

RESEARCH ARTICLE

Concordance between gene expression in peripheral whole blood and colonic tissue in children with inflammatory bowel disease

Nathan P. Palmer¹, Jocelyn A. Silvester²*, Jessica J. Lee², Andrew L. Beam¹, Inbar Fried¹, Vladimir I. Valtchinov^{1,3}, Fedik Rahimov⁴, Sek Won Kong⁵, Saum Ghodoussipour², Helen C. Hood², Athos Bousvaros², Richard J. Grand², Louis M. Kunkel⁴, Isaac S. Kohane¹

1 Department of Biomedical Informatics, Harvard Medical School, Boston, Massachusetts, United States of America, **2** Division of Gastroenterology and Nutrition, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America, **3** Center for Evidence Based Imaging, Brigham and Women's Hospital, Harvard Medical School, Massachusetts, United States of America, **4** Division of Genetics and Genomics, Boston Children's Hospital, Departments of Genetics and Pediatrics, Harvard Medical School, Boston, Massachusetts, United States of America, **5** Computational Health Informatics Program, Boston Children's Hospital, Harvard Medical School, Boston, Massachusetts, United States of America

* These authors contributed equally to this work.

* Jocelyn.silvester@childrens.harvard.edu



OPEN ACCESS

Citation: Palmer NP, Silvester JA, Lee JJ, Beam AL, Fried I, Valtchinov VI, et al. (2019) Concordance between gene expression in peripheral whole blood and colonic tissue in children with inflammatory bowel disease. PLoS ONE 14(10): e0222952. <https://doi.org/10.1371/journal.pone.0222952>

Editor: Nikhil Pai, McMaster University, CANADA

Received: April 10, 2019

Accepted: September 10, 2019

Published: October 16, 2019

Copyright: This is an open access article, free of all copyright, and may be freely reproduced, distributed, transmitted, modified, built upon, or otherwise used by anyone for any lawful purpose. The work is made available under the [Creative Commons CC0](https://creativecommons.org/licenses/by/4.0/) public domain dedication.

Data Availability Statement: All gene expression data are available in GEO under accession number GSE126124.

Funding: This work was supported by the National Institute of Diabetes and Digestive and Kidney Diseases (US) (T32 DK 07760 to JAS, K23 DK 119584 to JAS, MERIT Award R37 DK 32658 to R.J.G., R01 DK 061382 to R.J.G., and P30 DK 34854), the National Institutes of Health (MO1-RR002172, UL1 RR025758, UL1 TR001102, P30 HD 18655), Crohn's and Colitis Foundation of America (Career

Abstract

Background

Presenting features of inflammatory bowel disease (IBD) are non-specific. We hypothesized that mRNA profiles could (1) identify genes and pathways involved in disease pathogenesis; (2) identify a molecular signature that differentiates IBD from other conditions; (3) provide insight into systemic and colon-specific dysregulation through study of the concordance of the gene expression.

Methods

Children (8–18 years) were prospectively recruited at the time of diagnostic colonoscopy for possible IBD. We used transcriptome-wide mRNA profiling to study gene expression in colon biopsies and paired whole blood samples. Using blood mRNA measurements, we fit a regression model for disease state prediction that was validated in an independent test set of adult subjects (GSE3365).

Results

Ninety-eight children were recruited [39 Crohn's disease, 18 ulcerative colitis, 2 IBDU, 39 non-IBD]. There were 1,118 significantly differentially (IBD vs non-IBD) expressed genes in colon tissue, and 880 in blood. The direction of relative change in expression was concordant for 106/112 genes differentially expressed in both tissue types. The regression model from the blood mRNA measurements distinguished IBD vs non-IBD disease status in the

Development Award to J.J.L.), and Thrasher Research Fund (Early career award to J.J.L.). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: I have read the journal's policy and the authors of this manuscript have the following competing interests: AB has received research support from Prometheus, Janssen (ADAPT), and Abbvie (Envision), served on a data and safety monitoring board for Shire, royalties from Up-to-date, honoraria from Harvard University, CE outcomes, Alexion, Boston University, and Nutricia, performed medicolegal services for Fowler-White-Burnett and Cooney Scully Dowling and consulted for Best Doctors and Grand Rounds. RJG has received royalties from UpToDate. ISK has received speaking honoraria Takeda Pharmaceuticals. JAS has consulted for Takeda Pharmaceuticals and received research support from Biomedal S.L., Glutenostics LLC and Cour Pharmaceuticals. Nothing to disclose: ALB, IF, SG, HCH, LMK, SWK, J.J.L., NPP, FR, VIV.

independent test set with 80% accuracy using only 6 genes. The overlap of 5 immune and metabolic pathways in the two tissue types was significant ($p < 0.001$).

Conclusions

Blood and colon tissue from patients with IBD share a common transcriptional profile dominated by immune and metabolic pathways. Our results suggest that peripheral blood expression levels of as few as 6 genes (*IL7R*, *UBB*, *TXNIP*, *S100A8*, *ALAS2*, and *SLC2A3*) may distinguish patients with IBD from non-IBD.

Introduction

Crohn's disease (CD) and ulcerative colitis (UC) are chronic inflammatory disorders with presenting symptoms similar to those of other intestinal conditions—altered bowel habit, abdominal pain, gastrointestinal bleeding, anorexia and weight loss. Thus, diagnosis of inflammatory bowel disease (IBD) is often delayed, which may have deleterious consequences, particularly for children in whom malnutrition may retard growth[1]. Currently, diagnosis of IBD is based upon a compatible constellation of symptoms, laboratory blood and serum tests, imaging studies, and tissue biopsy, as well as a high index of clinical suspicion[2,3]. Various blood markers have been proposed as potential diagnostic surrogates, including markers of inflammation (e.g., ESR, CRP) and of immunity (ANCA, ASCA and OmpC), but the sensitivity and specificity of these tests are suboptimal[4]. Thus, many children who undergo invasive endoscopic procedures for a clinical suspicion of IBD do not have evidence of CD or UC.

Current hypotheses of IBD pathogenesis implicate genetic predisposition, dysregulated immune responses, environmental factors, and an altered microbiome[5]. As genome-wide analyses has become available, the polygenic nature of IBD has been confirmed in genome-wide association studies (GWAS)[6]. In contrast to GWAS studies, gene expression profiling provides insight into the in situ activity of cellular pathways affected. Several studies have investigated differential gene expression in intestinal biopsies from adults with CD, UC, and non-IBD patients using microarray profiling[7–10]. Most of these studies included patients with long-standing disease who had been exposed to a variety of therapeutic agents. Furthermore, these studies have primarily focused upon gene expression in the actively inflamed tissue, although a few have examined gene expression in peripheral blood[11–16].

We studied a cohort of pediatric patients with incipient IBD because they are treatment-naïve with fewer co-morbidities and lifestyle influences than commonly present in adults. The purpose of our study was to use gene expression analysis to identify a molecular signature of IBD in blood and in colon biopsies and to elucidate key gene pathways involved in disease pathogenesis. The concordance (or lack of it) of gene expression pathways across colon and blood expression in this treatment-naïve pediatric population would also inform us whether the dysregulated pathways were tissue-specific or systemically expressed. A subsidiary aim was to determine the extent to which gene expression in peripheral blood may differentiate children with untreated IBD from similarly symptomatic controls.

Materials and methods

Study population and acquisition of intestinal biopsies

Children aged 5–18 years undergoing diagnostic colonoscopy because the treating provider strongly suspected IBD were recruited prospectively at Boston Children's Hospital from

September 2008 to July 2010. A parent/guardian provided written informed consent and minors assented to study procedures. The study protocol was approved by the Boston Children's Hospital Committee on Clinical Investigation (CCI). Children with a prior diagnosis of IBD or another autoimmune disorder and those receiving corticosteroids, immune modulators or biologics were excluded. Demographic and clinical characteristics recorded on the day of diagnostic colonoscopy were used to calculate Pediatric Crohn's Disease Activity Index (PCDAI) scores[17] and Pediatric Ulcerative Colitis Activity Index (PUCAI) scores[18] for patients with CD and UC, respectively. Higher scores on these clinician-completed disease activity indices reflect greater symptoms (PCDAI 0 to 100; PUCAI 0 to 85).

Given conflicting data regarding a relationship between colonic biopsy location and gene expression[7,9], and to minimize potential bias, biopsy samples were obtained from the ascending colon or cecum. This is the most commonly involved colonic site for CD, and the plurality of pediatric-onset UC patients present with extensive colitis[19]. Paired endoscopic pinch biopsies were obtained from the most grossly affected area or randomly in patients with a normal appearing right colon. One biopsy was used for histologic examination by the Boston Children's Hospital Department of Pathology, and the other was immediately stabilized in RNAlater (Ambion, Inc) until RNA was extracted. Peripheral blood was drawn from an indwelling catheter at the time of endoscopy and RNA was extracted using the PAXgene Blood RNA system (PreAnalytiX, Hombrechtikon, Switzerland). Biotinylated cDNA libraries were hybridized to Affymetrix Human Gene 1.0 ST arrays (Affymetrix, Santa Clara, CA).

The diagnosis of IBD was established utilizing clinical, radiographic, and endoscopic findings based upon the Porto criteria[2]. The NASPGHAN algorithm was used to distinguish CD from UC[3]. Those who were found to have a condition other than IBD (e.g., infectious colitis) or normal ileo-colonic histology were designated as symptomatic controls.

Data analysis

Preprocessing and normalization. The Affymetrix Power Tools software package was used to generate probe set (transcript cluster) level measurements using the analysis workflow specification

“-a rma-bg,quant-norm.sketch = 0.usepm = true.bioc = true,pm-only,med-polish” to the apt-probeset-summarize program, corresponding to RMA background correction of the perfect match probes, followed by quantile normalization and median polish summarization. This procedure, and each subsequent step in the analysis, was carried out independently for the colon and blood samples (i.e., all colon samples were processed as a single batch, and all blood samples were processed as a single batch).

Affymetrix transcript identifiers were mapped to Entrez genes using the Bioconductor toolkit[20]. Where multiple transcripts mapped to the same gene, the mean expression intensity across all clusters was computed as a representative value for the gene. All subsequent analyses were performed on these gene-level expression values in the R programming environment[21] unless otherwise specified.

Statistical analysis

Principal components analysis (PCA) via singular value decomposition (SVD) was performed on a matrix of centered gene-level expression intensities for each sample. Gene-by-gene differential (between the disease groups) expression analysis was performed using a two-tailed Student's t-test with Bonferroni-adjustment for multiple comparisons and a corrected $p < 0.05$ was considered significant. The ConsensusPathDB tool[22] was used to map genes differentially expressed between the disease states to Gene Ontology (GO) Biological Process terms

(levels 2, 3, and 4)[23], and to the KEGG[24] and Reactome[25] databases with reference to the background genes represented on the microarray and significance of overlap was tested using a hypergeometric distribution.

Disease state analysis

Using the glmnet package, we fitted a 5-fold cross-validated Lasso (least absolute shrinkage and selection operator) regularized regression model to our pediatric blood mRNA measurements to predict IBD status. The utility of the derived model was tested on an independent set of mRNA measurements from IBD-diagnosed and healthy adults available in the NCBI's Gene Expression Omnibus (accession GSE3365)[12]. Because the two data sets were measured on different microarray platforms (Affymetrix GeneChip Human Gene 1.0 ST for the present study, Affymetrix Human Genome 133A for the test data set), the absolute mRNA quantitation values after within-dataset normalizations were not directly comparable and required further adjustment to ensure that the coefficients of a model learned on one data set would be applicable to the other. The expression intensities for each sample in both data sets were log transformed, zero-centered, and scaled to unit variance to achieve a notionally common scale, and finally exponentiated to preserve effect size for this analysis. Information about gene function was obtained through reviewing articles retrieved using the HGNC gene name as a search term in PubMed.

Results and discussion

Useable peripheral blood samples were obtained from 98 subjects (78 of whom also had available biopsy samples; Table 1). The median age at the time of diagnostic colonoscopy was 14.2 years (IQR 11.3–16.2). There was one subject with CD whose biopsy sample was obtained from the sigmoid colon as the colonoscopy was terminated prematurely due to fulminant disease. All other biopsies were obtained from the right colon.

Ultimately, 39 subjects were diagnosed with CD and 18 subjects were diagnosed with UC. Despite extensive evaluation, two remained IBD unclassified (IBDU). Diagnoses of the 39 symptomatic controls included infectious colitis, juvenile polyp, and functional intestinal disorders. Patients with CD and UC had predominantly ileocolonic disease and pancolitis, respectively, consistent with the most commonly reported phenotypes in children and adolescents[19,26]. Disease severity varied widely with PCDAI scores ranging from 5 to 65, and PUCAI scores ranging from 0 to 80. In general, patients with CD had more systemic inflammation, as evidenced by higher CRP, ESR and IL-6 levels.

Colonic gene expression in IBD

Colonic biopsy gene expression profiles of the 58 subjects with newly diagnosed IBD were compared to those of 22 symptomatic controls. Of 18,305 annotated genes represented on the Affymetrix GeneChip[®] Human Gene 1.0 ST array, 1,118 were significantly differentially expressed (539 upregulated, 579 downregulated) in IBD compared to symptomatic controls (S1 File). All genes were included in principal components analysis (PCA) of gene expression, an unsupervised learning procedure that identifies linear combinations of genes with coordinated expression patterns across the entire set of samples (irrespective of disease state). The first two principal components (accounting for 54.2% of total variance), showed strong separation of IBD from symptomatic controls, indicating that the disease is likely mediated by a small number of groups of genes whose expression is highly correlated (though there may be many genes in each of these few groups; Fig 1A). The relatively large first principal component

Table 1. Demographic and clinical characteristics of study patients.

	CD (n = 39)	UC (n = 18)	Non-IBD (n = 39)
Male (%)	25 (64)	11 (61)	18 (46)
Age, years [median (IQR)]	13.1 (10.8–15.9)	14.9 (12.7–16.2)	14.7 (11.3–16.3)
Disease Location (Paris classification)	L1: Ileal: 15 L2: Colonic: 4 L3: Ileocolonic: 15 L4: Upper tract: 18 (Upper tract only 4)	E1: Proctitis: 1 E2: Left-sided UC: 4 E4: Pancolitis: 13	-
PCDAI [median (IQR)]	35 (23–43)	-	-
PUCAI [median (IQR)]	-	50 (33–69)	-
CRP (mg/L) [median (IQR)]	26.5 (12.6–49)	6.8 (1–8.1)	5.4 (0.3–3.2)
ESR (mm/hr) [median (IQR)]	43 (24–60)	17.0 (8–34)	13 (7–24)
IL-6 (pg/mL) [median (IQR)]	9 (5.7–13.7)	4.4 (2.1–8.7)	2.0 (1.4–2.7)
Biopsy	n = 38	n = 18	n = 22
Affected cecum	8	-	-
Affected ascending colon	21	11	1
Affected sigmoid	1	-	-
Unaffected ascending colon	8	7	21

CD, Crohn’s disease; UC, ulcerative colitis; IBD, inflammatory bowel disease; PCDAI, Pediatric Crohn’s Disease Activity Index; PUCAI, Pediatric Ulcerative Colitis Activity Index; CRP, C-reactive protein; ESR, erythrocyte sedimentation rate; IL-6, interleukin-6; SD, standard deviation; IQR, interquartile range. There were 2 patients with IBDU who are not included in the table.

<https://doi.org/10.1371/journal.pone.0222952.t001>

also indicates that the disease-specific mRNA signature dominates all other cohort-wide transcriptional artifacts.

Notably, there were three symptomatic controls who appeared to be more similar to the IBD group. One (* in Fig 1) had isolated non-specific gastroduodenal inflammation on the index endoscopy. The subject had an appendectomy two months later and the treating clinician retrospectively applied a diagnosis of CD following review of records from another hospital. Another subject (** in Fig 1), who had eosinophilic colitis, had ongoing joint pain and was lost to follow-up prior to completion of rheumatologic evaluation. The third subject (***) in Fig 1) had neutrophilic esophagitis and there was loss of vascular markings in the rectosigmoid colon noted at endoscopy, but histopathologic examination and capsule endoscopy were normal.

Peripheral blood gene expression in IBD

There were 880 genes that were significantly differentially expressed (418 upregulated, 462 downregulated) in the peripheral blood of IBD patients (n = 58) compared to symptomatic controls (n = 39; S2 File). PCA performed on all genes expressed in the peripheral blood samples revealed that, similar to colon biopsy tissue, peripheral blood demonstrates a strong IBD-specific transcriptional profile in this pediatric treatment-naïve cohort with the first two principal components representing 47.6% of total variance (Fig 2). Again, there were some patients who appeared to be more similar to those in another group, one of whom presented with joint and abdominal pain with a normal endoscopy (# in Fig 2). Of the IBDU subjects, the one who progressed to a severe form of Crohn’s disease (* in Fig 2) was much more similar to others with IBD, whereas the one that recovered without therapy (** in Fig 2) was more similar to symptomatic controls.

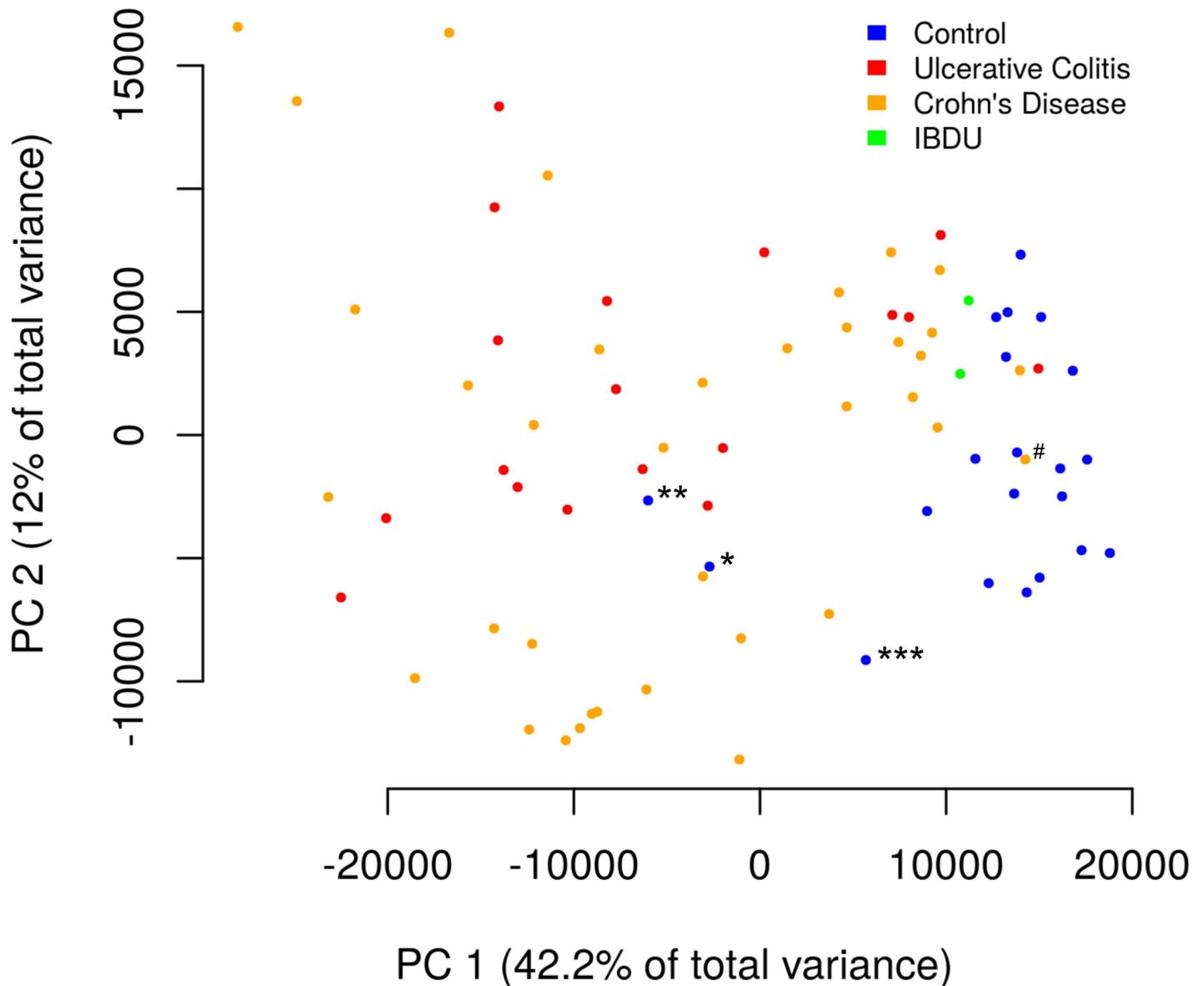


Fig 1. Principal components analysis of colon biopsy tissue samples from IBD and non-IBD patients. Each sample is represented as a point on the first two principal components of gene expression space, with the first principal component on the horizontal axis and the second principal component on the vertical axis. Samples from patients without IBD are shown as blue circles, while samples from patients with Crohn's disease are shown as orange circles, samples from patients with ulcerative colitis are shown as red circles, and samples from patients with IBDU are shown as lime green circles. The first principal component mostly separates the IBD from non-IBD samples. The subject labelled * was subsequently diagnosed with Crohn's disease and the subject labelled ** had significant rheumatologic symptoms but was lost to follow-up before a final diagnosis was made. Another subject (labelled ***) had neutrophilic esophagitis and there was loss of vascular markings in the rectosigmoid colon noted at endoscopy, but histopathologic examination and capsule endoscopy were normal. The subject labelled # presented with growth failure without intestinal symptoms.

<https://doi.org/10.1371/journal.pone.0222952.g001>

Colonic biopsy and peripheral blood gene expression

Altogether, 112 genes were differentially expressed in both colon and blood (S1 and S2 Files). The direction of differential expression was concordant in 92% (62 up-regulated, 41 down-regulated). The 9 genes with discordant differential expression (*BSG*, *CDH1*, *CYP4F12*, *DSC2*, *PADI2*, *PGM1*, *RNF10*, *SRPK1*, *VDR*) were each up-regulated in blood and down-regulated in colon.

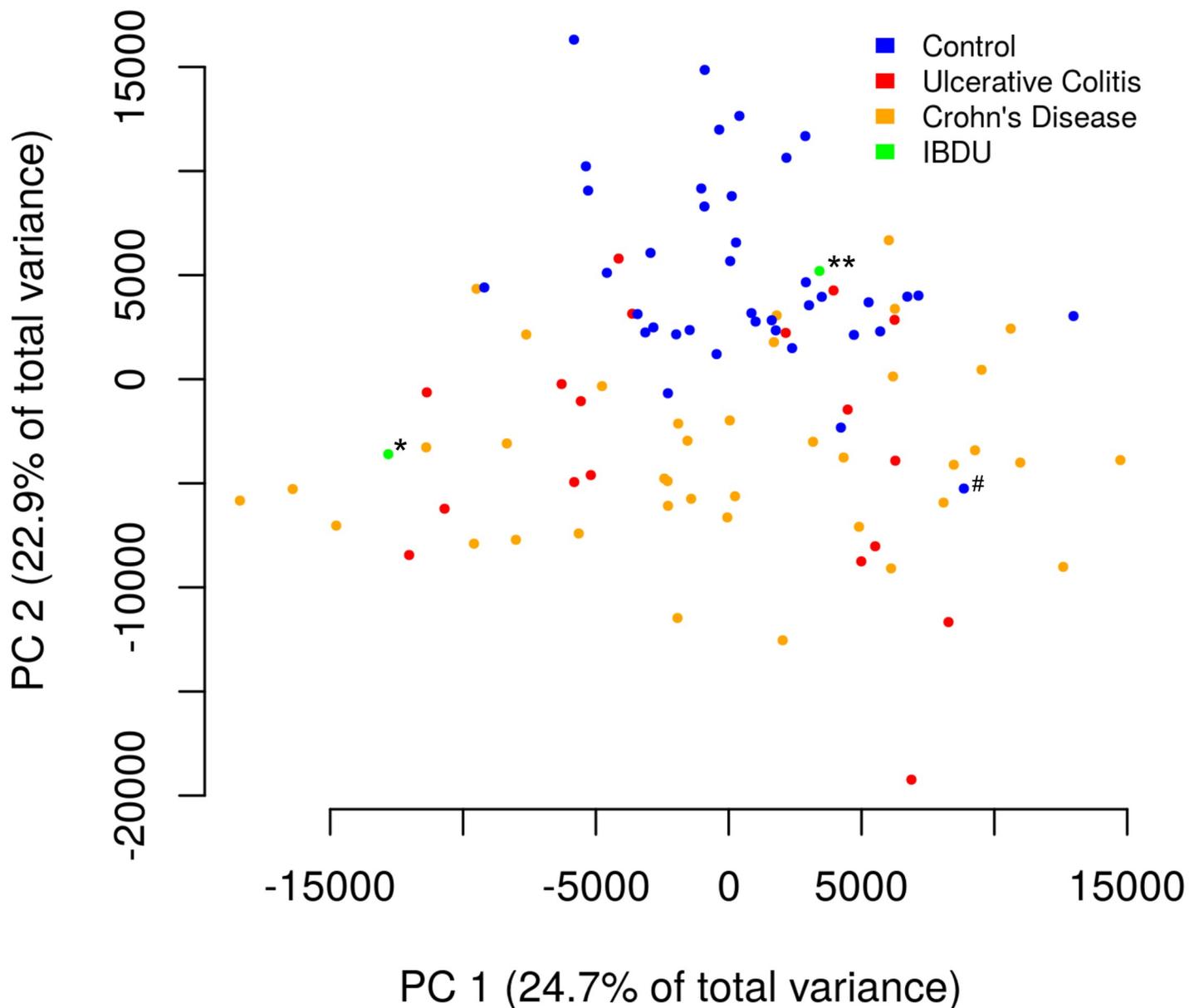


Fig 2. Principal components analysis of peripheral blood samples from IBD and non-IBD patients. Each sample is represented as a point on the first two principal components of gene expression space, with the first principal component on the horizontal axis and the second principal component on the vertical axis. Samples from patients without IBD are shown as blue circles, while samples from patients with Crohn's disease are shown as orange circles, samples from patients with ulcerative colitis are shown as red circles, and samples from patients with IBDU are shown as lime green circles. The first principal component mostly separates the IBD from non-IBD samples. Please refer to text for outcomes of IBDU patients.

<https://doi.org/10.1371/journal.pone.0222952.g002>

Gene Set Enrichment Analysis was performed to gain insight into the biological function of genes that were differentially expressed. After correcting for multiple hypothesis testing, the analysis identified 463 Gene Ontology (GO) sets and 67 pathways (KEGG, Reactome) that are significantly associated with the genes that were differentially expressed in colon (Fig 3A and S1 File). In addition, 282 GO sets, and 22 pathways were identified as being significantly (q -value < 0.05) associated with the genes that were differentially expressed in blood (Fig 3B and S2 File). There was an overlap of 97 GO terms and 5 pathways that were differentially overexpressed in both colon and blood. The pathways dysregulated in both tissue types were:

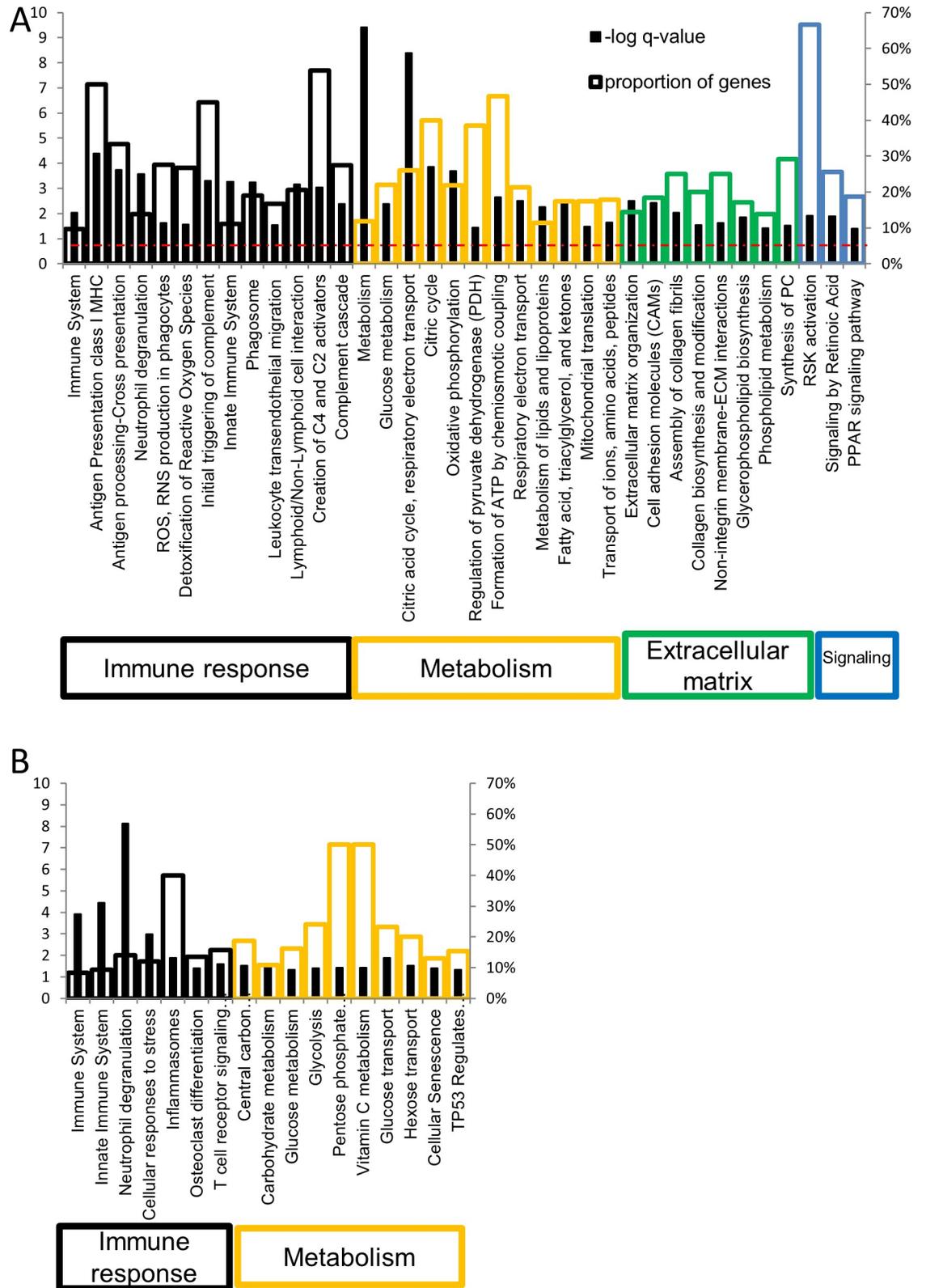


Fig 3. Gene sets (KEGG, Reactome) with expression levels significantly associated with IBD. (A) colon tissue. (B) peripheral blood. Solid black bars indicate the $-\log Q$ -value for each gene set (values on left axis) whereas the open colored bars indicate the proportion of genes in each set (values on right axis) whose expression levels differ significantly between those with IBD and those without IBD.

<https://doi.org/10.1371/journal.pone.0222952.g003>

Table 2. Genes with non-zero coefficients in linear model of gene expression in children with active IBD compared to symptomatic controls.

Transcript	Gene name	Coefficient	Biological functions
IL7R	Interleukin 7 receptor	-0.072	B and T cell regulation; V(D)J recombination
TXNIP	Thioredoxin-interacting protein	-0.043	Thioredoxin inhibitor; NLRP3 inflammasome activation
ALAS2	5'-aminolevulinate synthase 2	0.039	Mitochondrial heme synthesis in erythrocytes
S100A8	S100 calcium binding protein A8 (a.k.a., calprotectin L1L subunit)	0.026	Innate immunity; Cell migration; Reactive oxygen species generation
UBB	Ubiquitin B	0.021	Antigen processing and presentation; Regulation of gene expression; Stress response
SLC2A3	Solute carrier family 3 member 3 (a.k.a., GLUT3)	0.12	Glucose transport into cells

<https://doi.org/10.1371/journal.pone.0222952.t002>

neutrophil degranulation, immune system, innate immune system, osteoclast differentiation and glucose metabolism. This overlap was significant using the hypergeometric distribution ($p < 0.001$). Complete lists of the gene sets and pathways can be found in [S1](#) and [S2 Files](#).

Peripheral blood mRNA expression-based classification: IBD vs. symptomatic controls

This observed overlap between gene expression patterns in colon tissue and peripheral blood motivated us to ask whether the gene expression patterns in peripheral blood could be correlated with disease status. Therefore, we applied a Lasso regression model to the blood mRNA measurements. The predictor had non-zero coefficients for only 6 genes, *IL7R*, *UBB*, *TXNIP*, *S100A8*, *ALAS2*, and *SLC2A3* ([Table 2](#)). To establish the validity of these genes as markers of disease status, we tested the model using a publicly available independent microarray expression data set consisting of PBMC transcriptional profiles from 42 healthy adults, 59 adults with treated CD and 26 adults with treated UC. Our classifier achieved a positive predictive value (PPV) of 0.84 and negative predictive value (NPV) of 0.72 on the held-out test set. The differences in populations and chips used suggests robustness to technical measurement artifacts.

In this study, comparison of gene expression between children with untreated IBD and similarly symptomatic controls revealed a distinctive molecular signature of IBD with substantial overlap of the signatures in colon biopsies and in peripheral whole blood. In both tissue types, affected gene pathways represent dysregulated immune responses and metabolic derangement that are characteristic of IBD [27]. Colon biopsies also demonstrated pathways related to cell surface receptors, extracellular matrix remodeling and localization of immune cells, which reflects tissue specific changes of inflammatory colitis. The direction of differential gene expression (up- or down-regulated) was consistent for 103 of 112 genes differentially expressed in both whole blood and colon biopsies. The 9 genes that had discordant direction of change in expression were primarily those for glycoproteins and cell-adhesion molecules, several of which have been associated with colitis (*CDH1* [28], *PADI2* [29], *VDR* [30]). Differential expression of the vitamin D receptor (*VDR*) in colon and peripheral blood may partially explain discordant reports regarding the role of vitamin D in IBD [31]. Thus, while gene expression signatures of IBD in colon and whole blood may not be identical, their shared structure and composition lends credence to the notion that local perturbations in gene expression are reflected in changes in gene expression in circulating blood and its constituent immune cells [32]. Furthermore, we identified six genes (*IL7R*, *UBB*, *TXNIP*, *S100A8*, *ALAS2*, and *SLC2A3*) whose expression levels in peripheral blood were predictive of disease status ([Table 2](#)). The utility of this signature was validated using an independent test in which the gene signature had 80% accuracy to distinguish adult patients receiving treatment for active IBD from healthy adults.

The demonstration of similar perturbations of gene expression in peripheral blood and in colon tissue and the nodal position of these genes in the immune signaling network supports the concept that IBD is a disease of generalized systemic immune dysregulation. Polymorphisms in *IL7R*, which had the largest coefficient, are associated with UC susceptibility[28]. *IL7R* is under balancing selective pressures, and genetic variability correlates with pathogen diversity[33]. Ubiquitin (*UBB*), one of the most conserved proteins in eukaryotic organisms, was the only gene in the regression model that was not itself significantly differentially expressed in blood; however, there were significant differences in expression of other genes involved in ubiquitin signaling (e.g., *UBE2L6*, *USP7*, *PSMB8*, *PSME1*). It is noteworthy that adheroinvasive *E coli*, which have been implicated in IBD pathogenesis, upregulate the ubiquitin-proteasome system via the *NFκB* pathway[34]. Reduced expression of *TXNIP*, an antioxidant and *NFκB* inhibitor which is highly expressed in lymphocytes and intestinal epithelial cells[35] in IBD patients suggests that modulation of redox signaling may be an important therapeutic target. *S100A8* encodes one subunit of calprotectin, an acute phase reactant expressed primarily by activated granulocytes that may have prognostic value[36]. Recently, peripheral blood transcript levels of the closely related homologue *S100A12* were correlated with Mayo score in adults with UC[16], which both validates our findings and further supports the potential usefulness of gene expression in peripheral blood as a marker of local disease activity. The two other genes (*ALAS2* and *SLC2A3*) whose expression levels were predictive of disease status highlight the links between the immune system and metabolism[37]. *SLC2A3* encodes GLUT3, a glucose transporter that fuels activated immune cells[38]. Intriguingly, deletion of *SLC2A3* has been identified as a protective factor in rheumatoid arthritis with an effect size that is second only to HLA genes[39]. The role of metabolic pathways in regulating immune activation has been reviewed recently[40].

Differences between colon and blood provide further insights into the pathophysiology of IBD. Transmural infiltration of neutrophils is pathognomonic for IBD, and neutrophil activation pathways were the strongest signal in peripheral blood in the current study. In the colon, the gene expression signature also reflected local events, including antigen presentation, and epithelial barrier function and repair. Expression of transmembrane transport proteins reflects both the requirement to shuttle metabolic substrates between different cellular compartments and the physiology of secretory diarrhea.

Several of the most significantly differentially expressed genes have been identified as important therapeutic targets in IBD or other autoimmune diseases, whereas others suggest promising new targets for IBD (Table 3). These include the *IL12B1* receptor which is involved in signaling by *IL-12* and *IL-23*, the cytokines targeted by ustekinumab. Our study also corroborates the finding that Lipocalin-2 (*LCN2*; also known as *NGAL*) expression is increased in colon tissue in IBD[41]. An antimicrobial protein that inhibits growth of iron dependent bacteria, such as *E coli*, Lipocalin-2 has been proposed as a fecal marker of IBD activity[42].

Strengths of this study include use of paired blood and colon biopsy samples from treatment-naïve pediatric IBD patients to reduce the effects of prior treatment or lifestyle factors. Secondly, prospective study design allowed patients to be recruited prior to diagnosis using the same inclusion and exclusion criteria for IBD patients and symptomatic controls. Third, the classifier was validated in blood samples from an independent cohort of adult patients with gene expression profiling performed on a different platform[12]. The robustness of our classifier with this dataset is strong evidence that there is an underlying structure to gene expression in IBD.

A potential weakness of this study is the inclusion of diverse participants with varying clinical expression and phenotypes. More precise definition of IBD patients may facilitate identification of subgroups; however, UC and CD are similar at a gene expression level[15] and this is

Table 3. Selected differentially expressed genes with putative mechanistic role/potential therapeutic targets in IBD.

Gene	Gene Name	Tissue location and expression change in IBD compared to symptomatic controls	Biological function	Preclinical/clinical studies
<i>ADM</i>	Adrenomedullin	Blood (↑)	Vasodilator, angiogenesis factor and anti-inflammatory	Mucosal healing in 5/7 patients with intractable UC after 14 daily infusions
<i>FURIN</i>	Furin	Blood (↑)	Proprotein convertase (substrates: IGF-1, extracellular matrix metalloproteases, TGF-β, viruses, bacterial toxins)	Copy number variants affect penetrance of CTLA4 mutations associated with early onset severe crohn's disease.Reduced inflammation in murine collagen-induced arthritis
		Colon (↑)		
<i>AGTRAP</i>	Angiotensin II receptor type 1 associated protein	Blood (↑)	Inhibition of angiotensin II signaling	Renin-angiotensin system activation exacerbates colitis[43]. Angiotensin II type 1 receptor attenuated colitis in mice by inhibition of NFκB pathway[44]
		Colon (↑)		
<i>ST3GAL4; ST3GAL2</i>	Sialyltransferase	Blood (↑)	Glycosylation	Mucous barrier function; relationship to inflammation associated colorectal cancer; TNF induced.
		Colon (↑)		

<https://doi.org/10.1371/journal.pone.0222952.t003>

supported by the robustness of our classifier to identify patients with different IBD subtypes. While the use of symptomatic controls may obscure the signal of gastrointestinal symptoms, it does suggest that gene expression may be used to distinguish IBD from other causes of gastrointestinal symptoms.

Conclusions

Treatment-naïve inflammatory bowel disease is associated with characteristic gene expression profiles in colon tissue and in peripheral blood. Overlap between the gene expression signatures in the affected tissue and peripheral whole blood suggests that IBD affects gene expression both locally in the colon and in circulating immune cells in a characteristic and reproducible manner. With this in mind, we used our gene expression data from peripheral blood to develop a linear regression model. The six transcripts identified had 80% accuracy to predict IBD status in an independent cohort of adults with and without IBD. This result was unexpected given the differences in study populations and gene chips used, which supports the robustness of our findings that gene expression profiling of peripheral whole blood reflects tissue-specific changes in IBD. Additional larger studies of well-defined cohorts are needed to determine whether patterns of RNA expression in peripheral blood may be a useful adjunct to current markers of disease activity in patients with IBD.

Supporting information

S1 Table. Real-time quantitative PCR validation of selected microarray data.
(PDF)

S2 Table. Sex distribution.
(PDF)

S1 File. Genes significantly differentially expressed in colon biopsies from IBD patients and symptomatic controls.
(XLSX)

S2 File. Genes significantly differentially expressed in peripheral blood from IBD patients and symptomatic controls.
(XLSX)

Acknowledgments

The authors thank the patients and families at Boston Children's Hospital who participated in the study and the Rasmussen, Wolfman, Ward and MacInnes families whose donations to the Inflammatory Bowel Disease Center supported the study. We also thank Dr Robert Burakoff for valuable discussions at the start of the study and the physicians, nurses and other staff in the Gastrointestinal Procedure Unit at Boston Children's Hospital who assisted with sample collection.

Author Contributions

Conceptualization: Nathan P. Palmer, Jocelyn A. Silvester.

Data curation: Jocelyn A. Silvester.

Formal analysis: Nathan P. Palmer, Jocelyn A. Silvester.

Funding acquisition: Nathan P. Palmer, Jocelyn A. Silvester.

Investigation: Nathan P. Palmer, Jocelyn A. Silvester, Jessica J. Lee, Andrew L. Beam, Inbar Fried, Vladimir I. Valtchinov, Fedik Rahimov, Sek Won Kong, Saum Ghodoussipour, Helen C. Hood, Athos Bousvaros, Richard J. Grand, Louis M. Kunkel, Isaac S. Kohane.

Methodology: Nathan P. Palmer, Jocelyn A. Silvester.

Project administration: Nathan P. Palmer, Jocelyn A. Silvester.

Resources: Nathan P. Palmer, Jocelyn A. Silvester.

Supervision: Nathan P. Palmer, Jocelyn A. Silvester.

Writing – original draft: Jocelyn A. Silvester.

Writing – review & editing: Jocelyn A. Silvester.

References

1. Kappelman MD, Bousvaros A. Nutritional concerns in pediatric inflammatory bowel disease patients. *Mol Nutr Food Res*. 2008; 52: 867–874. <https://doi.org/10.1002/mnfr.200700156> PMID: 18324705
2. Escher JC. Inflammatory bowel disease in children and adolescents: Recommendations for diagnosis—The Porto criteria. *J Pediatr Gastroenterol Nutr*. 2005; 41: 1–7. <https://doi.org/10.1097/01.mpg.0000163736.30261.82> PMID: 15990620
3. Bousvaros A, Antonioli DA, Colletti RB, Dubinsky MC, Glickman JN, Gold BD, et al. Differentiating ulcerative colitis from Crohn disease in children and young adults: report of a working group of the North American Society for Pediatric Gastroenterology, Hepatology, and Nutrition and the Crohn's and Colitis Foundation of America. *J Pediatr Gastroenterol Nutr*. 2007; 44: 653–74. <https://doi.org/10.1097/MPG.0b013e31805563f3> PMID: 17460505
4. Mack DR, Langton C, Markowitz J, LeLeiko N, Griffiths a, Bousvaros a, et al. Laboratory values for children with newly diagnosed inflammatory bowel disease. *Pediatrics*. 2007; 119: 1113–1119. <https://doi.org/10.1542/peds.2006-1865> PMID: 17545378
5. Dalal SR, Chang EB. The microbial basis of inflammatory bowel diseases. *J Clin Invest*. 2014; 124: 4190–6. <https://doi.org/10.1172/JCI72330> PMID: 25083986
6. Uniken Venema WT, Voskuil MD, Dijkstra G, Weersma RK, Festen EA. The genetic background of inflammatory bowel disease: from correlation to causality. *J Pathol*. 2017; 241: 146–158. <https://doi.org/10.1002/path.4817> PMID: 27785786
7. Noble CL, Abbas a R, Cornelius J, Lees CW, Ho G-T, Toy K, et al. Regional variation in gene expression in the healthy colon is dysregulated in ulcerative colitis. *Gut*. 2008; 57: 1398–1405. <https://doi.org/10.1136/gut.2008.148395> PMID: 18523026
8. Olsen J, Gerds TA, Seidelin JB, Csillag C, Bjerrum JT, Troelsen JT, et al. Diagnosis of ulcerative colitis before onset of inflammation by multivariate modeling of genome-wide gene expression data. *Inflamm Bowel Dis*. 2009; 15: 1032–1038. <https://doi.org/10.1002/ibd.20879> PMID: 19177426

9. Wu F, Dassopoulos T, Cope L, Maitra A, Brant SR, Harris ML, et al. Genome-wide gene expression differences in Crohn's disease and ulcerative colitis from endoscopic pinch biopsies: Insights into distinctive pathogenesis. *Inflamm Bowel Dis*. 2007; 13: 807–821. <https://doi.org/10.1002/ibd.20110> PMID: 17262812
10. Haberman Y, Tickle TL, Dexheimer PJ, Kim MO, Tang D, Karns R, et al. Pediatric Crohn disease patients exhibit specific ileal transcriptome and microbiome signature. *J Clin Invest*. 2014; 124: 3617–3633. <https://doi.org/10.1172/JCI75436> PMID: 25003194
11. Kuo B, Bhasin M, Jacquart J, Scult MA, Slipp L, Riklin EIK, et al. Genomic and clinical effects associated with a relaxation response mind-body intervention in patients with irritable bowel syndrome and inflammatory bowel disease. *PLoS One*. 2015; 10: 1–26. <https://doi.org/10.1371/journal.pone.0123861> PMID: 25927528
12. Burczynski ME, Peterson RL, Twine NC, Zuberek KA, Brodeur BJ, Casciotti L, et al. Molecular classification of Crohn's disease and ulcerative colitis patients using transcriptional profiles in peripheral blood mononuclear cells. *J Mol diagnostics*. 2006; 8: 51–61. <https://doi.org/10.2353/jmoldx.2006.050079> PMID: 16436634
13. Kabakchiev B, Turner D, Hyams J, Mack D, Leleiko N, Crandall W, et al. Gene expression changes associated with resistance to intravenous corticosteroid therapy in children with severe ulcerative colitis. *PLoS One*. 2010; 5: 1–8. <https://doi.org/10.1371/journal.pone.0013085> PMID: 20941359
14. Burakoff R, Chao S, Perencevich M, Ying J, Friedman S, Makrauer F, et al. Blood-based biomarkers can differentiate ulcerative colitis from Crohn's disease and noninflammatory diarrhea. *Inflamm Bowel Dis*. 2011; 17: 1719–1725. <https://doi.org/10.1002/ibd.21574> PMID: 21744426
15. Van Lierop PPE, Swagemakers SM, De Bie CI, Middendorp S, Van Baarlen P, Samsom JN, et al. Gene expression analysis of peripheral cells for subclassification of pediatric inflammatory bowel disease in remission. *PLoS One*. 2013; 8: 1–8. <https://doi.org/10.1371/journal.pone.0079549> PMID: 24260248
16. Planell N, Masamunt MC, Leal RF, Rodríguez L, Esteller M, Lozano JJ, et al. Usefulness of Transcriptional Blood Biomarkers as a Non-invasive Surrogate Marker of Mucosal Healing and Endoscopic Response in Ulcerative Colitis. *J Crohns Colitis*. 2017; 11: 1335–1346. <https://doi.org/10.1093/ecco-jcc/jjx091> PMID: 28981629
17. Hyams JS, Ferry GD, Mandel FS, Gryboski JD, Kibort PM, Kirschner BS, et al. Development and validation of a pediatric Crohn's disease activity index. [Internet]. *Journal of pediatric gastroenterology and nutrition*. 1991. pp. 439–47. Available: <http://www.ncbi.nlm.nih.gov/pubmed/1678008> PMID: 1678008
18. Turner D, Otley AR, Mack D, Hyams J, de Bruijne J, Uusoue K, et al. Development, validation, and evaluation of a pediatric ulcerative colitis activity index: a prospective multicenter study. *Gastroenterology*. 2007; 133: 423–32. <https://doi.org/10.1053/j.gastro.2007.05.029> PMID: 17681163
19. Fumery M, Duricova D, Gower-Rousseau C, Annese V, Peyrin-Biroulet L, Lakatos PL. Review article: The natural history of paediatric-onset ulcerative colitis in population-based studies. *Aliment Pharmacol Ther*. 2016; 43: 346–355. <https://doi.org/10.1111/apt.13478> PMID: 26582737
20. Gentleman R, Carey V, Bates D, Bolstad B, Dettling M, Dudoit S, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004; 5: R80. <https://doi.org/10.1186/gb-2004-5-10-r80> PMID: 15461798
21. R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2016. Available: <http://www.r-project.org>
22. Kamburov A, Pentchev K, Galicka H, Wierling C, Lehrach H, Herwig R. ConsensusPathDB: Toward a more complete picture of cell biology. *Nucleic Acids Res*. 2011; 39: 712–717. <https://doi.org/10.1093/nar/gkq779>
23. The Gene Ontology Consortium, Ashburner M, Ball CA, Blake JA, Botstein D, Butler H, et al. Gene ontology: tool for the unification of biology. *The Gene Ontology Consortium*. *Nat Genet*. 2000; 25: 25–9. <https://doi.org/10.1038/75556> PMID: 10802651
24. Kanehisa M, Sato Y, Kawashima M, Furumichi M, Tanabe M. KEGG as a reference resource for gene and protein annotation. *Nucleic Acids Res*. 2016; 44: D457–D462. <https://doi.org/10.1093/nar/gkv1070> PMID: 26476454
25. Fabregat A, Sidiropoulos K, Garapati P, Gillespie M, Hausmann K, Haw R, et al. The reactome pathway knowledgebase. *Nucleic Acids Res*. 2016; 44: D481–D487. <https://doi.org/10.1093/nar/gkv1351> PMID: 26656494
26. De Bie CI, Paerregaard A, Kolacek S, Ruemmele FM, Koletzko S, Fell JME, et al. Disease phenotype at diagnosis in pediatric Crohn's disease: 5-year analyses of the EUROKIDS registry. *Inflamm Bowel Dis*. 2013; 19: 378–385. <https://doi.org/10.1002/ibd.23008> PMID: 22573581
27. Novak EA, Mollen KP. Mitochondrial dysfunction in inflammatory bowel disease. *Front cell Dev Biol*. 2015; 3: 62. <https://doi.org/10.3389/fcell.2015.00062> PMID: 26484345

28. Anderson CA, Boucher G, Lees CW, Franke A, D'Amato M, Taylor KD, et al. Meta-analysis identifies 29 additional ulcerative colitis risk loci, increasing the number of confirmed associations to 47. *Nat Genet.* 2011; 43: 246–252. <https://doi.org/10.1038/ng.764> PMID: 21297633
29. Cantarino N, Musulen E, Valero V, Peinado MA, Perucho M, Moreno V, et al. Downregulation of the Deiminase PADI2 Is an Early Event in Colorectal Carcinogenesis and Indicates Poor Prognosis. *Mol Cancer Res.* 2016; 14: 841–848. <https://doi.org/10.1158/1541-7786.MCR-16-0034> PMID: 27280713
30. Simmons JD, Mullighan C, Welsh KI, Jewell DP. Vitamin D receptor gene polymorphism: association with Crohn's disease susceptibility. *Gut.* 2000; 47: 211–214. <https://doi.org/10.1136/gut.47.2.211> PMID: 10896912
31. Del Pinto R, Pietropaoli D, Chandar AK, Ferri C, Cominelli F. Association Between Inflammatory Bowel Disease and Vitamin D Deficiency: A Systematic Review and Meta-analysis. *Inflamm Bowel Dis.* 2015; 21: 2708–2717. <https://doi.org/10.1097/MIB.0000000000000546> PMID: 26348447
32. Kohane IS, Valtchinov VI. Quantifying the white blood cell transcriptome as an accessible window to the multiorgan transcriptome. *Bioinformatics.* 2012; 28: 538–545. <https://doi.org/10.1093/bioinformatics/btr713> PMID: 22219206
33. Fumagalli M, Pozzoli U, Cagliani R, Comi GP, Riva S, Clerici M, et al. Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions. *J Exp Med.* 2009; 206: 1395–408. <https://doi.org/10.1084/jem.20082779> PMID: 19468064
34. Palmela C, Chevarin C, Xu Z, Torres J, Sevrin G, Hirten R, et al. Adherent-invasive *Escherichia coli* in inflammatory bowel disease. *Gut.* 2017; [gutjnl-2017-314903](https://doi.org/10.1136/gutjnl-2017-314903). <https://doi.org/10.1136/gutjnl-2017-314903> PMID: 29141957
35. Takahashi Y, Ishii Y, Murata A, Nagata T, Asai S. Localization of thioredoxin-interacting protein (TXNIP) mRNA in epithelium of human gastrointestinal tract. *J Histochem Cytochem.* 2003; 51: 973–6. <https://doi.org/10.1177/002215540305100713> PMID: 12810848
36. Kalla R, Kennedy NA, Ventham NT, Boyapati RK, Adams AT, Nimmo ER, et al. Serum Calprotectin: A Novel Diagnostic and Prognostic Marker in Inflammatory Bowel Diseases. *Am J Gastroenterol.* 2016; 111: 1796–1805. <https://doi.org/10.1038/ajg.2016.342> PMID: 27596694
37. Perl A. Review: Metabolic Control of Immune System Activation in Rheumatic Diseases. *Arthritis Rheumatol.* 2017; 69: 2259–2270. <https://doi.org/10.1002/art.40223> PMID: 28841779
38. Maratou E, Dimitriadis G, Kollias A, Boutati E, Lambadiari V, Mitrou P, et al. Glucose transporter expression on the plasma membrane of resting and activated white blood cells. *Eur J Clin Invest.* 2007; 37: 282–90. <https://doi.org/10.1111/j.1365-2362.2007.01786.x> PMID: 17373964
39. Veal CD, Reekie KE, Lorentzen JC, Gregersen PK, Padyukov L, Brookes AJ. A 129-kb deletion on chromosome 12 confers substantial protection against rheumatoid arthritis, implicating the gene SLC2A3. *Hum Mutat.* 2014; 35: 248–256. <https://doi.org/10.1002/humu.22471> PMID: 24178905
40. Buck MD, Sowell RT, Kaech SM, Pearce EL. Review Metabolic Instruction of Immunity. *Cell.* 2017; 169: 570–586. <https://doi.org/10.1016/j.cell.2017.04.004> PMID: 28475890
41. Csillag C, Haagen Nielsen O, Vainer B, Olsen J, Dieckgraefe BK, Hendel J, et al. Expression of the genes *dual oxidase 2*, *lipocalin 2* and *regenerating islet-derived 1 alpha* in Crohn's disease. *Scand J Gastroenterol.* 2007; 42: 454–463. <https://doi.org/10.1080/00365520600976266> PMID: 17454855
42. Thorsvik S, Damås J, Granlund A, Flo T, Østvik A, Sandvik A. Fecal neutrophil gelatinase-associated lipocalin (NGAL) is a promising biomarker for inflammatory bowel disease and NGAL is expressed in paneth cells. *Inflamm Bowel Dis.* 2016; 22: S44. <https://doi.org/10.1111/jgh.13598>
43. Shi Y, Liu T, He L, Dougherty U, Chen L, Adhikari S, et al. Activation of the Renin-Angiotensin System Promotes Colitis Development. *Sci Rep.* 2016; 6: 27552. <https://doi.org/10.1038/srep27552> PMID: 27271344
44. Mizushima T, Sasaki M, Ando T, Wada T, Tanaka M, Okamoto Y, et al. Blockage of angiotensin II type 1 receptor regulates TNF- α -induced MAdCAM-1 expression via inhibition of NF- κ B translocation to the nucleus and ameliorates colitis. *AJP Gastrointest Liver Physiol.* 2010; 298: G255–G266. <https://doi.org/10.1152/ajpgi.00264.2009> PMID: 19940029