Highlights in BioScience
ISSN:2682-4043
DOI:10.36462/H.BioSci.202202

Research Article

Open Access

Identification of hub genes and potential molecular mechanisms associated with inflammatory bowel diseases using meta-analysis of gene expression data

Khaled H. Mousa[*1,2], Ahmed E. Nassar[1,2]

[1] International Center for Agricultural Research in the Dry Areas (ICARDA), Cairo, Egypt.

[2] Genetics Department, Faculty of Agriculture, Mansoura University, Mansoura, Egypt.

*** To whom correspondence should be addressed: khaledhelmy444@gmail.com**

Editor: Hatem Zayed, *College of Health and Sciences, Qatar University, Doha, Qatar.*

Reviewer(s):

Santosh K Maurya, *Molecular Signaling & Drug Discovery Laboratory, Department of Biochemistry, Central University of Punjab, Bathinda, Punjab, India.*

Akhilesh Maurya, *Indian Institute of Information Technology Allahabad, Devghat, Jhalwa, Prayagraj, Uttar Pradesh 211015, India.*

Received: May 1, 2022

Accepted: July 1, 2022

Published: July 15, 2022

Citation: Mousa KH, Nassar AE . Identification of hub genes and potential molecular mechanisms associated with inflammatory bowel diseases using meta-analysis of gene expression data. 2022 July 15;5:bs202202

Data Availability Statement: All relevant data are within the paper and supplementary materials.

Funding: The authors have no support or funding to report.

Competing interests: The authors declare that they have no competing interests.

Abstract

Inflammatory bowel diseases (IBDs), which primarily include Crohn's disease (CD) and ulcerative colitis (UC), are chronic recurrent diseases of the gastrointestinal tract with increasing prevalence and incidence worldwide. In this study, we aimed to identify key factor genes that control the progression of inflammatory bowel disease, identify common and unique nodal genes, examine gene-protein interactions, assess current advances in the published literature on inflammatory bowel disease, and examine the impact of various biological pathways. Gene expression profiles were obtained from the Gene Expression Omnibus (GEO) database. We performed gene expression analysis to identify deferentially expressed genes. Subsequently, GO and KEGG pathway enrichment analyzes and protein-protein interaction network analyzes (PPI) of DEGs were performed. Text mining was used to examine the frequency of genes in the published IBD literature. Four GEO databases (GSE156044, GSE159751, GSE159008, and GSE102746) were downloaded from GEO databases. A total of 368 DEGs were identified. The results of GO term analysis showed that DEGs were mainly involved in the activity of cytokine receptors, integral components of the plasma membrane, and cytokine-mediated signaling. KEGG pathway analysis showed that DEGs were mainly enriched in bile secretion, mineral absorption, and cytokine-cytokine receptor interaction. The results of PPI analysis showed that about 10 genes were the key genes for the occurrence of CED. Text mining revealed the existence of 399 genes associated with CED. Our results suggest a possible link between CED and other diseases such as triple negative breast cancer (TNBC) and lung adenocarcinoma (LUAD), and provide new insights into the mechanisms of inflammatory bowel disease and new treatment targets.

Keywords: Inflammatory bowel diseases, Deferentially expressed genes, Hub genes, Text mining

Introduction

Inflammatory bowel disease (IBD), which includes Crohn's disease (CD) and ulcerative colitis (UC), is a chronic inflammatory disease of the gastrointestinal tract that is becoming increasingly common worldwide [1]. Inflammatory bowel diseases (IBDs) are the most common gastrointestinal (GI) diseases, affecting 0.3% to 0.5% of the world population [2]. Ulcerative colitis (CD) is confined to the mucosa of the colon, always involving the rectum, and may continuously extend to more proximal parts of the intestine [3]. Crohn's disease (CD), on the other hand, is a transmural, progressive inflammatory disease that can affect any part of the gastro-intestinal tract (GI). As a result, complications such as strictures, fistulas, and abscesses are more common in CD [4]. Patients with these conditions have symptoms such as bloating, abdominal pain, and altered bowel habits (diarrhea and/or constipation), and approximately 40% of IBD patients have an irritable bowel

Irritable Bowel Syndrome (IBS)-like symptoms [5]. IBD affects both children and adults, with 15-20% of patients diagnosed in childhood [6]. Inflammatory bowel disease is caused by a variety of agents, including genetic, immune system, microbial, and environmental factors [7].

Malnutrition is a common problem inIBD patients [8]. Malnutrition inIBD patients is multi-factorial and has been associated with malabsorption, caloric restriction, pharmacological treatment, loss of nutrients in the gastrointestinal tract, and increased energy consumption [9]. Inflammatory bowel disease (IBD) is a lifelong condition that has no cure. Patients suffering from IBD may present with symptoms of common mental disorders such as anxiety and depression [10]. The prevalence of CD and UC can be attributed to a variety of factors, including geographic location, poor nutrition, genetics, and an ineffective immune response [11]. Studies of familial clustering and twin pairs have provided evidence that genetic factors contribute to the development of IBD. Studies of discordant twin pairs with IBD have pointed to some genetically independent factors underlying pathogenesis [12]. Genetic research on inflammatory bowel disease (IBD) has identified genes and pathways in inflammatory pathology. [13]. Most available IBD therapies suppress the immune system, which increases the risk of infections and some cancers and does not benefit all patients [14]. To reduce the prevalence of IBD, the IBD community should focus on prevention of the disease [15]. Differential gene expression (DGE) analysis is one of the most common applications of RNA sequencing (RNA-seq) data. This method can be used to elucidate differentially expressed genes under two or more conditions, and it is used in many applications for the analysis of RNA-seq data. [16]. Using differential gene expression, nodal genes have been identified in Alzheimer's disease (AD) [17], gastric cancer (GC) [18], breast cancer (BC) [19], tuberculosis (TB) [20], lung adenocarcinoma [21], colon cancer (CRC) [22]. In this study, we used bioinformatics approaches to identify differentially expressed genes (DEGs) between normal and inflamed tissues. These DEGs were also analyzed, including GO and KEGG enrichment analysis of DEGs, construction of protein-protein interaction networks (PPI), identification of key genes associated with CED, and identification of genes common to Crohn's disease (CD) and ulcerative colitis (UC). The biological functions and key signaling pathways of these DEGs are discussed. The major genes discovered in previous work were reviewed.

Materials and Methods

Gene expression data

We downloaded the original gene expression profiles GSE156044, GSE159751, GSE159008 and GSE102746 from the National Center for Biotechnology Information (NCBI) Gene Expression Omnibus (GEO) (https://www.ncbi.nlm.nih.gov/geo/) [23]. A total of 40 normal tissues (control) and 31 inflamed tissues (patients) were downloaded. We excluded 27 inflamed tissues in GSE159008. We selected samples of normal tissues and inflamed tissues of different types of inflammatory bowel disease only.

Identification of DEGs

Raw data were subjected to differentially gene expression analysis using the R package limma. Gene expression was analyzed with the package limma in R and displayed as a heat map. Genes with P.Val 0.05 were considered as deferentially expressed.

Gene enrichment analysis

Gene set enrichment (GSE) is the most effective method for determining the underlying biological functions of various genes or proteins [24]. To determine the functions of overlapping DEGs, GO functional and KEGG pathway enrichment of proteins encoded by candidate genes were analyzed and these genes were annotated using the Enrichr online analysis database (available online: https://maayanlab.cloud/Enrichr/). Enrichr is a gene set search engine that can be used to query hundreds of thousands of annotated gene sets. This platform provides several methods for calculating gene set enrichment, and the results are visualized in several interactive ways. [25].

PPI network construction and analysis

Protein-protein interactions was assessed using the database STRING (Search Tool for the Retrieval of Interacting Genes; http://string-db.org/cgi/input.pl). Using the STRING database and the Cytoscape programme, a PPI network of DEGs was produced (version 3.9.0). Additionally, the Cytoscape program is built on a network. Each node represents a gene, protein, or other biological molecule, and the connections between nodes represent how these biological molecules interact. This model can be used to determine how proteins encoded by DEGs interact with one another and with other proteins in pathways in inflammatory bowel diseases (IBDs).

Text-mining

The importance of text mining in service management is growing with the expansion of access to Big Data via digital platforms that enable such services [26]. Text mining tools are commonly used to extract information about disease-related genes, proteins, molecular interactions, and signaling pathways [27]. Previous research has documented the use of these tools in studying regulatory mechanisms for various types of diseases, including inflammatory bowel disease [28]. In the current study, we obtained IBD-related abstracts from the PubMed database by searching for 'inflammatory bowel disease' using the esearch tool and downloading these data using the efetch tool, which is responsible for retrieving records in the desired format [29]. This tool is provided by Entrez Direct (EDirect) software [30]. We counted genes associated with IBDs from the published literature using Python, and these genes were illustrated using the R package. These genes had been studied extensively in the context of IBD.

Results

Identification of DEGs

A total of 93, 361, 907, and 1123 DEGs were identified from GSE156044, GSE159008, GSE102746, and GSE159751, respectively, according to differential gene expression analysis.

The heat map of DEGs expression is shown in **Figure 1**. Among the 2484 genes with adj.P.Val ≤ 0.04, 368 DEGs were identified, including 180 upregulated genes and 188 downregulated genes.

Biological annotation of DEGs in inflammatory bowel disease was performed using the Enrichr online database. Bars of the graph sorted by p-value ranking. **Figure 3** shows the ten most enriched pathways from GO and KEGG analyzes for enrichment of DEGs. GO Analysis of DEGs was divided into three functional groups, including molecular function (MF), cellular component (CC), and biological process (BP). At MF, these DEGs were significantly enriched in cytokine receptor activity, metal ion binding, CXCR chemokine receptor binding, icosatetraenoic acid binding, and arachidonic acid binding (**Figure 3c**). At CC, DEGs were enriched in the integral component of the plasma membrane, the lumen of secretory granules, the lumen of intercellular organelles, the RNA polymerase III complex, and the alpha-beta T-cell receptor (**Figure 3b**). At BP, DEGs were enriched in cytokine-mediated signaling, positive regulation of telomerase RNA localization in the Cajal body, neutrophil degranulation, neutrophil activation in immune response, and regulation of telomerase RNA localization in the Cajal body (**Figure 3a**). In KEGG signaling pathways, DEGs were enriched mainly in bile secretion, mineral absorption, cytokine-cytokine receptor interaction, alanine, aspartate, and glutamate metabolism, and JAK-STAT signaling (**Figure 3d**). These significantly enriched GO terms and KEGG signaling pathways will help us better understand the key molecules involved in the progression of IBD.

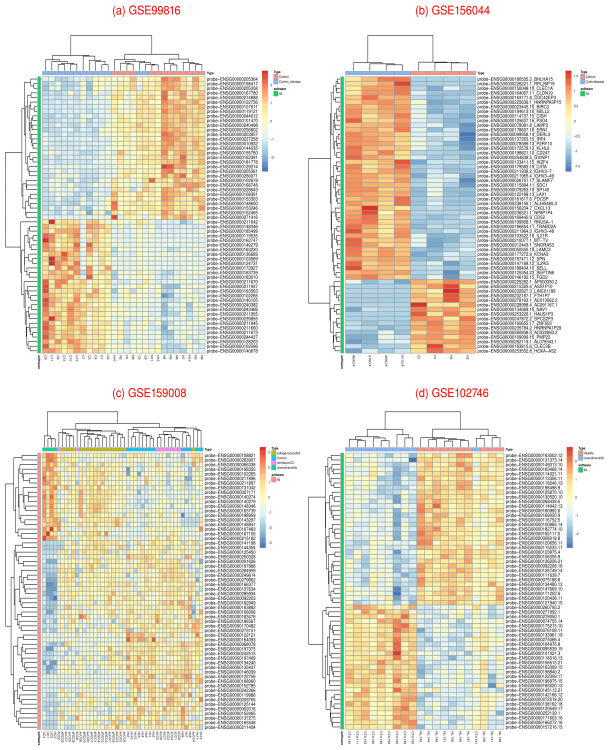

Figure 1. Heat maps of gene expression data. The gene expression profiles of the studied IBDs, (a) GSE99816(Crohn's disease), (b) GSE156044(Crohn's disease), (c) GSE159008(ulcerative colitis), (d) GSE102746 (ulcerative colitis) you can add references of every experiment.

PPI network construction and analysis

Genes associated with inflammatory bowel disease were subjected to PPI analysis. The PPI network grouped genes based on their interaction activity to discover genes associated with the studied diseases, as shown in **Figure 4**. As a result, we obtained 42 genes with significant interactions, 10 of which were associated with IBD. These genes were POLR2H, LCN2, TIMP1, CXCL1, MUC1, CSF3R, S100A9, SPI1, CFTR, and FCGR3B.

Text mining

We examined 42192 abstracts and obtained the 399 most frequent genes in the published literature associated with CED (**Figure 5**). Using the online analysis tool ShinyGO. Gene Ontology Enrichment analysis was performed according to biological processes (BP), cellular components (CC), and molecular functions (MF). In addition, the cutoff value for screening pathways and significant functionality was set at P 0.05. Table 1 shows the ten most enriched pathways of the Go enrichment analysis. KEGG pathway enrichment analysis was also performed using g:Profiler. A Venn diagram was used to classify the similarities and differences between the DEGs of the four datasets and the IBD-related genes identified by text mining (two genes).

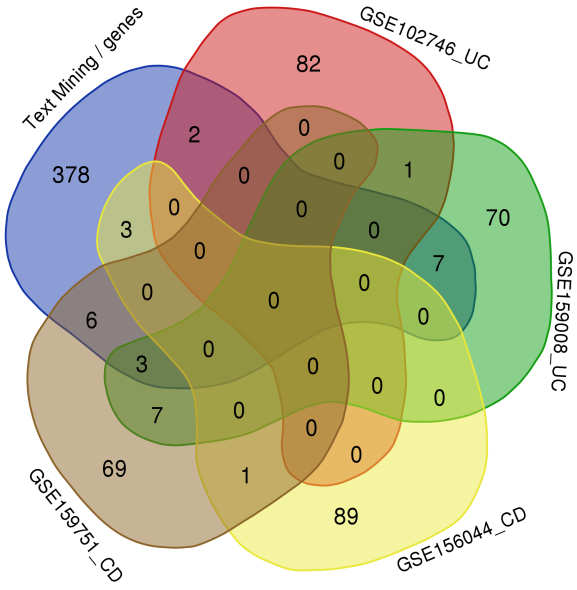

Figure 2. Venn digram, Similarities and differences between four datasets and 399 most frequent genes in previous publications related to IBD, identified by text mining.

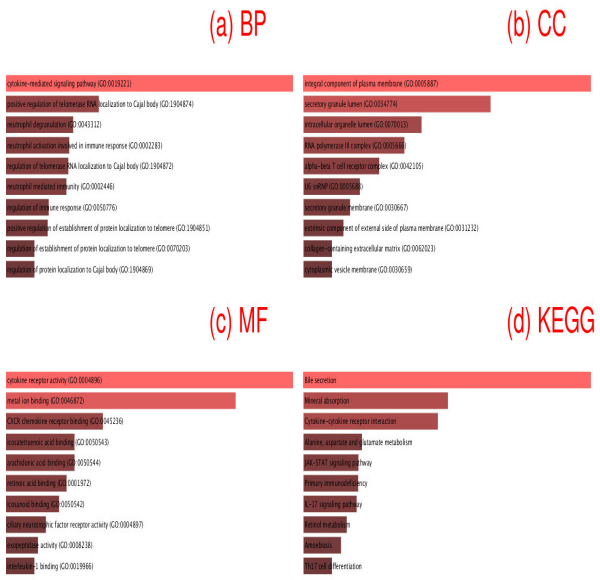

Figure 3. GO molecular function and KEGG pathways analysis of DEGs. Red bars represent the number of DEGs. Here only show the top 10 pathways: (a) Biological processes (BP); (b) Cellular components (CC); (c) Molecular function (MF); (d) KEGG pathway.

Discussion

Inflammatory bowel disease (IBD) is a chronic inflammatory disease of the gastrointestinal tract [31]. Ulcerative colitis and Crohn's disease are the main forms of inflammatory bowel disease [32]. Inflammatory bowel disease is a long-term chronic condition, with approximately 20% of IBD patients progressing to colorectal cancer [33]. Intestinal fibrosis is a common complication of inflammatory bowel disease, which is usually the result of chronic inflammation [34]. Genetic studies have identified more than 200 loci that regulate IBD risk [35]. In patients with IBD, alterations in various immune cells and neuroimmune signaling pathways have been identified in the lamina propria [36]. Therefore, it is important to investigate the molecular mechanisms of inflammation and inflammatory bowel disease development. Bioinformatics tools provide extensive technical support for the processing and analysis of gene expression profiles. As a result, bioinformatics analysis has been widely used in recent years to identify novel therapeutic targets and diagnostic markers [37]. In this study, four GEO datasets related to IBDs, GSE156044, GSE159008, GSE102746 and GSE159751, were examined using the Limma package in R statistical software and 368 DEGs were identified. The results of GO analysis, which included MF, CC, and BP, showed that these DEGs were mainly enriched in cytokine receptor activity, extracellular response, and cellular response to chemical stimuli. KEGG pathway enrichment analysis also showed that these DEGs were significantly enriched in bile secretion, cytokine-cytokine receptor interaction, mineral absorption, retinol metabolism, and JAK-STAT pathways. This gene set enrichment analysis provides insights into the molecular mechanism of IBD progression. We used Cytoscape to create a PPI network according to DEGs and identified 42 genes with significant interaction, 10 of which are associated with IBD. They are POLR2H, LCN2, TIMP1, CXCL1, MUC1, CSF3R, S100A9, SPI1, CFTR, and FCGR3B. POLR2H is overexpressed in organoids of rectal tumors. POLR2H is the gene encoding the common H subunit of RNA polymerases I, II and III [38]. LCN2 is secreted into the intestinal lumen at high levels and has a significant impact on controlling both the composition of the gut microbiome and host inflammation [39]. Short-term consumption of a high-fat diet (HFD) significantly increases intestinal Lcn2 expression and secretion into the intestinal lumen. Lcn2 deficiency accelerates the development of HFD-induced intestinal inflammation [40]. TIMP1 plays an important role in promoting tumorigenesis [41] and acts as a prognostic biomarker for ulcerative colitis-related colorectal cancer [42]. CXCL1 enhances myeloid cell-mediated immunosuppression during acute colitis and has been linked to tumorigenesis in various tumor types [43]. MUC1 is a heterodimeric protein that enhances the response to inflammation. Expression of MUC1 is associated with activation of inflammatory pathways and development of colitis [44]. CSF3R is an important regulator of proliferation and differentiation of myeloid cells. Mutations in the CSF3R gene have been found in patients with chronic neutrophilic leukemia (CNL) and acute myeloid leukemia (AML) [45]. Recent research has shown that mutations in the CSF3R gene play an oncogenic role in the development of hematologic malignancies [46]. S100A9 is a myeloid-related protein involved in inflammatory processes and various carcinogens [47]. SPI1 is a gene cluster containing 39 genes encoding T3SS-1. SPI1 genes are responsible for host cell invasion and host inflammatory response. T3SS-1 encoded by SPI1 provides effector proteins required for intestinal invasion and the development of enteritis [48]. CFTR is a protein located at the apical membrane of epithelial cells and has been associated with cystic fibrosis disease. Cystic fibrosis is a potentially fatal disease that causes severe damage to the lungs and digestive system [49]. FcY receptors (FcYRs), are expressed by six genes, one of which is FCGR3B [50]. FCGR3B increases the risk of inflammatory diseases such as systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA) [51].

Text mining is concerned with obtaining information about biological entities such as genes, proteins, and phenotypes, or more broadly, biological pathways [52]. Text mining was used to extract genes associated with CED from the PubMed database. We obtained the 399 most frequent genes in the published literature associated with CED. Comparison between the DEGs of the four datasets and the text mining results revealed 21 overlapping genes that had been intensively studied. By examining the PPI network of these 21 genes, we discovered four genes that interact highly with each other (PTPRC, LCN2, CXCL1, and S100A9). These genes suggest that there may be a potential link between them and other diseases. In our study, the results of text mining

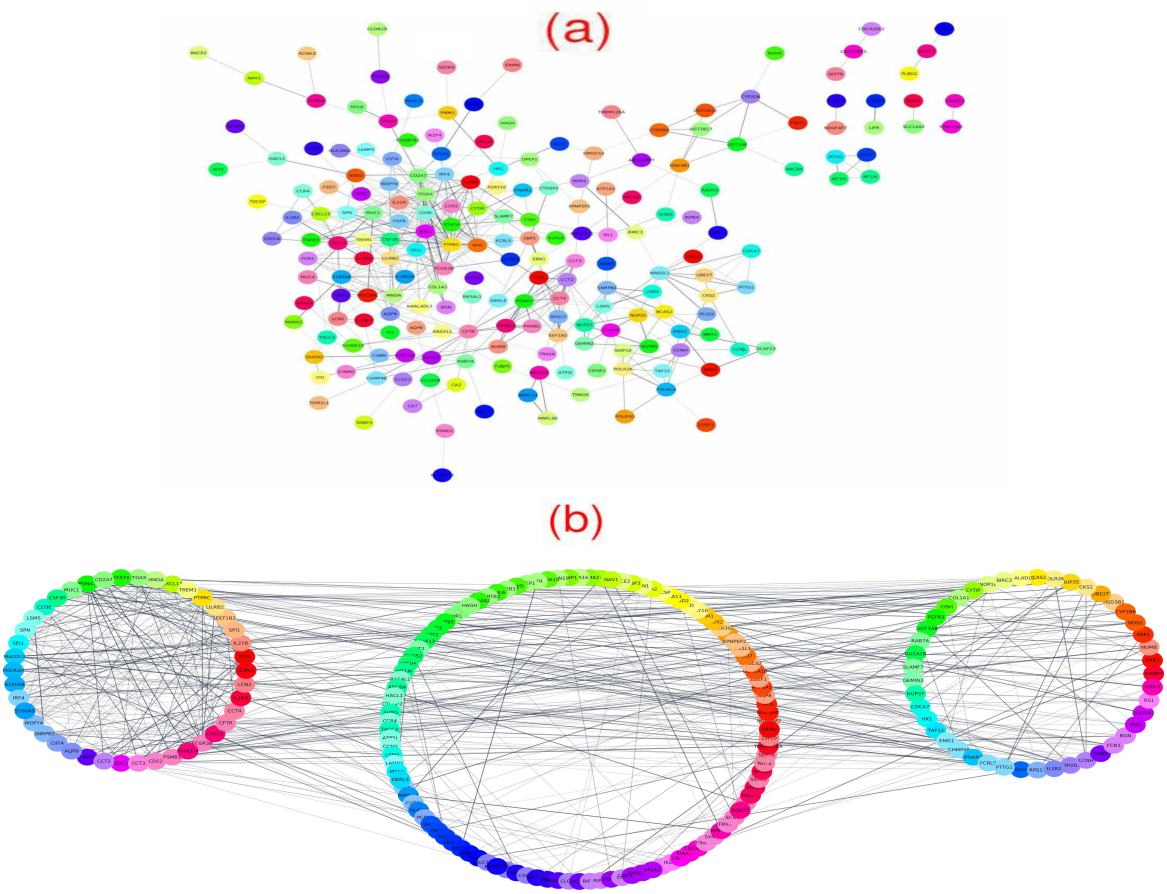

Figure 4. PPI network of significant genes associated with IBD diseases. (a) Protein-protein interaction network of DEGs. (b) Clustering of genes according to their interaction activity.

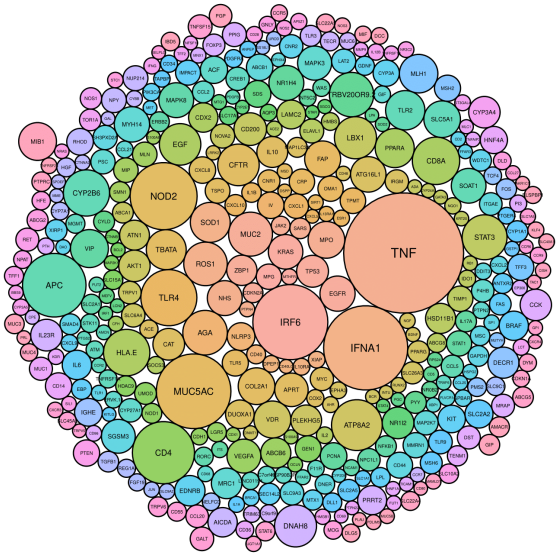

Figure 5. Text Mining of genes associated with IBDs in published literature. 399 most frequent genes in previous publications related to IBD, identified by text mining.

partially agreed with the results of DEG analysis, but they also provided some unique conclusions. PTPRC was associated with triple negative breast cancer (TNBC) [53]. PTPRC may be a potential prognostic marker in lung adenocarcinoma (LUAD), which could affect the immune function of T cells and other immune cells by participating in the regulation of tumor microenvironment (TME) immune activity [54]. LCN2 is an acute-phase response protein that is upregulated in CNS disease or injury and serves as a regulatory enhancer in neuroinflammation [55]. During lymphangiogenesis, CXCL1 enables tumor cells to migrate into lymphatic vessels, leading to lymph node metastasis [56]. A growing body of research suggests that S100A9, also known as myeloid-related proteins 8 and 14, plays an important role in inflammation and inflammation-related tissue damage [57]. S100A9 has been reported to be regulated by pathogen infection and to play a role in the progression of various cancers, including hepatitis B virus (HBV)-related hepatocellular carcinoma (HCC) [58]. In summary, the aim of this study was to describe IBD-related gene expression and systematically investigate the pathways and networks of IBD-related genes using bioinformatics to reveal beneficial targets and pathways in the pathogenesis of IBD.

References

1. John Gubatan, Steven Levitte, Akshar Patel, Tatiana Balabanis, Mike T Wei, and Sidhartha R Sinha. Artificial intelligence applications in inflammatory bowel disease: emerging technologies and future directions. World journal of gastroenterology, 27(17):1920, 2021.
2. Arnau Vich Vila, Floris Imhann, Valerie Collij, Soesma A Jankipersadsing, Thomas Gurry, Zlatan Mujagic, Alexander Kurilshikov, Marc Jan Bonder, Xiaofang Jiang, Ettje F Tigchelaar, et al. Gut microbiota composition and functional changes in inflammatory bowel disease and irritable bowel syndrome. Science translational medicine, 10(472):eaap8914, 2018.
3. Gabrielle Wark, Dorit Samocha-Bonet, Simon Ghaly, and Mark Danta. The role of diet in the pathogenesis and management of inflammatory bowel disease: a review. Nutrients, 13(1):135, 2021.
4. Amritpal Dhaliwal, Jonathan I Quinlan, Kellie Overthrow, Carolyn Greig, Janet M Lord, Matthew J Armstrong, and Sheldon C Cooper. Sarcopenia in inflammatory bowel disease: A narrativeoverview. Nutrients, 13(2):656, 2021.
5. Natalia Pedersen, Dorit Vedel Ankersen, Maria Felding, Henrik Wachmann, Zsuzsanna V´egh, Line Molzen, Johan Burisch, Jens Rikardt Andersen, and Pia Munkholm. Low-fodmap diet reduces irritable bowel symptoms in patients with inflammatory bowel disease. World journal of gastroenterology, 23(18):3356, 2017.
6. Alastair Forbes, Johanna Escher, Xavier H´ebuterne, Stanis-law Klek, Zeljko Krznaric, St´ephane Schneider, Raanan Shamir, Kalina Stardelova, Nicolette Wierdsma, Anthony E Wiskin, et al. Espen guideline: Clinical nutrition in inflammatory bowel disease. Clinical Nutrition, 36(2):321– 347, 2017.
7. Jordan E Axelrad, Ken H Cadwell, Jean-Frederic Colombel, and Shailja C Shah. Systematic review: gastrointestinal infection and incident inflammatory bowel disease. Alimentary pharmacology & therapeutics, 51(12):1222–1232, 2020.
8. Jing Liu, Xiaolong Ge, Chunhui Ouyang, Dongxu Wang, Xiaoqi Zhang, Jie Liang, Weiming Zhu, Qian Cao, et al. Prevalence of malnutrition, its risk factors, and the use of nutrition support in patients with inflammatory bowel disease. Inflammatory bowel diseases, 2022.
9. Alicja Ewa Ratajczak, Anna Maria Rychter, Agnieszka Zawada, Agnieszka Dobrowolska, and Iwona Krela-Ka´zmierczak. Do only calcium and vitamin d matter? micronutrients in the diet of inflammatory bowel diseases patients and the risk of osteoporosis. Nutrients, 13(2):525, 2021.
10. Brigida Barberio, Mohammad Zamani, Christopher J Black, Edoardo V Savarino, and Alexan- der C Ford. Prevalence of symptoms of anxiety and depression in patients with inflammatory bowel disease: a systematic review and meta-analysis. The Lancet Gastroenterology Hepatology, 2021.
11. Seyed Saeid Seyedian, Forogh Nokhostin, and Mehrdad Dargahi Malamir. A review of the di- agnosis, prevention, and treatment methods of inflammatory bowel disease. Journal of medicine and life, 12(2):113, 2019.
12. Juan Du, Jie Yin, Haojie Du, and Jiawei Zhang. Revisiting an expression dataset of discordant inflammatory bowel disease twin pairs using a mutation burden test reveals cyp2c18 as a novel marker. Frontiers in Genetics, 12:1020, 2021.
13. Daniel B Graham and Ramnik J Xavier. Pathway paradigms revealed from the genetics of in- flammatory bowel disease. Nature, 578(7796):527–539, 2020.
14. Benjamin M Scott, Cristina Gutierrez-Vazquez, Liliana M Sanmarco, Jessica A da Silva Pereira, Zhaorong Li, Agustin Plasencia, Patrick Hewson, Laura M Cox, Madelynn O'Brien, Steven K Chen, et al. Self-tunable engineered yeast probiotics for the treatment of inflammatory bowel disease. Nature Medicine, 27(7):1212–1222, 2021.
15. Gilaad G Kaplan and Joseph W Windsor. The four epidemiological stages in the global evolution of inflammatory bowel disease. Nature Reviews Gastroenterology Hepatology, 18(1):56–66, 2021.
16. Adam McDermaid, Brandon Monier, Jing Zhao, Bingqiang Liu, and Qin Ma. Interpretation of differential gene expression results of rna-seq data: review and integration. Briefings in bioinformatics, 20(6):2044–2054, 2019.
17. Min Zhu, Longfei Jia, Fangyu Li, and Jianping Jia. Identification of kiaa0513 and other hub genes associated with alzheimer disease using weighted gene coexpression network analysis. Frontiers in genetics, 11:981, 2020.
18. Yun Ji, Lu Gao, Can Zhang, Xu Sun, Liping Dai, Zhenyu Ji, Jianying Zhang, Zhida Zhang, Wei Cao, Yang Zhao, et al. Identification of the hub genes and prognostic indicators of gastric cancer and correlation of indicators with tumor-infiltrating immune cell levels. Journal of Cancer, 12(13):4025, 2021.
19. Ming Wu, Yujie Zhao, Nanxi Peng, Zuo Tao, and Bo Chen. Identification of chemoresistance- associated micrornas and hub genes in breast cancer using bioinformatics analysis. Investigational New Drugs, 39(3):705–712, 2021.

20. Tiancheng Zhang, Guihua Rao, and Xiwen Gao. Identification of hub genes in tuberculosis via bioinformatics analysis. Computational and Mathematical Methods in Medicine, 2021, 2021.
21. Shuaiqun Wang, Xiaoling Xu, and Wei Kong. Identification of hub genes associated with lung adenocarcinoma based on bioinformatics analysis. Computational and Mathematical Methods in Medicine, 2021, 2021.
22. Xuan Yang, Wangao Wei, Shisheng Tan, Linrui Guo, Song Qiao, Biao Yao, and Zi Wang. Identifi- cation and verification of hcar3 and insl5 as new potential therapeutic targets of colorectal cancer. World journal of surgical oncology, 19(1):1–12, 2021.
23. Emily Clough and Tanya Barrett. The gene expression omnibus database. In Statistical genomics, pages 93–110. Springer, 2016.
24. Dechao Bu, Haitao Luo, Peipei Huo, Zhihao Wang, Shan Zhang, Zihao He, Yang Wu, Lianhe Zhao, Jingjia Liu, Jincheng Guo, et al. Kobas-i: intelligent prioritization and exploratory visualization of biological functions for gene enrichment analysis. Nucleic acids research, 49(W1):W317–W325, 2021.
25. Zhuorui Xie, Allison Bailey, Maxim V Kuleshov, Daniel JB Clarke, John E Evangelista, Sherry L Jenkins, Alexander Lachmann, Megan L Wojciechowicz, Eryk Kropiwnicki, Kathleen M Jagodnik, et al. Gene set knowledge discovery with enrichr. Current protocols, 1(3):e90, 2021.
26. Sunil Kumar, Arpan Kumar Kar, and P Vigneswara Ilavarasan. Applications of text mining in services management: A systematic literature review. International Journal of Information Management Data Insights, 1(1):100008, 2021.
27. Wei Ma, Bei Shi, Fangkun Zhao, Yunfei Wu, and Feng Jin. Systematic analysis of breast atypical hyperplasia-associated hub genes and pathways based on text mining. European Journal of Cancer Prevention, 28(6):507, 2019.
28. Farhat Khan, Aleksandar Radovanovic, Takashi Gojobori, and Mandeep Kaur. Ibddb: a manually curated and text-mining-enhanced database of genes involved in inflammatory bowel disease. Database, 2021, 2021.
29. Jonathan Kans. Entrez direct: E-utilities on the unix command line. In Entrez Programming Utilities Help [Internet]. National Center for Biotechnology Information (US), 2022.
30. Vincent F Scalfani. Using ncbi entrez direct (edirect) for small molecule chemical information searching in a unix terminal. Journal of Chemical Education, 98(12):3904–3914, 2021.
31. Qingdong Guan. A comprehensive review and update on the pathogenesis of inflammatory bowel disease. Journal of Immunology Research, 2019, 2019.
32. Christopher Andrew Lamb, Nicholas A Kennedy, Tim Raine, Philip Anthony Hendy, Philip J Smith, Jimmy K Limdi, Bu'Hussain Hayee, Miranda CE Lomer, Gareth C Parkes, Christian Selinger, et al. British society of gastroenterology consensus guidelines on the management of inflammatory bowel disease in adults. Gut, 68(Suppl 3):s1–s106, 2019.
33. Wenna Yang, Ping Zhao, Xia Li, Lanping Guo, and Wenyuan Gao. The potential roles of nat- ural plant polysaccharides in inflammatory bowel disease: A review. Carbohydrate Polymers, 277:118821, 2022.
34. Geert D'Haens, Florian Rieder, Brian G Feagan, Peter DR Higgins, Julian Pan´es, Christian Maaser, Gerhard Rogler, Mark L¨owenberg, Robbert van der Voort, Massimo Pinzani, et al. Chal- lenges in the pathophysiology, diagnosis, and management of intestinal fibrosis in inflammatory bowel disease. Gastroenterology, 162(1):26–31, 2022.
35. Roberta Caruso, Bernard C Lo, and Gabriel Nu´n~ez. Host–microbiota interactions in inflammatory bowel disease. Nature Reviews Immunology, 20(7):411–426, 2020.
36. Ida Schoultz and Asa V Keita. Cellular and molecular therapeutic targets in inflammatory bowel disease—focusing on intestinal barrier function. Cells, 8(2):193, 2019.
37. Jiani Huang, Fang Wen, Wenjie Huang, Yingfeng Bai, Xiaona Lu, and Peng Shu. Identification of hub genes and discovery of promising compounds in gastric cancer based on bioinformatics analysis. Biomarkers in Medicine, 14(12):1069–1084, 2020.
38. Alba Costales-Carrera, Asunci´on Fern´andez-Barral, Pilar Bustamante-Madrid, Orlando Dom´ınguez, Laura Guerra-Pastri´an, Ram´on Cantero, Luis Del Peso, Aurora Burgos, Antonio Barb´achano, and Alberto Mun~oz. Comparative study of organoids from patient-derived normal and tumor colon and rectal tissue. Cancers, 12(8):2302, 2020.
39. Andreas Zollner, Andreas Schmiderer, Simon J Reider, Georg Oberhuber, Alexandra Pfister, Bernhard Texler, Christina Watschinger, Robert Koch, Maria Effenberger, Tim Raine, et al. Faecal biomarkers in inflammatory bowel diseases: Calprotectin versus lipocalin-2—a comparative study. Journal of Crohn's and Colitis, 15(1):43–54, 2021.
40. Xiaoxue Qiu, Marissa G Macchietto, Xiaotong Liu, You Lu, Yiwei Ma, Hong Guo, Milena Saqui- Salces, David A Bernlohr, Chi Chen, Steven Shen, et al. Identification of gut microbiota and microbial metabolites regulated by an antimicrobial peptide lipocalin 2 in high fat diet-induced obesity. International Journal of Obesity, 45(1):143–154, 2021.
41. Guohe Song, Shifeng Xu, Hong Zhang, Yupeng Wang, Chao Xiao, Tao Jiang, Leilei Wu, Tao Zhang, Xing Sun, Lin Zhong, et al. Timp1 is a prognostic marker for the progression and metastasis of colon cancer through fak-pi3k/akt and mapk pathway. Journal of Experimental Clinical Cancer Research, 35(1):1–12, 2016.
42. Ru Huang, Kaijing Wang, Lei Gao, and Wei Gao. Timp1 is a potential key gene associated with the pathogenesis and prognosis of ulcerative colitis-associated colorectal cancer. OncoTargets and therapy, 12:8895, 2019.
43. Zhen-Yu Liu, Ming Zheng, Yi-Ming Li, Xin-Yu Fan, Jian-Chao Wang, Zhu-Chun Li, Hai-Jiao Yang, Jing-Min Yu, Jian Cui, Jian-Li Jiang, et al. Rip3 promotes colitis-associated colorectal cancer by controlling tumor cell proliferation and cxcl1-induced immune suppression. Theranos- tics, 9(12):3659, 2019.
44. Wei Li, Ning Zhang, Caining Jin, Mark D Long, Hasan Rajabi, Yota Yasumizu, Atsushi Fushimi, Nami Yamashita, Masayuki Hagiwara, Rongbin Zheng, et al. Muc1-c drives stemness in progres- sion of colitis to colorectal cancer. JCI insight, 5(12), 2020.
45. Amanda Lance, Lawrence J Druhan, C Greer Vestal, Nury M Steuerwald, Alicia Hamilton, Mathew Smith, Andrea Price, Elise

Tjaden, Andee N Fox, and Belinda R Avalos. Altered ex- pression of csf3r splice variants impacts signal response and is associated with srsf2 mutations. Leukemia, 34(2):369–379, 2020.

46. Haijiao Zhang, Cody Coblentz, Kevin Watanabe-Smith, Sophie Means, Jasmine Means, Julia E Maxson, and Jeffrey W Tyner. Gain-of-function mutations in granulocyte colony–stimulating fac- tor receptor (csf3r) reveal distinct mechanisms of csf3r activation. Journal of Biological Chemistry, 293(19):7387–7396, 2018.
47. Fatemeh Shabani, Alireza Farasat, Majid Mahdavi, and Nematollah Gheibi. Calprotectin (s100a8/s100a9): a key protein between inflammation and cancer. Inflammation Research, 67(10):801–812, 2018.
48. Lixin Lou, Peng Zhang, Rongli Piao, and Yang Wang. Salmonella pathogenicity island 1 (spi-1) and its complex regulatory network. Frontiers in cellular and infection microbiology, 9:270, 2019.
49. Shanshan Kong and Weiqiang Zhang. The function of macromolecular complex of cftr-nherf2-lpa2 in inflammatory responses of intestinal epithelial cells. bioRxiv, page 186023, 2019.
50. Jonathan C Strefford, Malgorzata Nowicka, Chantal E Hargreaves, Cathy Burton, Andrew Davies, Rosalind Ganderton, Wolfgang Hiddemann, Chisako Iriyama, Wolfram Klapper, Kate V Latham, et al. Single-nucleotide fcγ receptor polymorphisms do not impact obinutuzumab/rituximab out- come in patients with lymphoma. Blood advances, 5(15):2935–2944, 2021.
51. Raheleh Rahbari, Luciana W Zuccherato, German Tischler, Belinda Chihota, Hasret Ozturk, Sara Saleem, Eduardo Tarazona-Santos, Lee R Machado, and Edward J Hollox. Understanding the genomic structure of copy-number variation of the low-affinity fcγ receptor region allows confirmation of the association of fcgr3b deletion with rheumatoid arthritis. Human mutation, 38(4):390–399, 2017.
52. David Westergaard, Hans-Henrik Stærfeldt, Christian Tønsberg, Lars Juhl Jensen, and Søren Brunak. A comprehensive and quantitative comparison of text-mining in 15 million full-text articles versus their corresponding abstracts. PLoS computational biology, 14(2):e1005962, 2018.
53. Ji-Yeon Kim, Hae Hyun Jung, Insuk Sohn, Sook Young Woo, Hyun Cho, Eun Yoon Cho, Jeong Eon Lee, Seok Won Kim, Seok Jin Nam, Yeon Hee Park, et al. Prognostication of a 13-immune-related-gene signature in patients with early triple-negative breast cancer. Breast Cancer Research and Treatment, 184(2):325–334, 2020.
54. Jie Wei, Dalang Fang, and Weijie Zhou. Ccr2 and ptprc are regulators of tumor microenvironment and potential prognostic biomarkers of lung adenocarcinoma. Annals of Translational Medicine, 9(18), 2021.
55. Xuankang Wang, Xin Li, Xiaoshuang Zuo, Zhuowen Liang, Tan Ding, Kun Li, Yangguang Ma, Penghui Li, Zhijie Zhu, Cheng Ju, et al. Photobiomodulation inhibits the activation of neurotoxic microglia and astrocytes by inhibiting lcn2/jak2-stat3 crosstalk after spinal cord injury in male rats. Journal of neuroinflammation, 18(1):1–20, 2021.
56. Jan Korbecki, Katarzyna Barczak, Izabela Gutowska, Dariusz Chlubek, and Irena Baranowska- Bosiacka. Cxcl1: Gene, promoter, regulation of expression, mrna stability, regulation of activity in the intercellular space. International Journal of Molecular Sciences, 23(2):792, 2022.
57. Boying Zhao, Renfu Lu, Jianjun Chen, Ming Xie, Xingji Zhao, and Lingwen Kong. S100a9 blockade prevents lipopolysaccharide-induced lung injury via suppressing the nlrp3 pathway. Respiratory Research, 22(1):1–11, 2021.
58. Liang Duan, Rui Wu, Xiuyu Zhang, Ding Wang, Yan You, Yunyuan Zhang, Lan Zhou, and Weixian Chen. Hbx-induced s100a9 in nf-κb dependent manner promotes growth and metastasis of hepatocellular carcinoma cells. Cell death disease, 9(6):1–14, 2018.