Journal of Advances in Medicine and Medical Research

33(3): 79-104, 2021; Article no.JAMMR.65206
ISSN: 2456-8899
(Past name: British Journal of Medicine and Medical Research, Past ISSN: 2231-0614, NLM ID: 101570965)

Detection of Unspecified Inflammatory Bowel Disease, Crohn's Disease and Ulcerative Colitis via Metabolite Analysis and Machine Learning

Elliot Kim[1], Valentina L. Kouznetsova[2,3] and Igor F. Tsigelny[2,3,4*]

[1]Williams College, Williamstown, Mass., USA.
[2]San Diego Supercomputer Center, UC San Diego, Calif., USA.
[3]BiAna, San Diego, Calif., USA.
[4]Department of Neurosciences, UC San Diego, Calif., USA.

Authors' contributions

This work was carried out in collaboration among all authors. Authors IFT and EK conceptualized the study. Authors VLK, IFT and EK proposed a set of methods to use. Author EK did machine-learning development and testing. Author VLK conducted the gene–metabolite networks studies. Authors IFT, VLK and EK analyzed the results of theoretical experiments. Authors EK and VLK wrote the article. Author IFT supervised the project. All authors read and approved the final manuscript.

Article Information

DOI: 10.9734/JAMMR/2021/v33i330823
Editor(s):
(1) Dr. Emin Umit Bagriacik, Gazi University, Turkey.
Reviewers:
(1) Masumeh Saeidi, Tehran University of Medical Sciences, Iran.
(2) Paolo Sossai, University of Camerino, Italy.
Complete Peer review History: http://www.sdiarticle4.com/review-history/65206

Original Research Article

Received 20 November 2020
Accepted 26 January 2021
Published 01 March 2021

ABSTRACT

The current standard of inflammatory bowel disease (IBD), especially, Crohn's disease (CD), diagnosis is set through an invasive endoscopy procedure. However, serum metabolites hold potential as useful biomarkers for non-invasive diagnosis and treatment of IBDs. The goal of this research was to elucidate the biomarkers including metabolites and genes related to IBDs, to show their distinguishing and common features, and to create a machine-learning (ML) model for recognition of each disease. We explored metabolic pathways and gene–metabolite networks related to unspecified-IBD (uIBD), Crohn's diseaseand ulcerative colitis (UC). P38 MAPK, ERK1/2, AMPK, and proinsulin were found to be closely related to the pathology of IBDs. The best performing ML model, trained on filtered disease-specific metabolite datasets, was able to predict metabolite class with 92.17% accuracy. Through examination of IBD-related serum, significant

**Corresponding author: E-mail: itsigeln@ucsd.edu, itsigeln@health.ucsd.edu;*

relationships between the inputted metabolites and certain metabolic and signaling pathways were found, which can be pinpointed and used to increase accuracy of disease diagnoses. Development of a ML model including metabolites and their chemical descriptors made it possible to achieve considerable accuracy of prediction of the IBDs. Our results elucidate a large variety of metabolites, genes, and pathways that could be used for better understanding of IBDs' molecular mechanisms.

Keywords: Inflammatory bowel disease; Crohn's disease; ulcerative colitis; metabolomics; metabolic networks; biomarkers; machine learning.

1. INTRODUCTION

Recent metabolomics research on the human gut microbiome has led to significant attention of bacteria species in the human gut and the respective metabolomic biomarkers and metabolic pathways, which may create conditions that foster the occurrence of certain diseases. We work within these lines to try and elucidate metabolites that may serve as useful biomarkers in the inflammatory bowel diseases (IBD): ulcerative colitis (UC), Crohn's disease (CD), and unspecified-IBD (uIBD). According to the Centers for Disease Control and Prevention, approximately 1.3% of US adults (3 million cases) were reported to being diagnosed with IBD—either Crohn's disease or ulcerative colitis—in 2015 [1]. Monitoring metabolic changes within the body, specifically those stemming from the gut microbiome, can be used to elucidate the inner workings of the many molecular processes present within the body to a greater degree. In our case, the pathology and mechanisms of IBD, along with how certain changes in the concentrations of metabolites produced in the gut, are leveraged as indicators of the presence of the disease.

Analysis of IBD presence and the related metabolites can be useful for specific diagnosis of the IBDs, which are comprised of two chronic relapsing diseases—Crohn's disease and ulcerative colitis—that inflame and damage the gastrointestinal tract (GIT). Crohn's disease most often occurs in the end or the small intestine but can occur in any part of the gastrointestinal tract. Ulcerative colitis occurs in the large intestine and the rectum. Common symptoms of these IBDs include persistent diarrhea, abdominal pain, rectal bleeding and bloody stools, weight loss, and fatigue [2]. Formerly, IBD has been associated with defective immune systems, certain diets, genetic susceptibility, and environmental factors [3], but in recent years, the gut microbiota has been increasingly suspected to have a major impact on the pathogenesis of IBD [4]. There are approximately 2000 bacteria species from 12 different phyla in the human gastrointestinal tract [5], and the commensalism between the microbiota has a significant role in not only the metabolism of food and production of energy, but also in the protection of the body against pathogens [6,7]. Prior research indicated that the microbiota and the associated metabolites affect the gut health [8]. The gut microbiota synthesizes these metabolites, which may then be used by the host; these metabolites play essential roles in the maintenance of the host's homeostatic health system. Recent research has also increased the number of metabolites that have functional roles in IBD pathogenesis [9].

Dysbiosis can occur in both the inflamed and non-inflamed parts of the gastrointestinal tract in IBD patients, and research has shown there are significant differences in the composition of the microbiota of IBD patients as compared to healthy patients [10]. It has also been shown that the richness and abundance of bacterial species decreases in patients with IBD [11]. The probiotic bacteria from the Firmicutes and Bacteroidetes phyla that are predominant in healthy human guts, have been found to be present at significantly lower levels or even depleted in the guts of IBD patients [12]. However, the pathogenic Proteobacteria phylum [13] and the pivotal Actinobacteria phylum, which are present in small quantities in health patients [14], are both elevated in the guts of IBD patients [15].

The current standard of IBD, specifically Crohn's disease diagnosis is the invasive endoscopy procedure. However, serum microbiota holds potential as a useful biomarker for non-invasive diagnosis and treatment of CD [16]. The profiling of the gut microbiota has become a useful diagnostic tool in IBD treatment, and with methods such as deep sequencing and Genome Analyzer map (GA-map) dysbiosis testing, dysbiosis in IBD patients and the composition of gut microbiota can be elucidated [17]. Certainly

the identification of IBD early on in its pathogenesis would be ideal, either via metabolic profiling of IBD-patients' plasma, serum, urine, or fecal samples. Treatments that are known to be effective for reducing inflammation and can help counteract the pathogenesis of IBD may be administered. Metabolic profiles from IBD patients can be integrated with data from healthy patients to identify specific IBD biomarkers or metabolites. As for the specific type of metabolic profiles, metabolites from the blood—plasma or serum—would present the metabolites that are most impactfully and directly involved in the disease pathogenesis. Yu and colleagues found that reproducibility is good with both plasma and serum profiles, but the higher metabolite concentrations in serum allow for more sensitive results in biomarker detection [18]. And thus, serum metabolomic profiling is a promising direction for IBD diagnosis and monitoring for its unique metabolite profiles [19].

1.1 IBD-Related Research

Lavelle and Sokol [20] conducted an extensive review that defines the classes of metabolites produced in the human gut microbiome that are altered in IBD-patients, as well as describes the pathophysiological evidence for those associations. These alterations in the composition and function of the human microbiota have been researched in numerous studies on IBD and contribute to the identification of possible targets for therapeutics.

A similar study provided an updated summary on the progress regarding research on the human gut microbiota and its influences on IBD pathogenesis [21]. Their sources from the PubMed database were composed of clinical studies as well as animal studies related to intestinal microbiomes and IBD. The general consensus of the previous research findings were that the biodiversity of probiotic-associated microbiota is decreased, while pathogenic microbiota is increased [21]. This brings up supported evidence for the significance that microbiome health and its associated metabolites have on human well-being, specifically intestinal permeability and the immune response [21].

Lai and coauthors used high-resolution mass spectrometry (HRMS) to conduct and untargeted LC/MS metabolomic profiling in Crohn's disease patients. Serum samples of both positive-tested and negative-tested patients were used for collection and profiling with state-of-the-art compound identification workflow [22]. Results of this study showed a distinct metabolic profile of Crohn's disease compared to that of the control, with most metabolites being downregulated [22]. Thus this study upholds the effectiveness that untargeted metabolomics has for biomarker development and analytical interpretation, reinforcing the value of biomarker research in the etiological inquiries for IBD.

Dawiskiba and colleagues performed proton nuclear magnetic resonance (NMR) spectroscopy to analyze, diagnose, and monitor serum and urine samples from CD and UC patients [23]. The study found certain metabolites that greatly contributed to distinguishing active-IBD from IBD in remission. N-acetylated compounds and phenylalanine were up-regulated in serum, while low-density lipoproteins and very-low-density lipoproteins were decreased in serum. As for urine samples, glycine concentrations were found to be increased and acetoacetate decreased [23]. They also elucidated metabolites that distinguished patients with active IBD from healthy control subjects. Leucine, isoleucine, 3-hydroxybutyric acid, N-acetylated compounds, acetoacetate, glycine, phenylalanine, and lactate were found to be increased in serum samples. Creatine, dimethyl sulfone, histidine, and choline and its derivatives were found to be decreased in serum samples [23]. Citrate, hippurate, trigonelline, taurine, succinate and 2-hydroxyisobutyrate were found to be decreased in urine samples [23]. The findings endorse the analytical capacity of NMR-based metabolomic research of serum and urine samples as a useful tool in distinguishing active IBD-patients from those with IBD in remission and from those with no IBD.

Daniluk and coauthors performed liquid chromatography and mass spectrometry in children to perform an untargeted metabolomics analysis to detect metabolic differences in serum metabolites between newly diagnosed and untreated pediatric CD and UC patients, in contrast to a healthy control group [24]. Using serum untargeted metabolomics, they were able to find that only one metabolite—lactosylceramide 18:1/16:0 (LacCer 18:1/16:0)—that significantly discriminated CD patients from UC patients [24].

Despite the numerous studies performed on CD, UC, and uIBD detection and biomarker discovery situated on plasma, serum, urine, and fecal

metabolomes, further metabolic profiling—filtered by specific needs of researchers, along with the assortments and advantages provided by the different types of samples—may yield new putative markers due to the ever-changing dynamic quality of each of the metabolomes. Thus, there is a compelling need to develop more reliable IBD markers and through the analysis of metabolite profiles, certain biological pathways can be identified, allowing for specific targeting of treatment for the disease. The goal of this research project was to analyze metabolites sets of uIBD, CD, and UC and to show their distinguishing and common features, and to create a machine-learning (ML) model that can distinguish them from each other and from healthy control subjects.

2. METHODS

2.1 Approach Overview

The programs used for metabolomics analysis are MetaboAnalyst 4.0 [25,26], Ingenuity® Pathway Analysis (IPA®) [27,28], PaDEL-descriptor [29,30], and Waikato Environment for Knowledge Analysis (WEKA) [31,32]. The flowchart of methods is shown in Fig. 1.

The study began with the selection of metabolites significantly related to uIBD, CD, and UC diseases, along with a set of random metabolites from HMDB [33,34], using metabolites' IDs (HMDB IDs) to represent a control group. The sets of metabolites used for testing and training were obtained from IBD patient serum samples from the publication of Scoville and colleagues [19]. The initial table [19] contains fold changes (FC) and *p*-values for metabolite expression in all groups. The metabolites were filtered by *p*-values. Those with *p*-values greater than −0.05 were deleted. The selected metabolites for uIBD, CD, and UC are presented in Supplementary Table S1, sorted by decreasing concentration. Selected metabolites were analyzed with MetaboAnalyst 4.0 [25,26] to find the relevant metabolic pathways for each disease. The significance of each metabolic

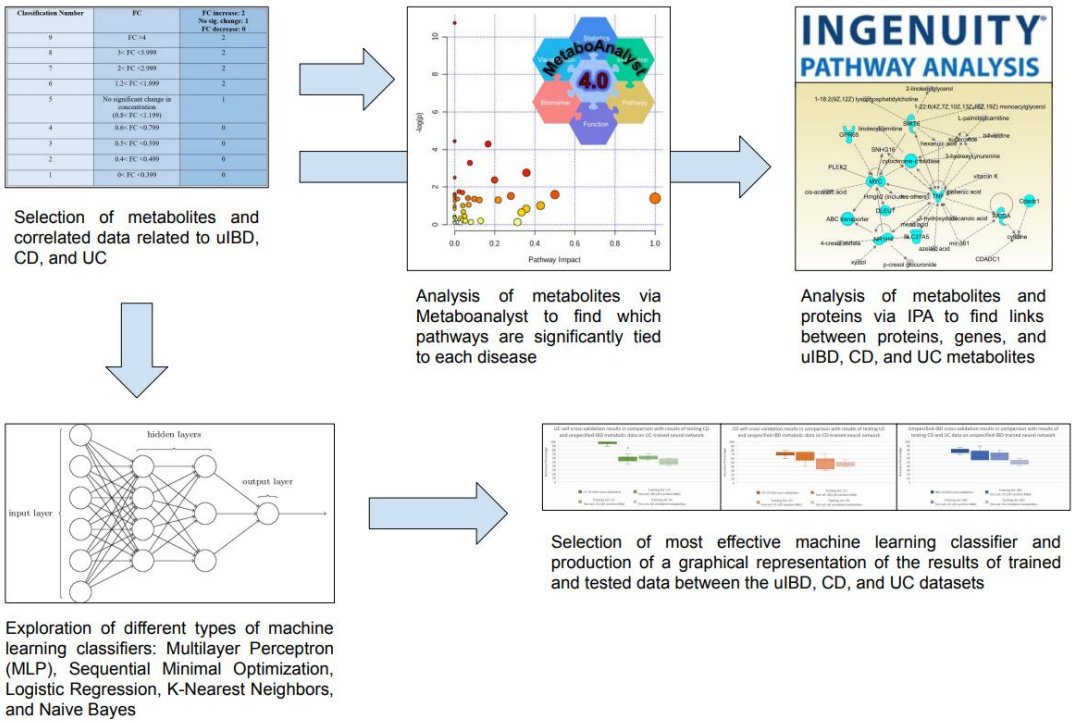


Fig. 1. Overview of the methods of the study. A selection of metabolites related to uIBD, CD, and UC were found and verified from public sources, then inputted into MetaboAnalyst, Ingenuity Pathway Analysis, and WEKA programs. Analysis of the metabolites and genes with these programs allowed for the elucidation of certain significant pathways tied to each disease category, as well as links between the metabolites and proteins. We were also able to create a metabolite-based model for recognition of each of the three disease categories

pathway for the diseases was then graphically displayed. Then the selected metabolites were analyzed with the Ingenuity Pathway Analysis (IPA) software [27,28]. IPA elucidated metabolite–gene networks, with additional metabolites and genes interacting with metabolites. We used machine-learning (ML) methods, specifically the Multilayer Perceptron (MLP) classifier in the Waikato Environment for Knowledge Analysis (WEKA) [31,32], using PaDEL program [29,30] for calculation and selection of metabolites' descriptors, to create a metabolite-based model for recognition of and differentiation between uIBD, CD, UC, and control.

2.2 MetaboAnalyst

MetaboAnalyst 4.0 (Xia Lab, McGill University, Montreal Quebec, Canada) [25,26] is a program for statistical, functional, and integrative analysis of metabolomics data. It allows user to perform exploratory statistical analysis, functional enrichment analysis, data integration and systems biology (biomarker analysis, pathway analysis, and network explorer), and data processing. The software accepts a large variety of metabolomics data input types, such as a list of gene/compound names, KEGG ID orthologs (KOs) [35], or Human Metabolite Database index numbers (HMDB ID) [33,34], to support integrative analysis with transcriptomics or metagenomics.

2.3 Ingenuity® Pathway Analysis

Ingenuity® Pathway Analysis (IPA®; QIAGEN Inc., Redwood City, Calif., USA) [27,28] is a dynamic genomics and metabolomics analysis tool that yields the significance of the inputted data and identifies new candidate biomarkers within the sample of biological systems. IPA is widely used in the scientific research community and is cited in thousands of articles for its analysis and interpretation of 'omics data.

2.4 Machine-Learning Analysis

ML analysis was performed on the uIBD, CD, and UC datasets with WEKA program environment [31,32]. WEKA is a workbench that supports several pattern-recognition methods. Attribute selection was performed with PaDEL, which introduced 1451 descriptor sets [29,30]. These PaDEL descriptors highly contribute to the construction of a machine-learning classifier, as they provide a vast amount of information needed to effectively narrow down the datasets. The InfoGain attribute evaluator, along with the ranker search method, was then used to rank the significance of all these descriptors by measuring the information gain with respect to class. After InfoGain filtration, the number of descriptors sets for each of the three datasets was reduced from 1451 descriptors per disease, down to 111 descriptors for uIBD metabolites, 147 descriptors for CD metabolites, and 34 descriptors for UC metabolites. The multilayer perceptron classifier (MLP) method was used to build three robust models, one per each disease, via three training sets and the corresponding testing sets. Other modeling methods and classifiers were attempted as well. The full data preprocessing and modeling workflow is presented in Fig. 2.

Each model was trained using the MLP classifier to detect metabolic data of unspecified-IBD (uIBD), CD, and UC, through 10-fold cross validation tests. To demonstrate the validity of these systems, we tested metabolic data from CD and UC on the trained uIBD system, tested uIBD and CD metabolic data on the trained UC system, and uIBD and UC metabolic data on the trained CD system. An overview of the workflow is shown in Fig. 3. As a second test to further demonstrate and conclude the validity of these systems, we tested 60 random human database metabolites and their corresponding data on each of the three disease categories. Model performance was evaluated by measuring accuracy, which here is understood as the total number of correctly classified instances over the total number of instances.

3. RESULTS

Unspecified-IBD (uIBD), CD, and UC metabolites along with FC-values, *p*-values, and other corresponding attributes were extracted from the work of Scoville and colleagues [19] were examined to select the best sets for use as biomarkers for the diagnosis of each disease. After filtration of the metabolites by filtering out those with *p*-values greater than −0.05 and fold changes between 0.8 and 1.2, we examined 140 metabolites for uIBD, 200 for CD, and 37 for UC. With the MetaboAnalyst and IPA tools, we studied the pathways related to each of the three disease categories that could be useful for diagnostics and therapeutics.

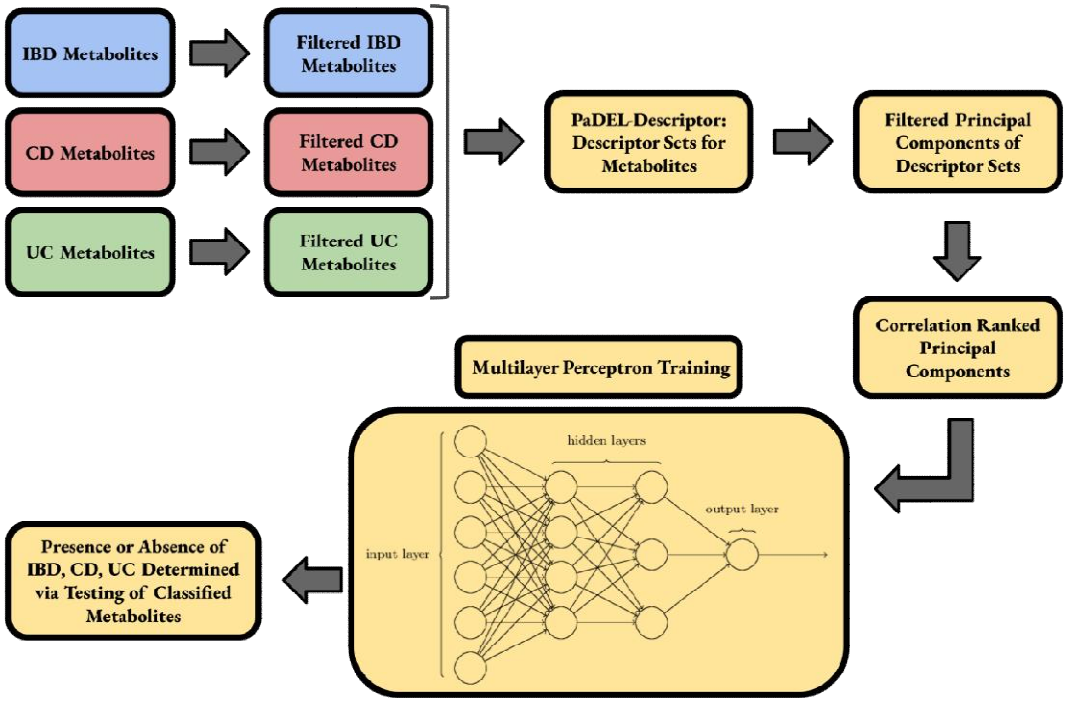


Fig. 2. Block-diagram of the full data preprocessing and modeling steps via machine learning. The starting, unfiltered metabolites were collected, then filtered by *p*-value and fold change attributes. Then with PaDEL descriptor sets, a more comprehensive dataset was formed, allowing for the elaborate training of the machine learning model via the multilayer perceptron classifier. Ultimately, the goal is to fortify the detection of uIBD, CD, and UC via machine learning models by testing and training classified metabolic datasets

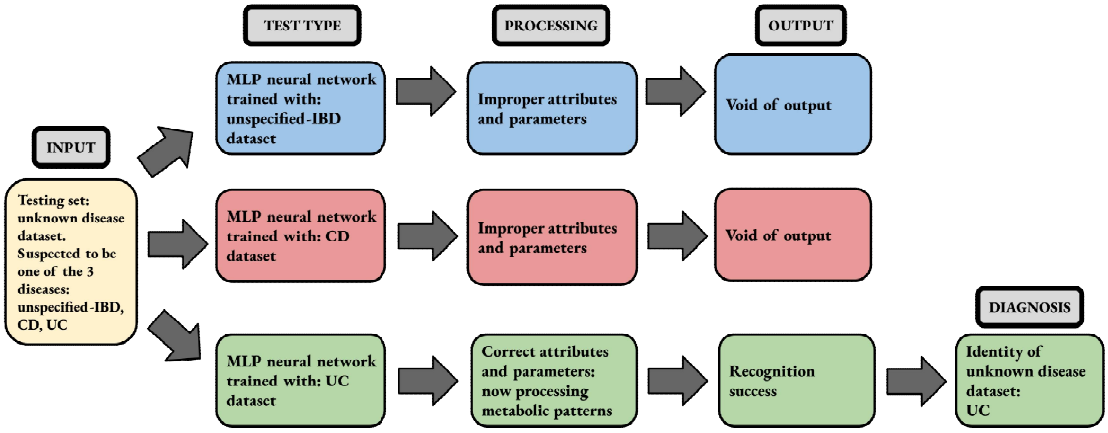


Fig. 3. An overview of the workflow of the machine-learning classifier development strategy to identify one disease from another

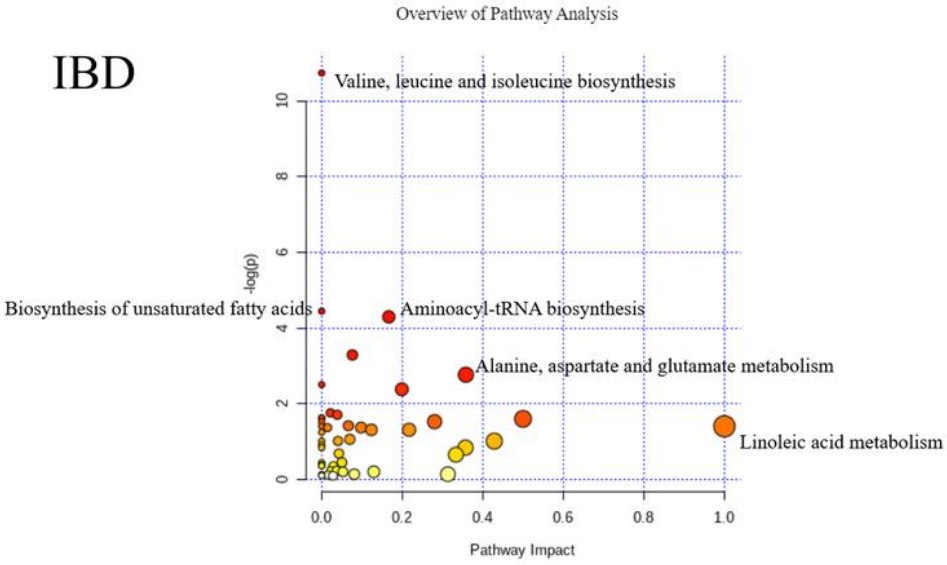
3.1 Metabolic Pathways

With the use of the pathway analysis provided by MetaboAnalyst, the most significant of the matched pathways corresponding to the inputted metabolites arranged by *p*-values is displayed on the Y-axis, and pathway impact values are displayed on the X-axis (the node colors are based on the significance, determined by *p*-value, and the size of the radii of the nodes are determined by their pathway impact values).

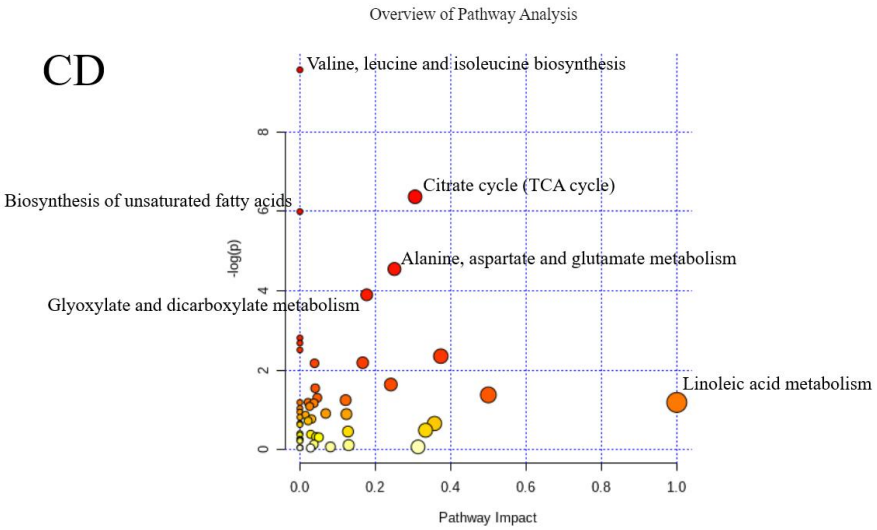
These *p*-values and corresponding impact values were generated and scaled to fit each specific pathway by MetaboAnalyst. Fig. 4a shows the pathways significantly related to uIBD: Valine, leucine, and isoleucine biosynthesis, Biosynthesis of unsaturated fatty acids, Aminoacyl-tRNA biosynthesis, Alanine, aspartate, and glutamate metabolism, and Linoleic acid metabolism. Fig. 4b shows the pathways significantly related to CD: Valine, leucine, and isoleucine biosynthesis, Biosynthesis of unsaturated fatty acids, Citrate cycle (TCA cycle), Alanine, aspartate, and glutamate metabolism, Glyoxylate and dicarboxylate metabolism, and Linoleic acid metabolism. Fig. 4c shows pathways significantly related to UC: Valine, leucine, and isoleucine biosynthesis, Aminoacyl-tRNA biosynthesis, Arginine biosynthesis, Glycine, serine, and threonine metabolism, Taurine and hypotaurine metabolism, and Linoleic acid metabolism.

These pathways are explained in detail in the following sections with respect to the diseases they were found active in.

(a)



(b)



(c)

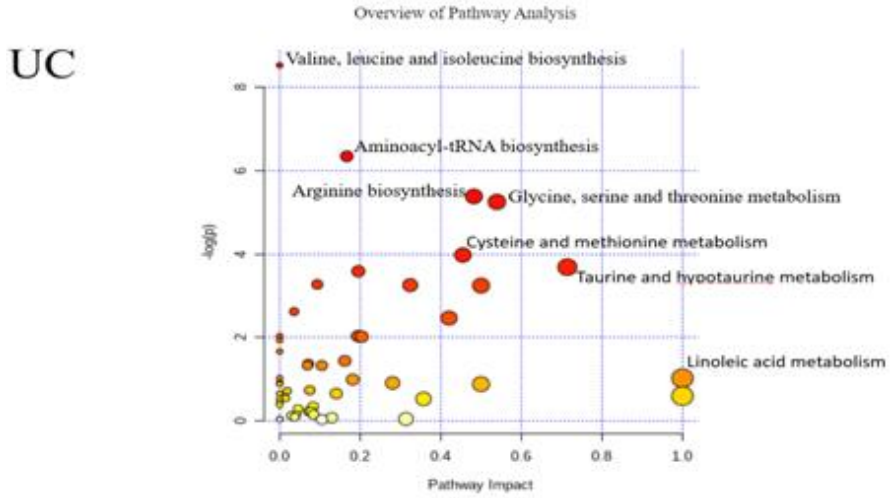


Fig. 4. Most significant of the pathways corresponding to the inputted metabolites arranged by *p*-values on the Y-axis, and pathway impact values on the X-axis (The node colors are based on the significance—determined by *p*-value—and the size of the radii of the nodes are determined by their pathway impact values) of (a) uIBD. Valine, leucine, and isoleucine biosynthesis, Biosynthesis of unsaturated fatty acids, Aminoacyl-tRNA biosynthesis, and Arginine biosynthesis. (b) CD. Valine, leucine, and isoleucine biosynthesis, Citrate cycle (TCA cycle), Biosynthesis of unsaturated fatty acids, Alanine, aspartate, and glutamate metabolism, and Glyoxylate and dicarboxylate metabolism. (c) UC. Valine, leucine, and isoleucine biosynthesis, Aminoacyl-tRNA biosynthesis, Arginine biosynthesis, Glycine, serine, and threonine metabolism, Cysteine and methionine metabolism, Taurine and hypotaurine metabolism, Glutathione metabolism, Primary bile acid biosynthesis, Arginine and proline metabolism, Phenylalanine, tyrosine, and tryptophan biosynthesis

3.1.1 Metabolic pathways of unspecified-ibd and previous findings

Several significant metabolic pathways involved in uIBD were elucidated via MetaboAnalyst software [25,26] and were related to the following pathways: Valine, leucine, and isoleucine biosynthesis, Biosynthesis of unsaturated fatty acids, Aminoacyl-tRNA biosynthesis, and Arginine biosynthesis.

Valine, leucine, and isoleucine biosynthesis: F. He and colleagues highlight the functions and signaling mechanisms of specific amino acids and their role in intestinal inflammation in IBD. Valine, leucine, and isoleucine are shown to have significant roles in intestinal inflammation based on their involvement with NF-κB, iNOS, MAPK, ACE2, GCN2, CaSR, and mTOR signaling pathways [36]. Valine, leucine and isoleucine biosynthesis is present in all three diseases, with proportionally high significance (as measured by *p*-value and indicative by the bold red circle) and proportionally low pathway impact to the pathogenesis of all three disease categories.

***Biosynthesis of unsaturated fatty acids*:** Abulizi and colleagues used 16S rRNA gene sequencing and mass-spectrometry-based relative quantification of the metaproteome to indicate that increased omega-6 polyunsaturated fatty acids in one's diet may lead to gut dysbiosis, and therefore is a risk factor for IBD in humans [37]. Biosynthesis of unsaturated fatty acids is notably present in uIBD and CD, with higher significance in uIBD than in CD, and equal pathway impact values in both diseases.

Aminoacyl-tRNA biosynthesis: A case study performed by Fagbemi and colleagues provides evidence that recessive mutations involved in cytosolic isoleucyl-tRNA synthetase lead to the pathogenesis of IBD [38]. Aminoacyl-tRNA biosynthesis is notably present in IBD and UC, with higher significance in IBD than in UC, and equal pathway impact values in both diseases.

Arginine biosynthesis: Morgan and colleagues observed alterations in the human gastrointestinal microbiota metabolism in both UC and CD. Amino-acid metabolism showed major perturbations, including the increase of arginine gene abundance [39].

3.1.2 Metabolic pathways of Crohn's disease and previous findings

Several significant metabolic pathways involved in uIBD were elucidated via MetaboAnalyst software [25,26]. Metabolites that exhibited fold changes under 0.8 and over 1.2 were related to the following pathways: Valine, leucine, and isoleucine biosynthesis, Citrate cycle (TCA cycle), Biosynthesis of unsaturated fatty acids, Alanine, aspartate, and glutamate metabolism, and Glyoxylate and dicarboxylate metabolism.

Valine, leucine, and isoleucine biosynthesis: Valine, leucine and isoleucine biosynthesis is present in all three diseases, with proportionally high significance (as measured by *p*-value and indicative by the bold red circle in Fig. 4) and proportionally low pathway impact to the pathogenesis of all three disease categories. Chiba and colleagues measured plasma-free amino-acid profiles in Crohn's disease patients. Via high-performance liquid chromatography, the fasting plasma concentrations of valine, leucine, and isoleucine amino acids were measured. Results showed significant correlations in all patients, between Crohn's disease activity index and concentrations of valine, leucine, and isoleucine, amongst other amino acids as well [40].

Citrate cycle (TCA cycle): Citrate cycle (TCA cycle) is notably present only in CD, with a similar significance value to the Biosynthesis of unsaturated fatty acids in the same disease, and a similar pathway impact value to Alanine, aspartate, and glutamate metabolism in the same disease. Weiser and coauthors found several pathways, including the TCA cycle, that are related to signaling of G-protein coupled receptors, which, through migration and accumulation within the inflamed tissues, are highly expressed in monocytes and macrophages with crucial roles in the pathogenesis of CD [41].

Biosynthesis of unsaturated fatty acids: Liu and colleagues used gas chromatography to elucidate significant changes in the metabolic levels in the synthesis of long-chain polyunsaturated fatty acids and indicated that impaired fatty acid desaturation contributes to chronic inflammation in CD [42]. Biosynthesis of unsaturated fatty acids is notably present in uIBD and CD, with higher significance in uIBD than in CD, and equal pathway impact values in both diseases.

Alanine, aspartate, and glutamate metabolism: Diab and colleagues performed integrative pathway analysis to elucidate significantly altered metabolic pathways, with alanine, aspartate, and glutamate metabolism included. This pathway consists of N-acetyl-L-aspartic acid, L-asparagine, L-glutamine, L-glutamic acid, gamma-aminobutyric acid, fumaric acid, succinic acid, and holds a *p*-value of 0.014 and impact value of 0.53 [43].

Glyoxylate and dicarboxylate metabolism: Q. He and coauthors analyzed functional changes associated with CD and microbiota alterations via pairwise comparisons, to elucidate the enrichment of genes in pathways involved in glyoxylate and dicarboxylate metabolism [44].

3.1.3 Metabolic pathways of ulcerative colitis and previous findings

Several significant metabolic pathways involved in uIBD were elucidated via MetaboAnalyst: Valine, leucine, and isoleucine biosynthesis, Aminoacyl-tRNA biosynthesis, Arginine biosynthesis, Glycine, serine, and threonine metabolism, Cysteine and methionine metabolism, Taurine and hypotaurine metabolism, Glutathione metabolism, Primary bile acid biosynthesis, Arginine and proline metabolism, and Phenylalanine, tyrosine, and tryptophan biosynthesis.

Valine, leucine, and isoleucine biosynthesis: Wang and colleagues studied UC-related urine samples and found that valine, leucine, and isoleucine biosynthesis were significantly related to the disease [45]. Valine, leucine and isoleucine biosynthesis is present in all three diseases, with proportionally high significance (as measured by *p*-value and indicative by the bold red circle in Fig. 4) and proportionally low pathway impact to the pathogenesis of all three disease categories.

Aminoacyl-tRNA biosynthesis: Filimoniuk and colleagues found that aminoacyl-tRNA biosynthesis is one of the metabolic pathways that is most altered in patients with inflammatory bowel diseases [46]. Aminoacyl-tRNA

biosynthesis is notably present in IBD and UC, with higher significance in IBD than in UC, and equal pathway impact values in both diseases.

Arginine biosynthesis: Coburn and coauthors found diminished tissue L-arginine in UC patients, likely attributable to a decrease in its cellular uptake. Together with the decreased ARG1 expression, there is a pattern of dysregulated L-Arg availability and metabolism in UC [47].

Glycine, serine, and threonine metabolism: Gu and colleagues used GC-MS-based metabolomics analysis to compare colitis mice to healthy control mice, and found that amino acids including glycine, serine, and threonine concentrations were increased in the disease-positive mice [48]. Glycine, serine and threonine metabolism is notably present only in UC with significance and pathway impact values similar to that of Arginine biosynthesis in the same disease.

Cysteine and methionine metabolism: Cysteine and methionine metabolism is notably present only in UC, with significance slightly lower than that of Arginine biosynthesis in the same disease, and similar pathway impact value as Arginine biosynthesis in the same disease. Zhang and coauthors used the Mann–Whitney U test to find alterations of sulfur and cysteine/methionine metabolism pathways in the mucosal-luminal interface microbiome of IBD (CD and UC) patients [49].

Taurine and hypotaurine metabolism: Taurine and hypotaurine metabolism is notably present only in UC, with significance values similar to that of Cysteine and methionine metabolism, and a pathway impact value somewhat higher than that of Cysteine and methionine metabolism. Kolho and colleagues analyzed blood and stool samples from pediatric patients with UC and found that a large proportion of the observed metabolic pathways were altered when compared to controls, with taurine and hypotaurine metabolism being the most highly enriched pathways [50]. When related to the blood inflammatory marker ESR, taurine was amongst the UC-related metabolites with the strongest correlations. Furthermore, the relation between fecal calprotectin—a marker of intestinal inflammation—and fecal metabolites showed that taurine is one of the most significant metabolites associated with the level of inflammation [50].

Glutathione metabolism: Holmes and colleagues collected endoscopic biopsies of colon mucosa from normal subjects, from macroscopically normal tissue of patients with either inactive or active UC, and from inflamed tissue of patients with active UC. The mucosal contents were analyzed via liquid chromatography and the oxidized glutathione content of the mucosa was found to have significant positive correlations with disease severity among UC patients [51].

***Primary bile acid biosynthesis*:** Miettinen and coauthors found that diarrhea in UC patients is not caused by the excessive fecal loss of bile salts, but rather by the decreased absorption and water and electrolyte retention due to the damaged colonic mucosa [52].

Arginine and proline metabolism: Arginine biosynthesis is notably present only in UC, with significance and pathway impact values similar to that of Glycine, serine and threonine metabolism in the same disease. Schicho and colleagues studied the differences in serum and plasma metabolite levels of UC and CD patients versus healthy controls, and found that in both UC and CD patients, there is noticeable alteration in amino acid metabolism in serum and plasma, including the increased levels of arginine and proline [53].

Phenylalanine, tyrosine, and tryptophan biosynthesis: Nikolaus and coauthors used IVD-CE certified high-performance liquid chromatography kit (ClinRep® Complete Kit for the analysis of phenylalanine, tyrosine and tryptophan; RECIPE Chemicals + Instruments GmbH, Munich, Germany) and found significantly increased tryptophan concentrations and tryptophan metabolism associated with the activities of both CD and UC [54].

3.2 Integrative Analysis of Networks

Sets of metabolites involved with uIBD, CD, and UC were separately submitted to the IPA program [27,28]. With the extensive database of interactions in the IPA software, numerous sets of gene–metabolite networks were elucidated, along with specific interactions between each of them (Fig. 5). Notable links in the networks are further described below.

3.2.1 uIBD networks

The first uIBD network (Fig. 5a) includes a set of submitted metabolites, indicated in gray, and generated metabolites, in white, which interact with proteins highlighted in blue. Some notable interactions are between the metabolites and

ERK1/2, AMPK, and proinsulin. Previous research shows that sustained activation of ERK1/2 is observed in IBD [55]. AMPK has been indicated to be a key enzyme involved in the inflammatory bowel diseases [56], due to its ameliorative effects that work to mitigate 'leaky gut' symptoms and improve epithelial barrier functions [56]. Furthermore, the inhibition or inactivation of AMPK may lead to a decrease in the proinsulin production, the precursor to insulin, which would present effects similar to that of insulin resistance. Insulin resistance has previously been associated with chronic inflammation, a significant characteristic of IBD [57].

The second uIBD network (Fig. 5b) includes metabolites that interact with EGFR and PARP1. Previous findings show that EGFR is frequently expressed in IBD-related research [58]. Inhibition of PARP-1 may reduce some of the major triggers for apoptosis and decrease production of AP1 [59], which has been shown to be linked to inflammation [60]. Thus, the targeting of the PARP-1 pathway may be a viable approach to IBD therapeutics. Additionally, the calcium ion metabolite appears to be a major contributor to many sub-networks in this second network, with direct links to both EGFR and PARP1.

The third uIBD network (Fig. 5c) includes metabolites that interact with APP and TNF. APP has shown strong links to Alzheimer's disease, but research also suggests that APP may influence susceptibility towards gut inflammatory diseases [61]. Evidence showing a link between TNF and IBD have been reported in previous publications in which IBD-patients showed increased levels of TNF in their serum samples [62]. Additionally, the metabolites—cyclic AMP and nitric oxide—appear to be a major contributor to many sub-networks in this third network, with direct links to both APP and TNF.

The extensive, protein-heavy fourth uIBD network (Fig. 5d) shows some metabolite-to-protein interactions, but predominantly depicts protein-to-protein interactions. The notable metabolite—which was found by the IPA tool—cytokine, has direct links to many of the proteins in the network, including ERK, MAPK, Akt, and JNK, amongst numerous others. This extensive series of links are displayed in blue colored lines. As previously stated, past findings suggest a correlation between sustained activation of MAPK/ERK and IBD [55]. Furthermore, inflammatory cytokines have been implicated in activation of both the Akt and Jnk signaling pathways [63,64].

3.2.2 CD networks

The first CD network (Fig. 5e) includes metabolites that interact with proinsulin, ERK1/2, and Ldh-complex. Past research on proinsulin and ERK 1/2 are described above. Additionally, ERK1/2 activation may significantly impact the symptom of diarrhea in patients with CD [65]. In this network, citric acid, kynurenic acid, succinic acid, and L-malic acid are directly linked to and un-reciprocatively act upon the Ldh-complex. These links are displayed in blue colored lines. This may cause an overproduction of mentioned complex, with no regulation to limit production, which could potentially lead to the pathogenesis of CD. Extensive research regarding the effect of the Ldh-complex on IBD is not yet readily present, but there are findings that strongly suggest a positive correlation between Ldh-complex concentrations and pathogenesis of diseases [66].

The second CD network (Fig. 5f) includes metabolites that interact with AMPK, insulin, and Akt, amongst numerous other proteins. Literature has shown that the mTORC1 protein holds a central role in autophagy regulation, and it has been suggested that AMPK is involved in the inhibition of mTORC1 [67]. It has also been found that patients with CD display increased insulin secretion levels caused by an enhanced beta cell function [68]. This increased secretion of insulin may even override the insulin resistance caused by the chronic inflammatory state of the disease [68]. Lastly, inflammatory cytokines have been implicated in activation of the Akt signaling pathway [63].

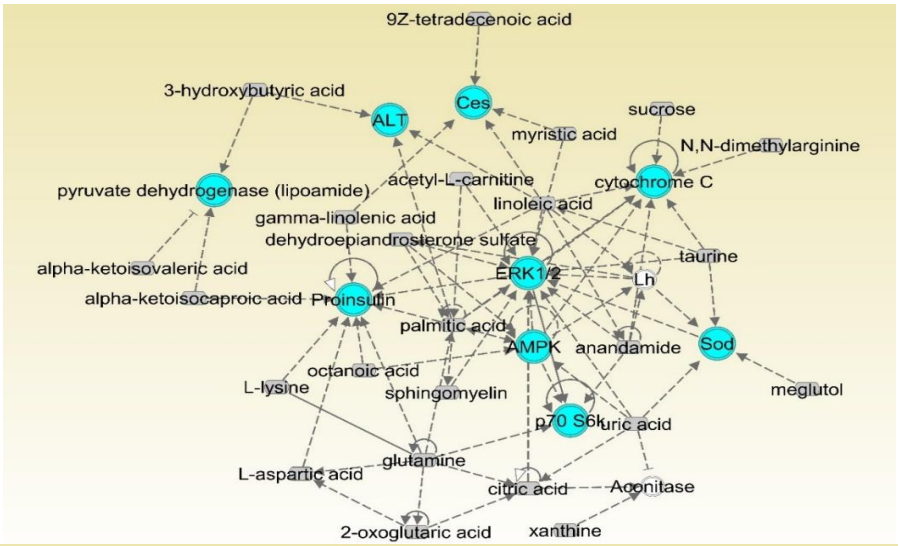
The third CD network (Fig. 5g) includes metabolites that interact with EGFR and FAH protein complexes. Past research on EGFR is described above [58]. In this network, it is shown that the EGFR protein acts upon the uridine metabolite, which then acts upon the FAH protein, as well as the UDP-D-glucose and UDP-N-acetylglucosamine metabolites. These links are displayed in blue colored lines. There is no regulation of the FAH protein; it is activated by uridine and has no inhibitors or regulators of any kind. This under-regulation could potentially lead to the pathogenesis of CD. Also, notable, many metabolites and protein complexes in this network activate or bind to D-glucose, including EGFR.

The fourth CD network (Fig. 5h) includes metabolites that interact with TNF and MYC protein complexes. Previous findings suggest that IBD-patients show increased levels of TNF in their serum samples [62]. As for the effect of MYC in CD, it has been found that in active CD-patients, the down-expression of c-MYC in patients' epithelium may result in attenuated cell proliferation, which therefore suggest that it could contribute to mucosal ulceration [69].

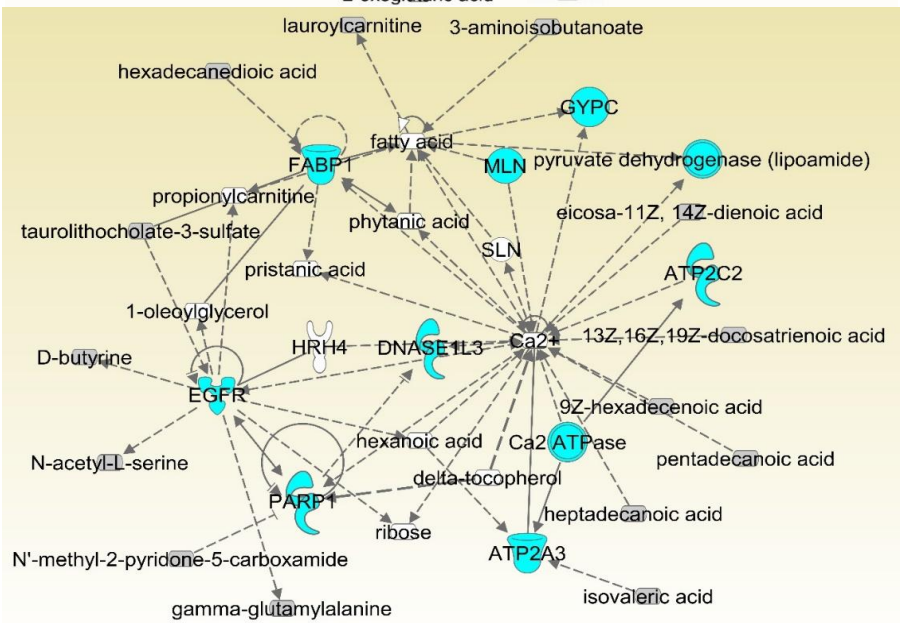
3.2.3 UC networks

The first UC network (Fig. 5i) includes a set of submitted metabolites, indicated in gray, and generated metabolites, in white, which interact with P38 MAPK and ERK1/2, both which activate the Akt pathway. Previous findings suggest a link between P38 MAPK and ERK1/2 to inflammation [60]. Also present in this network is AMPK; AMPK inactivation has been previously found to be increased by insulin absorption and is suggested to be an etiological factor in intestinal dysfunctions [70].

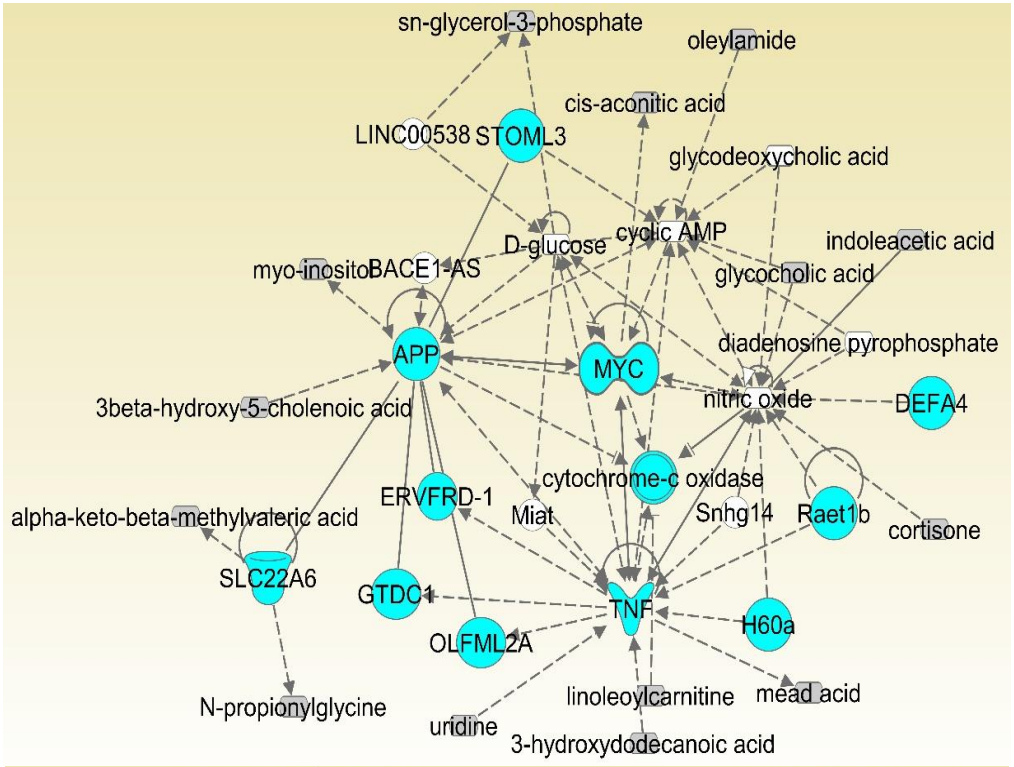
The second UC network (Fig. 5j) includes notable interactions between metabolites and the pro-inflammatory cytokine interleukin IL6, which has growing evidence that it plays a crucial part in intestinal inflammation, a defining characteristic of IBDs [71]. The metabolites also interact with EGFR, which previous research shows is frequently expressed in UC [58].



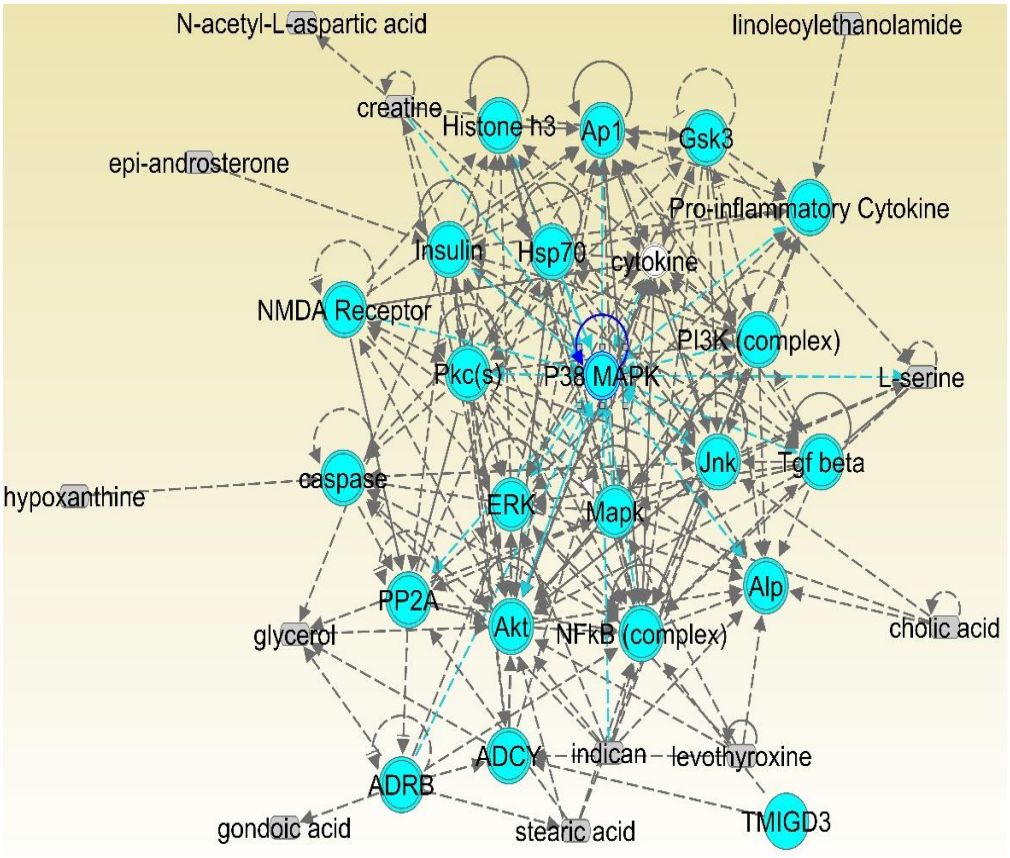
(a)



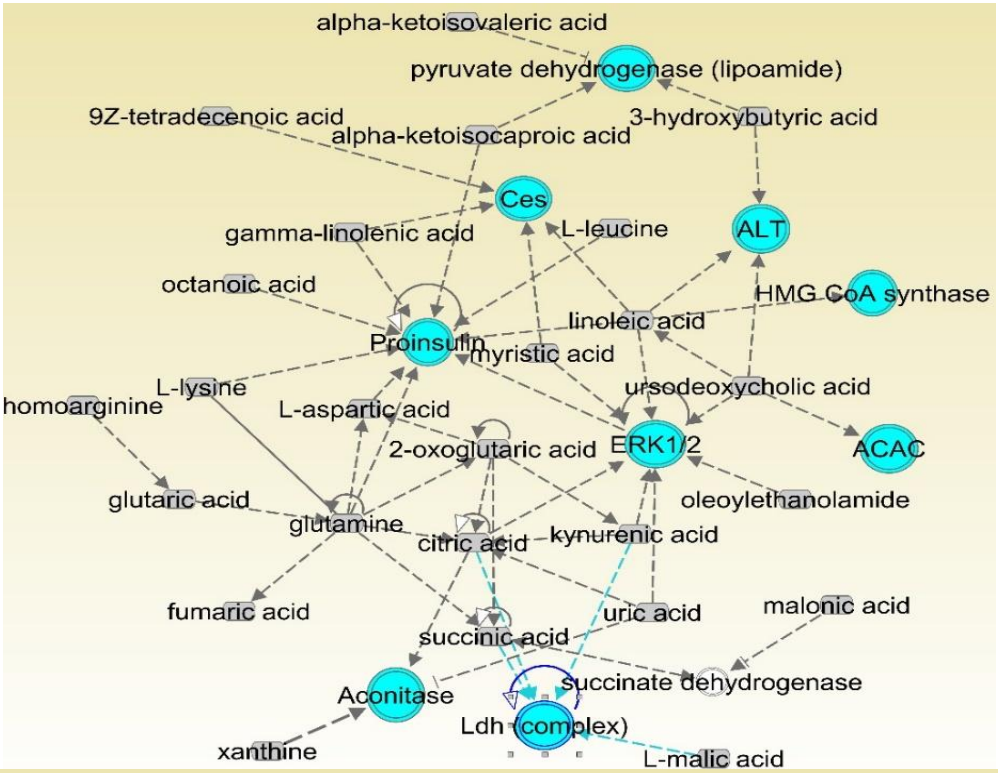
(b)


sn-glycerol-3-phosphate
oleylamide
cis-aconitic acid
LINC00538
STOML3
glycodeoxycholic acid
D-glucose
cyclic AMP
indoleacetic acid
myo-inositol
BACE1-AS
glycocholic acid
APP
MYC
diadenosine pyrophosphate
3beta-hydroxy-5-cholenoic acid
nitric oxide
DEFA4
cytochrome-c oxidase
ERVFRD-1
alpha-keto-beta-methylvaleric acid
Miat
Snhg14
Raet1b
cortisone
SLC22A6
GTDC1
TNF
H60a
OLFML2A
N-propionylglycine
linoleoylcarnitine
mead acid
uridine
3-hydroxydodecanoic acid

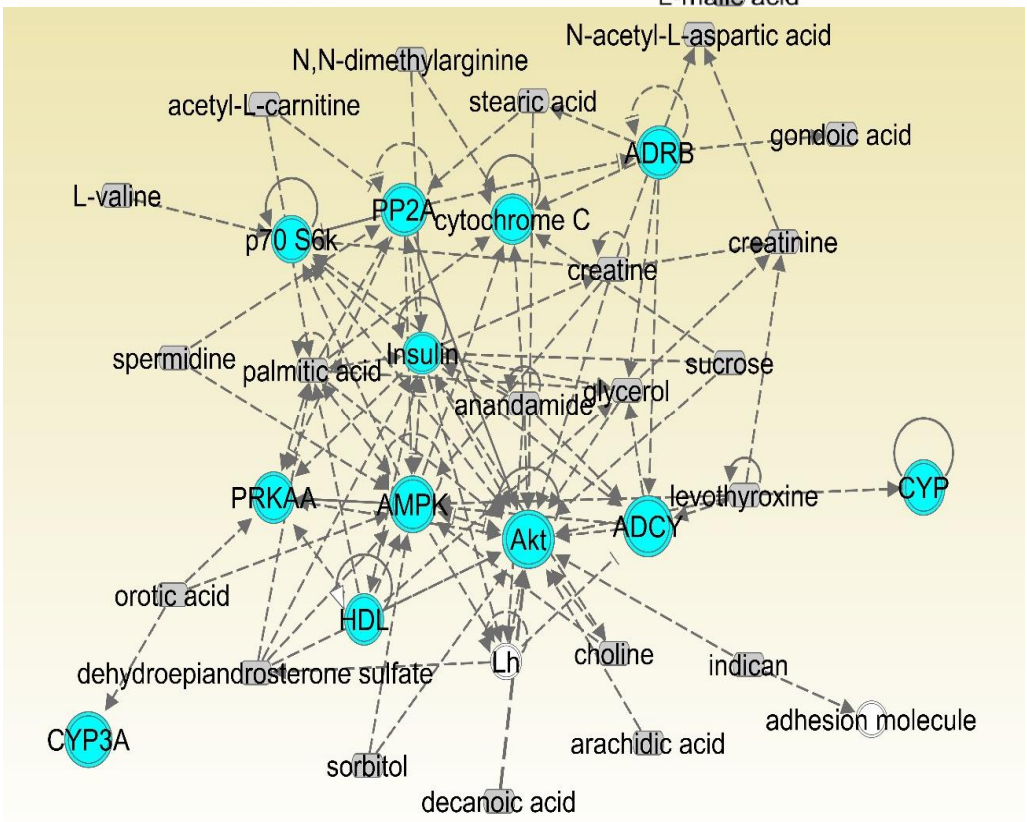
(c)


N-acetyl-L-aspartic acid
linoleoylethanolamide
creatine
Histone h3
Ap1
Gsk3
epi-androsterone
Pro-inflammatory Cytokine
Insulin
Hsp70
cytokine
NMDA Receptor
PI3K (complex)
Pkc(s)
P38 MAPK
L-serine
caspase
Jnk
Tgf beta
hypoxanthine
ERK
Mapk
Alp
PP2A
Akt
NFkB (complex)
glycerol
cholic acid
ADCY
indican
levothyroxine
ADRB
gondoic acid
stearic acid
TMIGD3

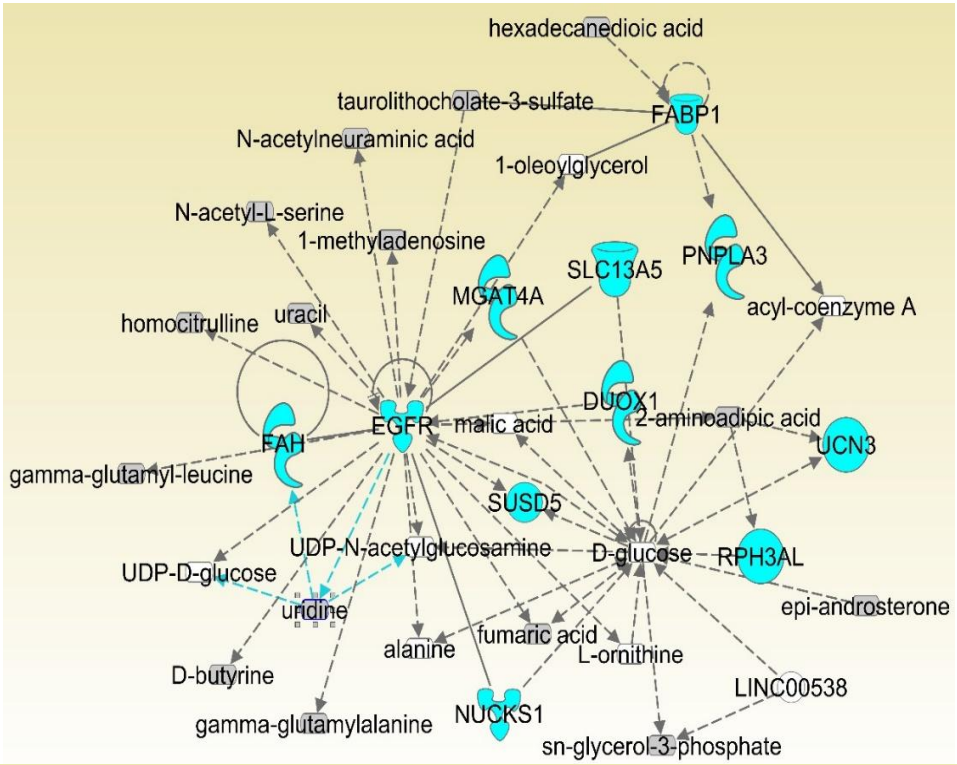
(d)


alpha-ketoisovaleric acid
pyruvate dehydrogenase (lipoamide)
9Z-tetradecenoic acid
alpha-ketoisocaproic acid
3-hydroxybutyric acid
Ces
gamma-linolenic acid
L-leucine
ALT
octanoic acid
HMG CoA synthase
linoleic acid
Proinsulin
myristic acid
L-lysine
homoarginine
L-aspartic acid
ursodeoxycholic acid
2-oxoglutaric acid
ERK1/2
ACAC
glutaric acid
oleoylethanolamide
glutamine
citric acid
kynurenic acid
fumaric acid
uric acid
malonic acid
succinic acid
Aconitase
succinate dehydrogenase
Ldh (complex)
xanthine
L-malic acid

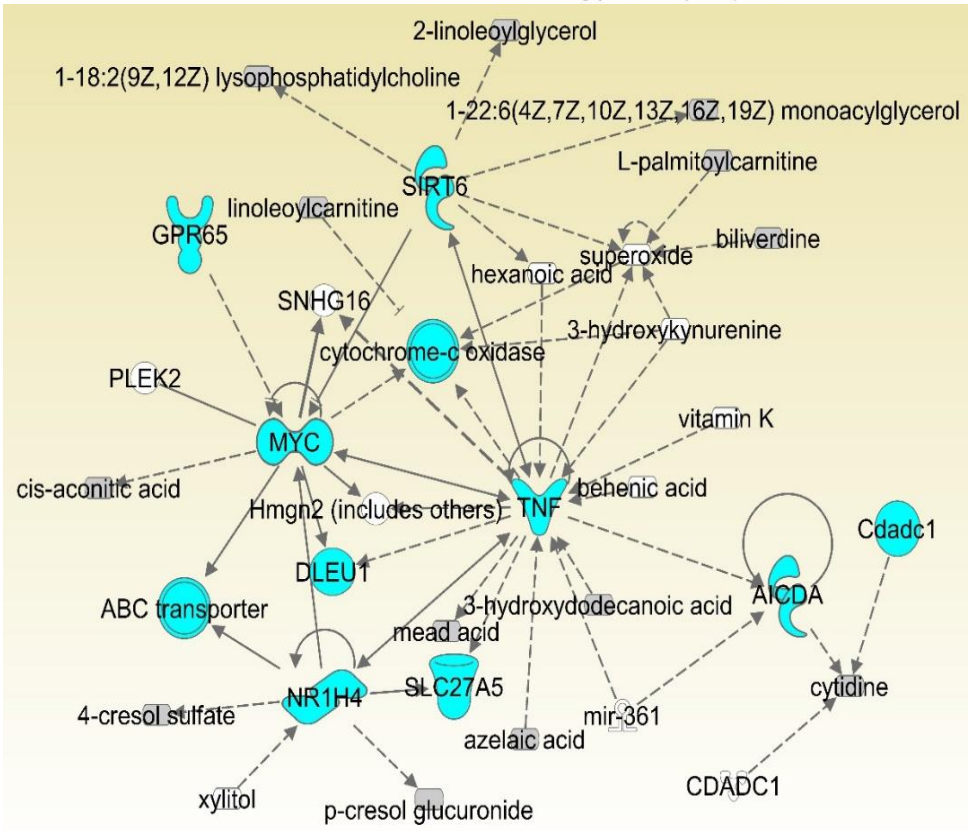
(e)


N-acetyl-L-aspartic acid
N,N-dimethylarginine
acetyl-L-carnitine
stearic acid
ADRB
gondoic acid
L-valine
PP2A
cytochrome C
p70 S6k
creatinine
creatine
spermidine
palmitic acid
Insulin
sucrose
glycerol
anandamide
PRKAA
AMPK
levothyroxine
CYP
ADCY
Akt
orotic acid
HDL
choline
indican
dehydroepiandrosterone sulfate
Lh
adhesion molecule
CYP3A
arachidic acid
sorbitol
decanoic acid

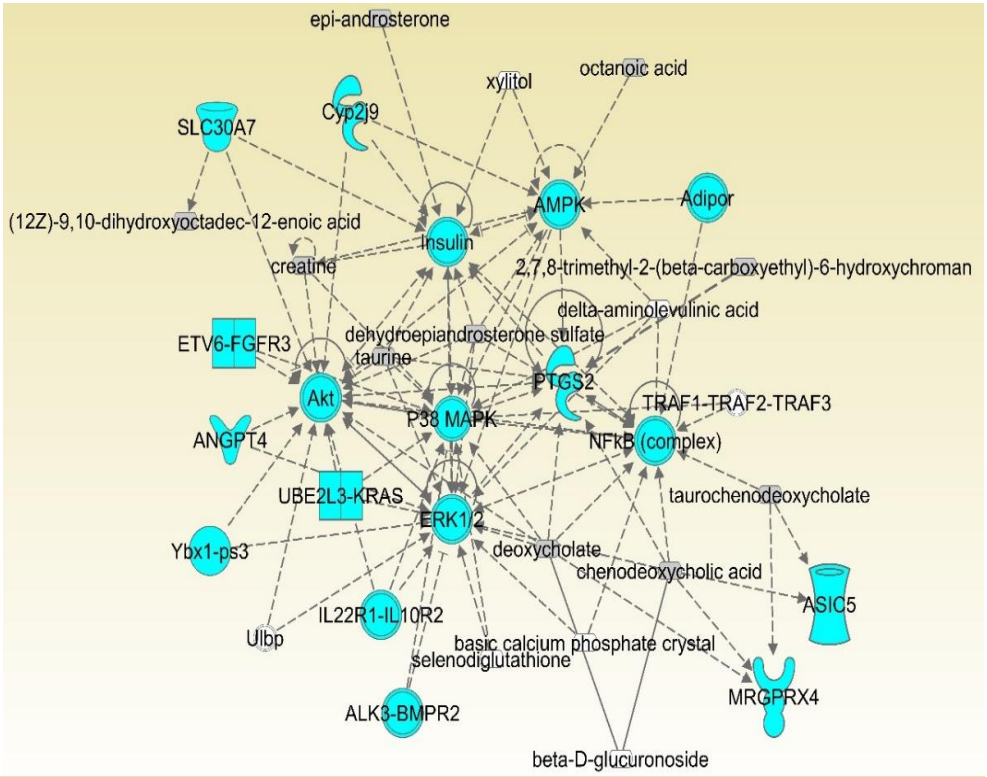
(f)


hexadecanedioic acid
taurolithocholate-3-sulfate
FABP1
N-acetylneuraminic acid
1-oleoylglycerol
N-acetyl-L-serine
1-methyladenosine
PNPLA3
SLC13A5
MGAT4A
acyl-coenzyme A
homocitrulline
uracil
DUOX1
EGFR
malic acid
2-aminoadipic acid
FAH
UCN3
gamma-glutamyl-leucine
SUSD5
UDP-N-acetylglucosamine
D-glucose
RPH3AL
UDP-D-glucose
epi-androsterone
uridine
alanine
fumaric acid
L-ornithine
D-butyrine
LINC00538
NUCKS1
gamma-glutamylalanine
sn-glycerol-3-phosphate

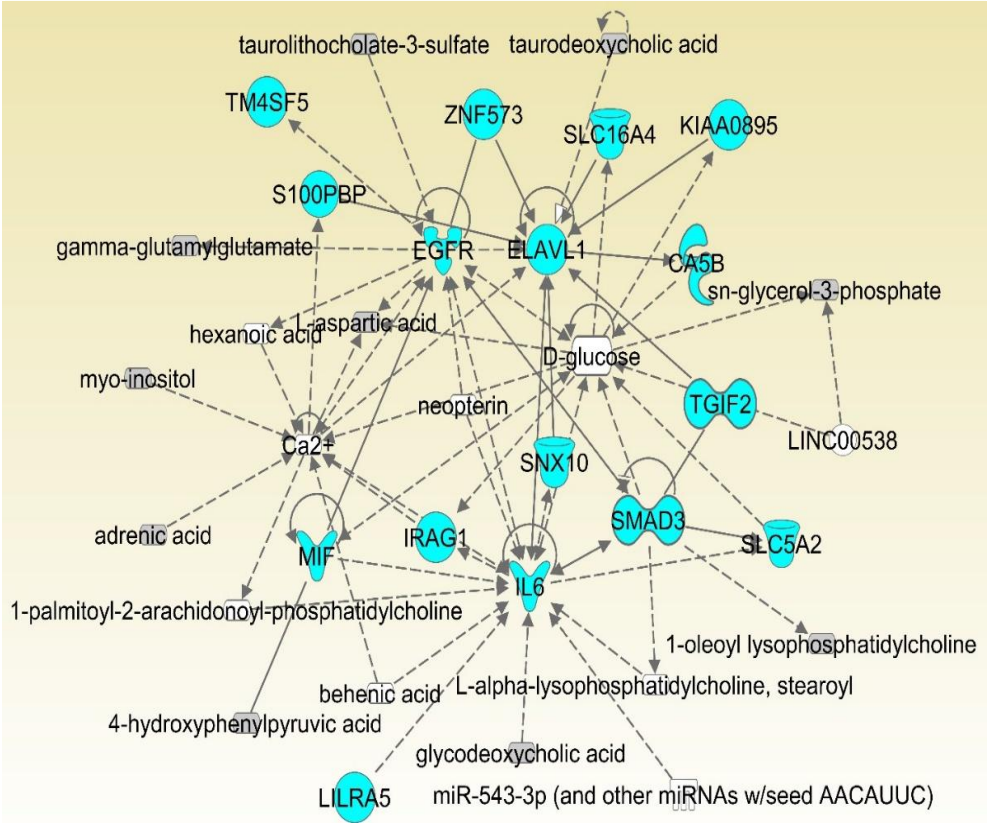
(g)


2-linoleoylglycerol
1-18:2(9Z,12Z) lysophosphatidylcholine
1-22:6(4Z,7Z,10Z,13Z,16Z,19Z) monoacylglycerol
L-palmitoylcarnitine
SIRT6
linoleoylcarnitine
GPR65
biliverdine
superoxide
hexanoic acid
SNHG16
3-hydroxykynurenine
cytochrome-c oxidase
PLEK2
vitamin K
MYC
cis-aconitic acid
behenic acid
Hmgn2 (includes others)
TNF
Cdadc1
DLEU1
ABC transporter
3-hydroxydodecanoic acid
AICDA
mead acid
cytidine
NR1H4
SLC27A5
4-cresol sulfate
mir-361
azelaic acid
xylitol
CDADC1
p-cresol glucuronide

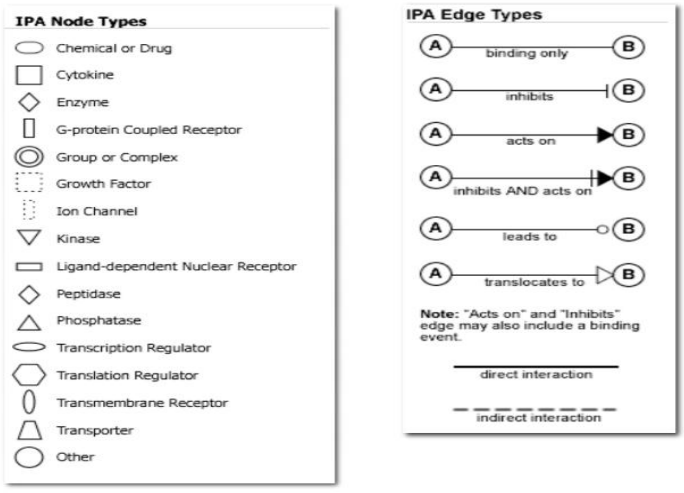
(h)


epi-androsterone
xylitol
octanoic acid
Cyp2j9
SLC30A7
AMPK
Adipor
(12Z)-9,10-dihydroxyoctadec-12-enoic acid
Insulin
creatine
2,7,8-trimethyl-2-(beta-carboxyethyl)-6-hydroxychroman
delta-aminolevulinic acid
ETV6-FGFR3
dehydroepiandrosterone sulfate
taurine
PTGS2
Akt
P38 MAPK
TRAF1-TRAF2-TRAF3
ANGPT4
NFkB (complex)
UBE2L3-KRAS
taurochenodeoxycholate
ERK1/2
Ybx1-ps3
deoxycholate
chenodeoxycholic acid
ASIC5
IL22R1-IL10R2
Ulbp
basic calcium phosphate crystal
selenodiglutathione
MRGPRX4
ALK3-BMPR2
beta-D-glucuronoside

(i)


taurolithocholate-3-sulfate
taurodeoxycholic acid
TM4SF5
ZNF573
SLC16A4
KIAA0895
S100PBP
gamma-glutamylglutamate
EGFR
ELAVL1
CA5B
sn-glycerol-3-phosphate
hexanoic acid
L-aspartic acid
D-glucose
myo-inositol
neopterin
TGIF2
Ca2+
LINC00538
SNX10
SMAD3
adrenic acid
IRAG1
SLC5A2
MIF
IL6
1-palmitoyl-2-arachidonoyl-phosphatidylcholine
1-oleoyl lysophosphatidylcholine
behenic acid
L-alpha-lysophosphatidylcholine, stearoyl
4-hydroxyphenylpyruvic acid
glycodeoxycholic acid
LILRA5
miR-543-3p (and other miRNAs w/seed AACAUUC)

(j)



(k)

Fig. 5. The interactomes of genes and metabolites for uIBD, CD, and UC. (a) First uIBD network. (b) Second uIBD network. (c) Third uIBD network. (d) Fourth uIBD network. (e) First CD network. (f) Second CD network. (g) Third CD network. (h) Fourth CD network. (i) First UC network. (j) Second UC network. (k) IPA Legend [28]. Note: submitted metabolites are colored in gray and generated metabolites are colored white

3.3 Machine-Learning Analysis via MLP Classifier

3.3.1 Combination of three disease-recognition modules

The use of machine-learning techniques to facilitate the elucidation of patterns in biomedical data has been on the rise. Using machine learning (ML) for biomedical purposes has become popular and continues to solve many problems in biomarker-based diagnostics, as well as in drug discovery and therapy. The metabolites analyzed in this study were characterized using a set of PaDEL descriptors in order to create generalized classifiers that would be compatible with a variety of possible datasets derived from different metabolomics platforms [29,30]. The original sets of descriptors were preprocessed with InfoGain filtration in order to concentrate the information gain present in the data, eliminating any noise contributed by redundant or insignificant variance patterns. We filtered out *p*-values greater than −0.05 and fold changes between 0.8 and 1.2, and examined the remaining 140 metabolites for uIBD, 200 for CD, and 37 for UC, each with their sets of filtered PaDEL descriptors. Using WEKA [31,32], we created a set of disease-recognition modules for uIBD, CD, and UC. We explored many different types of ML classification techniques including Naïve Bayes, Logistic, and SGD, but only MLP consistently showed high accuracy on the test results. The results of the tests are as follows.

For the uIBD recognition module, four sets of tests were performed: (1) a 10-fold cross-validation; (2) trials of testing CD metabolites on the trained uIBD module; (3) trials of testing UC metabolites on the trained uIBD module; and (4) testing random Human Metabolite Database (HMDB) [33,34] metabolites on the trained uIBD module. The results are as expected, with the cross-validation test yielding the average accuracy percentage at 77.48%. The results of the other tests and a visual comparison of the four sets of tests are shown in Fig. 6a.

For the CD module, four sets of tests were performed: (1) a 10-fold cross-validation; (2) trials of testing uIBD metabolites on the trained CD module; (3) trials of testing UC metabolites on the trained CD module; and (4) testing random HMDB metabolites on the trained CD module. The results are as expected, with the cross-validation test yielding the highest accuracy percentage at around 80%. The results of the other tests and a visual comparison of the four sets of tests are shown in Fig. 6b.

For the UC module, four sets of tests were performed: (1) a 10-fold cross-validation; (2) trials of testing CD metabolites on the trained UC module; (3) trials of testing uIBD metabolites on the trained UC model; and (4) testing random HMDB metabolites on the trained UC module. The results are as expected, with the cross-validation test yielding the highest accuracy

percentage of the cross validations of uIBD and CD, at 97.30%. The results of the other tests and a visual comparison of the four sets of tests are shown in Fig. 6c.

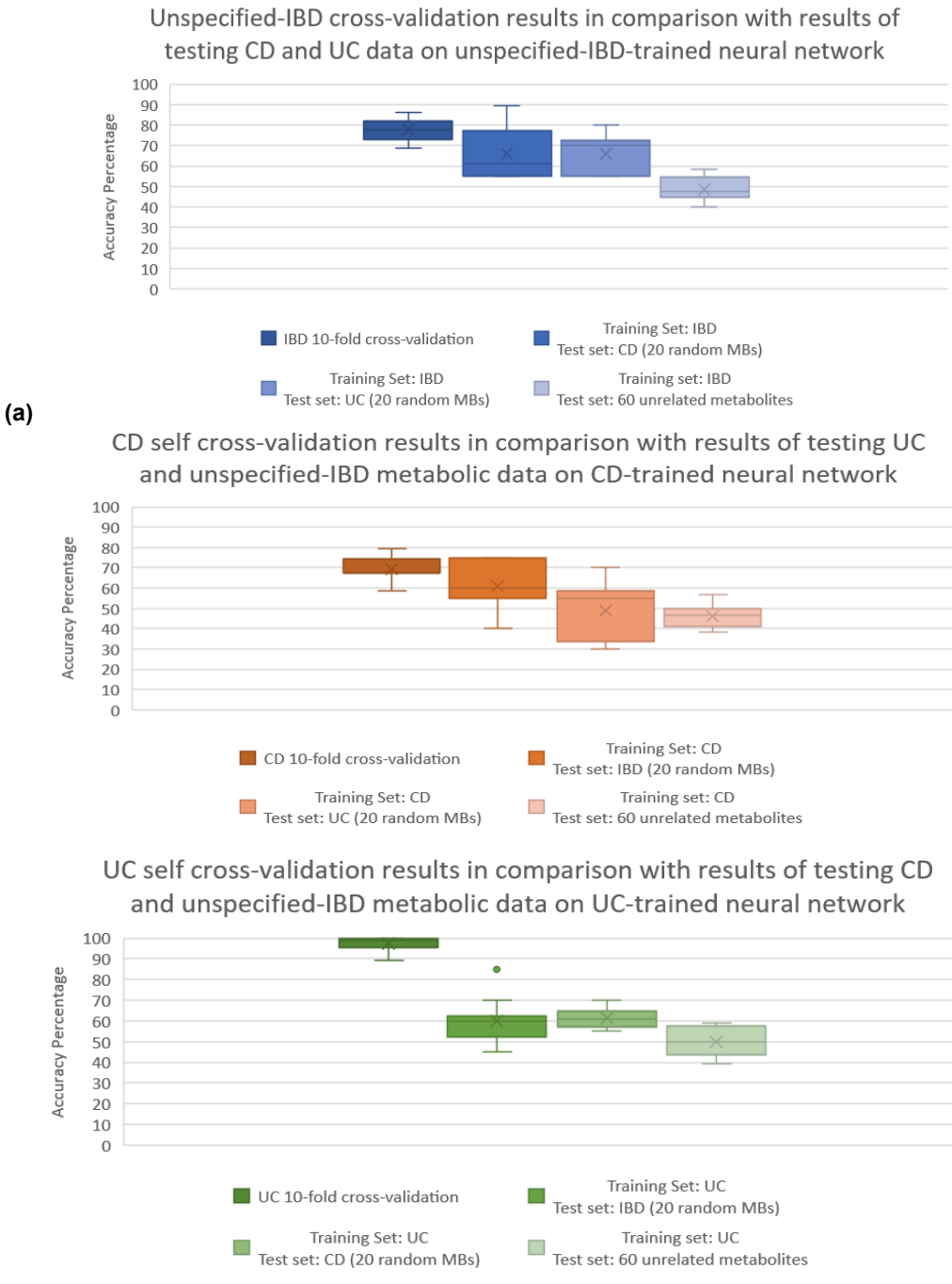
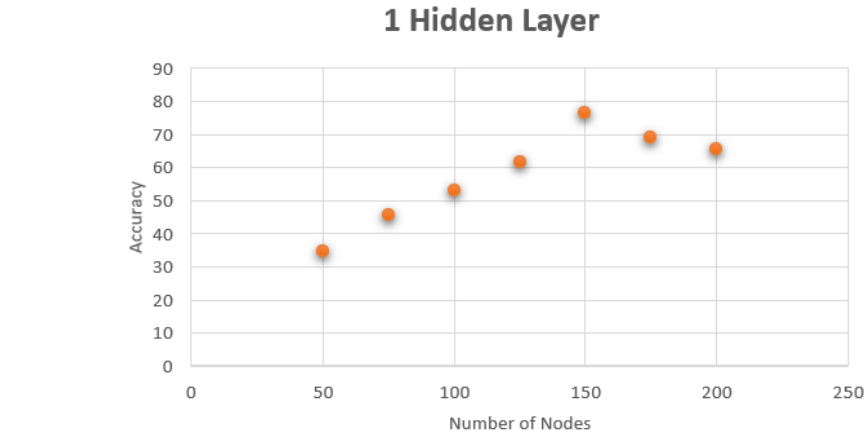


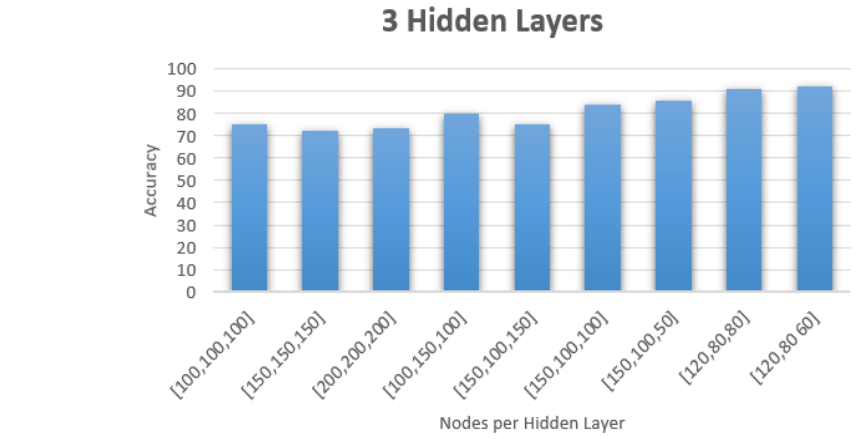
Fig. 6. Performance of each metabolite set examined is visually displayed. (a) A box plot visual for the four sets of tests run for the uIBD dataset. (b) A box plot visual for the four sets of tests run for the CD dataset. (c) A box plot visual for the four sets of tests run for the UC dataset

3.3.2 Multidisease classifier

We then ran a multiclass disease classification to build a model that differentiates between metabolic data of uIBD, CD, UC, and control patients, using a multilayer perceptron. In the first few tests, the accuracy did not go past approximately 78%. To overcome this hurdle, we experimented and found that a certain number of hidden layers and nodes that worked well with this dataset. We began trials with one hidden layer with on-hundred nodes, resulting in around 70% accuracy. Then we ran multiple trials with one hidden layer, varying the number of nodes. Out of the wide range of nodes we tested, 150 nodes yielded the highest accuracy—of 76.57%. We then tested different numbers of hidden layers, each with 100 nodes per layer to begin with. Only three hidden layers proved to be a viable number. We tested the number of nodes per each hidden layer. The options were plenty; the nodes could be in a descending order, ascending order, staggered, etc. After numerous rounds of testing, we found that one-hundred-twenty nodes in the first hidden layer, then eighty in the second hidden layer, then sixty in the third hidden layer proved to be the most well-matched for our dataset. This was indeed proven when the test resulted in 92.17% accuracy in correctly classifying the metabolic data to the respective classes. To test the validity of this system, we tested random HMDB metabolites on the system and the system did not recognize the dataset, as expected. The progress and results of this are showed in Fig. 7.



(a)



(b)

Fig. 7. The progress and end results of the multiclass disease classification or the control and three disease datasets—uIBD, CD, and UC. (a) One-hidden-layer trials in the experimentation process. (b) Three-hidden-layer trials resulting in highest accuracies

To further test the reliability of the multidisease classifier model, we ran additional tests with submissions of varying numbers of metabolites from each disease, all randomly selected using a random number generator. If this multidisease model were truly effective, these additional tests would result in similar percentages to the test with a submission of all the metabolites, which, as stated above, resulted in 92.17% accuracy. Multiple trials of five different numbers of metabolites were conducted. The intervals were five metabolites, ten metabolites, twenty metabolites, fifty metabolites, and one-hundred metabolites from each disease. Each set number of metabolites was tested ten times, with each trial consisting of a new array of metabolites determined by a random number generator. The resulting accuracies from inputting the test sets to the multidisease classifier are as follows. The testing sets with five metabolites resulted in an average of 65.85% accuracy. The testing sets with ten metabolites resulted in an average accuracy of 72.46%. The testing sets with twenty metabolites resulted in an average accuracy of 82.92%. The testing sets with fifty metabolites resulted in an average accuracy of 89.12%. The testing sets with a hundred metabolites resulted in an average accuracy of 87.45%. It is obvious that the best accuracy can be achieved with 50 or more input metabolites set, but quite reasonable accuracy can be achieved even with 20 metabolites set. In real life often, the test sets can have a limited number of metabolites especially if other than mass-spectroscopy methods of analysis are used. Thus, it could be inferred that a testing sets must consist of at least twenty metabolites to yield viable testing accuracies. The results of experiments are illustrated in Fig. 8.

A few supplementary tests were run as well. In order to support that this model is isolated for uIBD, CD, and UC data, and can effectively detect only their presence and not some other disease, we submitted metabolite datasets of other diseases to the multidisease classifier model. The selected diseases were: bladder cancer [72], breast cancer [73], liver disease [74], and Alzheimer's disease [75]. After the input of these diseases' metabolites into the multidisease classifier model, the produced results were as expected. Each testing set was tested five times, and the average accuracies are as follows. The testing set of bladder cancer metabolites resulted in 46.72% average accuracy. The testing set for breast cancer metabolites resulted in an average of 49.46%. Showing slightly higher results, the testing set for liver disease metabolites resulted in an average of 58.30%. Finally, the Alzheimer's disease metabolite test set resulted in average of 45.81%. These results are similar to the tested we conducted with random HMDB metabolites, as displayed in Fig. 6a-c. The results of these tests are depicted in Fig. 9. The low, close-to-chance accuracy percentages of these other disease tests support the efficacy of the multidisease classifier model.

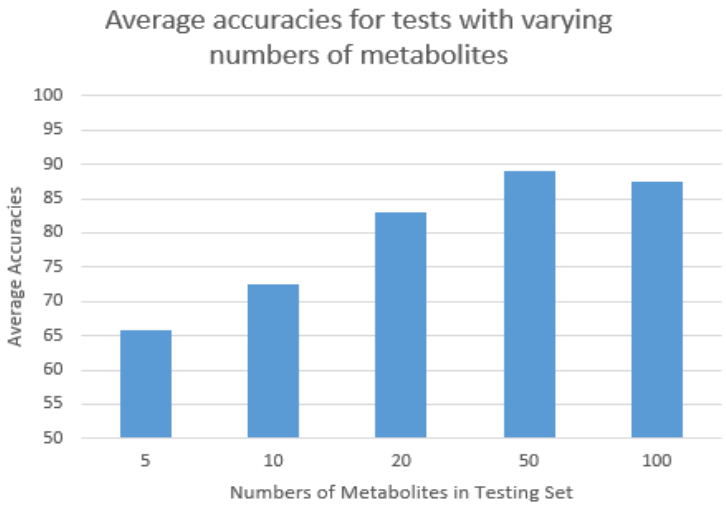


Fig. 8. Average accuracies of testing trials with varying intervals of metabolites

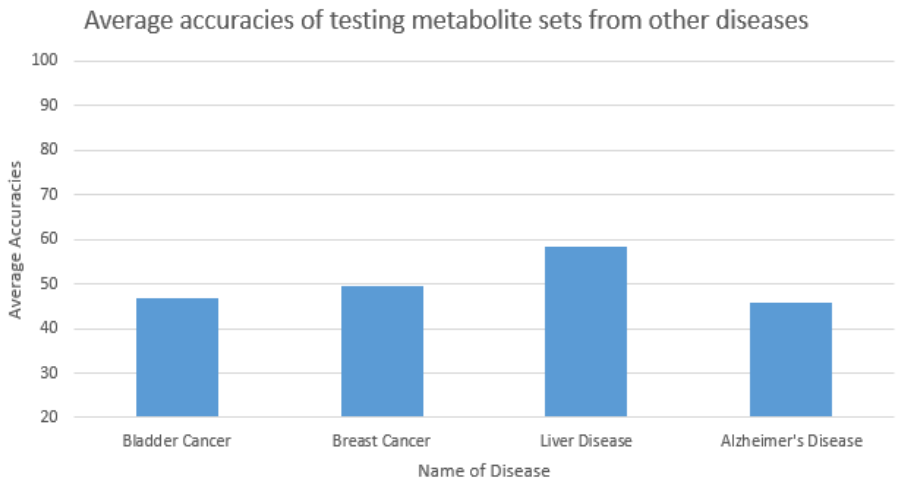


Fig. 9. Graph of the average accuracies from the input of metabolite sets from the four other diseases—bladder cancer, breast cancer, liver disease, and Alzheimer's disease

Table 1. Hard voting scheme. Four of twelve possibilities shown. Other eight were omitted due to repetitiveness

uIBD rec	UC rec	CD rec	Multi disease Classifier	Hard-Voting Consensus
yes	no	no	uIBD	uIBD
yes	no	no	UC	none
yes	no	no	CD	none
no	**yes**	no	uIBD	none
no	**yes**	no	UC	UC
no	**yes**	no	CD	none
no	no	**yes**	uIBD	none
no	no	**yes**	UC	none
no	no	**yes**	CD	CD
yes	**yes**	no	uIBD	uIBD
yes	**yes**	no	UC	UC
yes	**yes**	no	CD	none
no	**yes**	**yes**	uIBD	none
no	**yes**	**yes**	UC	UC
no	**yes**	**yes**	CD	CD
yes	no	**yes**	uIBD	uIBD
yes	no	**yes**	UC	none
yes	no	**yes**	CD	CD
yes	**yes**	**yes**	uIBD	uIBD
yes	**yes**	**yes**	UC	UC
yes	**yes**	**yes**	CD	CD

To add, we found IBD-related metabolites from another public source to test on our models. Specifically, we found CD metabolites from the study done on Crohn's disease metabolites by Jansson and colleagues in 2009 [76]. We used PaDEL Descriptors to get the descriptors for these metabolites and then tested them on our CD recognition model, multi-disease classifier model, and also on the other individual recognition models for good measure. The individual CD model resulted in 96.15%, and the same accuracy resulted for the multi-disease classifier. Thus, this fortifies the viability of these models. As expected, the uIBD and UC individual recognition models did not recognize the CD metabolic data.

The final prediction was decided through majority voting for classification. We used a hard-voting scheme to finalize these results, as a

consensus disease predictor. The hard-voting scheme was applied to each of the three individual disease models—uIBD, CD, and UC recognition models—as well as the multi disease model. We calculated all the possible scenarios of results from these models, and then calculated the majority vote of each of all twelve possibilities. Each possible prediction result—uIBD, CD, UC, and healthy—were able to be predicted through hard voting on the four models. The hard-voting scheme is shown in Table 1.

4. CONCLUSION

Analysis of metabolites can be an effective method in the detection of various diseases. Some significant relationships between the inputted metabolites and metabolic signaling pathways were shown using the metabolomics data, allowing for the determination and exploration of hubs in these networks and opening the opportunity to look for new treatment approaches.

We can create a more robust tool for analysis to discover novel biomarkers and/or compounds to be used for targeted therapy by integrating metabolomic and genomic data. For example, the absence of genes and metabolites in Crohn's disease patients but their presence in ulcerative colitis patients signifies the potential that metabolomics has in the field of disease diagnostics. Our results suggest a large variety of metabolites, genes, and pathways that could be utilized in further studies and possibly lead to the selection of inflammatory bowel disease biomarker sets.

In this study we elucidated metabolites and genes significantly related to the uIBD, CD, and UC. The significant metabolic pathways and the corresponding proteins and metabolites involved are described. Data sets containing descriptors packed with high information content concerning their potential in machine-learning system applications were used, specifically for the diagnosis of uIBD, CD, and UC.

5. LIMITATIONS OF STUDY

Limited amount of metabolite profiles of IBD patients available from public sources thus far.

CONSENT

It is not applicable.

ETHICAL APPROVAL

As per international standard or university standard written ethical approval has been collected and preserved by the authors.

ACKNOWLEDGEMENTS

Study by Scoville and colleagues on serum metabolites related to the IBDs.

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Dahlhamer JM, Zammitti EP, Ward BW, Wheaton AG, Croft JB. Prevalence of inflammatory bowel disease among adults aged ≥18 years — United States; 2015. MMWR Morb Mortal Wkly Rep. 2016;651166–1169. DOI:http://dx.doi.org/10.15585/mmwr.mm6542a3
2. CDC—What is inflammatory bowel disease (IBD)? - Inflammatory Bowel Disease - Division of Population Health. Centers for Disease Control and Prevention. Available:http://cdc.gov/ibd/what-is-IBD.htm Accessed 2 June 2020.
3. Inflammatory bowel disease (IBD). Mayo Clinic. Available:https://www.mayoclinic.org/diseases-conditions/inflammatory-bowel-disease/symptoms-causes/syc-20353315 Accessed 2 June 2020.
4. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M et al. Diversity of the human intestinal microbial flora. Science 2005:308;1635–1638. DOI: 10.1126/science.1110591
5. Thursby E, Juge N. Introduction to the human gut microbiota. Biochem J 2017, 474(11), 1823–1836. DOI:10.1042/BCJ20160510
6. Hooper LV. Bacterial contributions to mammalian gut development. Trends Microbiol 2004:12:129–134. DOI: 10.1016/j.tim. 2004.01. 001
7. Chow J, Lee SM, Shen Y, Khosravi A, Mazmanian SK. Host-bacterial symbiosis in health and disease. Adv Immunol 2010:107:243–274.

DOI: 10.1016/B978-0-12-381300-8.00008-3

8. Groen RN, de Clercq NC, Nieuwdorp M, Hoenders H, Groen AK. Gut microbiota, metabolism and psychopathology: A critical review and novel perspectives. Crit Rev Clin Lab Sci 2018, 55, 283–293.
DOI: 10.1080/10408363.2018.1463507
9. Storr M, Vogel HJ, Schicho R. Metabolomics: Is it useful for inflammatory bowel diseases? Curr Opin Gastroenterol 2013;29:378–383
DOI: 10.1097/MOG.0b013e328361f488
10. Hirano A, Umeno J, Okamoto Y, Shibata, H, Ogura Y, Moriyama T et al. Comparison of the microbial community structure between inflamed and non-inflamed sites in patients with ulcerative colitis. J Gastroenterol Hepatol 2018;33:1590–1597.
DOI: 10.1111/jgh.14129
11. Gong D, Gong X, Wang L, Yu X, Dong Q. Involvement of reduced microbial diversity in inflammatory bowel disease. Gastroenterol Res Pract. 2016;6951091.
DOI: 10.1155/2016/6951091
12. Eckburg PB, Bik EM, Bernstein CN, Purdom E, Dethlefsen L, Sargent M et al. Diversity of the human intestinal microbial flora. Science 2005, 308, 1635–1638.
DOI: 10.1126/science.111059124
13. Rinninella E, Raoul P, Cintoni M, Franceschi F, Migziano GAD, Gasbarrini A, Mele MC. What is the healthy gut microbiota composition? A changing ecosystem across age, environment, diet, and diseases. Microorganisms 2019;7(1):14.
DOI: 10.3390/microorganisms7010014
14. Binda C, Lopetuso LR, Rizzatti G, Gibiino G, Cennamo V, Gasbarrini A. Actinobacteria: A relevant minority for the maintenance of gut homeostasis. Dig Liver Dis. 2018;50(5):421–428.
DOI:10.1016/j.dld.2018.02.012
15. Bamola VD, Ghosh A, Kapardar RK, Banwari L, Cheema S, Sarma P, Chaudhry R. Gut microbial diversity in health and disease: Experience of healthy Indian subjects, and colon carcinoma and inflammatory bowel disease patients. Microb Ecol Health Dis. 2017;28 (1):1322447.
DOI:10.1080/16512235.2017.1322447
16. Tedjo DI, Smolinska A, Savelkoul PH, Masclee AA, van Schooten FJ, Pierik MJ, et al. The fecal microbiota as a biomarker for disease activity in Crohn's disease. Sci Rep. 2016;6(1):35216.
DOI:10.1038/srep35216
17. Casén C, Vebø HC, Sekelja M, Hegge FT, Karlsson MK, Ciemniejewska E et al. Deviations in human gut microbiota: A novel diagnostic test for determining dysbiosis in patients with IBS or IBD. Aliment Pharmacol Ther. 2015;42: 71–83.
DOI: 10.1111/apt.13236
18. Yu Z, Kastenmüller G, He Y, Belcredi P, Möller G, Prehn C et al. Differences between human plasma and serum metabolite profiles. PloS One. 2011;6(7): e21230.
DOI: 10.1371/journal.pone.0021230
19. Scoville EA, Allaman MM, Brown CT, Motley AK, Horst SN, Williams CS et al. Alterations in lipid, amino acid, and energy metabolism distinguish Crohn's disease from ulcerative colitis and control subjects by serum metabolomic profiling. Metabolomics. 2018;14(1):17.
DOI: 10.1007/s11306-017-1311-y
20. Lavelle A, Sokol H. Gut microbiota-derived metabolites as key actors in inflammatory bowel disease. Nat Rev Gastroenterol Hepatol. 2020;17:223–237.
DOI: 10.1038/s41575-019-0258-z
21. Dong LN, Wang M, Guo J, Wang JP. Role of intestinal microbiota and metabolites in inflammatory bowel disease. Chin Med J (Engl). 2019;132(13):1610–1614.
DOI: 10.1097/CM9.0000000000000290
22. Lai Y, Xue J, Liu CW, Gao B, Chi L, Tu P et al. Serum metabolomics identifies altered bioenergetics, signaling cascades in parallel with exposome markers in Crohn's disease. Molecules. 2019;24(3): 449.
DOI: 10.3390/molecules24030449
23. Dawiskiba T, Deja S, Mulak A, Ząbek A, Jawień E, Pawełka D et al. Serum and urine metabolomic fingerprinting in diagnostics of inflammatory bowel diseases. World J Gastroenterol. 2014; 20(1):163–174.
DOI: 10.3748/wjg.v20.i1.163
24. Daniluk, U, Daniluk J, Kucharski R, Kowalczyk T, Pietrowska K, Samczuk P et al. Untargeted metabolomics and inflammatory markers profiling in children with Crohn's disease and ulcerative colitis—A preliminary study. Inflamm Bowel Dis. 2019;25(7):1120–1128.
DOI: 10.1093/ibd/izy402

25. Chong J, Wishart DS, Xia J. Using MetaboAnalyst 4.0 for comprehensive and integrative metabolomics data analysis. Curr Protoc Bioinformatics. 2019;68(1): e86-128.
DOI: 10.1002/cpbi.86
26. MetaboAnalyst—statistical, functional, and integrative analysis of metabolomics data. Available:http://www.metaboanalyst.ca/
Accessed on 15 September 2020.
27. Krämer A, Green J, Pollard J. Tugendreich S Jr. Causal analysis approaches in Ingenuity Pathway Analysis. Bioinformatics 2014;30(4);523–530.
DOI: 10.1093/bioinformatics/btt703
28. QIAGEN Ingenuity Pathway Analysis (QIAGEN IPA).
Available:https://digitalinsights.qiagen.com/products-overview/discovery-insights-portfolio/analysis-and-visualization/qiagen-ipa/
Accessed 17 September 2020.
29. Yap CW. PaDEL-Descriptor: An open source software to calculate molecular descriptors and fingerprints. J Comput Chem. 2011;32(7):1466–1474.
DOI: 10.1002/jcc.21707
30. PaDEL-Descriptor.
Available:http://www.yapcwsoft.com/dd/padeldescriptor/
Accessed 18 September 2020.
31. Frank E, Hall MA, Witten IH. The WEKA Workbench. In: Data mining: Practical machine learning tools and techniques. 4th ed., Witten IH, Frank E, Hall MA, Pal CJ, eds. Morgan kaufmann: Cambridge mass, USA. 2016;553–572.
32. WEKA: The workbench for machine learning.
Available:https://www.cs.waikato.ac.nz/ml/weka/index.html
Accessed 18 September 2020.
33. Wishart DS, Feunang YD, Marcu A, Guo, AC, Liang K et al. HMDB 4.0—The Human Metabolome Database for 2018. Nucleic Acids Res. 2018;46(D1):D608–D617.
DOI: 10.1093/nar/gkx1089
34. HMDB—The Human Metabolite Database. Welcome to HMDB Version 4.0.
Available:https://hmdb.ca/
Accessed 18 September 2020
35. KO (KEGG ORTHOLOGY) Database.
Available:https://www.genome.jp/kegg/ko.html
Accessed 18 September 2020
36. He F, Wu C, Li P, Li N, Zhang D, Zhu Q et al. Functions and Signaling pathways of amino acids in intestinal inflammation. Biomed Res Int. 2018;9171905.
DOI: 10.1155/2018/9171905.9171905
37. Abulizi N, Quin C, Brown K, Chan YK, Gill SK, Gibson DL. Gut mucosal proteins and bacteriome are shaped by the saturation index of dietary lipids. Nutrients 2019;11(2):418.
DOI: 10.3390/nu11020418
38. Fagbemi A, Newman WG, Tangye SG, Hughes SM, Cheesman E, Arkwright PD. Refractory very early-onset inflammatory bowel disease associated with cytosolic isoleucyl–tRNA synthetase deficiency: A case report. World J Gastroenterol. 2020, 26(15):1841–1846.
DOI: 10.3748/wjg.v26.i15.1841
39. Morgan XC, Tickle TL, Sokol H, Gevers D, Devaney KL, Ward DV et al. Dysfunction of the intestinal microbiome in inflammatory bowel disease and treatment. Genome Biol. 2012;13(9):R79.
DOI:10.1186/gb-2012-13-9-r79
40. Chiba T, Suzuki K, Matsumoto T. Plasma-free amino acid profiles in Crohn's disease: Relationship with the Crohn disease activity index. Clin Med Insights Gastroenterol.2018;11:117955221879 1173.
DOI: 10.1177/1179552218791173
41. Weiser M, Simon JM, Kochar B, Tovar A, Israel JW, Robinson A et al. Molecular classification of Crohn's disease reveals two clinically relevant subtypes. Gut 2018;67(1):36–42.
DOI: 10.1136/gutjnl-2016-312518
42. Liu R, Qiao S, Shen W, Lin Z, Guo Z, Ge Y et al. Disturbance of fatty acid desaturation mediated by FADS2 in mesenteric adipocytes contributes to chronic inflammation of Crohn's disease. J Crohns Colitis. 2020;jjaa086.
DOI: 10.1093/ecco-jcc/jjaa086
43. Diab J, Hansen T, Goll R, Stenlund H, Jensen E, Moritz T et al. Mucosal metabolomic profiling and pathway analysis reveal the metabolic signature of Ulcerative Colitis. Metabolites. 2019;9(12): 291.
DOI: 10.3390/metabo9120291
44. He Q, Gao Y, Jie Z, Yu X, Laursen JM, Xiao L et al. Two distinct metacommunities characterize the gut microbiota in Crohn's disease patients. Gigascience. 2017;6(7): 1–11.
DOI: 10.1093/gigascience/gix050

45. Wang D, Ma X, Guo S, Wang Y, Li T, Zou D00 et al. Effect of Huangqin Tang on urine metabolic profile in rats with ulcerative colitis based on UPLC-Q-Exactive Orbitrap MS. Evid Based Complement Alternat Med. 2020;1874065. DOI: 10.1155/2020/1874065
46. Filimoniuk A, Daniluk U, Samczuk P, Wasilewska N, Jakimiec P, Kucharska M et al. Metabolomic profiling in children with inflammatory bowel disease. Adv Med Sci 2020;65(1):65–70.
DOI: 10.1016/j.advms.2019.12.009
47. Coburn LA, Horst SN, Allaman MM, Brown, CT, Williams CS, Hodges ME, et al. L-arginine availability and metabolism is altered in ulcerative colitis. Inflamm Bowel Dis. 2016;22(8):1847–1858.
DOI: 10.1097/MIB.0000000000000790
48. Gu X, Song Y, Chai Y, Lu F, Gonzalez FJ, Fan G, Qi Y. GC-MS metabolomics on PPARα-dependent exacerbation of colitis. Mol Biosyst 2015;11(5):1329–1337.
DOI: 10.1039/c5mb00048c
49. Zhang X, Deeke SA, Ning Z, Starr AE, Butcher J, Li J et al. Metaproteomics reveals associations between microbiome and intestinal extracellular vesicle proteins in pediatric inflammatory bowel disease. Nat Commun. 2018;9(1):2873.
DOI: 10.1038/s41467-018-05357-4
50. Kolho KL, Pessia A, Jaakkola T, de Vos W. MD, Velagapudi V. Faecal and serum metabolomics in paediatric inflammatory bowel disease. J Crohns Colitis. 2017; 11(3):321–334.
DOI: 10.1093/ecco-jcc/jjw158
51. Holmes EW, Yong SL, Eiznhamer D, Keshavarzian A. Glutathione content of colonic mucosa: Evidence for oxidative damage in active ulcerative colitis. Dig Dis Sci 1998;43(5):1088–1095.
DOI:10.1023/a:1018899222258
52. Miettinen TA. The role of bile salts in diarrhoea of patients with ulcerative colitis. Gut. 1971;12(8):632–635.
DOI: 10.1136/gut.12.8.632
53. Schicho R, Shaykhutdinov R, Ngo J, Nazyrova A, Schneider C, Panaccione R et al. Quantitative metabolomic profiling of serum, plasma, and urine by(1)H NMR spectroscopy discriminates between patients with inflammatory bowel disease and healthy individuals. J Proteome Res. 2012;11(6):3344–3357.
DOI: 10.1021/pr300139q
54. Nikolaus S, Schulte B, Al-Massad N, Thieme F, Schulte DM, Bethge J et al. Increased Tryptophan metabolism is associated with activity of inflammatory bowel diseases. Gastroenterology 2017;153(6):1504–1516.e2.
DOI: 10.1053/j.gastro.2017.08.028
55. Waetzig GH, Seegert D, Rosenstiel P, Nikolaus S, Schreiber S. p38 mitogen-activated protein kinase is activated and linked to TNF-alpha signaling in inflammatory bowel disease. J Immunol. 2002;168(10):5342–5351.
DOI: 10.4049/jimmunol.168.10.5342
56. Sun X, Yang Q, Rogers C, Du M, Zhu MJ. AMPK improves gut epithelial differentiation and barrier function via regulating Cdx2 expression. Cell Death Differ 2017;24(5):819–831.
DOI: 10.1038/cdd.2017.14
57. Shanahan F. The gut microbiota—A clinical perspective on lessons learned. Nat Rev Gastroenterol Hepatol. 2012;9(10):609–614.
DOI: 10.1038/nrgastro.2012.145
58. Svrcek M, Cosnes, J, Tiret E, Bennis M, Parc Y, Fléjou JF. Expression of epidermal growth factor receptor (EGFR) is frequent in inflammatory bowel disease (IBD)-associated intestinal cancer. Virchows Arch. 2007;450(2):243–244.
DOI: 10.1007/s00428-006-0338-7
59. Zingarelli B, Hake PW, Burroughs TJ, Piraino G, O'Connor M, Denenberg A. Activator protein-1 signalling pathway and apoptosis are modulated by poly(ADP-ribose) polymerase-1 in experimental colitis. Immunology. 2004;113(4):509–517.
DOI: 10.1111/j.1365-2567.2004.01991.x
60. Karin M. The regulation of AP-1 activity by mitogen-activated protein kinases. J Biol Chem. 1995;270(28):16483–16486.
DOI:10.1074/jbc.270.28.16483
61. Puig KL, Lutz BM, Urquhart SA, Rebel AA, Zhou X, Manocha GD et al. Overexpression of mutant amyloid-β protein precursor and presenilin 1 modulates enteric nervous system. J Alzheimers Dis. 2015;44(4):1263–1278.
DOI: 10.3233/JAD-142259
62. Perše M, Unkovič A. The role of TNF in the pathogenesis of inflammatory bowel disease. In: Biological Therapy for Inflammatory Bowel Disease. Leal RF, Torriani T, eds.

Available:https://www.intechopen.com/books/biological-therapy-for-inflammatory-bowel-disease/the-role-of-tnf-in-the-pathogenesis-of-inflammatory-bowel-disease
Accessed 18 September 2020.
DOI: /10.5772/intechopen.84375

63. Dahan S, Roda G, Pinn D, Roth-Walter F, Kamalu O, Martin AP, Mayer L. Epithelial: Lamina propria lymphocyte interactions promote epithelial cell differentiation. Gastroenterology. 2008;134(1):192–203.
DOI: 10.1053/j.gastro.2007.10.022
64. Roy PK, Rashid F, Bragg J, Ibdah JA. Role of the JNK signal transduction pathway in inflammatory bowel disease. World J Gastroenterol. 2008;14(2):200–202.
DOI: 10.3748/wjg.14.200
65. Broom OJ, Widjaya B, Troelsen J, Olsen J, Nielsen OH. Mitogen activated protein kinases: A role in inflammatory bowel disease? Clin Exp Immunol 2009;158(3):272–280.
DOI: 10.1111/j.1365-2249.2009.04033.x
66. Ferriero R, Nusco E, De Cegli R, Carissimo A, Manco G, Brunetti-Pierri N. Pyruvate dehydrogenase complex and lactate dehydrogenase are targets for therapy of acute liver failure. J Hepatol. 2018;69(2):325–335.
DOI: 10.1016/j.jhep.2018.03.016
67. Hooper KP, Barlow PG, Stevens C, Henderson P. Inflammatory bowel disease drugs: A focus on autophagy. J Crohns Colitis. 2017;11(1):118–127.
DOI: 10.1093/ecco-jcc/jjw127
68. Bregenzer N, Hartmann A, Strauch U, Schölmerich J, Andus T, Bollheimer LC. Increased insulin resistance and beta cell activity in patients with Crohn's disease. Inflamm Bowel Dis. 2006;12(1):53–56.
DOI:10.1097/01.mib.0000195975.97673.f5
69. Sipos F, Firneisz G, Műzes G. Therapeutic aspects of c-MYC signaling in inflammatory and cancerous colonic diseases. World J Gastroenterol. 2016:22(35):7938–7950.
DOI: 10.3748/wjg.v22.i35.7938
70. Sun X, Zhu MJ. AMP-activated protein kinase: A therapeutic target in intestinal diseases. Open Biol. 2017;7(8):170 104.
DOI: 10.1098/rsob.170104
71. Atreya R, Neurath MF. Involvement of IL-6 in the pathogenesis of inflammatory bowel disease and colon cancer. Clin Rev Allergy Immunol. 2005;28(3):187–196.
DOI: 10.1385/CRIAI:28:3:187
72. McDunn JE, Perichon R, Neri B, Wittmann B. Biomarkers for bladder cancer and methods using the same. US Patent. Pub. No.: US 2015/0065366 A1, 3/5/2015.
Available:https://patentimages.storage.googleapis.com/42/d3/37/b454c930e66f77/US201 50065366A1.pdf
Accessed 8 August 2020
73. His M, Viallon V, Dossus L, Gicquiau A, Achaintre D, Scalbert A et al. Prospective analysis of circulating metabolites and breast cancer in EPIC. BMC Med. 2019;17:178.
DOI: 10.1186/s12916-019-1408-4
74. Shahzad K, Lopreiato V, Liang Y, Trevisi E, Osorio JS, Xu C, Loor JJ. Hepatic metabolomics and transcriptomics to study susceptibility to ketosis in response to prepartal nutritional management. J Anim Sci Biotechnol. 2019;10:96.
DOI:10.1186/s40104-019-0404-z
75. Czech C, Berndt P, Busch K, Schmitz O, Wiemer J, Most V et al. Metabolite profiling of Alzheimer's disease cerebrospinal fluid. PloS One. 2012;7(2):e31501.
DOI: /10.1371/journal.pone.0031501
76. Jansson J, Willing B, Lucio M, Fekete A, Dicksved J, Halfvarson J et al. Metabolomics reveals metabolic biomarkers of Crohn's disease. PloS One 2009;4(7):e6386.
DOI: 10.1371/journal.pone.0006386

Peer-review history:
The peer review history for this paper can be accessed here:
http://www.sdiarticle4.com/review-history/65206