinicopathologic characteristics between the discovery cohort and the validation cohort.

| Clinicopathological Features | Overall | Discovery cohort | Validation cohort | P value |
|---|---|---|---|---|
| | (n =1452) | (n = 966) | (n = 486) | |
| Sex | | | | 0.859 |
| Male | 795 (54.7) | 531 (55.0) | 264 (54.3) | |
| Female | 657 (45.3) | 435 (45.0) | 222 (45.7) | |
| Age (years) | | | | **0.019** |
| <65 | 790 (54.4) | 504 (52.2) | 286 (58.8) | |
| ≥65 | 662 (45.6) | 462 (47.8) | 200 (41.2) | |
| Tumour Location | | | | 0.615 |
| Right Colon | 328 (22.6) | 222 (23.0) | 106 (21.8) | |
| Left Colon | 445 (30.6) | 288 (29.8) | 157 (32.3) | |
| Rectum | 679 (46.8) | 456 (47.2) | 223 (45.9) | |
| Histologic type | | | | 0.955 |
| Grade I Adenocarcinoma | 45 (3.1) | 14 (2.9) | 31 (3.2) | |
| Grade II Adenocarcinoma | 1056 (72.7) | 356 (73.3) | 700 (72.5) | |
| Grade III Adenocarcinoma | 131 (9.0) | 45 (9.3) | 86 (8.9) | |
| Mucinous Adenocarcinoma | 220 (15.2) | 71 (14.6) | 149 (15.4) | |
| pT stage | | | | 0.353 |
| T1 | 87 (6.0) | 52 (5.4) | 35 (7.2) | |
| T2 | 246 (16.9) | 158 (16.4) | 88 (18.1) | |
| T3 | 810 (55.8) | 543 (56.2) | 267 (54.9) | |
| T4 | 309 (21.3) | 213 (22.0) | 96 (19.8) | |
| pN stage | | | | 0.533 |
| N0 | 849 (58.5) | 564 (58.4) | 285 (58.6) | |
| N1 | 364 (25.1) | 249 (25.8) | 115 (23.7) | |
| N2 | 239 (16.5) | 153 (15.8) | 86 (17.7) | |
| pTNM stage | | | | 0.48 |
| I | 266 (18.3) | 169 (17.5) | 97 (20.0) | |
| II | 583 (40.2) | 395 (40.9) | 188 (38.7) | |
| III | 603 (41.5) | 402 (41.6) | 201 (41.4) | |
| Perineural/Vascular invasion | | | | 0.59 |
| No | 1332 (91.7) | 883 (91.4) | 449 (92.4) | |
| Yes | 120 (8.3) | 83 (8.6) | 37 (7.6) | |
| NPTMs before surgery | | | | 0.167 |
| 0 | 1062 (73.1) | 696 (72.0) | 366 (75.3) | |
| 1 | 300 (20.7) | 213 (22.0) | 87 (17.9) | |
| 2 | 90 (6.2) | 57 (5.9) | 33 (6.8) | |
| NPTMs after surgery | | | | 0.83 |
| 0 | 1346 (92.7) | 897 (92.9) | 449 (92.4) | |
| 1 | 86 (5.9) | 55 (5.7) | 31 (6.4) | |
| 2 | 20 (1.4) | 14 (1.4) | 6 (1.2) | |

NPTMs, the number of positive tumour markers.
Statistically significant values are in bold.

TABLE 2 Associations of NPTMs with clinicopathological characteristics in stage I-III CRC patients after radical resection.

| Clinicopathological Features | NPTMs before surgery | | | P value | NPTMs after surgery | | | P value |
|---|---|---|---|---|---|---|---|---|
| | 0 (n = 1062) | 1 (n = 300) | 2 (n = 90) | | 0 (n = 1346) | 1 (n = 86) | 2 (n = 20) | |
| Sex | | | | 0.263 | | | | 0.44 |
| Male | 590 (55.6) | 163 (54.3) | 42 (46.7) | | 739 (54.9) | 43 (50) | 13 (65) | |
| Female | 472 (44.4) | 137 (45.7) | 48 (53.3) | | 607 (45.1) | 43 (50) | 7 (35) | |
| Age (years) | | | | 0.649 | | | | **0.011** |
| <65 | 570 (53.7) | 169 (56.3) | 51 (56.7) | | 747 (55.5) | 34 (39.5) | 9 (45) | |
| ≥65 | 492 (46.3) | 131 (43.7) | 39 (43.3) | | 599 (44.5) | 52 (60.5) | 11 (55) | |
| Tumour Location | | | | **0.003** | | | | 0.642 |
| Right Colon | 231 (21.8) | 67 (22.3) | 30 (33.3) | | 302 (22.4) | 19 (22.1) | 7 (35) | |
| Left Colon | 308 (29) | 112 (37.3) | 25 (27.8) | | 417 (31) | 24 (27.9) | 4 (20) | |
| Rectum | 523 (49.2) | 121 (40.3) | 35 (38.9) | | 627 (46.6) | 43 (50) | 9 (45) | |
| Histologic type | | | | **< 0.001** | | | | **0.025** |
| Grade I Adenocarcinoma | 43 (4) | 2 (0.7) | 0 (0) | | 43 (3.2) | 2 (2.3) | 0 (0) | |
| Grade II Adenocarcinoma | 791 (74.5) | 200 (66.7) | 65 (72.2) | | 987 (73.3) | 59 (68.6) | 10 (50) | |
| Grade III Adenocarcinoma | 80 (7.5) | 38 (12.7) | 13 (14.4) | | 112 (8.3) | 13 (15.1) | 6 (30) | |
| Mucinous Adenocarcinoma | 148 (13.9) | 60 (20) | 12 (13.3) | | 204 (15.2) | 12 (14) | 4 (20) | |
| pT stage | | | | **< 0.001** | | | | 0.079 |
| T1 | 82 (7.7) | 3 (1) | 2 (2.2) | | 80 (5.9) | 7 (8.1) | 0 (0) | |
| T2 | 214 (20.2) | 26 (8.7) | 6 (6.7) | | 237 (17.6) | 8 (9.3) | 1 (5) | |
| T3 | 580 (54.6) | 183 (61) | 47 (52.2) | | 751 (55.8) | 48 (55.8) | 11 (55) | |
| T4 | 186 (17.5) | 88 (29.3) | 35 (38.9) | | 278 (20.7) | 23 (26.7) | 8 (40) | |
| pN stage | | | | **< 0.001** | | | | **< 0.001** |
| N0 | 678 (63.8) | 136 (45.3) | 35 (38.9) | | 798 (59.3) | 47 (54.7) | 4 (20) | |
| N1 | 244 (23) | 87 (29) | 33 (36.7) | | 344 (25.6) | 16 (18.6) | 4 (20) | |
| N2 | 140 (13.2) | 77 (25.7) | 22 (24.4) | | 204 (15.2) | 23 (26.7) | 12 (60) | |
| pTNM stage | | | | **< 0.001** | | | | **0.009** |
| I | 237 (22.3) | 23 (7.7) | 6 (6.7) | | 252 (18.7) | 14 (16.3) | 0 (0) | |
| II | 441 (41.5) | 113 (37.7) | 29 (32.2) | | 546 (40.6) | 33 (38.4) | 4 (20) | |
| III | 384 (36.2) | 164 (54.7) | 55 (61.1) | | 548 (40.7) | 39 (45.3) | 16 (80) | |
| Perineural/Vascular invasion | | | | **0.041** | | | | **0.01** |
| No | 986 (92.8) | 266 (88.7) | 80 (88.9) | | 1242 (92.3) | 75 (87.2) | 15 (75) | |
| Yes | 76 (7.2) | 34 (11.3) | 10 (11.1) | | 104 (7.7) | 11 (12.8) | 5 (25) | |

NPTMs, the number of positive tumour markers.
Statistically significant values are in bold.

of lipopolysaccharide in intestinal carcinogenesis (22), and Li et al.' s study on mice suggested that the secretory protein cathepsin K can be used as a new predictive biomarker for CRC (23). In addition, previous studies have confirmed the potential prognostic value of absolute quantification of free circulating DNA (24)and long non-coding RNA plasmacytoma variant translocation 1 (25) in CRC patients as biomarkers. Serum CEA and CA19-9 are common and cost-effective biomarkers in clinical practice and they are instrumental in predicting the prognosis of CRC, holding significant value in both pre- and post-operation (26, 27). However, previous researches have primarily paid attention to the prognostic significance of preoperative STMs (28–30), with little attention given to postoperative CEA and CA19-9. Study has demonstrated that in terms of predicting survival duration, combined tumour markers assessments hold an advantage over single marker tests (31). In recent years, NPTMs, introduced as a

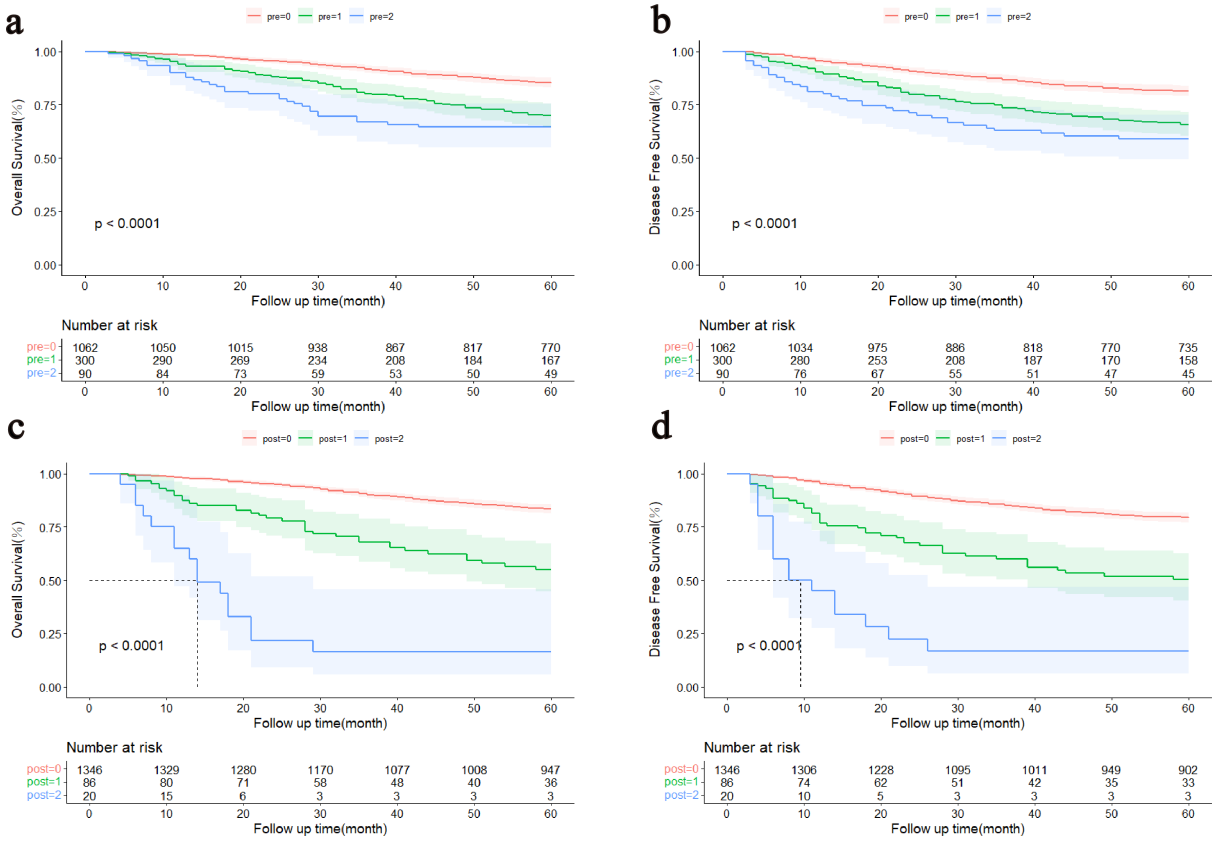



FIGURE 2
Kaplan–Meier curves of the patients with stage I-III colorectal cancer. **(A)** Association between the NPTMs before surgery and OS. **(B)** Association between the NPTMs before surgery and DFS. **(C)** Association between the NPTMs after surgery and OS. **(D)** Association between NPTMs after surgery and DFS. OS, overall survival; DFS, disease-free survival; NPTMs, the number of positive tumour markers; pre, the number of positive tumour markers before surgery; post, the number of positive tumour markers after surgery.

**TABLE 3** Univariate and multivariate COX analysis of clinicopathological characteristics concerning overall survival of CRC patients in the discovery cohort.

| Clinicopathological Features | Univariable analysis | | | Multivariable analysis | | |
|---|---|---|---|---|---|---|
| | HR | 95%CI | P value | HR | 95%CI | P value |
| **Sex** | | | | | | |
| Male | Reference | | | | | |
| Female | 0.646 | 0.471 - 0.887 | **0.007** | 0.583 | 0.421 - 0.807 | **0.001** |
| **Age (years)** | | | | | | |
| <65 | Reference | | | | | |
| ≥65 | 2.045 | 1.490 - 2.807 | **0.000** | 2.424 | 1.741 - 3.374 | **0.000** |
| **Tumour Location** | | | | | | |
| Right Colon | Reference | | | | | |
| Left Colon | 0.553 | 0.358 - 0.854 | **0.008** | | | 0.218 |
| Rectum | 0.863 | 0.603 - 1.234 | 0.420 | | | 0.449 |
| **Histologic type** | | | | | | |
| Grade I Adenocarcinoma | Reference | | | | | |

*(Continued)*

TABLE 3 Continued

| Clinicopathological Features | Univariable analysis | | | Multivariable analysis | | |
|---|---|---|---|---|---|---|
| | HR | 95%CI | P value | HR | 95%CI | P value |
| **Histologic type** | | | | | | |
| Grade II Adenocarcinoma | 4.530 | 0.632 - 32.463 | 0.133 | | | 0.372 |
| Grade III Adenocarcinoma | 12.624 | 1.717 - 92.783 | **0.013** | | | 0.091 |
| Mucinous Adenocarcinoma | 6.979 | 0.955 - 51.026 | 0.056 | | | 0.199 |
| **pN stage** | | | | | | |
| N0 | Reference | | | | | |
| N1 | 2.301 | 1.572 - 3.368 | **0.000** | 2.262 | 1.538 - 3.326 | **0.000** |
| N2 | 5.195 | 3.600 - 7.496 | **0.000** | 4.627 | 3.148 - 6.801 | **0.000** |
| **NPTMs before surgery** | | | | | | |
| 0 | Reference | | | | | |
| 1 | 2.114 | 1.508 - 2.965 | **0.000** | 1.573 | 1.105 - 2.239 | **0.012** |
| 2 | 3.611 | 2.247 - 5.803 | **0.000** | 2.550 | 1.495 - 4.348 | **0.001** |
| **NPTMs after surgery** | | | | | | |
| 0 | Reference | | | | | |
| 1 | 2.901 | 1.814 - 4.640 | **0.000** | 1.789 | 1.078 - 2.967 | **0.024** |
| 2 | 12.447 | 6.521 - 23.761 | **0.000** | 4.187 | 2.038 - 8.602 | **0.000** |

NPTMs, the number of positive tumour markers.
Statistically significant values are in bold.

TABLE 4 Univariate and multivariate COX analyses of clinicopathological characteristics concerning disease-free survival of CRC patients.

| Clinicopathological Features | Univariable analysis | | | Multivariable analysis | | |
|---|---|---|---|---|---|---|
| | HR | 95%CI | P value | HR | 95%CI | P value |
| **Sex** | | | | | | |
| Male | Reference | | | | | |
| Female | 0.820 | 0.620 - 1.084 | 0.163 | 0.739 | 0.556 - 0.984 | **0.038** |
| **Age (years)** | | | | | | |
| <65 | Reference | | | | | |
| ≥65 | 1.937 | 1.457 - 2.573 | **0.000** | 2.235 | 1.661 - 3.007 | **0.000** |
| **Tumour Location** | | | | | | |
| Right Colon | Reference | | | | | |
| Left Colon | 0.589 | 0.397 - 0.873 | **0.008** | | | 0.287 |
| Rectum | 0.908 | 0.655 - 1.260 | 0.564 | | | 0.268 |
| **Histologic type** | | | | | | |
| Grade I Adenocarcinoma | Reference | | | | | |
| Grade II Adenocarcinoma | 1.423 | 0.526 - 3.848 | 0.487 | | | 0.867 |
| Grade III Adenocarcinoma | 3.289 | 1.159 - 9.335 | **0.025** | | | 0.305 |
| Mucinous Adenocarcinoma | 2.072 | 0.740 - 5.799 | 0.165 | | | 0.620 |

*(Continued)*

TABLE 4 Continued

| Clinicopathological Features | Univariable analysis | | | Multivariable analysis | | |
|---|---|---|---|---|---|---|
| | HR | 95%CI | P value | HR | 95%CI | P value |
| pN stage | | | | | | |
| N0 | Reference | | | | | |
| N1 | 1.833 | 1.302 - 2.582 | **0.001** | 1.856 | 1.312 - 2.627 | **0.000** |
| N2 | 4.281 | 3.090 - 5.930 | **0.000** | 4.012 | 2.843 - 5.663 | **0.000** |
| NPTMs before surgery | | | | | | |
| 0 | | | | | | |
| 1 | 1.826 | 1.339 - 2.490 | **0.000** | 1.431 | 1.036 - 1.978 | **0.030** |
| 2 | 2.962 | 1.896 - 4.629 | **0.000** | 2.232 | 1.365 - 3.650 | **0.001** |
| NPTMs after surgery | | | | | | |
| 0 | Reference | | | | | |
| 1 | 2.713 | 1.755 - 4.195 | **0.000** | 1.676 | 1.050 - 2.673 | **0.030** |
| 2 | 8.743 | 4.607 -16.594 | **0.000** | 2.986 | 1.478 - 6.031 | **0.002** |

NPTMs, the number of positive tumour markers.
Statistically significant values are in bold.

novel reference index, has exhibited profound prognostic potential (11, 15). Given their significant clinical value, this study categorized patients according to NPTMs, assessed the prognostic significance of combined STMs detection before and after surgery, and subsequently developed clinical prediction models.

In this study, we found that the increased NPTMs before and after surgery were associated with a poor prognosis of CRC. Furthermore, NPTMs before surgery was closely related to the TNM stage and tumour location, consistent with prior findings (11). Studies reported that patients with normal STMs after surgery possessed a notably better prognosis compared with patients with abnormal STMs (13, 15). This research also supported this result. Compared with patients with normal postoperative CEA and CA19-9, patients with both tumour markers positive postoperatively had approximately a 4.2-fold increased risk of death and a 3.0-fold increased risk of recurrence. We also discovered that both preoperative and postoperative positive CEA and CA19-9 were more likely to occur in population with higher pTNM stage, higher pN stage, and those with neural/vascular invasion. For these patients, a more intensive follow-up strategy should be implemented.

The role of circulating tumour DNA (ct-DNA) in predicting the prognosis of colorectal cancer has garnered widespread attention (32, 33). Study has reported a correlation between CT-DNA in tumour cells and residual microcancer cells, but its clinical application remains limited due to its high costs (34). Conversely, tumour marker detection is affordable and easy to operate. Konishi et al. reported that patients with elevated postoperative CEA faced a higher hazard of early recurrence, especially within the first year after radical surgery (13). Sonoda et al. found that elevation of CEA post-surgery is independently correlated with an unfavorable prognosis in stage II-III CRC (14). In this study, we found that NPTMs before and after surgery were independent prognostic factors for OS and DFS in patients with stage I-III colorectal cancer. The elevation of tumour marker levels postoperatively may suggest the presence of unrecognized residual minute cancer cells at the time of surgery or in postoperative radiological examinations, which raises the possibility of relapse (35, 36). Therefore, in clinical practice, it is essential not only to perform combined tumour markers testing before surgery for colorectal cancer patients but also to pay attention to postoperative combined tumour markers testing. Patients with positive tumour markers might benefit from comprehensive treatment and require followed-up regularly.

TNM stage is commonly used for prognosis prediction and assessment, but its ability to predict patient outcomes may be limited (37). Nomogram is a powerful graphical prediction tool that illustrates the likelihood of a specific event occurring based on multiple variables (17). Compared with the TNM stage, the nomogram is more intuitive and easier to understand. Moreover, it can incorporate more risk factors, significantly enhancing the accuracy of prediction. Previous studies have developed CRC-related survival prediction models based on STMs (38, 39), but they were limited in sample size and did not focus on the prognostic significance of post-surgical STMs. This might have restricted their predictive accuracy to some extent. Therefore, this study constructed two more comprehensive clinical prediction models based on NPTMs to help clinicians predict and evaluate the prognosis. Both nomograms included five variables: sex, age at surgery, pN stage, NPTMs before and after surgery. The pN stage and NPTMs after surgery had significant effects on total scores of two models.

The C-index results indicated that the discriminative ability of the two models is significantly superior to the TNM stage. Risk stratification analysis revealed that the predictive models for OS and DFS exhibited commendable discrimination proficiency. Further,

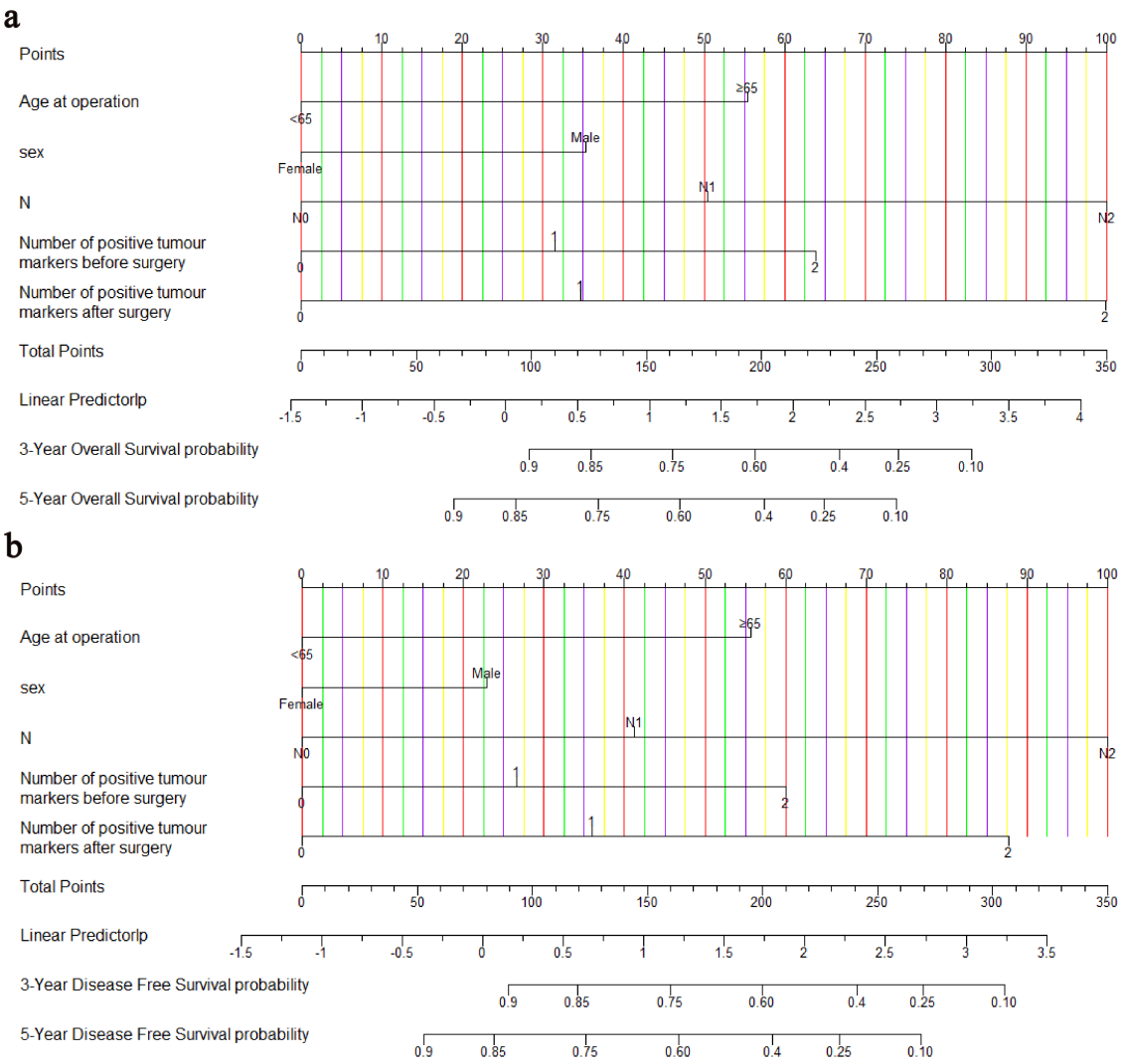



FIGURE 3
Nomograms for predicting OS **(A)** and DFS **(B)**. OS, overall survival; DFS, disease-free survival.

the results of DCA underscored that our models exhibited superior performance in clinical decision-making compared with TNM stage. The calibration curves for both groups also confirmed the strong concordance between the predictive model and the actual outcomes. NRI and IDI are two statistical indicators used to assess the enhanced performance of predictive models. Through their comparative analysis, we can discern the differential performance of various models and select the optimal one (40, 41). Within our investigation, both NRI and IDI metrics indicated that the novel models had superior accuracy and discriminatory ability in forecasting 3-year and 5-year OS and DFS for CRC patients. To summarize, both prognostic models exhibited strong predictive efficacy and clinical applicability, and can be utilized in clinical settings to forecast the prognosis of stage I-III CRC patients.

This study had several limitations. First of all, both the discovery cohort and validation cohort were established through random grouping, which could lead to imbalances at baseline. For instance, distinct variations in clinicopathological characteristics were observed in both cohorts (age, $P < 0.05$). Secondly, for the constructed model, the validation cohort performed better than the discovery cohort in some aspect. Thirdly, since stage IV patients receive different treatment methods from stage I–III patients, we did not include stage IV patients in our study, so the nomograms cannot be applied to stage IV CRC patients. Finally, Some factors that might be associated with prognosis, such as BRAF and KRAS mutation status and nutritional condition, were not considered in our study. Future studies should incorporate more valuable prognostic factors and conduct external validation.

## 5 Conclusion

NPTMs, both preoperatively and postoperatively, were closely related to the prognosis of stage I-III colorectal cancer patients. Compared with the AJCC 8th TNM stage, the nomograms based on preoperative and postoperative CEA and CA19-9 demonstrated

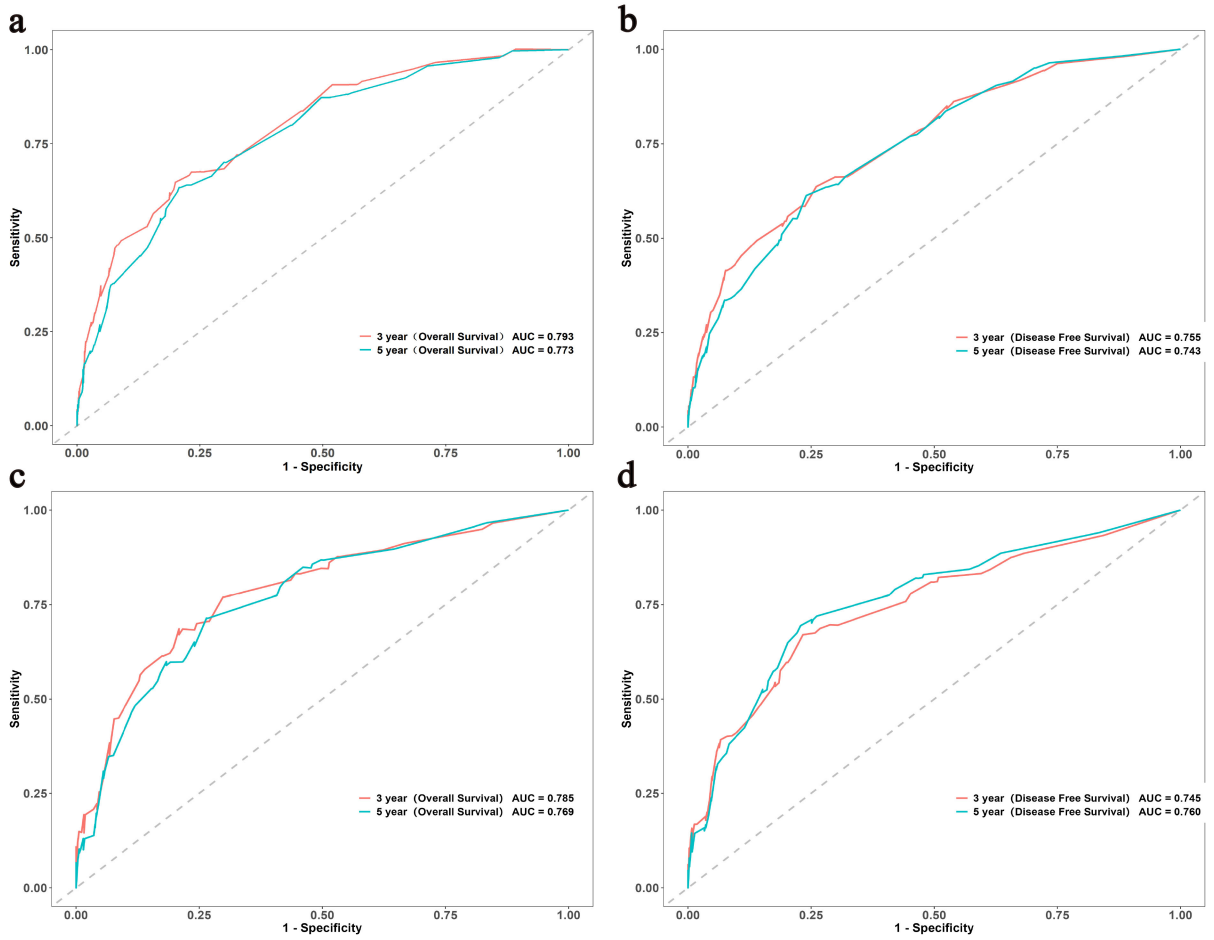



FIGURE 4
The ROC curves of the OS and DFS in both the discovery **(A, B)** and validation **(C, D)** cohorts. ROC, Receiver Operating Characteristic; OS, overall survival; DFS, disease-free survival.

TABLE 5 Comparison between nomograms and pTNM stage in C-index, NRI, and IDI.

| Index | Discovery cohort | | | Validation cohort | | |
|---|---|---|---|---|---|---|
| | Estimate | 95% CI | P value | Estimate | 95% CI | P value |
| NRI (vs.pTNM stage) | | | | | | |
| For 3-year OS | 0.519 | 0.389 - 0.653 | | 0.464 | 0.298 - 0.627 | |
| For 5-year OS | 0.515 | 0.378 - 0.646 | | 0.472 | 0.308 - 0.635 | |
| IDI (vs.pTNM stage) | | | | | | |
| For 3-year OS | 0.131 | 0.087 - 0.197 | 0.000 | 0.172 | 0.095 - 0.273 | 0.000 |
| For 5-year OS | 0.117 | 0.083 - 0.172 | 0.000 | 0.140 | 0.088 - 0.217 | 0.000 |
| C-index (OS) | | | | | | |
| The nomogram | 0.760 | 0.724 - 0.796 | | 0.772 | 0.723 - 0.821 | |
| The pTNM stage | 0.655 | 0.618 - 0.691 | | 0.665 | 0.619 - 0.711 | |
| Change | 0.105 | 0.064 - 0.142 | 0.000 | 0.107 | 0.056 - 0.165 | 0.000 |

*(Continued)*

TABLE 5 Continued

| Index | Discovery cohort | | | Validation cohort | | |
|---|---|---|---|---|---|---|
| | Estimate | 95% CI | P value | Estimate | 95% CI | P value |
| NRI (vs.pTNM stage) | | | | | | |
| For 3-year DFS | 0.481 | 0.344 - 0.587 | | 0.423 | 0.243 - 0.648 | |
| For 5-year DFS | 0.444 | 0.307 - 0.555 | | 0.406 | 0.240 - 0.580 | |
| IDI (vs.pTNM stage) | | | | | | |
| For 3-year DFS | 0.101 | 0.071 - 0.149 | 0.000 | 0.126 | 0.080 - 0.206 | 0.000 |
| For 5-year DFS | 0.098 | 0.066 - 0.145 | 0.000 | 0.109 | 0.061 - 0.186 | 0.000 |
| C-index (DFS) | | | | | | |
| The nomogram | 0.724 | 0.689 - 0.759 | | 0.748 | 0.701 - 0.795 | |
| The pTNM stage | 0.633 | 0.600 - 0.667 | | 0.676 | 0.635 - 0.716 | |
| Change | 0.090 | 0.051 - 0.132 | 0.000 | 0.072 | 0.034 - 0.112 | 0.000 |

superior predictive capability and clinical applicability, offering more precise prognosis for colorectal cancer patients. The results of this study suggest that preoperative and postoperative CEA and CA199, are crucial in predicting patients' prognosis, and both clinicians and patients should be aware of the importance of the preoperative and postoperative testing of these two tumour markers. Therefore, it is not recommended to ignore the testing of these markers for various reasons. Further validation of the

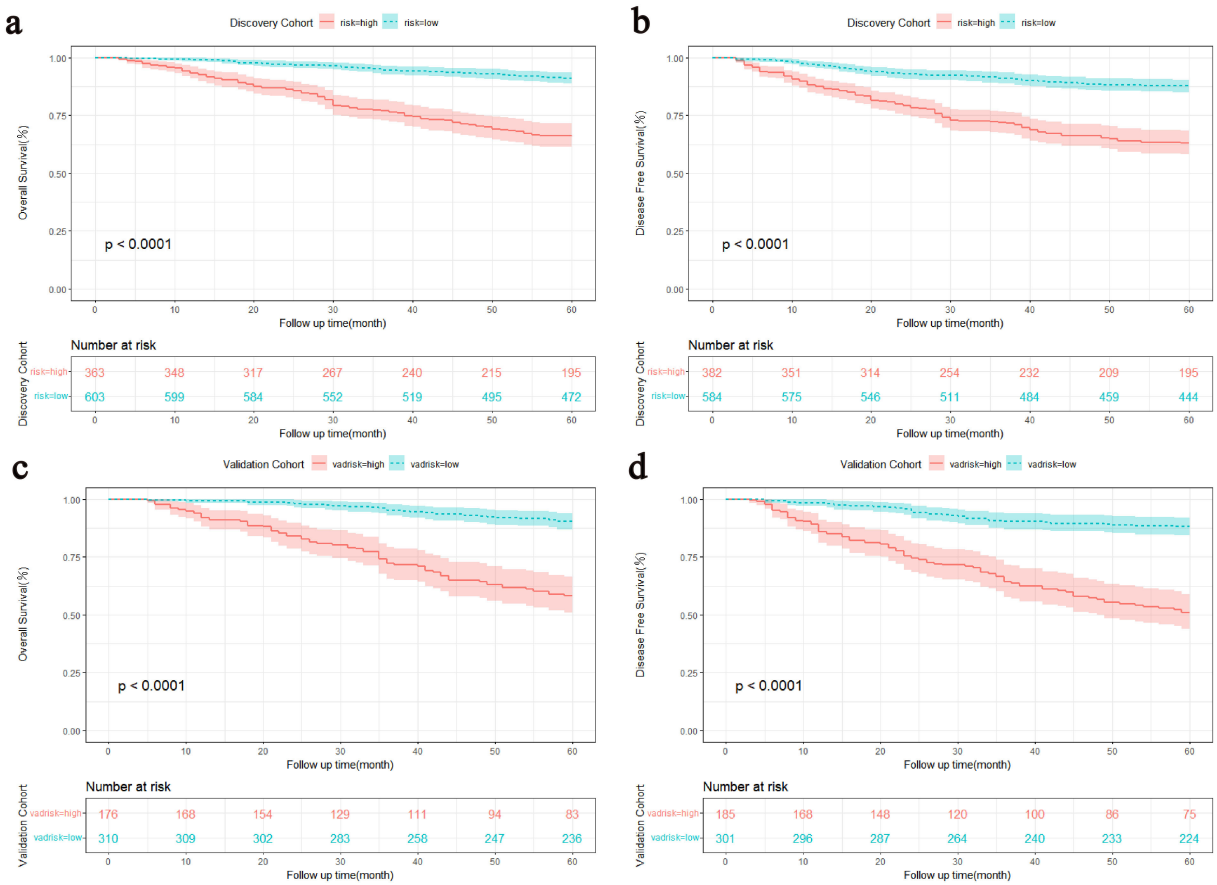



FIGURE 5
Kaplan-Meier survival curves stratified by the median of the total score of the discovery cohort. **(A, C)** OS curves for the discovery and validation cohorts. **(B, D)** DFS curves for the discovery and validation cohorts. OS, Overall survival; DFS, Disease-free survival.

nomograms in different cohorts is needed to enhance their generalizability.

## Data availability statement

The data presented in this study are available on reasonable request from the corresponding author. The data are not publicly available due to privacy.

## Ethics statement

The studies involving humans were approved by the Ethics Committee of Xinhua Hospital Affiliated to Shanghai Jiaotong University School of Medicine. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

XD: Formal analysis, Methodology, Writing – original draft, Data curation. YL: Data curation, Formal analysis, Methodology, Writing – original draft. HW: Data curation, Writing – original draft. ZD: Data curation, Writing – review & editing. YC: Data curation, Writing – review & editing. YL: Funding acquisition, Supervision, Writing – review & editing, Conceptualization, Data curation. SH: Conceptualization, Supervision, Writing – review & editing, Data curation.

## Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. The work was supported by grants from the National Natural Science Foundation of China (82073056) and Shanghai Pujiang Talent Program (19PJ1407600).

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1402847/full#supplementary-material

SUPPLEMENTARY FIGURE 1
The calibration curves for forecasting the survival of the discovery cohort and validation cohort. **(A, C)** Calibration curves for the OS of the discovery and validation cohorts. **(B, D)** Calibration curves for the DFS of the discovery and validation cohorts. OS, overall survival; DFS, disease-free survival. A model's predicted probability or score is represented on the x-axis. This is the model's estimate of how likely an event is to occur. In addition, the y-axis shows the rate of event occurrence, also in the range 0 to 1. This is the proportion of events that occur in real data. Using the calibration curve, we can plot the relationship between the predicted probabilities of the model and the actual observations. A dotted line represents the ideal calibration line of the theory, which is 45 degrees diagonally. When the calibration curve coincides with this line, the model makes perfect predictions. Error lines (yellow and blue) on the calibration curve indicate its uncertainty. Usually, error lines represent confidence intervals.

SUPPLEMENTARY FIGURE 2
DCA of the nomogram and the AJCC 8th TNM stage for the survival prediction. **(A, C)** DCA for 3-year OS in the discovery and validation cohorts. **(B, D)** DCA for 5-year OS in the discovery and validation cohorts. **(E, G)** DCA for 3-year DFS in the discovery and validation cohorts. **(F, H)** DCA for 5-year DFS in the discovery and validation cohorts. DCA, Decision Curve Analysis; OS, overall survival; DFS, disease-free survival. By quantifying the net benefit at different threshold probabilities, DCA helps evaluate the clinical utility of a model. As benchmarks, we used curves representing full treatment, denoting maximum clinical benefits, and no treatment, denoting no clinical benefits.

## References

1. Siegel RL, Miller KD, Goding Sauer A, Fedewa SA, Butterly LF, Anderson JC, et al. Colorectal cancer statistics, 2020. *CA Cancer J Clin*. (2020) 70:145–64. doi: 10.3322/caac.21601

2. Hashiguchi Y, Muro K, Saito Y, Ito Y, Ajioka Y, Hamaguchi T, et al. Japanese Society for Cancer of the Colon and Rectum (JSCCR) guidelines 2019 for the treatment of colorectal cancer. *Int J Clin Oncol*. (2020) 25:1–42. doi: 10.1007/s10147-019-01485-z

3. Steele SR, Chang GJ, Hendren S, Weiser M, Irani J, Buie WD, et al. Practice guideline for the surveillance of patients after curative treatment of colon and rectal cancer. *Dis colon rectum*. (2015) 58:713–25. doi: 10.1097/DCR.0000000000000410

4. Weitz J, Koch M, Debus J, Höhler T, Galle PR, Büchler MW. Colorectal cancer. *Lancet*. (2005) 365:153–65. doi: 10.1016/s0140-6736(05)17706-x

5. Wang D, Rayani S, Marshall JL. Carcinoembryonic antigen as a vaccine target. *Expert Rev Vaccines*. (2008) 7:987–93. doi: 10.1586/14760584.7.7.987

6. Argiles G, Tabernero J, Labianca R, Hochhauser D, Salazar R, Iveson T, et al. Localised colon cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann oncology: Off J Eur Soc Med Oncol*. (2020) 31:1291–305. doi: 10.1016/j.annonc.2020.06.022

7. Duffy MJ, van Dalen A, Haglund C, Hansson L, Holinski-Feder E, Klapdor R, et al. Tumour markers in colorectal cancer: European Group on Tumour Markers (EGTM) guidelines for clinical use. *Eur J Cancer (Oxford England: 1990)*. (2007) 43:1348–60. doi: 10.1016/j.ejca.2007.03.021

8. Kouri M, Pyrhönen S, Kuusela P. Elevated CA19-9 as the most significant prognostic factor in advanced colorectal carcinoma. *J Surg Oncol*. (1992) 49:78–85. doi: 10.1002/jso.2930490204

9. Li C, Zhao K, Zhang D, Pang X, Pu H, Lei M, et al. Prediction models of colorectal cancer prognosis incorporating perioperative longitudinal serum tumor markers: a

retrospective longitudinal cohort study. *BMC Med*. (2023) 21:63. doi: 10.1186/s12916-023-02773-2

10. Lin JX, Wang W, Lin JP, Xie JW, Wang JB, Lu J, et al. Preoperative tumor markers independently predict survival in stage III gastric cancer patients: should we include tumor markers in AJCC staging? *Ann Surg Oncol*. (2018) 25:2703–12. doi: 10.1245/s10434-018-6634-z

11. Lin JP, Lin JX, Ma YB, Xie JW, Yan S, Wang JB, et al. Prognostic significance of pre- and post-operative tumour markers for patients with gastric cancer. *Br J cancer*. (2020) 123:418–25. doi: 10.1038/s41416-020-0901-z

12. You W, Sheng N, Yan L, Chen H, Gong J, He Z, et al. The difference in prognosis of stage II and III colorectal cancer based on preoperative serum tumor markers. *J Cancer*. (2019) 10:3757–66. doi: 10.7150/jca.31660

13. Konishi T, Shimada Y, Hsu M, Tufts L, Jimenez-Rodriguez R, Cercek A, et al. Association of preoperative and postoperative serum carcinoembryonic antigen and colon cancer outcome. *JAMA Oncol*. (2018) 4:309–15. doi: 10.1001/jamaoncol.2017.4420

14. Sonoda H, Yamada T, Matsuda A, Ohta R, Shinji S, Yokoyama Y, et al. Elevated serum carcinoembryonic antigen level after curative surgery is a prognostic biomarker of stage II-III colorectal cancer. *Eur J Surg Oncol*. (2021) 47:2880–7. doi: 10.1016/j.ejso.2021.05.041

15. Zheng Z, Wang X, Lu X, Huang Y, Chi P. Prognostic significance of carcinoembryonic antigen combined with carbohydrate antigen 19-9 following neoadjuvant chemoradiotherapy in patients with locally advanced rectal cancer. *Colorectal Dis*. (2021) 23:2320–30. doi: 10.1111/codi.15694

16. Shan J, Gu B, Shi L, Wang X, Ye W, Zhou W, et al. Prognostic value of CEA and CA19-9 in patients with local advanced rectal cancer receiving neoadjuvant chemoradiotherapy, radical surgery and postoperative chemotherapy. *Transl Cancer Res*. (2021) 10:88–98. doi: 10.21037/tcr-20-2269

17. Iasonos A, Schrag D, Raj GV, Panageas KS. How to build and interpret a nomogram for cancer prognosis. *J Clin Oncol*. (2008) 26:1364–70. doi: 10.1200/jco.2007.12.9791

18. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol*. (2015) 16:e173–80. doi: 10.1016/s1470-2045(14)71116-7

19. Nicholson BD, Shinkins B, Pathiraja I, Roberts NW, James TJ, Mallett S, et al. Blood CEA levels for detecting recurrent colorectal cancer. *Cochrane Database Syst Rev*. (2015) 2015:Cd011134. doi: 10.1002/14651858.CD011134.pub2

20. Moretto R, Rossini D, Conca V, Lonardi S, Rasola C, Antoniotti C, et al. CEA increase as a marker of disease progression after first-line induction therapy in metastatic colorectal cancer patients. A pooled analysis of TRIBE and TRIBE2 studies. *Br J Cancer*. (2021) 125:839–45. doi: 10.1038/s41416-021-01483-x

21. Wu L, Huang P, Wang F, Li D, Xie E, Zhang Y, et al. Relationship between serum CA19-9 and CEA levels and prognosis of pancreatic cancer. *Ann Transl Med*. (2015) 3:328. doi: 10.3978/j.issn.2305-5839.2015.11.17

22. Sahoo DK, Borcherding DC, Chandra L, Jergens AE, Atherly T, Bourgois-Mochel A, et al. Differential transcriptomic profiles following stimulation with lipopolysaccharide in intestinal organoids from dogs with inflammatory bowel disease and intestinal mast cell tumor. *Cancers*. (2022) 14:3525. doi: 10.3390/cancers14143525

23. Li R, Zhou R, Wang H, Li W, Pan M, Yao X, et al. Gut microbiota-stimulated cathepsin K secretion mediates TLR4-dependent M2 macrophage polarization and promotes tumor metastasis in colorectal cancer. *Cell Death Differentiation*. (2019) 26:2447–63. doi: 10.1038/s41418-019-0312-y

24. Cassinotti E, Boni L, Segato S, Rausei S, Marzorati A, Rovera F, et al. Free circulating DNA as a biomarker of colorectal cancer. *Int J Surg*. (2013) 11:S54–7. doi: 10.1016/s1743-9191(13)60017-5

25. Ogunwobi OO, Mahmood F, Akingboye A. Biomarkers in colorectal cancer: current research and future prospects. *Int J Mol Sci*. (2020) 21:5311. doi: 10.3390/ijms21155311

26. Zhao J, Zhao H, Jia T, Yang S, Wang X. Combination of changes in CEA and CA199 concentration after neoadjuvant chemoradiotherapy could predict the prognosis of stage II/III rectal cancer patients receiving neoadjuvant chemoradiotherapy followed by total mesorectal excision. *Cancer Manage Res*. (2022) 14:2933–44. doi: 10.2147/cmar.S377784

27. Paku M, Uemura M, Kitakaze M, Miyoshi N, Takahashi H, Mizushima T, et al. Clinical significance of preoperative and postoperative serum CEA and carbohydrate antigen 19-9 levels in patients undergoing curative resection of locally recurrent rectal cancer. *Dis colon rectum*. (2023)-) 66:392–400. doi: 10.1097/dcr.0000000000002655

28. Beom SH, Shin SJ, Kim CG, Kim JH, Hur H, Min BS, et al. Clinical significance of preoperative serum carcinoembryonic antigen within the normal range in colorectal cancer patients undergoing curative resection. *Ann Surg Oncol*. (2020) 27:2774–83. doi: 10.1245/s10434-020-08256-5

29. Kim CG, Ahn JB, Jung M, Beom SH, Heo SJ, Kim JH, et al. Preoperative serum carcinoembryonic antigen level as a prognostic factor for recurrence and survival after curative resection followed by adjuvant chemotherapy in stage III colon cancer. *Ann Surg Oncol*. (2017) 24:227–35. doi: 10.1245/s10434-016-5613-5

30. Vural S, Muhtaroğlu A, Uygur FA. The relationship between preoperative CEA and CA19-9 status and patient characteristics and lymph node involvement in early-stage colon cancer. *Eur Rev Med Pharmacol Sci*. (2023) 27:4563–9. doi: 10.26355/eurrev_202305_32462

31. Toyoda H, Kumada T, Tada T, Niinomi T, Ito T, Kaneoka Y, et al. Prognostic significance of a combination of pre- and post-treatment tumor markers for hepatocellular carcinoma curatively treated with hepatectomy. *J Hepatol*. (2012) 57:1251–7. doi: 10.1016/j.jhep.2012.07.018

32. Luo H, Zhao Q, Wei W, Zheng L, Yi S, Li G, et al. Circulating tumor DNA methylation profiles enable early diagnosis, prognosis prediction, and screening for colorectal cancer. *Sci Trans Med*. (2020) 12:eaax7533. doi: 10.1126/scitranslmed.aax7533

33. Zhou H, Zhu L, Song J, Wang G, Li P, Li W, et al. Liquid biopsy at the frontier of detection, prognosis and progression monitoring in colorectal cancer. *Mol cancer*. (2022) 21:86. doi: 10.1186/s12943-022-01556-2

34. Reece M, Saluja H, Hollington P, Karapetis CS, Vatandoust S, Young GP, et al. The use of circulating tumor DNA to monitor and predict response to treatment in colorectal cancer. *Front Genet*. (2019) 10:1118. doi: 10.3389/fgene.2019.01118

35. Yamamoto H, Murata K, Fukunaga M, Ohnishi T, Noura S, Miyake Y, et al. Micrometastasis volume in lymph nodes determines disease recurrence rate of stage II colorectal cancer: A prospective multicenter trial. *Clin Cancer Res*. (2016) 22:3201–8. doi: 10.1158/1078-0432.CCR-15-2199

36. Bhatti I, Patel M, Dennison AR, Thomas MW, Garcea G. Utility of postoperative CEA for surveillance of recurrence after resection of primary colorectal cancer. *Int J Surg (London England)*. (2015) 16:123–8. doi: 10.1016/j.ijsu.2015.03.002

37. Mo S, Zhou Z, Li Y, Hu X, Ma X, Zhang L, et al. Establishment and validation of a novel nomogram incorporating clinicopathological parameters into the TNM staging system to predict prognosis for stage II colorectal cancer. *Cancer Cell Int*. (2020) 20:285. doi: 10.1186/s12935-020-01382-w

38. Chen L, Ma X, Dong H, Qu B, Yang T, Xu M, et al. Construction and assessment of a joint prediction model and nomogram for colorectal cancer. *J Gastrointest Oncol*. (2022) 13:2406–14. doi: 10.21037/jgo-22-917

39. Kuang J, Gong Y, Xie H, Yan L, Huang S, Gao F, et al. The prognostic value of preoperative serum CA724 for CEA-normal colorectal cancer patients. *PeerJ*. (2020) 8:e8936. doi: 10.7717/peerj.8936

40. Wang Z, Cheng Y, Seaberg EC, Becker JT. Quantifying diagnostic accuracy improvement of new biomarkers for competing risk outcomes. *Biostatistics (Oxford England)*. (2020) 23:666–82. doi: 10.1093/biostatistics/kxaa048

41. Deng X, Hou H, Wang X, Li Q, Li X, Yang Z, et al. Development and validation of a nomogram to better predict hypertension based on a 10-year retrospective cohort study in China. *Elife*. (2021) 10:e66419. doi: 10.7554/eLife.66419

frontiers | Frontiers in Medicine

TYPE Original Research
PUBLISHED 21 November 2024
DOI 10.3389/fmed.2024.1487443



OPEN ACCESS

EDITED BY
Romy Monika Heilmann,
Leipzig University, Germany

REVIEWED BY
Barbara Adamik,
Wroclaw Medical University, Poland
Georgia Damoraki,
National and Kapodistrian University of Athens, Greece

*CORRESPONDENCE
Bing Wei
✉ weibing202408@126.com

RECEIVED 28 August 2024
ACCEPTED 05 November 2024
PUBLISHED 21 November 2024

CITATION
Ye X, Wang J, Hu L, Zhang Y, Li Y, Xuan J, Han S, Qu Y, Yang L, Yang J, Wang J and Wei B (2024) The diagnostic and prognostic value of soluble ST2 in Sepsis.
*Front. Med.* 11:1487443.
doi: 10.3389/fmed.2024.1487443



# The diagnostic and prognostic value of soluble ST2 in Sepsis

Xinghua Ye, Jia Wang, Le Hu, Ying Zhang, Yixuan Li, Jingchao Xuan, Silu Han, Yifan Qu, Long Yang, Jun Yang, Junyu Wang and Bing Wei*

Emergency Medicine Clinical Research Center, Beijing Key Laboratory of Cardiopulmonary Cerebral Resuscitation, Beijing Chao-Yang Hospital, Capital Medical University, Beijing, China

**Objective:** To determine the diagnostic and prognostic value of soluble suppression of tumorigenicity 2 (sST2) in patients with sepsis.

**Methods:** A total of 113 critically ill patients were enrolled at the emergency department of Beijing Chaoyang Hospital Jing Xi Branch. Venous blood levels of sST2 were measured using the AFIAS-6 dry fluorescence immunoassay analyzer. Based on Sepsis 3.0 criteria, patients were categorized into a sepsis group (76 cases) and a non-sepsis group (37 cases). The sepsis group was further divided into non-survivors (38 cases) and survivors (38 cases) based on 28-day survival outcomes. The vital signs, blood gas analysis, routine blood tests, liver and kidney function tests, procalcitonin (PCT), C-reactive protein (CRP), sST2, left ventricular ejection fraction (LVEF), and other basic characteristics of the patients were recorded. Further, the SOFA, qSOFA and APACHE II scores of each patient were calculated. Statistical analysis was performed using SPSS 25.0, including logistic regression and ROC curve analysis to assess prognostic factors.

**Results:** The serum sST2 levels in the sepsis group ($125.00 \pm 60.32$ ng/mL) were significantly higher than in the non-sepsis group ($58.55 \pm 39.03$ ng/mL) ($p < 0.05$). The SOFA score ($8.08 \pm 2.88$), APACHE II score ($18.00 \pm 4.72$), blood sST2 levels ($168.06 \pm 36.75$ ng/mL) and lactic acid levels ($2.89 \pm 3.28$) in the non-survivor group were significantly higher than the survivor group ($p < 0.05$). Multiple logistic regression analysis showed that sST2, SOFA score, APACHE II score and lactic acid levels were independent risk factors for poor prognosis in patients with sepsis. The ROC curve analysis of the above indexes showed no significant differences between the AUC of sST2 (0.912) and the SOFA score (0.929) ($z = 0.389$, $p = 0.697$), or the APACHE II score (0.933) ($z = 0.484$, $p = 0.627$). However, there was a significant difference between the AUC of sST2 (0.912) and lactic acid levels (0.768) ($z = 2.153$, $p = 0.030$).

**Conclusion:** Blood levels of sST2 show a clinically diagnostic and prognostic value in sepsis. Further, sST2 shows a similar predictive ability as the SOFA and APACHE II scores in determining the prognosis of sepsis patients. However, sST2 has a higher predictive ability than lactic acid levels in determining prognosis in sepsis.

KEYWORDS

APACHE II score, diagnosis, lactic acid, prognosis, Sepsis, SOFA score, soluble ST2

# Introduction

Sepsis, a life-threatening condition characterized by a dysregulated host response to infection, remains a significant contributor to mortality rates in intensive care units worldwide, affecting approximately 20–30% of individuals hospitalized in these critical care settings (1). Despite advances in critical care, the early diagnosis and accurate prognosis of sepsis continue to pose significant challenges (2). Accurate biomarkers that can facilitate early detection and provide reliable prognostic information are urgently needed to improve patient outcomes.

Interleukin 1 receptor-like 1, also known as Suppression of tumorigenicity 2 (ST2), is a member of the interleukin-1 (IL-1) receptor family. In recent years, ST2 has attracted attention as a new marker in heart failure and inflammation (3). ST2, a specific receptor for IL-33 within the IL-1 receptor family, plays a crucial role in immune regulation and systemic inflammatory responses (4–6). It exists in four isoforms: transmembrane ST2L, soluble sST2, truncated ST2v, and long variant ST2LV. sST2 (soluble ST2) competitively binds to IL-33, preventing its interaction with membrane-bound ST2 and inhibiting subsequent signaling. In cases of severe infections, sST2 functions as a negative regulator by binding to IL-33, thereby contributing to immunosuppression (7, 8). Several studies have reported that the IL-33 / ST2 signaling pathway is crucial in various inflammatory diseases, cancer, and heart diseases (9–11). However, only a few studies have investigated the role of ST2 in sepsis. Given the pivotal role of inflammation in the pathophysiology of sepsis, the potential utility of soluble ST2 (sST2) as a diagnostic and prognostic biomarker warrants thorough investigation. This study analyzed the blood levels of sST2 in acute and critically ill patients. Further, the study also explored the diagnostic and prognostic role of sST2 in sepsis patients. By shedding light on the diagnostic and prognostic value of sST2 in sepsis, this study contributes to the broader effort to improve outcomes in this challenging and often fatal condition.

# Materials and methods

## Study population

A total of 120 sepsis patients were screened from December 2020 to April 2021. Of these, 7 patients were excluded: 5 due to missing or incomplete data and 2 who refused treatment in the emergency room. Consequently, 113 acute and critically ill patients were prospectively enrolled in the emergency department at the Beijing Chaoyang Hospital Jing Xi Branch during this period. Inclusion criteria were: (a) patients aged ≥18 years and (b) patients with a diagnosis of infectious diseases during the admission period from December 2020 to April 2021. Exclusion criteria included: (a) age < 18 years, (b) missing or incomplete patient data, and (c) refusal to be managed in the emergency department. The flowchart of the patient screening process is presented in Figure 1. The patients included 53 males and 60 females, aged between 33 and 94 years. Sepsis and septic shock (hereinafter referred to as "sepsis 3.0") were diagnosed based on the international consensus on the definition of sepsis published by the European Society of Critical Care Medicine in 2016 (12). The patients were then classified into the sepsis group ($n = 76$) and the non-sepsis group ($n = 37$). Patients in the sepsis group were further subdivided based on the outcome after 28 days into the non-survivor group (38 cases) and the survivor group (38 cases). Routine diagnostic tests were conducted, and treatment was optimized based on the outcomes of the tests. Data on the vital signs, routine blood tests, liver and kidney function tests, blood gas analysis, C-reactive protein (CRP), procalcitonin (PCT), and cardiopulmonary function were recorded. Collected vital signs included body temperature, heart rate, respiratory rate, and mean arterial pressure (MAP). Routine blood tests comprised white blood cell count (WBC), hemoglobin level (HB), hematocrit (HCT), platelet count (PLT), and liver function tests such as aspartate aminotransferase (AST), alanine aminotransferase (ALT), total bilirubin (TBIL), and albumin (ALB). Kidney function tests included blood urea nitrogen (BUN) and creatinine (CR) measured from blood serum samples. Blood gas analysis included pH, partial pressure of oxygen (PaO2), partial pressure of carbon dioxide (PaCO2), and lactate level, all determined from arterial blood samples. Cardiopulmonary function parameters included ejection fraction (EF) (%) and oxygenation index, which were measured using echocardiography and blood gas analysis, respectively. These parameters were measured and assessed immediately upon the patient's arrival in the emergency room. The SOFA and APACHE II scores were subsequently calculated based on the collected data. sST2 was measured within 72 h of admission. Further, patient's survival was followed up for 28 days. This study obtained the informed consent of all patients and their families, signed the informed consent form, and was approved by the ethics committee of Beijing Chao Yang Hospital, Capital Medical University (number: 2020-6-17-2).

## ST2 detection by immunofluorescence

Venous blood was collected within 2 h of the patient appearing in the emergency department. The samples were collected into purple capped tubes lined with K2-EDTA anticoagulant. ST2 was analyzed using an automated immunofluorescence immunoassay system (AFIAS) immune analyzer (Model: AFIAS-6, Origin: Korea) and AFIAS ST2 Kit (REF: SMFP-70, Origin: Korea). All methods were performed in accordance with the relevant guidelines and regulations in the methods section to this effect.

## Statistical analysis

Statistical analysis was conducted using the statistical software SPSS 25.0. The normal distribution of data was assessed using the nonparametric Kolmogorov–Smirnov test. Data were expressed as mean ± standard deviation ($\bar{x} \pm s$) for normally distributed data or median and interquartile range for not normally distributed data.

Abbreviations: ALB, Albumin; ALT, Alanine Aminotransferase; APACHE II, Acute Physiology and Chronic Health Evaluation II Score; AST, Aspartate Aminotransferase; BNP, B-type Natriuretic Peptide; BUN, Blood Urea Nitrogen; CR, Creatinine; CRP, C-Reactive Protein; EF, Ejection Fraction; HGB, Hemoglobin; HCT, Hematocrit; LAC, Lactic Acid; LVEF, Left Ventricular Ejection Fraction; PaCO2, Partial Pressure of Carbon Dioxide; PaO2, Partial Pressure of Oxygen; PLT, Platelet Count; PCT, Procalcitonin; sST2, Soluble suppression of tumorigenicity 2; SOFA, Sequential Organ Failure Assessment; TBIL, Total Bilirubin; TnI, Troponin I; WBC, White Blood Cell Count.

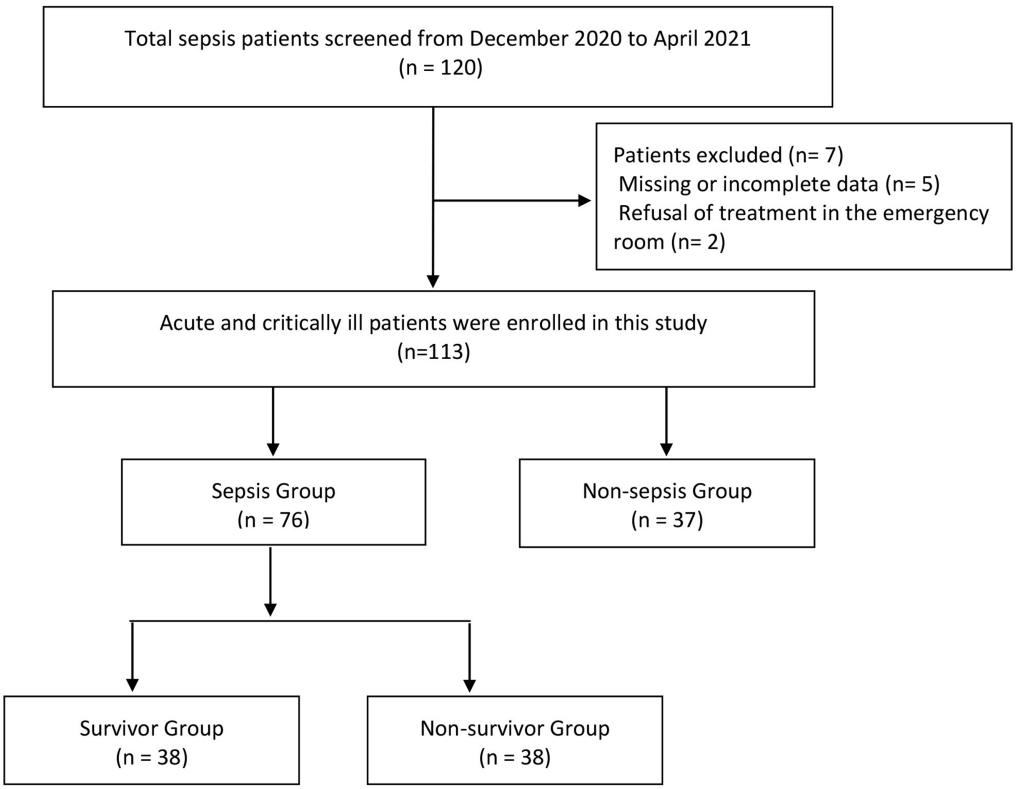



FIGURE 1
The flowchart of the patient screening process.

Differences in qualitative parameters between groups were assessed using two independent sample *t*-test (for normally distributed data). In contrast, the Mann Whitney U test was used for comparisons between groups (for not normally distributed data). On the other hand, one-way ANOVA was used for comparison between multiple groups. Categorical variables were expressed as numbers, and the data were analyzed using the chi-square test. Correlation between variables was conducted using Spearman correlation coefficients. The logistic regression model was used to analyze the prognostic factors. The receiver operating characteristic curve (ROC curve) was plotted to evaluate factors affecting patient prognosis. Statistically significant differences were considered at a $p$-value $< 0.05$.

# Results

## Comparison of the general information

There were 76 patients in the sepsis group, with an average age of 80.75 years, including 33 males and 43 females. However, there were 37 patients in the non-sepsis group, with an average age of 74.38 years, including 20 males and 17 females. There was no statistically significant difference in gender and age between the sepsis group and the non-sepsis group (all $p > 0.05$) (Table 1). Vital signs showed no significant difference in heart rate, mean arterial pressure, and body temperature between the sepsis and non-sepsis groups (all $p > 0.05$). The respiratory rate was significantly higher in the sepsis group ($p = 0.021$). Laboratory findings revealed no significant differences in white blood cell counts and platelet counts between the two groups (all $p > 0.05$). In contrast, hemoglobin levels were significantly reduced in the sepsis group ($p = 0.014$). Additionally, liver enzyme levels (AST and ALT) were comparable across both groups, with all $p$-values greater than 0.05. The baseline characteristics of the sepsis group and the non-sepsis group are summarized in Table 1.

## Comparison of sST2 values between the sepsis group and the non-sepsis group

The sST2 in venous blood of patients in the sepsis group was higher than in the non-sepsis group, with a statistically significant difference (Table 1). The Spearman correlation analysis showed that sST2 was positively correlated with the SOFA score ($r = 0.539$, $p \leq 0.001$) and APACHE II score ($r = 0.482$, $p \leq 0.001$).

## Prognosis prediction in sepsis patients using sST2 and other laboratory parameters

There were no statistically significant differences in age, gender, brain natriuretic peptide (BNP), troponin I (TnI), LVEF, and length of hospital stay between the non-survivor group and the survivor group (all $p > 0.05$) (Table 2). However, the non-survivor group's SOFA score, APACHE II score, lactic acid, PCT, CRP, and sST2 levels were significantly higher than the survivor group (all $p < 0.05$).

TABLE 1 Comparison of baseline characteristics between the sepsis group and non-sepsis group.

| Detection indexes | Sepsis group ($n = 76$) | Non-sepsis group ($n = 37$) | *p*-value |
|---|---|---|---|
| Age (years) | 80.75 ± 8.90 | 74.38 ± 8.95 | 0.782 |
| Male/Female | 33/43 | 20/17 | 0.65 |
| SOFA score | 5.71 ± 3.35 | 2.22 ± 2.32 | 0.005 |
| APACHE II score | 14.34 ± 5.46 | 9.68 ± 4.53 | 0.282 |
| Temperature (°C) | 36.53 ± 0.44 | 36.47 ± 0.37 | 0.441 |
| Heart rate (Bpm) | 90.34 ± 21.07 | 88.43 ± 20.50 | 0.649 |
| Respiratory rate (breaths/min) | 26.70 ± 8.90 | 22.65 ± 8.40 | 0.021 |
| Mean arterial pressure (mmHg) | 92.16 ± 14.56 | 92.06 ± 14.37 | 0.973 |
| pH | 7.43 ± 0.09 | 7.44 ± 0.07 | 0.35 |
| PaCo2 (mmHg) | 42.97 ± 11.88 | 40.05 ± 7.18 | 0.108 |
| PaO2 (mmHg) | 92.28 ± 47.19 | 95.00 ± 37.95 | 0.742 |
| Lactic acid (mmol/l) | 2.03 ± 2.49 | 1.43 ± 1.14 | 0.079 |
| White blood cells ($10^9$/L) | 9.51 ± 4.75 | 9.34 ± 4.26 | 0.845 |
| Platelets ($10^9$/L) | 208.78 ± 69.01 | 213.01 ± 98.04 | 0.792 |
| Hematocrit (%) | 40.52 ± 27.87 | 32.27 ± 6.80 | 0.084 |
| Hemoglobin (g/L) | 104.83 ± 21.83 | 120.14 ± 33.28 | 0.014 |
| Aspartate aminotransferase (U/L) | 37.54 ± 49.30 | 40.75 ± 43.13 | 0.724 |
| Alanine aminotransferase (U/L) | 30.12 ± 37.39 | 28.53 ± 32.80 | 0.818 |
| Total bilirubin (μmol/L) | 16.56 ± 11.02 | 16.46 ± 8.21 | 0.958 |
| Albumin (g/L) | 27.38 ± 6.15 | 28.53 ± 32.80 | 0.014 |
| Blood urea nitrogen (mg/dL) | 16.27 ± 13.69 | 10.78 ± 11.12 | 0.025 |
| Creatinine (μmol/L) | 115.88 ± 110.40 | 100.16 ± 170.76 | 0.612 |
| Sodium (mEq/L) | 140.61 ± 9.00 | 138.07 ± 6.50 | 0.09 |
| Potassium (mEq/L) | 4.03 ± 0.71 | 4.02 ± 0.62 | 0.899 |
| PCT (ng/ml) | 2.01 ± 4.59 | 0.90 ± 2.75 | 0.095 |
| CRP (mg/l) | 49.22 ± 44.76 | 37.22 ± 44.55 | 0.555 |
| BNP (pg/ml) | 584.52 ± 772.13 | 474.27 ± 741.99 | 0.52 |
| TnI (ng/ml) | 0.15 ± 0.64 | 1.36 ± 4.87 | ≤0.001 |
| LVEF (%) | 60.50 ± 11.33 | 60.24 ± 12.54 | 0.239 |
| Hospital stay (days) | 14.64 ± 9.58 | 11.97 ± 8.30 | 0.135 |
| sST2 (ng/ml) | 125.00 ± 60.32 | 58.55 ± 39.03 | ≤0.001 |

Values are expressed as the mean ± standard deviation. APACHE II, acute physiology and chronic health evaluation II; BNP, B-type natriuretic peptide; CRP, C-reactive protein; LVEF, left ventricular ejection fraction; PCT, procalcitonin; SOFA, sequential organ failure assessment; sST2, soluble suppression of tumorigenicity 2; TnI, troponin I.

TABLE 2 Comparison of the detection indexes between the non-survivor group and the survivor group of sepsis patients.

| Detection indexes | Non-survivor group ($n = 38$) | Survivor group ($n = 38$) | *p-value* |
|---|---|---|---|
| Male/Female | 13/25 | 20/18 | 0.108 |
| Age (years) | 81.61 ± 7.93 | 79.89 ± 9.81 | 0.406 |
| SOFA score | 8.08 ± 2.88 | 3.34 ± 1.71 | ≤0.001 |
| APACHE II score | 18.00 ± 4.72 | 10.68 ± 3.26 | ≤0.001 |
| Lactic acid (mmol/l) | 2.89 ± 3.28 | 1.16 ± 0.52 | 0.010 |
| PCT (ng/ml) | 3.65 ± 6.07 | 0.36 ± 0.64 | 0.001 |
| CRP (mg/l) | 66.85 ± 45.38 | 31.59 ± 36.93 | ≤0.001 |
| BNP (pg/ml) | 666.78 ± 697.17 | 502.26 ± 841.74 | 0.356 |
| TnI (ng/ml) | 0.23 ± 0.90 | 0.08 ± 0.08 | 0.305 |
| LVEF (%) | 60.13 ± 8.74 | 60.87 ± 13.55 | 0.779 |
| Hospital stay (days) | 16.45 ± 10.85 | 12.84 ± 7.87 | 0.101 |
| sST2 (ng/ml) | 168.06 ± 36.75 | 81.93 ± 47.08 | ≤0.001 |

Values are expressed as the mean ± standard deviation. APACHE II, acute physiology and chronic health evaluation II; BNP, B-type natriuretic peptide; CRP, C-reactive protein; LVEF, left ventricular ejection fraction; PCT, procalcitonin; SOFA, sequential organ failure assessment; sST2, soluble suppression of tumorigenicity 2; TnI, Troponin I.

Multivariate logistic regression analysis of the statistically significant prognostic factors in the univariate analysis showed that sST2, SOFA score, APACHE II score, and lactic acid levels were independent prognostic factors for sepsis (Table 3). Analysis of the ROC curve showed that the area under the curve (AUC) of the sST2 and SOFA score (0.912 vs. 0.929) ($z = 0.389$, $p = 0.697$), and the area under the curve of the sST2 and Apache II score were not statistically significant (0.912 vs. 0.933) ($z = 0.484$, $p = 0.627$) (Figure 2 and Table 4). However, the AUC of sST2 and lactic acid levels was statistically significant (0.912 vs. 0.768) ($z = 2.153$, $p = 0.030$). sST2 showed a sensitivity, specificity, positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (+LR), and a negative likelihood ratio (−LR) of 97.4%, 76.3%, 80.4%, 96.7%, 4.11, and 0.03, respectively, in predicting the prognosis of sepsis. The SOFA score had a sensitivity, specificity, PPV, NPV, +LR, and −LR of 86.8%, 81.6%, 82.5%, 86.1%, 4.71, and 0.16, respectively, in predicting the prognosis of sepsis. The APACHE II score had a sensitivity, specificity, PPV, NPV, +LR, and −LR of 89.5%, 89.5%, 89.5%, 89.5%, 8.5, and 0.12, respectively, in predicting the prognosis of sepsis. The lactic acid levels had a sensitivity, specificity, PPV, NPV, +LR, and −LR of 71.1%, 73.7%, 73.0%, 71.8%, 2.7 and 0.39, respectively, in predicting the prognosis of sepsis. In summary, sST2 demonstrated prognostic and predictive ability comparable to the SOFA and APACHE II scores in sepsis, and it showed higher predictive ability than lactic acid levels. Moreover, ROC curve analysis revealed that the combination of SOFA with sST2 achieved the highest AUC of 0.973, indicating superior distinguishing ability for predicting outcomes in sepsis (Figure 3). This was closely followed by the combination of APACHE-II with sST2, with an AUC of 0.964. However, the results indicated that the combinations of SOFA with sST2 and APACHE-II with sST2 had similar distinguishing abilities, as there was no statistically significant difference between their AUCs ($z = 0.496$, $p = 0.620$) (Figure 3).

## Discussion

Sepsis is a life-threatening organ dysfunction caused by an imbalance in the body's response to infection, leading to septic shock or multiple organ dysfunction (13). Sepsis is a medical emergency that

TABLE 3 Multivariate logistic regression analysis of the factors affecting prognosis in sepsis patients.

| Detection indexes | Standard error | Wald | Sig. | EXP (B) | 95% CI of EXP(B) | |
|---|---|---|---|---|---|---|
| | | | | | Lower limit | Upper limit |
| SOFA score | 0.529 | 6.334 | 0.046 | 0.349 | 0.124 | 0.984 |
| APACHE II score | 0.237 | 3.888 | 0.049 | 0.626 | 0.394 | 0.997 |
| Lactic acid | 1.850 | 3.954 | 0.047 | 0.025 | 0.001 | 0.949 |
| PCT | 1.094 | 0.003 | 0.959 | 0.945 | 0.111 | 8.061 |
| CRP | 0.017 | 0.085 | 0.771 | 1.005 | 0.973 | 1.038 |
| sST2 | 0.029 | 4.337 | 0.037 | 0.941 | 0.889 | 0.996 |

APACHE II, acute physiology and chronic health evaluation II; CRP, C-reactive protein; Lactic acid, lactic acid; PCT, procalcitonin; sST2, soluble suppression of tumorigenicity 2; SOFA, sequential organ failure assessment.

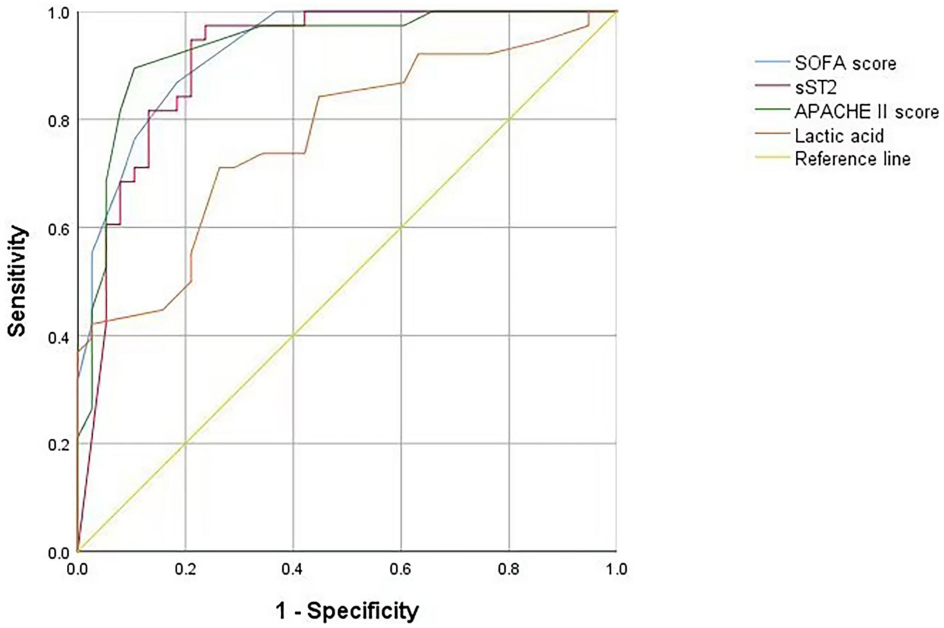



FIGURE 2
ROC curve of the SOFA score, sST2, APACHE II score and lactic acid levels on the prediction of mortality. ROC curve, receiver operating characteristic curve; SOFA score, sequential organ failure assessment score. sST2, soluble suppression of tumorigenicity 2; APACHE II, acute physiology and chronic health evaluation II.

presents as an acute and severe disease. It is associated with high mortality, which can be as high as 40% (14). The occurrence and development of sepsis involve complex immune mechanisms (15, 16). Sepsis is characterized by an inflammatory storm in the early stages and persistent immunosuppression in later stages. Further, it is characterized by reduced innate and acquired immune response and reduced ability for pathogen clearance, resulting in secondary opportunistic infections by pathogenic bacteria or viruses and severe complications (17).

ST2 is a specific receptor of IL-33 in the IL-1 family. IL-33 / ST2 signaling pathway plays an important role in the systemic inflammatory response and immune regulation (4–6, 18). ST2 includes four isoforms, ST2L, sST2, ST2v, and ST2LV. sST2 is a soluble ST2 that can competitively bind to IL-33, inhibiting its biological activity and signal transduction. In severe infection, sST2 acts as a negative regulator and combines with IL-33, thus participating in immunosuppression (7, 8). In the present study, the sST2 levels in the venous blood were higher in the sepsis group than in the non-sepsis group. This finding indicates that sST2 can be used as a diagnostic index in sepsis.

Moreover, in the sepsis group, the blood levels of sST2 were significantly higher in the non-survivor group than in the survivor group, suggesting that the blood levels of sST2 have a high predictivity ability in determining the prognosis of sepsis patients. Higher blood levels of sST2 were positively correlated with a poor prognosis. Furthermore, patients with high SOFA and APACHE II scores also had high blood levels of sST2, with a poorer prognosis, consistent with other studies (19, 20). Therefore, blood levels of sST2 in patients with sepsis can be used as clinical indicators to predict prognosis.

APACHE II scoring system has been widely used in ICUs since its inception in 1985 (21–24). It is of clinical significance as it can objectively evaluate the severity of the patient's condition, guide the monitoring and treatment plans, and evaluate treatment outcomes

TABLE 4 Diagnostic parameters of the SOFA score, sST2, APACHE II score and lactic acid levels.

| Detection indexes | AUC | Cut-off | Sensitivity (%) | Specificity (%) | PPV (%) | NPV (%) | +LR | -LR |
|---|---|---|---|---|---|---|---|---|
| sST2 | 0.912 | 103.055 | 97.4 | 76.3 | 80.4 | 96.7 | 4.11 | 0.03 |
| SOFA score | 0.929 | 4.5 | 86.8 | 81.6 | 82.5 | 86.1 | 4.71 | 0.16 |
| APACHE II score | 0.933 | 13.5 | 89.5 | 89.5 | 89.5 | 89.5 | 8.5 | 0.12 |
| Lactic acid | 0.768 | 1.5 | 71.1 | 73.7 | 73.0 | 71.8 | 2.7 | 0.39 |

APACHE II, acute physiology and chronic health evaluation II; Lactic acid, lactic acid; SOFA, sequential organ failure assessment; sST2, soluble suppression of tumorigenicity 2.

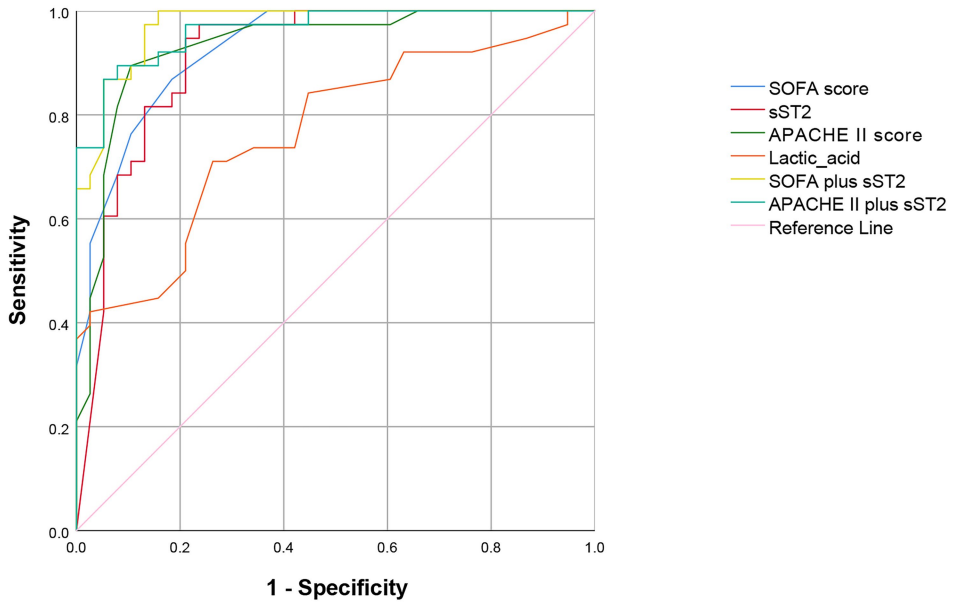



FIGURE 3
ROC curve showing the predictive value of combining SOFA with sST2 and APACHE II with sST2 for mortality prediction. ROC curve, receiver operating characteristic curve; SOFA score, sequential organ failure assessment score. sST2, soluble suppression of tumorigenicity 2; APACHE II, acute physiology and chronic health evaluation II.

(25). Furthermore, it can be used to predict and accurately assess the quality of care in ICU settings.

The SOFA score was described by the European Society of Intensive Care Medicine in 1944. The score aims to describe the occurrence and development of multiple organ dysfunction syndromes (MODS) and to evaluate the incidence rate (23, 24). The SOFA score is based on objective, simple, easy-to-obtain, reliable, and specific continuous variables in evaluating multiple organ dysfunction (26). Patient source, disease type, demographic characteristics, and the treatment administered do not influence these variables. The SOFA score can distinguish the degree of multiple organ dysfunction or failure of a single organ (26).

Lactic acid is a metabolite of anaerobic glycolysis in the human body. Under normal circumstances, levels of lactic acid exceeding 2 mmol/L overwhelm the capacity for liver clearance (27, 28). The dynamic monitoring of blood lactate levels is clinically significant in diagnosing lactic acidosis (29). Increased blood lactate levels can be used to evaluate disease severity and prognosis (27, 28).

The ROC curve analysis showed that sST2, SOFA, and APACHE II scores and the lactic acid levels had a prognostic, predictive ability in sepsis, consistent with previous studies. The sST2 showed similar prognostic and predictive ability with the SOFA and APACHE II scores. However, sST2 had a higher prognostic predictive ability than lactic acid levels. In conclusion, blood levels of sST2 can be used as clinical indices for the diagnosis and prognosis of sepsis.

Our study has several limitations. First, while sST2 demonstrates significant diagnostic and prognostic utility in sepsis, its performance must be validated through more extensive prospective cohort studies with a more diverse patient population. Second, the single-center design may limit the generalizability of our results to other clinical settings. Therefore, multi-center studies are warranted to corroborate the findings across various clinical environments. Third, although the 28-day follow-up period helps assess short-term outcomes, it may not capture the long-term prognostic significance of sST2. Further research is needed to explore this aspect. Fourth, we treated sepsis and septic shock as a homogeneous entity. While our primary focus was to assess the prognostic value of soluble ST2 across the full spectrum of sepsis, this approach may mask differences in outcomes associated with these distinct clinical phenotypes. Future studies should consider applying the Sepsis-3 criteria to provide deeper insights into the differential roles of soluble ST2 in sepsis and septic shock. Finally, while the study compared sST2 with SOFA, APACHE II, and lactic acid levels, additional comparisons with other biomarkers, such as procalcitonin and C-reactive protein, could provide further insights.

## Conclusion

In conclusion, this study demonstrates that sST2 has significant diagnostic and prognostic value in sepsis. The predictive ability of sST2 is comparable to established scoring systems like SOFA and APACHE II, which are widely used for determining sepsis prognosis. Notably, sST2 demonstrates superior predictive capability compared to lactic acid levels for sepsis outcomes, suggesting that sST2 could be a more reliable indicator for identifying patients at higher risk of poor prognosis. These findings support the potential incorporation of sST2 into routine clinical practice for more accurate diagnosis and prognosis of sepsis. Further research is needed to validate these results and explore the practical applications of sST2 in sepsis management.

## Data availability statement

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

## Ethics statement

The studies involving humans were approved by the studies involving humans were approved by Institution Ethics Committee of Beijing Chao Yang Hospital, Capital Medical University (number: 2020-6-17-2). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

XY: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Visualization, Writing – original draft, Writing – review & editing. JiW: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. LH: Writing – review & editing, Writing – original draft, Visualization, Validation, Resources, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. YZ: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. YL: Conceptualization, Formal analysis, Investigation, Methodology, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. JX: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Resources, Software, Validation, Writing – original draft, Writing – review & editing. SH: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Resources, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. YQ: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing. LY: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. JY: Conceptualization, Formal analysis, Investigation, Methodology, Project administration, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. JuW: Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Resources, Validation, Visualization, Writing – original draft, Writing – review & editing. BW: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

## Funding

The author(s) declare that no financial support was received for the research, authorship, and/or publication of this article.

## Acknowledgments

Funding from the Shijingshan District medical key support specialty construction foundation is gratefully acknowledged.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## References

1. Zampieri FG, Bagshaw SM, Semler MW. Fluid therapy for critically ill adults with Sepsis: a review. *JAMA*. (2023) 329:1967–80. doi: 10.1001/jama.2023.7560

2. Santacroce E, D'Angerio M, Ciobanu AL, Masini L, Lo Tartaro D, Coloretti I, et al. Advances and challenges in Sepsis management: modern tools and future directions. *Cells*. (2024) 13:13. doi: 10.3390/cells13050439

3. Kakkar R, Lee RT. The IL-33/ST2 pathway: therapeutic target and novel biomarker. *Nat Rev Drug Discov*. (2008) 7:827–40. doi: 10.1038/nrd2660

4. Homsak E, Gruson D. Soluble ST2: a complex and diverse role in several diseases. *Clinica Chimica Acta*. (2020) 507:75–87. doi: 10.1016/j.cca.2020.04.011

5. Kotsiou OS, Gourgoulianis KI, Zarogiannis SG. IL-33/ST2 Axis in organ fibrosis. *Front Immunol*. (2018) 9:2432. doi: 10.3389/fimmu.2018.02432

6. Griesenauer B, Paczesny S. The ST2/IL-33 Axis in immune cells during inflammatory diseases. *Front Immunol*. (2017) 8:475. doi: 10.3389/fimmu.2017.00475

7. Fattori V, Hohmann MSN, Rossaneis AC, Manchope MF, Alves-Filho JC, Cunha TM, et al. Targeting IL-33/ST2 signaling: regulation of immune function and analgesia. *Expert Opin Ther Targets*. (2017) 21:1141–52. doi: 10.1080/14728222.2017.1398734

8. Altara R, Ghali R, Mallat Z, Cataliotti A, Booz GW, Zouein FA. Conflicting vascular and metabolic impact of the IL-33/sST2 axis. *Cardiovasc Res*. (2018) 114:1578–94. doi: 10.1093/cvr/cvy166

9. He Z, Song J, Hua J, Yang M, Ma Y, Yu T, et al. Mast cells are essential intermediaries in regulating IL-33/ST2 signaling for an immune network favorable to mucosal healing in experimentally inflamed colons. *Cell Death Dis*. (2018) 9:1173. doi: 10.1038/s41419-018-1223-4

10. Li J, Shen D, Tang J, Wang Y, Wang B, Xiao Y, et al. IL33 attenuates ventricular remodeling after myocardial infarction through inducing alternatively activated macrophages ethical standards statement. *Eur J Pharmacol*. (2019) 854:307–19. doi: 10.1016/j.ejphar.2019.04.046

11. Milovanovic M, Volarevic V, Radosavljevic G, Jovanovic I, Pejnovic N, Arsenijevic N, et al. IL-33/ST2 axis in inflammation and immunopathology. *Immunol Res*. (2012) 52:89–99. doi: 10.1007/s12026-012-8283-9

12. Singer M, Deutschman CS, Seymour CW, Shankar-Hari M, Annane D, Bauer M, et al. The third international consensus definitions for Sepsis and septic shock (Sepsis-3). *JAMA*. (2016) 315:801–10. doi: 10.1001/jama.2016.0287

13. Cecconi M, Evans L, Levy M, Rhodes A. Sepsis and septic shock. *Lancet*. (2018) 392:75–87. doi: 10.1016/S0140-6736(18)30696-2

14. Rudd KE, Johnson SC, Agesa KM, Shackelford KA, Tsoi D, Kievlan DR, et al. Global, regional, and national sepsis incidence and mortality, 1990-2017: analysis for the global burden of disease study. *Lancet*. (2020) 395:200–11. doi: 10.1016/S0140-6736(19)32989-7

15. van der Poll T, van de Veerdonk FL, Scicluna BP, Netea MG. The immunopathology of sepsis and potential therapeutic targets. *Nat Rev Immunol*. (2017) 17:407–20. doi: 10.1038/nri.2017.36

16. Font MD, Thyagarajan B, Khanna AK. Sepsis and septic shock - basics of diagnosis, pathophysiology and clinical decision making. *Med Clin North Am*. (2020) 104:573–85. doi: 10.1016/j.mcna.2020.02.011

17. Johnston GR, Webster NR. Cytokines and the immunomodulatory function of the vagus nerve. *Br J Anaesth*. (2009) 102:453–62. doi: 10.1093/bja/aep037

18. Villarreal DO, Weiner DB. Interleukin 33: a switch-hitting cytokine. *Curr Opin Immunol*. (2014) 28:102–6. doi: 10.1016/j.coi.2014.03.004

19. Wei Y, Xiao P, Wu B, Chen F, Shi X. Significance of sTREM-1 and sST2 combined diagnosis for sepsis detection and prognosis prediction. *Open Life Sci*. (2023) 18:20220639. doi: 10.1515/biol-2022-0639

20. Hoogerwerf JJ, Tanck MWT, van Zoelen MAD, Wittebole X, Laterre P-F, van der Poll T. Soluble ST2 plasma concentrations predict mortality in severe sepsis. *Intensive Care Med*. (2010) 36:630–7. doi: 10.1007/s00134-010-1773-0

21. Godinjak A, Iglica A, Rama A, Tančica I, Jusufović S, Ajanović A, et al. Predictive value of SAPS II and APACHE II scoring systems for patient outcome in a medical intensive care unit. *Acta Medica Academica*. (2016) 45:97–103. doi: 10.5644/ama2006-124.165

22. Rivers E, Nguyen B, Havstad S, Ressler J, Muzzin A, Knoblich B, et al. Early goal-directed therapy in the treatment of severe sepsis and septic shock. *N Engl J Med*. (2001) 345:1368–77. doi: 10.1056/NEJMoa010307

23. Basile-Filho A, Lago AF, Menegueti MG, Nicolini EA, Rodrigues LAB, Nunes RS, et al. The use of APACHE II, SOFA, SAPS 3, C-reactive protein/albumin ratio, and lactate to predict mortality of surgical critically ill patients: a retrospective cohort study. *Medicine*. (2019) 98:e16204. doi: 10.1097/MD.0000000000016204

24. Feng Q, Ai Y-H, Gong H, Wu L, Ai M-L, Deng S-Y, et al. Characterization of Sepsis and Sepsis-associated encephalopathy. *J Intensive Care Med*. (2019) 34:938–45. doi: 10.1177/0885066617719750

25. Tian Y, Yao Y, Zhou J, Diao X, Chen H, Cai K, et al. Dynamic APACHE II score to predict the outcome of intensive care unit patients. *Front Med*. (2022) 8:744907. doi: 10.3389/fmed.2021.744907

26. Lambden S, Laterre PF, Levy MM, Francois B. The SOFA score-development, utility and challenges of accurate assessment in clinical trials. *Crit Care*. (2019) 23:374. doi: 10.1186/s13054-019-2663-7

27. Okorie ON, Dellinger P. Lactate: biomarker and potential therapeutic target. *Crit Care Clin*. (2011) 27:299–326. doi: 10.1016/j.ccc.2010.12.013

28. Pan J, Peng M, Liao C, Hu X, Wang A, Li X. Relative efficacy and safety of early lactate clearance-guided therapy resuscitation in patients with sepsis: a meta-analysis. *Medicine*. (2019) 98:e14453. doi: 10.1097/MD.0000000000014453

29. Liu W, Peng L, Hua S. Clinical significance of dynamic monitoring of blood lactic acid, oxygenation index and C-reactive protein levels in patients with severe pneumonia. *Exp Ther Med*. (2015) 10:1824–8. doi: 10.3892/etm.2015.2770

frontiers | Frontiers in Oncology

TYPE Original Research
PUBLISHED 05 December 2024
DOI 10.3389/fonc.2024.1418283



OPEN ACCESS

EDITED BY
Kun Qian,
Shanghai Jiao Tong University, China

REVIEWED BY
Semra Bulbuloglu,
Istanbul Aydın University, Türkiye
Massimo Tusconi,
University of Cagliari, Italy

*CORRESPONDENCE
Jipeng Li
jipengli1974@aliyun.com
Jun Zhu
junzhu@fmmu.edu.cn

†These authors have contributed equally to this work

RECEIVED 16 April 2024
ACCEPTED 18 November 2024
PUBLISHED 05 December 2024

CITATION
Qian L, Song J, Zhang X, Qiao Y, Tan Z, Li S, Zhu J and Li J (2024) Elucidating the causal relationship between 486 genetically predicted blood metabolites and the risk of gastric cancer: a comprehensive Mendelian randomization analysis. *Front. Oncol.* 14:1418283. doi: 10.3389/fonc.2024.1418283



# Elucidating the causal relationship between 486 genetically predicted blood metabolites and the risk of gastric cancer: a comprehensive Mendelian randomization analysis

Lei Qian[1†], Jiawei Song[2†], Xiaoqun Zhang[3†], Yihuan Qiao[4], Zhaobang Tan[5], Shisen Li[5], Jun Zhu[6*] and Jipeng Li[1,5*]

[1]Department of Experiment Surgery, Xijing Hospital, Fourth Military Medical University, Xi'an, China, [2]School of Clinical Medicine, Xi'an Medical University, Xi'an, China, [3]Department of Pharmacy, Shaanxi Provincial Hospital of Chinese Medicine, Xi'an, China, [4]Department of Digestive Surgery, Honghui Hospital, Xi'an Jiaotong University, Xi'an, China, [5]Department of Gastrointestinal Surgery, Xijing Hospital, Fourth Military Medical University, Xi'an, China, [6]Department of Digestive Diseases, Xijing Hospital, Fourth Military Medical University, Xi'an, China

**Background:** Previous epidemiological studies have yielded inconclusive results regarding the causality between blood metabolites and the risk of gastric cancer (GC). To address this shortcoming, we conducted a two-sample Mendelian randomization (MR) study, combined with metabolomics techniques, to elucidate the causality between 486 genetically predicted blood metabolites and GC.

**Methods:** MR analysis and metabolomics techniques such as ultra-high performance liquid chromatography/tandem mass spectrometry (UPLC-MS/MS) and gas chromatography/tandem mass spectrometry (GC-MS/MS) technologies were employed to assess the causality of 486 genetically predicted blood metabolites on the risk of GC. The genome-wide association study (GWAS) summary data for 486 blood metabolites from 7,824 individuals. The GWAS summary data for GC (ebi-a-GCST90018849) were obtained from the IEU Open GWAS project, including 1,029 GC cases and 474,841 controls. Primary causality estimates were obtained using inverse variance weighting (IVW), supplemented with the weighted median, MR-Egger, weighted mode, and simple mode. In addition, we conducted sensitivity analyses (including Cochran's Q, MR-Egger intercept, MR-PRESSO, and leave-one-out tests), Steiger's test, linked disequilibrium score regression, and multivariate MR (MVMR) to improve the assessment of causality between GC and blood metabolite. Finally, we recruited a total of 11 patients diagnosed with gastric cancer from the First Affiliated Hospital of Air Force Military Medical University between September and October 2024. The control group comprised 11 healthy individuals. Serum samples were collected from both groups for the evaluation of blood-related metabolite expression levels using advanced techniques such as

ultra-performance liquid chromatography-tandem mass spectrometry (UPLC-MS/MS) and gas chromatography-mass spectrometry (GC-MS/MS).

**Results:** The MVMR analysis revealed a significant association between genetically predicted elevated levels of tryptophan (odds ratio [OR] = 0.523, 95% confidence interval [CI] = 0.313–0.872, p = 0.013), nonadecanoate (19:0) (odds ratio [OR] = 0.460, 95% confidence interval [CI] = 0.225–0.943, p = 0.034), and erythritol (odds ratio [OR] = 0.672, 95% confidence interval [CI] = 0.468–0.930, p = 0.016) with a decreased risk of gastric cancer. Based on metabolomic techniques such as UPLC-MS/MS and GC-MS/MS analyses, it has been demonstrated that the expression levels of tryptophan, nonadecanoate (19:0), and erythritol are reduced in patients with gastric cancer. This finding aligns with the results obtained from our MR analysis and provides further confirmation regarding the protective role of tryptophan, nonadecanoate (19:0), and erythritol against gastric cancer.

**Conclusions:** These findings indicate that three blood metabolites are causally related to GC and provide new perspectives for combining genomics and metabolomics to study the mechanisms of metabolite-mediated GC development.

KEYWORDS

**blood metabolites, gastric cancer, Mendelian randomization, causality, genomewide association study**

## 1 Introduction

Gastric cancer (GC), primarily characterized as an adenocarcinoma, ranks as fifth most prevalent malignancy and the third leading cause of cancer-related mortality globally in 2020 (1). For early-stage GC, endoscopic mucosal dissection is the main therapeutic approach, boasting an impressive 5-year postoperative survival rate of 92.6% (2). For patients with stage I/II GC who undergo laparoscopic or open distal gastrectomy, the survival rate is also commendable but slightly lower, ranging from 73% to 76% (3). However, it's crucial to note that the overall 5-year survival rate for GC patients, particularly those diagnosed at advanced stages, remains suboptimal, with a median survival of less than one year (4). This disparity underscores the critical need for early detection and intervention to not only enhance survival rates but also significantly reduce healthcare costs. By focusing on prevention and early treatment, we can potentially alleviate the economic burden of GC on both patients and healthcare systems.

Currently, the diagnosis of GC relies primarily on endoscopic and biopsy-based procedures. Although reliable, these methods have drawbacks, such as high financial cost, invasiveness, potential complications, and limited testing resources, which may discourage patient compliance and make them unsuitable for widespread screening initiatives. An ideal alternative would be noninvasive blood tests. Currently used gastrointestinal tumor markers include glycan antigen 199 (CA199) (5) and carcinoembryonic antigen (CEA) (6). Unfortunately, although highly specific, they have low sensitivity and significant rates of false-negative and false-positive results. This calls for further research, possibly in areas such as blood metabolomics, to identify novel biomarkers indicative of GC and facilitate early detection and treatment.

Metabolomics can identify cancer biomarkers and determinants of tumorigenesis by detecting changes in relevant metabolites over the course of disease progression (7). Ikeda et al. found pronounced differences in the serum metabolic profiles of individuals with gastrointestinal malignancies, including esophageal, gastric, and colorectal cancers, compared with those of healthy volunteers (8). Specifically, changes in 3-hydroxypropionic and pyruvic acid levels were found to be sufficiently discriminative to differentiate gastric, esophageal, and colorectal cancers, exceeding the sensitivity and specificity of conventional biomarkers such as CA199 and CEA (8). However, the scientific landscape is currently characterized by a paucity of comprehensive investigations to establish a causal relationship between blood metabolites and GC. Translating these metabolic discoveries into pathophysiological mechanisms and innovative therapeutic strategies remains challenging. Therefore, there is a need for a comprehensive analysis of the interplay between genetic elements and circulating blood metabolites in the etiology of GC.

Mendelian randomization (MR) has emerged as a key methodology in epidemiological research. It derives putative causal relationships between environmental exposure and health outcomes by using distinctive single nucleotide polymorphisms (SNPs) as instrumental variables (9). MR exploits genetic variability to simulate the construct of randomized controlled trials (RCTs). The use of independent genome-wide association study (GWAS) datasets provides the flexibility to independently assess SNPs associated with both exposure and outcome, thereby facilitating two-sample analysis. This technique provides compelling evidence for a causal relationship between disparate phenotypes. By carefully exploring the potential causal relationships between genetic predisposition to disease and various biological traits (e.g., blood metabolomics), MR paves the way for the identification of relevant disease-related biomarkers (10).

Numerous studies have investigated the association between various exposures and GC using magnetic resonance imaging. These investigations have predominantly focused on single exposures or prevalent exposure factors, including body mass index (11), interleukin-6 (12), vitamin D (13), lifestyle patterns such as smoking and alcohol consumption (14), sleep habits (15), and immunoproteins (16). However, investigations of blood metabolites related to GC are scarce. Given the vague nature of the causal link between blood metabolites and GC, we used a two-sample MR methodology, combined with metabolomics techniques, to investigate the causal dynamics between 486 human blood metabolites and GC.

# 2 Materials and methods

## 2.1 Study design

In MR research, the integrity of conclusions depends on three key assumptions (Figure 1). (1)Correlation assumption:a robust and statistically significant association between the SNP and exposure of interest. (2)Independence assumption: the SNP is free of any correlation with potential confounding variables. (3)Exclusivity assumption: the influence of an SNP on outcomes is exclusively mediated by exposure, ruling out any unaccounted pathways. Based on these principles, our methodology included the selection of high-quality comprehensive datasets from accessible GWASs. This allowed us to obtain appropriate instrumental variables (IVs) for the MR analyses, which were critical for identifying the relationships between an array of 486 blood metabolites and susceptibility to GC.

## 2.2 GWAS data sources

A blood metabolite profiling dataset was obtained from the Comprehensive Metabolomics GWAS repository (https://metabolomics.helmholtz-muenchen.de/gwas/). This dataset comprises a diverse European cohort of 7,824 individuals, including 1,768 participants from the KORA F4 study in Germany and 6,056 from the UK Twin Study (17). Genome-wide association and high-throughput metabolomic studies have revealed approximately 2.1 million SNPs and 486 different metabolites. Of these, 309 metabolites were identified and characterized. The identified metabolites were systematically classified into the following eight categories based on their chemical properties: amino acids, carbohydrates, cofactors, vitamins, energy substrates, lipids, nucleotides, peptides, and xenobiotics. These 486 metabolites are compiled in Supplementary Table S1, with the designation "X-" indicating those with yet-to-be-determined chemical properties.

The GWAS data for GC (ebi-a-GCST90018849) were obtained from the IEU Open GWAS Project (https://gwas.mrcieu.ac.uk/). The detailed attributes of the consolidated GC data are described in Supplementary Table S2.

## 2.3 Selection of IVs

In accordance with the fundamental tenets of MR analysis, we meticulously delineated a set of criteria for distinguishing the IVs associated with the 486 metabolites. In keeping with the axiom of relevance, our selection protocol strictly adhered to the established genome-wide significance boundary, setting the threshold to a robust p-value ($P < 5 \times 10\text{-}8$). Recognizing the subtleties inherent in the genetic underpinnings of certain metabolites, for which only a few SNPs were uncovered, a more permissive threshold was adopted ($P < 1 \times 10\text{-}5$) (18, 19).

In a concerted effort to mitigate the confounding intricacies of linkage disequilibrium, a judicious clustering strategy was used, encompassing a swath of 500 kilobase pairs augmented by a correlation coefficient ceiling of 0.01, to isolate SNPs of sovereign genetic locations. In addition, we removed SNPs carrying mismatched alleles or palindromic sequences, which are indicative of genotyping inaccuracies. Simultaneously, we excluded SNPs that had a statistical synergy with the outcome variable or were absent from the outcome cohort, thus preserving the integrity of the assay.

The veracity of each SNP as an IV was judged through the prism of the F statistic (18), given by the following formula, where " $N$" represents the cohort size and " $R^2$ " is the proportional variance attributed to the SNPs within the exposome profile.

$$F = \frac{R^2 \times (N - 2)}{1 - R^2}$$

The calculation of " $R^2$ " involved the following formula, where " $EAF$ " represents the frequency of the effect allele, " $\beta$ " is the regression coefficient explaining the magnitude of the SNP-exposure association, and " $SD$ " is the standard deviation.

$$R^2 = \frac{2 \times EAF \times (1 - EAF) \times \beta^2}{SD^2}$$

To mitigate bias associated with weak instrumental variables, we excluded SNPs with $F < 10$. Subsequently, we identified and

extracted the SNPs for our exposure of interest from the outcome data while excluding those that were significantly related to the outcome ($p < 1 \times 10-5$). The SNPs that survived the rigorous selection process are shown in Supplementary Table S2 (20).

## 2.4 Univariate MR analysis

In this study, a quintet of MR assays was used, with the predominant analysis using the inverse variance-weighted (IVW) paradigm for its statistical power. This approach synthesizes Wald ratio calculations for each SNP-outcome conjugation, providing a composite causal estimate (21). In addition, causal associations between a compendium of 486 metabolites and GC susceptibility were assessed using odds ratios (ORs) embedded within 95% confidence intervals (CI).

To strengthen the robustness and credibility of our MR conclusions, additional checks were performed using MR-Egger and weighted median (WM) evaluations. An MR-Egger inspection was implemented to detect and correct the putative pleiotropic effects, thereby providing more reliable estimates. The use of the weighted median method yields robust causal inferences, mitigates Type I errors, and enhances the detection of authentic effects even when a preponderance of the input comes from potentially compromised IVs. Harmonization of the WM and MR-Egger results ($P < 0.05$) with those of the IVW method, both in terms of the trajectory and magnitude of effect, was essential for confirming the validity of these findings.

MR scatter plots were generated to visualize the hypothesized causal relationship between the identified metabolites and GC risk.

## 2.5 Sensitivity analysis

To unravel the intricacies of SNP heterogeneity, we used Cochran's Q test within the IVW and MR-Egger frameworks. A P-value < 0.05 served as the arbiter of significant heterogeneity. It is worth noting that this statistical test highlights differences in IV effect sizes. In addition, the MR-Egger intercept coupled with the MR-PRESSO analysis tools was used to expose the spectrum of horizontal pleiotropy, with statistical significance determined using a P-value < 0.05 (22). The robustness of our inferential scaffold was tested using careful leave-one-out (LOO) analysis (23). This rigorous technique ensures that the influence of any single SNP does not unduly bias the overarching determination of causality.

## 2.6 Metabolic pathway analysis

To elucidate the underlying biological mechanisms through which prominent blood metabolites influence GC susceptibility, we expanded our analysis to include metabolic pathway exploration. We performed Kyoto Encyclopedia of Genes and Genomes pathway enrichment analysis using MetaboAnalyst (version 5.0; https://www.metaboanalyst.ca/).

## 2.7 Genetic correlation and direction validation

In the context of dissecting genetic correlations between determinants and clinical outcomes, MR estimation could potentially bias the interpretation of causality (24). To circumvent the confounding complications introduced by the coinheritance of significant metabolites and GC risk, we adopted the Linkage Disequilibrium Score Regression (LDSC) methodology. Additionally, we used the Steiger test to assess the potential for reverse causation, which is a critical step in determining whether genetic variants have a stronger association with the determinant than with the consequence (25). Within the confines of the Steiger framework, a P-value of less than 0.05 is statistically significant, supporting the primacy of genetic instruments in modulating the determinant, thus strengthening our primary hypothesis.

## 2.8 Multivariate Mendelian randomization analysis

Given the interrelationships among the salient metabolites that emerged as statistically significant, we conducted a series of multivariate MR (MVMR) analyses to elucidate the distinct causal contributions of multiple metabolite exposure to GC risk. Our baseline MVMR analysis used an IVW strategy, and we refined our investigation using the MR-PRESSO method to identify and correct for potential genetic-level heterogeneity and the confounding effects of outliers. This meticulous approach refines precision and strengthens the integrity of causal inferences.

## 2.9 Statistical analysis

Each MR study used the "TwoSampleMR" (version 0.4.22) software package for R (version 4.1.2) as the computational framework. LDSC was performed using LDSC software (version 1.0.1). The criterion for statistical significance was set at $P < 0.05$. The magnitude and direction of the causal associations between variables were quantified using ORs and their respective 95% CIs.

## 2.10 Metabolomic analysis

From September to October 2024, 11 preoperative blood samples from GC patients admitted to the First Affiliated Hospital of Air Force Military Medical University and 11 blood samples from healthy controls were selected for analysis. All samples were stored at -80°C. This study was approved by the Ethics Committee of the First Affiliated Hospital of Air Force Military Medical University (approval number KY20222083-F-1).

Subsequently, an untargeted metabolomic analysis was conducted utilizing ultra-high performance liquid chromatography/tandem mass spectrometry (UPLC-MS/MS) and gas chromatography/tandem mass spectrometry (GC-MS/MS) technologies based on the

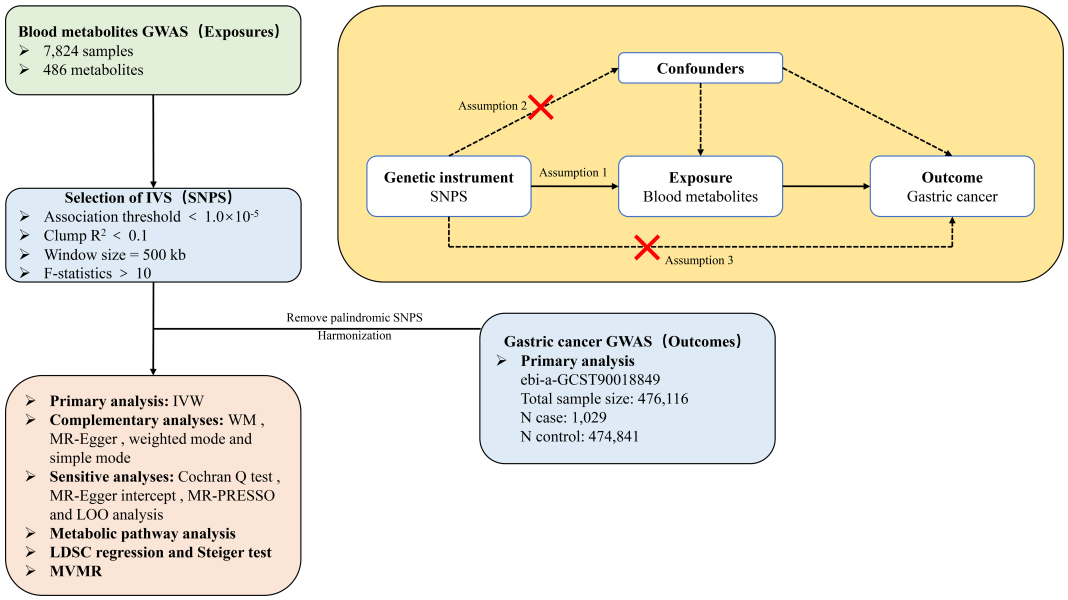



FIGURE 1
Overview of the research workflow.

HD4 high-resolution accurate mass analysis platform. The analysis focused on baseline serum metabolites, including tryptophan, nonadecanoate (19:0), and erythritol. All metabolite data were transformed using a logarithm and normalized on a batch basis. For further details regarding the specific metabolomic analysis methodology, please refer to reference (1, 26).

A comprehensive quality control and management system was implemented throughout the experiment to ensure accurate and consistent identification of the true chemical components and to eliminate any potential interference due to misattribution, background noise or system artifacts. The stability of the instrument's performance was evaluated by calculating the relative standard deviation (RSD) of the internal standards introduced to each sample prior to injection into the mass spectrometer.

# 3 Results

## 3.1 IVs for exposures

After an exhaustive and methodological selection process, 486 serum metabolites were selected for evaluation using MR analysis. The number of SNPs associated with these metabolites ranged from 3 to 503. Metabolites with the identifiers #00577, #32322, #33188, #34453, and #37459 were distinguished using the sparsest array of genetic tools, each of which was underpinned by only three correlated SNPs. Conversely, metabolite #33178 was located at the apex, with a substantial endowment of 496 SNPs, conferring genetic instrumentation. The F-statistics for all SNPs involved in the correlation analyses uniformly exceeded the threshold of 10, heralding the robust statistical power of the selected IVS and mitigating the risk of bias that could come from weak instruments. For a more detailed view of the IV data, please see Supplementary Table S2.

## 3.2 Primary analysis

Using IVW analysis, we identified 17 metabolites with potential relevance to GC etiology. Of these, 10 were characterized, whereas the remaining seven were not. These 17 metabolites had compelling associations with susceptibility to GC (Figure 2), and spanned a diverse spectrum of chemical classifications, such as amino acids, peptides, lipids, nucleotides, and xenobiotics, including tryptophan (OR = 0.523, 95% CI = 0.313–0.872, P = 0.013), tyrosine (OR = 2.489, 95% CI = 1.173–5.283, P = 0.018), C-glycosyltryptophan (OR = 0. 500, 95% CI = 0.263–0.951, P = 0.035), serine (OR = 1.622, 95% CI = 1.014–2.724, P = 0.044), gamma-glutamylmethionine (OR = 1.709, 95% CI = 1.093–2.671, P = 0. 019), X-13431-nonanoylcarnitine (OR = 0.784, 95% CI = 0.635–0.967, P = 0.023), nonadecanoate (19:0) (OR = 0.460, 95% CI = 0.225–0.943, P = 0.034), guanosine (OR = 0. 779, 95% CI = 0.607–0.999, P = 0.049), paraxanthine (OR = 0.778, 95% CI = 0.691–0.979, P = 0.032), and erythritol (OR = 0.672, 95% CI = 0.486 –0.930, P = 0.016) (Figure 3).

Within the framework of the IVW method, the concordance of the results derived from the MR-Egger method and weighted median estimations underscored the robustness of the associations between these metabolites and GC risk (Table 1).

## 3.3 Sensitivity analysis

To substantiate the robustness of our findings, we conducted a comprehensive series of sensitivity analyses, including Cochran's Q test, MR-Egger intercept test, MR-PRESSO, and LOO analysis. Cochran's Q test showed no significant heterogeneity, confirming the uniformity of the dataset. In addition, the MR-Egger intercept test showed no statistical evidence of horizontal pleiotropy (Table 1).

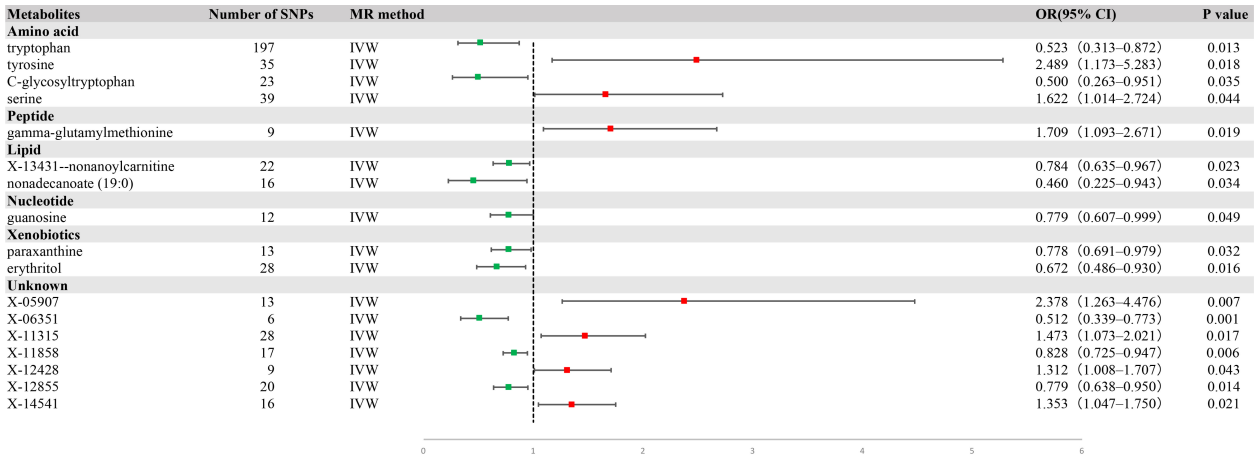



FIGURE 2
A forest plot of the causal effect of blood metabolites on gastric cancer (GC) risk from univariate Mendelian randomization with inverse variance weighting (IVW).

In the LOO analysis, the systematic exclusion and subsequent recalculation of MR estimates for each SNP in isolation confirmed the stability of our findings, indicating that no SNP introduced a consequential bias (Figure 4). The MR-PRESSO test, a sentinel test for outlier SNPs that potentially induce heterogeneity, did not reveal any significant differences (Supplementary Table S3).

## 3.4 Metabolic pathway analysis

Using insights from the 10 established metabolites, we uncovered a quintet of metabolic pathways potentially integral to GC (Table 2): aminoacyl-tRNA biosynthesis; phenylalanine, tyrosine, and tryptophan biosynthesis; ubiquinone and other terpenoid quinone biosynthesis; phenylalanine metabolism; and caffeine metabolism. These pathways may provide the foundation for the biological edifice within which GC emerges.

Notably, L-tyrosine was a recurrent participant in the first four enumerated pathways, L-tryptophan was critical to the aminoacyl-tRNA biosynthesis pathway, and 1,7-dimethylxanthine was exclusive to the caffeine metabolism pathway. These results suggest a potential direct involvement in malignant transformation processes that characterize gastric carcinogenesis and invite more exhaustive investigative efforts.

## 3.5 Evaluation of genetic correlation and directionality

Our results indicated a lack of statistically significant genetic correlation, underscoring the elusive nature of the genetic underpinnings that may link these metabolites to GC. Specifically, the regression coefficients (Rg) for tryptophan, tyrosine, guanosine, serine, nonadecanoate (19:0), and X-13431-nonanoylcarnitine were −0.0728, −0.0399, −0.3576, 0. 0674, −0.0276, and −0.1312, respectively, paired with the standard errors (Se) that underscore the imprecision of these estimates (0.0814, 0.1810, 0.2496, 0.2646, 0.1689, and 0.2297, respectively). None reached statistical significance, with p-values exceeding 0.05 (0.3709, 0.8255, 0.1518, 0.7988, 0.8701, and 0.5680, respectively). This finding suggests that the current cohort size was insufficient to detect a clear genetic association. The SNP heritability estimates for these metabolites ranged from 0.0725 (serine) to 0.9757 (tryptophan) (Supplementary Table S4).

We also applied the Steiger test to the cohort of 10 recognized metabolites to identify potential reverse causal vectors. Substantial results from the Steiger test rejected the hypothesis of an inverse effect, whereby GC perturbed the levels of these circulating metabolites. The available evidence (Supplementary Table S5) does not lend credence to such inverse dynamics.

## 3.6 MVMR analysis

To delineate the potential causal relationship between the selected metabolites and GC incidence, we performed MVMR analysis using the IVW method. Simultaneously, we screened the indicators of genetic instrument heterogeneity using the MR-PRESSO approach. Converging evidence from both the IVW and MR-PRESSO analyses suggested that the genetic proxies for tryptophan, nonadecanoate (19:0), and erythritol harbored direct and independent causal links to GC susceptibility, devoid of the confounding effects of other metabolites considered in our investigation (Figure 5, Supplementary Table S6).

## 3.7 The relative content of tryptophan, nonadecanoate (19:0) and erythritol in human blood samples

The relative content of tryptophan, nonadecanoate (19:0) and erythritol was determined by untargeted metabolomics analysis of blood samples collected from hospitals (Figure 6). The results

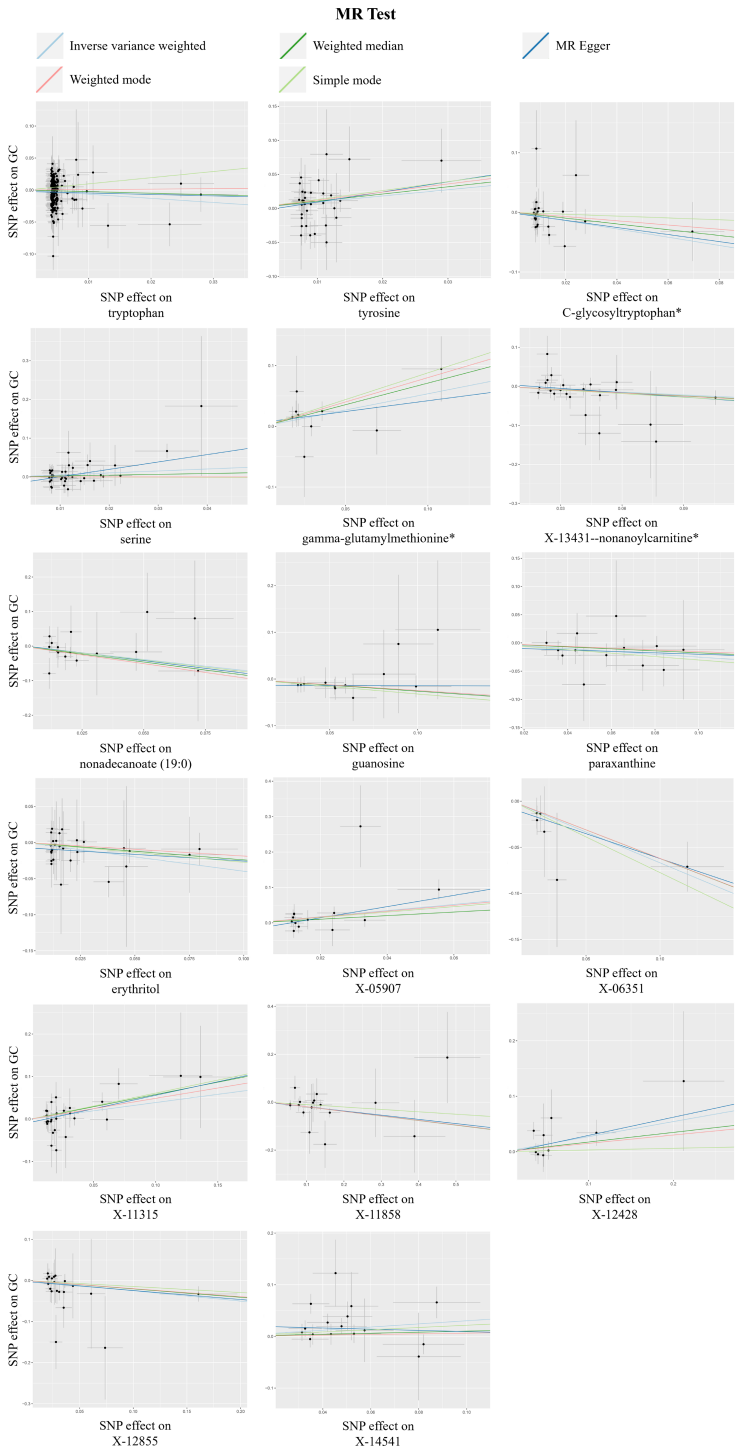



FIGURE 3
A scatterplot of the significant causal relationship (P < 0.05) between blood metabolites and gastric cancer (GC).

demonstrated that the levels of these three substances in the blood of GC patients were markedly diminished in comparison to those observed in healthy control groups ($p < 0.05$). This finding is consistent with the results of our MR analysis, which provides further confirmation of the role of tryptophan, nonadecanoate (19:0) and erythritol as protective factors against GC.

## 4 Discussion

Based on primary analyses utilizing Inverse Variance Weighting (IVW), weighted median approaches, and MR-Egger regression, along with sensitivity analysis, we identified 17 metabolites that are causally associated with gastric cancer (GC). Among these, 10 are

10.3389/fonc.2024.1418283

TABLE 1 Detection of causal relationships between 17 blood metabolites and GC risk using two MR models and tests for heterogeneity and horizontal pleiotropy.

| Metabolites | Number of SNPs | MR analysis | | | Heterogeneity | | Pleiotropy | |
|---|---|---|---|---|---|---|---|---|
| | | Method | OR (95% CI) | P-value | Q | P | Intercept | p |
| **Amino acid** | | | | | | | | |
| Tryptophan | 197 | ME | 0.798 (0.211– 3.020) | 0.740 | 211.685 | 0.196 | −0.00234672261246323 | 0.501 |
| | | WM | 0.781 (0.320– 1.907) | 0.587 | | | | |
| Tyrosine | 35 | ME | 4.419 (0.214–91.054) | 0.343 | 36.552 | 0.307 | −0.00562161200575098 | 0.703 |
| | | WM | 2.872 (0.934– 8.837) | 0.066 | | | | |
| C-glycosyltryptophan* | 23 | ME | 0.553 (0.165–1.851) | 0.347 | 15.287 | 0.808 | −0.00158207855094606 | 0.850 |
| | | WM | 0.616 (0.252–1.508) | 0.289 | | | | |
| Serine | 39 | ME | 6.728 (1.889–23.967) | 0.006 | 0.850 | 0.857 | −0.0187715047558943 | 0.025 |
| | | WM | 1.241 (0.560– 2.747) | 0.595 | | | | |
| **Peptide** | | | | | | | | |
| Gamma-glutamylmethionine* | 9 | ME | 1.426 (0.550–3.699) | 0.489 | 4.973 | 0.663 | 0.007 | 0.687 |
| | | WM | 2.037 (1.137–3.651) | 0.017 | | | | |
| **Lipid** | | | | | | | | |
| X-13431–nonanoylcarnitine* | 22 | ME | 0.689 (0.447–1.062) | 0.107 | 16.186 | 0.705 | 0.006 | 0.513 |
| | | WM | 0.762 (0.569–1.021) | 0.068 | | | | |
| nonadecanoate (19:0) | 16 | ME | 0.412 (0.060–2.847) | 0.384 | 10.669 | 0.712 | 0.002 | 0.905 |
| | | WM | 0.397 (0.142–1.108) | 0.078 | | | | |
| **Nucleotide** | | | | | | | | |
| Guanosine | 12 | ME | 0.989 (0.498–1.964) | 0.975 | 1.527 | 0.999 | −0.0132003086022429 | 0.481 |
| | | WM | 0.770 (0.562–1.057) | 0.106 | | | | |
| **Xenobiotics** | | | | | | | | |
| Paraxanthine | 13 | ME | 0.883 (0.440–1.768) | 0.731 | 4.723 | 0.944 | −0.007517064025338 | 0.714 |
| | | WM | 0.839 (0.610–1.154) | 0.281 | | | | |
| Erythritol | 28 | ME | 0.832 (0.491–1.409) | 0.500 | 11.851 | 0.992 | −0.00721226588820352 | 0.323 |
| | | WM | 0.786 (0.496–1.246) | 0.306 | | | | |
| **Unknown** | | | | | | | | |
| X-05907 | 13 | ME | 4.891 (1.487–16.090) | 0.024 | 12.980 | 0.295 | −0.0176623721801645 | 0.194 |
| | | WM | 1.662 (0.699– 3.951) | 0.250 | | | | |
| X-06351 | 6 | ME | 0.583 (0.334–1.017) | 0.130 | 0.943 | 0.918 | −0.00823932365242328 | 0.539 |
| | | WM | 0.537 (0.303–0.951) | 0.033 | | | | |
| X-11315 | 28 | ME | 1.886 (1.059–3.359) | 0.041 | 22.996 | 0.633 | −0.00754445825407327 | 0.325 |
| | | WM | 1.782 (1.119–2.839) | 0.015 | | | | |
| X-11858 | 17 | ME | 0.839 (0.607–1.160) | 0.305 | 13.384 | 0.573 | −0.00150060089220439 | 0.932 |
| | | WM | 0.826 (0.687–0.992) | 0.041 | | | | |
| X-12428 | 9 | ME | 1.388 (0.744–2.590) | 0.337 | 6.630 | 0.468 | −0.00338631534292095 | 0.850 |
| | | WM | 1.190 (0.824–1.719) | 0.353 | | | | |

*(Continued)*

TABLE 1 Continued

| Metabolites | Number of SNPs | MR analysis | | | Heterogeneity | | Pleiotropy | |
|---|---|---|---|---|---|---|---|---|
| | | Method | OR (95% CI) | P-value | Q | P | Intercept | p |
| Unknown | | | | | | | | |
| X-12855 | 20 | ME | 0.803 (0.606–1.063) | 0.143 | 13.474 | 0.763 | −0.00233762913162488 | 0.768 |
| | | WM | 0.816 (0.624–1.067) | 0.137 | | | | |
| X-14541 | 16 | ME | 0.889 (0.405–1.952) | 0.774 | 19.105 | 0.161 | 0.021 | 0.287 |
| | | WM | 1.108 (0.802–1.531) | 0.535 | | | | |

GC, gastric cancer; ME, MR-Egger.

well-documented, including tryptophan, tyrosine, C-glycosyltryptophan, serine, gamma-glutamylmethionine, X-13431-nonanoylcarnitine, nonadecanoate (19:0), guanosine, paraxanthine, and erythritol. Notably, After adjusting for relevant covariates using multivariate MR analysis, the associations of tryptophan, nonadecanoate (19:0), and erythritol with GC risk remained significant. Furthermore, the results of UPLC-MS/MS andvGC-MS/MS showed that compared with the healthy control group, the blood content of these three substances in GC patients was significantly reduced ($p < 0.05$). This finding is consistent with our MR Analysis and further confirms the role of tryptophan, palmitate (19:0) and erythritol as protective factors for GC. To our knowledge, this is the inaugural MR study to systematically investigate the prospective causal interactions between circulating metabolites and GC risk, highlighting the potential of these metabolites as biomarkers for both screening and therapeutic intervention.

The conclusions drawn from our rigorous MVMR analysis suggested that increased tryptophan, nonadecanoate (19:0), and erythritol levels were associated with a decreased risk of GC progression. This suggestion is supported by literature that implicates tryptophan metabolism as a potential antagonist in oncogenesis. Some tryptophan derivatives are implicated in orchestrating immune responses and limiting neoplastic proliferation. One example is the reduction of glutathione peroxidase 2, which triggers an increase in kynurenine, a tryptophan byproduct. This increase leads to the accumulation of reactive oxygen species via the tryptophan metabolic pathway, impeding the progression and metastatic potential of gastric malignancies (27). Nevertheless, the influence of tryptophan metabolism on carcinogenesis is not unambiguous, with bifurcated pathways toward either oncogenic or tumor-suppressive roles that depend on many elements, including distinct metabolic trajectories, neoplastic typology, tumor microenvironment, and host immune constitution (28).

Nonadecanoate (19:0), a long-chain fatty acid ester, was recently identified as the predominant constituent of essential oil derived from the fruits of Pistacia terebinthus, which exhibits promising antineoplastic properties against lung carcinoma cell lines (29). Unfortunately, the literature on the mechanistic insights related to this monomeric component is scarce.

Erythritol, a tetrahydric alcohol sugar synthesized endogenously from glucose via the pentose phosphate pathway in human cells, is available through dietary channels as a synthetic sweetener. Research has shown that erythritol may exert a critical influence on cerebral oncogenesis and the modulation of hydrogen peroxide, with its action depending on its concentration in biological systems (30).

We also identified a cadre of established metabolites that provide protection against gastric carcinoma. In particular, C-glycosyltryptophan, a metabolite within tryptophan metabolism, has historically been used as an index of renal function (31, 32), with empirical associations suggesting an increase in the infectious and inflammatory burden (33). Its role against GC is supported by correlations with cardiovascular and thyroid pathologies (34, 35). X-13431-nonanoylcarnitine represents a research gap, and its functional parity with recognized acylcarnitines is based on its structural cognate. The physiological and pathophysiological implications of acylcarnitines are manifold, such as influencing the sequelae of myocardial ischemia, glucose homeostasis, and inflammatory processes (36).

In this study, we found an association between elevated levels of tyrosine, serine, and gamma-glutamylmethionine in blood metabolites and an increased risk of GC progression. Tyrosine is involved in gluconeogenesis and ketogenesis, linking energy, lipid, and glucose metabolism. Disorders of tyrosine metabolism have been identified as biomarkers for hepatocellular carcinoma and gastroesophageal malignancies, and alterations in metabolism and related pathways play a key role in cancer development and progression (37, 38). Serine, which is essential for the rapid growth of tumor cells, contributes to the proliferation of colon cancer by providing single-carbon units. In addition, there is evidence that glycine supplementation may alter serine metabolism in tumor cells and that serine deprivation may inhibit tumor growth by affecting lipid metabolism pathways, particularly those involving palmitoyltransferases (39–41). Moreover, a negative correlation between plasma gamma-glutamylmethionine levels and the risk of lethal prostate cancer progression has been observed (42).

Our study has several strengths. First, it adopted an innovative approach by integrating metabolomics and genomics, which differs from previous MR analyses that focused only on single or conventional exposure factors. Moreover, by employing rigorous MR analysis and Steiger's analysis, we effectively addressed inherent limitations of traditional observational studies, such as reverse causality and confounding biases. Finally, we validated specific metabolite levels in the blood of patients with gastric cancer using

metabolomics techniques, further supporting the reliability of the results obtained by Mendelian analysis.

Nevertheless, our study has certain limitations. First, there were a limited number of available SNPs for the exposures at the genome-wide level. To counteract this, we made a deliberate decision to moderately adjust the p-value thresholds in our MR analysis. Nonetheless, it is crucial to emphasize that the F-statistic value for all SNPs we selected surpassed 10, which is a reassuring indicator of the strength and reliability of our instrumental variables. Second, our study did not account for potential confounding factors known to influence gastric cancer incidence, such as smoking and alcohol consumption. Nevertheless, the

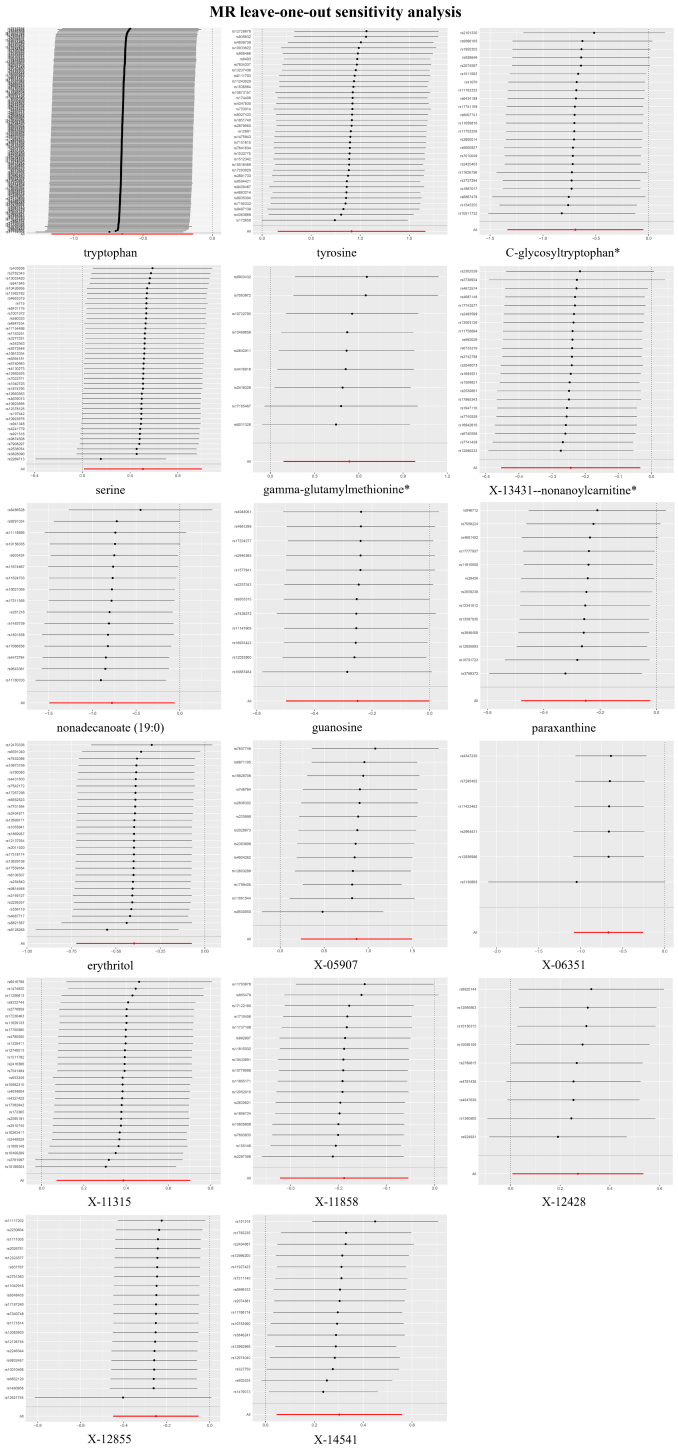



FIGURE 4
Leave-one-out plots for the causal association between blood metabolites and gastric cancer (GC).

TABLE 2 Significant metabolic pathways involved in the pathogenesis of gastric cancer (GC).

| Metabolic pathways | Involved metabolites | P-value |
|---|---|---|
| Aminoacyl-tRNA biosynthesis | L-Tryptophan/ L-Tyrosine | 0.005417 |
| Phenylalanine, tyrosine and tryptophan biosynthesis | L-Tyrosine | 0.010293 |
| Ubiquinone and other terpenoid-quinone biosynthesis | L-Tyrosine | 0.023046 |
| Phenylalanine metabolism | L-Tyrosine | 0.025582 |
| Caffeine metabolism | 1,7-Dimethylxanthine | 0.025582 |

intercept test from the MR-Egger method yielded P-values above 0.05, indicating that the SNPs associated with the metabolites we selected are not pleiotropic. In other words, these SNPs are unlikely to influence outcomes through pathways unrelated to the metabolites of interest. Despite this, we recognize the importance of considering these confounding factors in future research to gain a more comprehensive understanding of the causal links between blood metabolites and gastric cancer development. Finally,our analysis identified several metabolites as potential risk predictors for GC; however, these metabolites remained uncharacterized. Detailed studies on their molecular structures and functions may reveal novel biomarkers or therapeutic targets, thereby advancing the field of GC research.

In conclusion, our investigation sheds light on the potential causal associations between 10 known metabolites and GC through primary analysis. In addition, MVMR analysis andvmetabolomics techniques suggested that 3 metabolites affect the progression of GC. Our findings highlight the importance of GC in mediating the interplay between metabolites and GC, thereby opening new avenues for research on the etiology of GC, particularly its intersection with environmental factors.

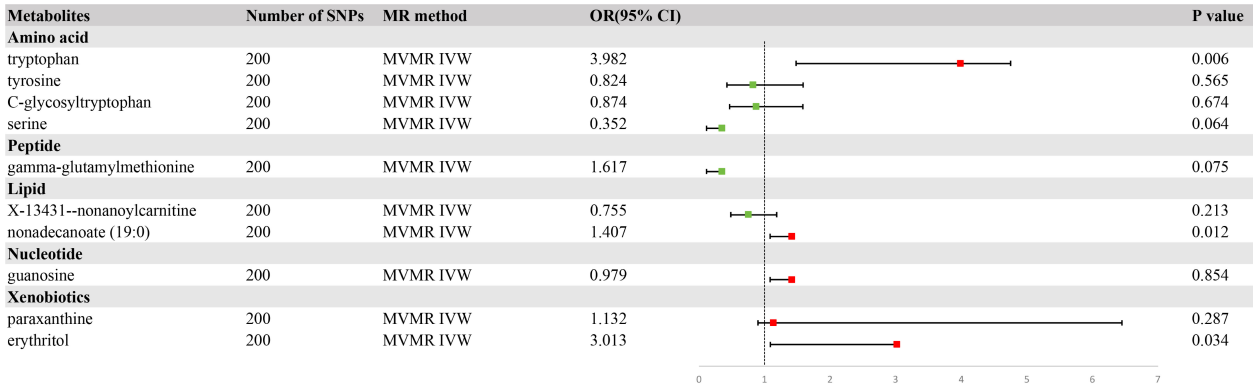



FIGURE 5
A forest plot of the causal effects of blood metabolites on gastric cancer (GC) risk from multivariate Mendelian randomization with inverse variance weighting (IVW).

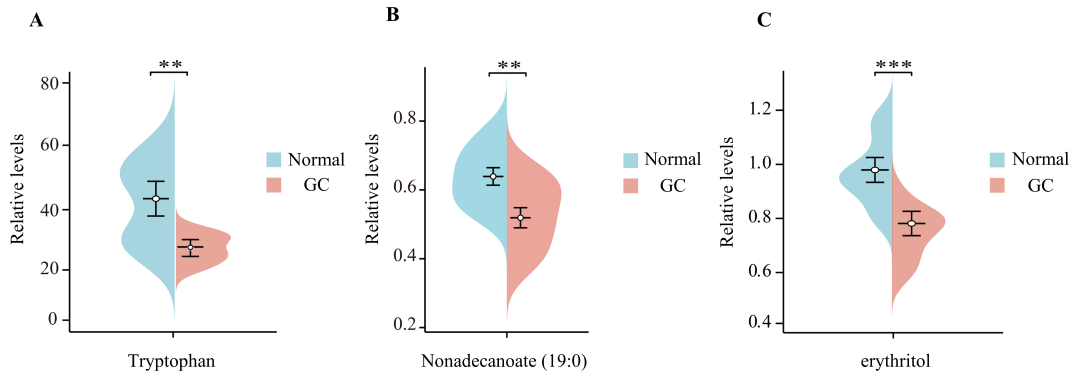



FIGURE 6
The beanplots of the relative amounts of tryptophan **(A)**, nonadecanoate (19:0) **(B)**, and erythritol **(C)** in blood samples from patients with gastric cancer (GC) compared to those from healthy controls '*' means $p < 0.05$; '**' means $p < 0.01$; '***' means $p < 0.001$.

## Data availability statement

The original contributions presented in the study are included in the article/Supplementary Material. Further inquiries can be directed to the corresponding author.

## Ethics statement

This study was performed in accordance with the ethical standards as laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards. However, ethics approval was not applicable as this study used publicly available genome-wide association study (GWAS) data. Ethical approval, consent to participate, and participant permission were available in the original GWAS research. The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

## Author contributions

LQ: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft. JS: Conceptualization, Investigation, Methodology, Writing – original draft. XZ: Conceptualization, Investigation, Methodology, Writing – original draft. YQ: Formal analysis, Methodology, Validation, Writing – original draft. ZT: Formal analysis, Investigation, Writing – original draft. SL: Formal analysis, Investigation, Writing – original draft. JZ: Conceptualization, Funding acquisition, Writing – review & editing. JL: Conceptualization, Funding acquisition, Supervision, Writing – review & editing.

## Funding

The author(s) declare that financial support was received for the research, authorship, and/or publication of this article. This study was supported by the National Natural Science Foundation of China (NSFC) (Grant No. 82172781) and the Key Research and Development Program in the Field of Social Development in Shaanxi Province (Grant No. S2023-YF-YBSF-1673).

## Acknowledgments

Genetic instruments for the studied exposures were obtained from the Metabolomics GWAS Server. Genetic association estimates for GC were obtained from published GWASs. We thank all participants and investigators for sharing their data.

## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note



## Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fonc.2024.1418283/full#supplementary-material

## References

1. Smyth EC, Nilsson M, Grabsch HI, van Grieken NC, Lordick F. Gastric cancer. *Lancet*. (2020) 396:635–48. doi: 10.1016/S0140-6736(20)31288-5

2. Suzuki H, Oda I, Abe S, Sekiguchi M, Mori G, Nonaka S, et al. High rate of 5-year survival among patients with early gastric cancer undergoing curative endoscopic submucosal dissection. *Gastric Cancer*. (2016) 19:198–205. doi: 10.1007/s10120-015-0469-0

3. Huang C, Liu H, Hu Y, Sun Y, Su X, Cao H, et al. Laparoscopic vs open distal gastrectomy for locally advanced gastric cancer: five-year outcomes from the CLASS-01 randomized clinical trial. *JAMA Surg*. (2022) 157:9–17. doi: 10.1001/jamasurg.2021.5104

4. Ajani JA, D'Amico TA, Bentrem DJ, Chao J, Cooke D, Corvera C, et al. Gastric cancer, version 2.2022, NCCN clinical practice guidelines in oncology. *J Natl Compr Canc Netw*. (2022) 20:167–92. doi: 10.6004/jnccn.2022.0008

5. Zeng P, Li H, Chen Y, Pei H, Zhang L. Serum CA199 levels are significantly increased in patients suffering from liver, lung, and other diseases. *Prog Mol Biol Transl Sci*. (2019) 162:253–64. doi: 10.1016/bs.pmbts.2018.12.010

6. Hao C, Zhang G, Zhang L. Serum CEA levels in 49 different types of cancer and noncancer diseases. *Prog Mol Biol Transl Sci*. (2019) 162:213–27. doi: 10.1016/bs.pmbts.2018.12.011

7. Rinschen MM, Ivanisevic J, Giera M, Siuzdak G. Identification of bioactive metabolites using activity metabolomics. *Nat Rev Mol Cell Biol*. (2019) 20:353–67. doi: 10.1038/s41580-019-0108-4

8. Ikeda A, Nishiumi S, Shinohara M, Yoshie T, Hatano N, Okuno T, et al. Serum metabolomics as a novel diagnostic approach for gastrointestinal cancer. *BioMed Chromatogr*. (2012) 26:548–58. doi: 10.1002/bmc.1671

9. The MR-Base platform supports systematic causal inference across the human phenome - PubMed. Available online at (Accessed April 16, 2024).

10. Integrative analysis of the plasma proteome and polygenic risk of cardiometabolic diseases - PubMed. Available online at (Accessed April 16, 2024).

11. Lee S, Park SK. Ethnic-specific associations between body mass index and gastric cancer: a Mendelian randomization study in European and Korean populations. *Gastric Cancer*. (2024) 27:19–27. doi: 10.1007/s10120-023-01439-5

12. Yang Z, Guo L, Sun Y, Huang Y, Li J, Lin Y, et al. Investigation of the causal relationship between Interleukin-6 signaling and gastrointestinal tract cancers: A Mendelian randomization study. *Dig Liver Dis*. (2024) 56:679–86. doi: 10.1016/j.dld.2023.08.040

13. Chen B, Diallo MT, Ma Y, Wang D. The association of vitamin D and digestive system cancers: a comprehensive Mendelian randomization study. *J Cancer Res Clin Oncol*. (2023) 149:13155–62. doi: 10.1007/s00432-023-05140-z

14. Yuan S, Chen J, Ruan X, Sun Y, Zhang K, Wang X, et al. Smoking, alcohol consumption, and 24 gastrointestinal diseases: Mendelian randomization analysis. *Elife*. (2023) 12:e84051. doi: 10.7554/eLife.84051

15. Yuan S, Mason AM, Titova OE, Vithayathil M, Kar S, Chen J, et al. Morning chronotype and digestive tract cancers: Mendelian randomization study. *Int J Cancer*. (2023) 152:697–704. doi: 10.1002/ijc.34284

16. IDDF2023-ABS-0105 Causal relationships between immunological proteins and common intestinal diseases: a bidirectional two-sample mendelian randomization study | Gut. Available online at: https://gut.bmj.com/content/72/Suppl_1/A198.1 (Accessed April 16, 2024).

17. An atlas of genetic influences on human blood metabolites - PubMed. Available online at (Accessed April 16, 2024).

18. Cai J, He L, Wang H, Rong X, Chen M, Shen Q, et al. Genetic liability for prescription opioid use and risk of cardiovascular diseases: a multivariable Mendelian randomization study. *Addiction*. (2022) 117:1382–91. doi: 10.1111/add.15767

19. Cai J, Li X, Wu S, Tian Y, Zhang Y, Wei Z, et al. Assessing the causal association between human blood metabolites and the risk of epilepsy. *J Transl Med*. (2022) 20:437. doi: 10.1186/s12967-022-03648-5

20. Gill D, Brewer CF, Monori G, Trégouët D-A, Franceschini N, Giambartolomei C, et al. Effects of genetically determined iron status on risk of venous thromboembolism and carotid atherosclerotic disease: A Mendelian randomization study. *J Am Heart Assoc*. (2019) 8:e012994. doi: 10.1161/JAHA.119.012994

21. Pierce BL, Burgess S. Efficient design for Mendelian randomization studies: subsample and 2-sample instrumental variable estimators. *Am J Epidemiol*. (2013) 178:1177–84. doi: 10.1093/aje/kwt084

22. Bowden J, Del Greco MF, Minelli C, Davey Smith G, Sheehan NA, Thompson JR. Assessing the suitability of summary data for two-sample Mendelian randomization analyses using MR-Egger regression: the role of the I2 statistic. *Int J Epidemiol*. (2016) 45:1961–74. doi: 10.1093/ije/dyw220

23. Burgess S, Thompson SG. Interpreting findings from Mendelian randomization using the MR-Egger method. *Eur J Epidemiol*. (2017) 32:377–89. doi: 10.1007/s10654-017-0255-x

24. O'Connor LJ, Price AL. Distinguishing genetic correlation from causation across 52 diseases and complex traits. *Nat Genet*. (2018) 50:1728–34. doi: 10.1038/s41588-018-0255-0

25. Orienting the causal relationship between imprecisely measured traits using GWAS summary data - PubMed. Available online at (Accessed April 16, 2024).

26. Lim JE, Huang J, Weinstein SJ, Parisi D, Männistö S, Albanes D. Serum metabolomic profile of hair dye use. *Sci Rep*. (2023) 13:3776. doi: 10.1038/s41598-023-30590-3

27. Xu H, Hu C, Wang Y, Shi Y, Yuan L, Xu J, et al. Glutathione peroxidase 2 knockdown suppresses gastric cancer progression and metastasis via regulation of kynurenine metabolism. *Oncogene*. (2023) 42:1994–2006. doi: 10.1038/s41388-023-02708-4

28. Yang Q, Hao J, Chi M, Wang Y, Xin B, Huang J, et al. Superior antitumor immunotherapy efficacy of kynureninase modified CAR-T cells through targeting kynurenine metabolism. *Oncoimmunology*. (2022) 11:2055703. doi: 10.1080/2162402X.2022.2055703

29. Chemical characterization and evaluation of anticancer activity of Pistacia terebinthus linn (Accessed April 16, 2024).

30. Abstract LB183: Erythritol acts as tumor enhancer and suppressor depending on concentrations in brain tumor cell lines | Cancer Research | American Association for Cancer Research. Available online at: https://aacrjournals.org/cancerres/article/81/13_Supplement/LB183/670147/Abstract-LB183-Erythritol-acts-as-tumor-enhancer (Accessed April 16, 2024).

31. Hocher B, Adamski J. Metabolomics for clinical use and research in chronic kidney disease. *Nat Rev Nephrol*. (2017) 13:269–84. doi: 10.1038/nrneph.2017.30

32. Takahira R, Yonemura K, Yonekawa O, Iwahara K, Kanno T, Fujise Y, et al. Tryptophan glycoconjugate as a novel marker of renal function. *Am J Med*. (2001) 110:192–7. doi: 10.1016/s0002-9343(00)00693-8

33. Lustgarten MS, Fielding RA. Metabolites associated with circulating interleukin-6 in older adults. *J Gerontol A Biol Sci Med Sci*. (2017) 72:1277–83. doi: 10.1093/gerona/glw039

34. Huang J, Weinstein SJ, Moore SC, Derkach A, Hua X, Liao LM, et al. Serum metabolomic profiling of all-cause mortality: A prospective analysis in the alpha-tocopherol, beta-carotene cancer prevention (ATBC) study cohort. *Am J Epidemiol*. (2018) 187:1721–32. doi: 10.1093/aje/kwy017

35. Zhang X, Zhou J, Xie Z, Li X, Hu J, He H, et al. Exploring blood metabolites and thyroid disorders: a bidirectional mendelian randomization study. *Front Endocrinol (Lausanne)*. (2023) 14:1270336. doi: 10.3389/fendo.2023.1270336

36. McCoin CS, Knotts TA, Adams SH. Acylcarnitines–old actors auditioning for new roles in metabolic physiology. *Nat Rev Endocrinol*. (2015) 11:617–25. doi: 10.1038/nrendo.2015.129

37. Wang J, Qiao Y, Sun H, Chang H, Zhao H, Zhang S, et al. Decreased SLC27A5 suppresses lipid synthesis and tyrosine metabolism to activate the cell cycle in hepatocellular carcinoma. *Biomedicines*. (2022) 10:234. doi: 10.3390/biomedicines10020234

38. Wiggins T, Kumar S, Markar SR, Antonowicz S, Hanna GB. Tyrosine, phenylalanine, and tryptophan in gastroesophageal Malignancy: a systematic review. *Cancer Epidemiol Biomarkers Prev*. (2015) 24:32–8. doi: 10.1158/1055-9965.EPI-14-0980

39. Fransen M, Lismont C, Walton P. The peroxisome-mitochondria connection: how and why? *Int J Mol Sci*. (2017) 18:1126. doi: 10.3390/ijms18061126

40. Cairns RA, Harris IS, Mak TW. Regulation of cancer cell metabolism. *Nat Rev Cancer*. (2011) 11:85–95. doi: 10.1038/nrc2981

41. Yang M, Vousden KH. Serine and one-carbon metabolism in cancer. *Nat Rev Cancer*. (2016) 16:650–62. doi: 10.1038/nrc.2016.81

42. Wang Y, Jacobs EJ, Carter BD, Gapstur SM, Stevens VL. Plasma metabolomic profiles and risk of advanced and fatal prostate cancer. *Eur Urol Oncol*. (2021) 4:56–65. doi: 10.1016/j.euo.2019.07.005

# Frontiers in Genetics

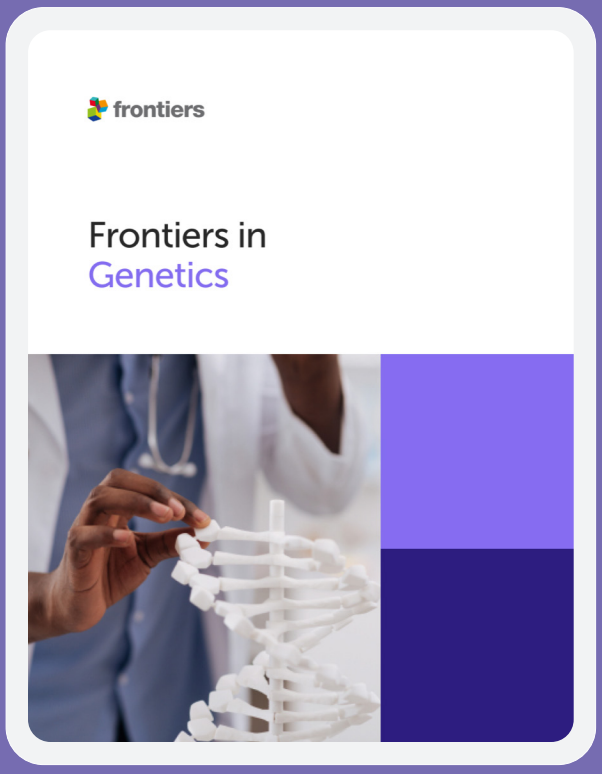
**Highlights genetic and genomic inquiry relating to all domains of life**

The most cited genetics and heredity journal, which advances our understanding of genes from humans to plants and other model organisms. It highlights developments in the function and variability of the genome, and the use of genomic tools.

---

## Discover the latest Research Topics

See more →




**Frontiers**

Avenue du Tribunal-Fédéral 34
1005 Lausanne, Switzerland
frontiersin.org

**Contact us**

+41 (0)21 510 17 00
frontiersin.org/about/contact