OXFORD

Original Article

Redefining the Practical Utility of Blood Transcriptome Biomarkers in Inflammatory Bowel Diseases

Jerzy Ostrowski,[a,b] Michalina Dabrowska,[a] Izabella Lazowska,[c] Agnieszka Paziewska,[b] Aneta Balabas,[a] Anna Kluska,[a] Maria Kulecka,[b] Jakub Karczmarski,[a] Filip Ambrozkiewicz,[a] Magdalena Piatkowska,[a] Krzysztof Goryca,[a] Natalia Zeber-Lubecka,[b] Jaroslaw Kierkus,[d] Piotr Socha,[d] Michal Lodyga,[e] Maria Klopocka,[f] Barbara Iwanczak,[g] Katarzyna Bak-Drabik,[h] Jaroslaw Walkowiak,[i] Piotr Radwan,[j] Urszula Grzybowska-Chlebowczyk,[k] Bartosz Korczowski,[l] Teresa Starzynska,[m] Michal Mikula[a]

aDepartment of Genetics, Maria Sklodowska-Curie Institute – Oncology Centre, Warsaw, Poland bDepartment of Gastroenterology and Hepatology, Medical Center for Postgraduate Education, Warsaw, Poland cDepartment of Pediatric Gastroenterology and Nutrition, Medical University of Warsaw, Warsaw, Poland dDepartment of Gastroenterology, Hepatology and Feeding Disorders, Children's Memorial Health Institute, Warsaw, Poland eDepartment of Internal Medicine and Gastroenterology with IBD Subdivision, Central Clinical Hospital of the Ministry of the Interior, Warsaw, Poland fVascular Diseases and Internal Medicine, Nicolaus Copernicus University in Torun, Collegium Medicum, Bydgoszcz, Poland gDepartment of Pediatrics, Gastroenterology and Nutrition, Wroclaw Medical University, Wroclaw, Poland hDepartment of Pediatrics, School of Medicine with the Division of Dentistry in Zabrze, Medical University of Silesia, Katowice, Poland iDepartment of Pediatric Gastroenterology & Metabolic Diseases, Poznan University of Medical Sciences, Poznan, Poland jDepartment of Gastroenterology, Medical University of Lublin, Lublin, Poland kDepartment of Pediatrics, School of Medicine in Katowice, Medical University of Silesia, Katowice, Poland lMedical College, University of Rzeszow, Rzeszow, Poland mDepartment of Gastroenterology, Pomeranian Medical University, Szczecin, Poland

Corresponding author: Jerzy Ostrowski, MD, PhD; Cancer Center-Institute, Roentgena 5, 02-781 Warsaw, Poland. Tel.: +48 225462575; e-mail: jostrow@warman.com.pl

Abstract

Background and Aims: The study investigates the practical utility of whole-blood gene expression profiling to diagnose inflammatory bowel diseases [IBDs].

Methods: The discovery cohorts included 102 and 51 paediatric IBD patients and controls, and 95 and 46 adult IBD patients and controls, respectively. The replication cohorts included 447 and 76 paediatric IBD patients and controls, and 271 and 108 adult IBD patients and controls, respectively. In the discovery phase, RNA samples extracted from whole peripheral blood were analysed using RNA-Seq, and the predictive values of selected biomarkers were validated using quantitative polymerase chain reaction [qPCR].

Results: In all, 15 differentially expressed transcripts [adjusted $p \leq 0.05$] were selected from the discovery sequencing datasets. The receiver operating characteristic curves and area under the curve [ROC-AUC] in replication analyses showed high discriminative power [AUC range, 0.91–0.98] for 11 mRNAs in paediatric patients with active IBD. By contrast, the AUC-ROC values ranged from

0.63 to 0.75 in comparison among inactive paediatric IBDs and active/inactive adult IBDs, indicating a lack of discriminative power. The best multi-mRNA diagnostic classifier showed moderate discriminative power [AUC = 0.81] for paediatric inactive IBD, but was not able to discriminate active or inactive adult IBD patients from controls. The AUC-ROC values did not confirm an ability of the mRNAs abundances to discriminate between active ulcerative colitis and active Crohn's disease in paediatric or adult populations.

Conclusions: This study identifies and validates blood transcriptional biomarkers that could be used in clinical settings as diagnostic predictors of IBD clinical activity in paediatric, but not adult, IBD patients.

Key Words: RNA-Seq; whole-blood gene expression, biomarker; inflammatory bowel disease

1. Introduction

Peripheral blood cells share >80% of their transcriptome with that of other tissues, and profiling of gene expression in these cells is employed in descriptive and comparative analyses of autoimmune and inflammatory diseases.[1–3] Crohn's disease [CD] and ulcerative colitis [UC] are chronic disorders that result from altered activation of intestinal immunopathological processes in response to the intestinal microbiota.[4] Whereas the onset of both of these inflammatory bowel diseases [IBDs] can occur from early childhood to beyond the sixth decade of life, differences in the polygenic architecture of paediatric- and adult-onset IBD[5] may indicate a relationship between genetically attributable risk and pathogenic variances in children and adults. Compared with adult-onset IBDs, paediatric IBDs are typically characterized by a more extensive disease course, a change in disease location over time, and a more frequent family history of IBD. Patients diagnosed between the ages of 20 and 30 years have a relatively less variable phenotype, and those diagnosed after the age of 60 years often have a mild disease severity.[5]

Previous studies revealed that gene expression profiles obtained from whole-blood cells may differentiate CD and UC from non-IBD conditions, and active from inactive CD.[6] However, the practical utilities of measuring whole-blood expression levels in IBD patients have not been resolved. To date, only small populations of IBD patients have been examined,[6–10] the results from discovery phases of the studies have not been replicated, and candidate biomarkers used in the studies have been selected from microarray-based datasets. In the current study, candidate mRNA biomarkers of IBD were screened using RNA sequencing, and the replication study was performed using polymerase chain reaction [PCR]-based testing. Both phases of the study used relatively large cohorts of paediatric and adult patients with defined clinical activities of IBD.

2. Methods

2.1. Ethics statement

All procedures involving human participants were performed in accordance with the ethical standards of the institutional and/or national research committees, and with the 1964 Helsinki Declaration and its later amendments, or comparable ethical standards. The study was approved by the Ethics Committee of the Medical Center for Postgraduate Education, Warsaw, Poland [decision 69/PW/2011 granted on 16 May 2011]. Informed consent was obtained from all participants in the study.

2.2. Subjects

A total of 915 patients with IBD and 280 healthy control individuals were recruited at the gastroenterology departments of various Polish hospitals. Criteria to participate in the study included a confirmed CD or UC diagnosis. There were no exclusion criteria except indeterminate colitis or concomitant infections. Overall, 488 patients [303 children aged 17 years and 185 adults] and 427 patients [245 children and 182 adults] were diagnosed with CD and UC, respectively. Clinical characteristics and biochemical parameters available from the enrolled patients at the time of blood collection for gene expression profiling are summarized in Supplementary Table S1, available as Supplementary data at *ECCO-JCC* online. IBD was diagnosed by experienced gastroenterologists during a standard diagnostic work-up, using the Porto criteria modified in accordance with the recommendations of the European Crohn's and Colitis Organisation [ECCO] in children, and according to ECCO guidelines in adults. Patients were recruited during a course of hospital treatment or during a scheduled visit at outpatient departments. The Crohn's disease activity index [CDAI], the ulcerative colitis activity index [UCAI], and their paediatric versions [PCDAI/PUCAI] were determined to estimate disease severity.[11–13] The patients were then assigned to two subgroups: patients in remission or with mild IBD were considered to have inactive disease, and those with moderate or severe IBD were considered to have active disease. Before inclusion most patients were given mesalazine, but for majority of them the blood samples were collected before additional medications [immunosupressants, glucocorticoids, biological therapy] were ordered. The control group of children consisted of surgical, orthopaedic, or ophthalmological patients who did not suffer from inflammation or intestinal diseases. The control group of adults were healthy individuals who were mainly recruited at cancer screening programmes.

2.3. RNA extraction

Peripheral blood was collected using the Tempus RNA Isolation Kit [Thermo Fisher Scientific], and total RNA was isolated according to the manufacturer's instructions. RNA quality and quantity were analysed using a NanoDrop spectrophotometer, and samples with A260/A280 ratios of 1.8–2.1 were assessed further using an Agilent 2100 Bioanalyzer.

2.4. RNA sequencing, reads processing, and statistical analysis

Preparation and sequencing of the RNA-Seq libraries was performed as described previously.[14] Signal processing and base calling were conducted with Torrent Suite version 5.0.4. Reads were mapped to

the hg19 AmpliSeq Transcriptome version 1 genome. Read counts per gene were obtained with HTSeq-count version 0.6[15] using default parameters. Normalization and differential gene expression estimations were performed using DESeq2, using default parameters and options.[16] A gene was considered differentially expressed when the adjusted *p*-value was less than 0.05.

2.5. Quantitative RT-PCR

Nine genes that were statistically significantly differentially expressed and had absolute fold changes [FCs] greater than 2 [as estimated by DESeq2] between controls and patients with the active form of disease [regardless of age] were verified by quantitative RT-PCR [qRT-PCR]. Six additional genes were also chosen for verification based on the adult cohort data only. The qRT-PCR analyses were performed as described previously,[17] using TaqMan chemistry. The geometrical mean expression level of the *RPLP0* and *UBC* mRNAs was used as a normalization factor. Gene expression levels were calculated using the ΔΔCt method, and the results were analysed using the Mann-Whitney U-test; *p* ≤0.05 was considered statistically significant.

2.6. Data accessibility statement

The RNA sequencing datasets generated during the study were deposited in European Nucleotide Archive under the PRJEB28822 accession number.

3. Results

3.1. The discovery step

To screen for candidate biomarkers, we used the Ion AmpliSeq Transcriptome Human Gene Expression Kit to sequence 293 RNA samples extracted from whole peripheral blood. Of these, 51 and 46 samples were from paediatric and adult patients with CD, 51 and 49 samples were from paediatric and adult patients with UC, and 50 and 46 samples were from control children and adults, respectively. In total, 31 patients in each paediatric IBD subgroup and seven patients in each adult IBD subgroup had active disease [with a score above 30] at blood collection. Table 1 shows a summary of the main epidemiological variables for the discovery cohorts.

On average, the RNA-Seq analyses generated 11 827 252 reads per sample, of which 88% were on target and mapped to the reference genome. The RNA-Seq data were combined according to patient age, diagnosis, and clinical activity [remission or mid-active versus active disease], and the pair-wise comparisons identified 148 and 111 differentially expressed genes [DEGs] (false discovery rate [FDR] ≤0.05; FC >2) between paediatric IBD patients and controls, and between adult IBD patients and controls, respectively [Table 2

and Supplementary Table S2, available as Supplementary data at *ECCO-JCC* online]. None of the mRNAs was able to differentiate between CD and UC in paediatric patients, and only 36 mRNAs [with FCs ranging from 0.7 to 1.3] were able to differentiate between these IBDs in adult patients.

To search for biomarkers with diagnostic utility, we selected 15 mRNAs that displayed the highest levels of differential expression between active IBD patients and healthy controls; all but one [KLRF1] were upregulated in IBD. This list included nine mRNAs [*S100A12*, *OPLAH*, *ATP9A*, *ANOS1*, *FCGR1A*, *ITGB4*, *UTS2R*, *MMP9*, and *COX6B2*] that showed at least a 2-fold difference between the active IBD and control groups and adjusted *p*-values less than 0.05 in both the paediatric and adult populations, and six mRNAs [*ANXA3*, *CACNA1E*, *GALNT14*, *IL18R1*, *KLRF1*, and *PFKFB3*] that showed the highest FCs between the active IBD and control groups in the adult population. The expression levels of these mRNAs were confirmed by qRT-PCR using the same RNA samples as those used in the RNA-Seq analyses. The expression levels of all 15 genes were statistically significantly different [adjusted *p* ≤0.05] between paediatric patients with active IBD and the corresponding controls [Tables 3 and 4], and similar results were obtained for the active UC and CD paediatric subgroups [data not shown]. By contrast, none of these genes was able to differentiate between control children and paediatric patients with inactive IBD. In the adult population, the expression levels of all but one [*ATP9A*] from the set of nine genes, and three [*GALNT14*, *KLRF1*, and *PFKFB3*] from the set of six genes, were significantly different between the active IBD and healthy control groups. In addition, only one gene [*KLRF1*] was differentially expressed between healthy controls and adults with inactive IBD.

3.2. The replication study

Next, we assessed the diagnostic potential of the 15 genes described above using newly recruited IBD patients and controls. The replication cohorts included 391 CD patients [252 children and 139 adults], 327 UC patients [195 children and 132 adults], and 184 controls [76 children and 108 adults]. Of these, 53 and 25 of the paediatric patients and 34 and 27 of the adult patients had active CD and active UC, respectively, at blood collection. Table 5 shows a summary of the main epidemiological variables for the replication cohorts.

The transcript levels were determined by qRT-PCR, and the FCs between the IBD subgroups and control groups were compared. In the paediatric population, the levels of all 15 mRNAs were significantly different [adjusted *p* ≤0.05] between the control group and the active UC or active CD subgroups. Furthermore, the expression

Table 1. Summary of the main epidemiological variables for the discovery cohorts.

	Demographics		Medication			
	Sex; [female/male]	Age; range [median, years]	5-ASAs	Immunosupressants	Glucocorticoids	Biological therapy
Paediatric CD	30/21	2–17 [13]	90.2%	45.1%	39.2%	11.8%
Paediatric UC	32/19	1–17 [15]	96.1%	23.5%	0	29.4%
Paediatric controls	28/22	1–17 [8]	0	0	0	0
Adult CD	23/23	19–69 [34]	86.9%	60.9%	26.1%	32.6%
Adult UC	28/21	21–66 [36]	97.9%	22.4%	26.5%	12.2%
Adult controls	26/20	38–62 [46]	0	0	0	0

ASA, aminosalicylic acid; UC, ulcerative colitis; CD, Crohn's disease.

levels of 13 of the 15 mRNAs examined were significantly different between the paediatric control group and the paediatric inactive UC or inactive CD subgroup [Table 6]. In the adult population, the expression level of *UTS2R* differed significantly between the control group and the group comprising patients with either active or inactive UC, and the expression levels of 10 and nine mRNAs differed significantly between the healthy adults and those with active or inactive CD, respectively [Table 7]. Gene expression levels did not correlate with patients' age, with a Kendall's tau coefficient lower than 0.1.

Next, the diagnostic potentials of the 15 mRNAs described above were assessed using receiver operating characteristic [ROC] curves and area under the curve [AUC] analyses. ROC curves enable assessment of the relationship between the sensitivity and specificity of a biomarker over various cut-offs.[10] AUC-ROC values greater than 0.8 were assumed to represent moderate [good] discriminative power, and those greater than 0.9 were assumed to represent high [excellent] discriminative power between the analysed groups. In the paediatric population, nine genes [*ANOS1*, *ANXA3*, *CACNA1E*, *GALNT14*, *ITGB4*, *MMP9*, *OPLA*, *PFKFB3*, and *S100A12*] showed high discriminative power [AUC range, 0.91–0.98] between the control group and each active IBD subgroup [active UC or active CD] [Figure 1]. In addition, *ATP9A* showed high discriminative power between the control and active UC groups, and *FCGR1B* showed high discriminative power between the control and active CD groups. By contrast, whereas several genes were significantly differentially expressed between paediatric patients with inactive IBDs and control children, the AUC-ROC values of these genes ranged from 0.63 to 0.75, and therefore did not confirm their discriminative properties. To determine whether the diagnostic ability of an mRNA signature is higher than those of the single mRNAs, we calculated AUC-ROC values for combinations of mRNAs with the best discriminatory powers. Using linear models of normalized expression values and a stepwise inclusion approach, we found that the best multi-mRNA diagnostic classifier

that differentiated control children from those with inactive UC [*ITGB4*, *MMP9*, *UTS2R*, *KLRF1*, and *GALANT14*] and inactive CD [*ANOS1*, *CACNA1E*, *MMP9*, and *UTS2R*] had an AUC value of 0.81, and therefore displayed only moderate discriminative power. Furthermore, the AUC-ROC values did not confirm the ability of gene expression profiling to discriminate between active UC and active CD in paediatric patients, even though statistically significant differences in mRNA levels were observed between the two groups [Table 6].

Unexpectedly, none of the genes that were significantly differentially expressed between the adult IBD and healthy adults groups in pair-wise comparisons showed sufficient discriminatory power in the ROC-AUC analyses, when either single mRNAs [Table 7] or combinations of mRNAs were analysed [not shown]. In adults, the best multi-mRNA diagnostic classifiers for active UC [*FCR1B*, *KLRF1*, *CACNA1E*, *ANXA3*, and *GALANT14*] and active CD [*KLRF1*, *CACNA1E*, *ANXA3*, and *GALANT14*] had AUC values of 0.78 and 0.77, respectively. As seen for the paediatric population, the AUC-ROC values did not confirm the ability of gene expression profiling to discriminate between active UC and active CD in adults [Table 7].

4. Discussion

Endoscopic/radiological examination followed by pathological evaluation, supported by laboratory testing of C-reactive protein [CRP] levels, the erythrocyte sedimentation rate, and faecal calprotectin [among others], is the 'gold standard' for diagnosis of IBD. The severity of IBD can vary from mild to severe, and its diagnosis can be delayed, especially for early-onset IBD when the clinical symptoms are ambiguous, non-specific, or indolent. Biomarkers have the potential to be useful for IBD diagnosis, identification of disease subtypes, and prognosis to therapeutic adjustment, and might therefore complement the use of clinical parameters. Candidate biomarkers are typically identified by high-throughput methods using either microarray or next-generation sequencing [NGS] technologies, and are subsequently validated by standard molecular methods. Microarrays represent a non-linear model of the relationship between readouts and actual amounts of mRNAs in the samples analysed. Due to the multiple steps involved in microarray hybridization and imaging, the results of such analyses are strictly dependent on the statistics included in the data processing. NGS can provide linear relationships between relative numbers of sequenced transcript fragments and true gene expression levels[18]; however, this method is hampered by its bias for identifying particular subsets of transcripts and the unknown minimum size of the sequenced transcriptome required for sufficient coverage of low-abundance mRNAs. Although neither of these technologies is optimal for measuring gene expression at the global level, they are both commonly employed for discovery of candidate biomarkers.[19]

Table 2. Summary of numbers of differentially expressed genes in different comparisons.

Comparison	DGE	FC >2.0	% DEG
Paediatric active IBD vs control children	5249	148	30.35
Paediatric inactive IBD vs control children	2637	11	18.19
CD children vs UC children	0	0	0
Adult active IBD vs control adults	1627	101	11.08
Adult inactive IBD vs control adults	4203	0	30.28
CD adults vs UC adults	35	0	0.24

DGE, number of differentially expressed genes; FC, fold change; %DEG, percentage of DEGs among tested genes; IBD, inflammatory bowel disease; CD, Crohn's disease; UC, ulcerative colitis.

Table 3. Differences in mRNAs levels from the nine gene sets selected by RNA sequencing and validated by qRT-PCR in the training cohorts.

Comparisons	ITGB4	MMP9	COX6B2	UTS2R	FCGR1B	OPLA	S100A12	ATP9A	ANOS1
Paediatric active IBD vs control children	**4.80E-06**	**1.40E-04**	**3.00E-04**	**4.10E-06**	**4.10E-06**	**4.70E-04**	**3.30E-06**	**5.90E-04**	**3.30E-06**
Paediatric inactive IBD vs control children	8.80E-01	6.30E-01	7.30E-01	7.30E-01	7.30E-01	7.40E-01	5.10E-01	6.30E-01	5.00E-01
Adult active IBD vs control adults	**1.50E-03**	**3.20E-03**	**6.00E-04**	**6.00E-04**	**1.80E-05**	**2.30E-04**	**1.70E-03**	3.50E-01	**5.50E-05**
Adult inactive IBD vs control adults	3.40E-01	4.30E-01	4.50E-01	**5.60E-03**	8.20E-01	4.30E-01	4.30E-01	5.90E-01	7.90E-01

In bold fonts are the statistically significant results with adjusted *p*-value <0.05.

IBD, inflammatory bowel disease.

Table 4. Differences in mRNAs levels from the six gene sets selected by RNA sequencing and validated by qRT-PCR in the training cohorts.

Comparison	ANXA3	CACNA1E	GALNT14	IL18R1	KLRF1	PFKFB3
Paediatric active IBD vs control children	**5.51E-05**	**1.46E-05**	**6.56E-06**	**0.003566**	**0.020675**	**0.000209**
Paediatric inactive IBD vs control children	0.996873	0.996873	0.996873	0.996873	0.996873	0.996873
Adult active IBD vs control adults	0.076794	0.195521	**0.046505**	0.878352	**9.40E-06**	**0.046505**
Adult inactive IBD vs control adults	0.529562	0.973081	0.529562	0.529562	**1.49E-05**	0.529562

In bold fonts are the statistically significant results with adjusted *p*-value <0.05.
IBD, inflammatory bowel disease.

Table 5. Summary of the main epidemiological variables for the replication cohorts.

	Demographics		Medication			
	Sex; [female/male]	Age; range [median, years]	5-ASAs	Immunosuppressants	Glucocorticoids	Biological therapy
Paediatric CD	102/150	2–17 [15]	90.1%	45.2%	39.3%	11.9%
Paediatric UC	107/88	1–17 [15]	97.4%	22.6%	0	30.3%
Paediatric controls	32/44	1–17 [8]	0	0	0	0
Adult CD	86/53	18–70 [29]	87.8%	33.1%	20.9%	8.6%
Adult UC	78/54	18–73 [35]	95.4%	23.5%	24.2%	18.2%
Adult controls	64/44	43–64 [58]	0	0	0	0

ASA, aminosalicylic acid; UC, ulcerative colitis; CD, Crohn's disease.

Blood cells suspended in a fluid matrix connect the entire biological system of an organism and constitute the first line of the immune defence system; hence, blood samples are often used as a surrogate for traditional tissue specimens in clinical diagnoses.[3] Changes in the whole-blood transcriptome are associated with injuries and a wide range of diseases, including autoimmune, inflammatory, infectious, psychiatric, cardiovascular, neurological, and neoplastic disorders, as well as various environmental stresses.[20]

Microarray studies have led to the development of a noninvasive test for the diagnosis of IBD which is based on expression profiles of peripheral leukocytes.[21] In 2008, Alsobrook *et al.* reported that the peripheral blood expression levels of a classifier set of six genes [*BLCAP*, *UBE2G1*, *GPX1*, *RAP1A*, *CALM3*, and *NONO*] are able to distinguish IBD patients from healthy controls with accuracy, sensitivity, and specificity rates of 84%, 89%, and 75%, respectively.[22] To date, a few other studies have confirmed that expression profiling of blood samples provides a noninvasive method of distinguishing clinically active from inactive IBD,[6] endoscopically active UC patients from patients in clinical remission,[8] active CD from CD in remission,[7] and paediatric IBD patients in clinical remission from healthy controls.[23] In sum, IBD-related blood gene expression in the previous studies has only been examined in small populations, and analyses of the DEGs generated results that could not unambiguously differentiate healthy from diseased samples. The only exception was the study of Burakoff *et al.*[7] describing four separate gene panels which discriminated either CD or UC patients with either mild or moderate-to-severe severity from other categories with an ROC–AUC ranging between 0.89 and 0.99. Our RNA-Seq-based study performed on 293 RNA samples determined a detectable expression of 19 out of 22 of those genes, but only one, PGM1, showed a moderate [AUC = 0.81] discriminative power between the control and active CD groups.

To re-examine the practical utility of whole-blood gene expression profiling in IBD diagnosis, we identified 15 potential biomarker transcripts [*S100A12*, *OPLAH*, *ATP9A*, *ANOS1*, *FCGR1A*, *ITGB4*, *UTS2R*, *MMP9*, *COX6B2*, *ANXA3*, *CACNA1E*, *GALNT14*, *IL18R1*, *KLRF1*, and *PFKFB3*] by sequencing RNA samples isolated from a total of 293 IBD patients and control individuals. These mRNAs were subsequently validated by qRT-PCR analyses of RNA samples from 902 newly recruited IBD patients and controls. Our findings confirmed that measurement of blood mRNA levels has a diagnostic potential for paediatric patients with active IBD, but not for paediatric patients with inactive IBD, or adults with active or inactive disease. Furthermore, although we found that analysing combinations of mRNAs has a moderate power to discriminate between paediatric inactive IBD patients and control children, we could not confirm that the accuracy of such multi-classifier mRNA signatures is higher than those of single mRNAs in adult patients. Furthermore, the blood mRNA signatures were unable to distinguish UC from CD in the adult or paediatric populations. Among 15 transcripts analysed, several encode proteins functionally connected with inflammatory processes and immune response. The S100A12, also known as calgranulin C, has recently emerged as marker of inflammation to predict IBD better than calprotectin from stool samples.[24–26] The FCGR1A encodes the neutrophils' expressed CD64 receptor, and its abundance is positively correlated with immune inflammation syndromes including IBD.[27,28] MMP-9 is one of major metalloproteases activated in intestinal tissues of patients with active IBD,[29] and serum MMP-9 levels could aid differentiation of active UC from active CD.[30] IL18R1 encodes IL18 receptor, essential for IL18-mediated signal transduction. IL18 is instrumental for controlling the outgrowth of bowel microbiota; however during instances of severe inflammation, exacerbated IL-18 expression causes a loss of goblet cells, ultimately depleting mucosal barrier function leading to colitis in a mouse model.[31] The KLRF1 encodes C-type lectin-like activating receptor NKp80, which is expressed on all NK cells in the peripheral blood. It recognizes activation-induced C-type lectin [AICL] that is upregulated on activated monocytes and NK cells, and NKp80–AICL interaction promotes NK cell-mediated control of monocytes.[32] The KLRF1 was the only transcript with expression downregulated in IBD when compared with healthy controls.

Table 6. Comparisons of mRNA levels analysed by qRT-PCR in the paediatric replication cohorts.

| mRNA | Ulcerative colitis vs control children | | | | Crohn's disease vs control children | | | | Active ulcerative colitis vs active Crohn's disease | |
| | Active | | Inactive | | Active | | Inactive | | | |
	AUC	Adj. *p*-value	AUC	Adj. *p*-value	AUC	Adj. *p*-value	AUC	Adj. *p*-value	AUC	Adj. *p*-value
ANOS1	**0.96**	**1.9E-11**	0.69	**1.6E-05**	**0.97**	**2.47E-18**	0.75	**3.49E-08**	0.62	**0.001611**
ANXA3	**0.98**	**1.13E-11**	0.66	**0.000316**	**0.95**	**2.21E-17**	0.69	**9.99E-06**	0.55	0.198678
ATP9A	**0.92**	**6.12E-10**	0.55	0.223911	**0.82**	**1.25E-09**	0.58	0.074931	0.51	0.853082
CACNA1E	**0.96**	**1.9E-11**	0.68	**3.71E-05**	**0.96**	**3.95E-18**	0.74	**1.07E-07**	0.61	**0.00308**
COX6B2	**0.88**	**3.93E-08**	0.63	**0.003564**	**0.87**	**1.92E-12**	0.62	**0.007121**	0.53	0.3854
FCGR1B	**0.90**	**9.3E-09**	0.63	**0.003013**	**0.95**	**2.2E-17**	0.72	**4.97E-07**	0.66	**6.75E-06**
GALNT14	**0.98**	**1.13E-11**	0.66	**0.000222**	**0.95**	**1.69E-17**	0.68	**2.94E-05**	0.55	0.198678
IL18R1	**0.83**	**1.86E-06**	0.57	0.105793	**0.83**	**4.49E-10**	0.50	0.999028	0.57	0.088585
ITGB4	**0.94**	**2.62E-10**	0.71	**5.14E-06**	**0.94**	**8.15E-17**	0.69	**1.23E-05**	0.52	0.50386
KLRF1	**0.83**	**1.05E-06**	0.69	**1.9E-05**	0.71	**4.77E-05**	0.68	**3.01E-05**	0.52	0.558309
MMP9	**0.97**	**1.13E-11**	0.71	**5.14E-06**	**0.97**	**2.47E-18**	0.73	**1.61E-07**	0.56	0.198678
OPLAH	**0.96**	**2.03E-11**	0.69	**1.6E-05**	**0.96**	**5.68E-18**	0.68	**2.71E-05**	0.54	0.313489
PFKFB3	**0.98**	**1.13E-11**	0.63	**0.003564**	**0.95**	**2.2E-17**	0.66	**0.000149**	0.55	0.198678
S100A12	**0.97**	**1.69E-11**	0.67	**9.36E-05**	**0.96**	**5.56E-18**	0.70	**3.19E-06**	0.56	0.171036
UTS2R	**0.81**	**5.74E-06**	0.70	**1.26E-05**	0.63	**0.013751**	0.70	**8.87E-06**	0.55	0.198678

In bold fonts are the statistically significant results with adjusted *p*-value <0.05 and AUC values >0.8.

AUC, area under the curve; adj, adjusted.

Table 7. Comparisons of mRNA levels analysed by qRT-PCR in the paediatric replication cohorts.

| mRNA | Ulcerative colitis vs control adults | | | | Crohn's disease vs control adults | | | | Active ulcerative colitis vs active Crohn's disease | |
| | Active | | Inactive | | Active | | Inactive | | | |
	AUC	Adj. *p*-value	AUC	Adj. *p*-value	AUC	Adj. *p*-value	AUC	Adj. *p*-value	AUC	Adj. *p*-value
ANOS1	0.58	0.11068	0.56	0.82618	0.71	**0.001198**	0.64	**0.00072**	0.55	0.0880
ANXA3	0.54	0.51191	0.49	0.93717	0.69	**0.003**	0.60	**0.00884**	0.57	**0.0287**
ATP9A	0.58	0.11068	0.48	0.93717	0.54	0.571025	0.62	**0.00373**	0.52	0.6237
CACNA1E	0.51	0.89914	0.54	0.93717	0.71	**0.001198**	0.61	**0.00373**	0.62	**0.0001**
COX6B2	0.51	0.89914	0.51	0.93717	0.64	**0.024824**	0.55	0.21516	0.55	0.0997
FCGR1B	0.51	0.90343	0.56	0.82618	0.71	**0.001198**	0.62	**0.00198**	0.64	**0.0001**
GALNT14	0.53	0.58935	0.52	0.93717	0.64	**0.024729**	0.60	**0.01125**	0.56	0.0577
IL18R1	0.57	0.15341	0.53	0.93717	0.53	0.655729	0.54	0.29748	0.51	0.7161
ITGB4	0.50	0.94155	0.50	1.00000	0.65	**0.017966**	0.50	0.99496	0.52	0.5403
KLRF1	0.62	0.00507	0.60	0.58882	0.71	**0.001198**	0.74	**0.00000**	0.63	**0.0001**
MMP9	0.50	0.94155	0.48	0.93717	0.60	0.104351	0.57	0.06547	0.57	**0.0287**
OPLA	0.55	0.27203	0.55	0.82618	0.55	0.399874	0.50	0.99496	0.56	0.0577
PFKFB3	0.57	0.15341	0.62	0.46185	0.59	0.140337	0.54	0.32333	0.55	0.0880
S100A12	0.54	0.51191	0.56	0.82618	0.70	**0.001519**	0.61	**0.00458**	0.58	**0.0206**
UTS2R	0.69	**1.7E-06**	0.76	**0.00079**	0.66	**0.015584**	0.65	**0.00030**	0.56	0.0709

In bold fonts are the statistically significant results with adjusted *p*-value <0.05.

AUC, area under the curve; adj, adjusted.

IBDs are heterogeneous chronic disorders characterized by a succession of variable severity of relapses and remissions, which in turn relates to variable therapeutic decisions.[33] In this study, the severity of IBD was assessed using the CDAI or UCAI and their paediatric equivalents [PCDAI/PUCAI]. However, although these indexing systems are commonly used due to the simplicity of collecting data for calculating the severity scores, they predominantly rely on symptomatology without consideration of other aspects of disease severity, such as inflammatory activity and structural damage. Patients with severe symptoms may have mild inflammation, and disease severity does not always reflect the true disease activity. Furthermore, IBD may co-exist with other non-inflammatory chronic intestinal conditions, such as irritable bowel syndrome.[33] Consequently,

our method of assigning the severity of IBD as mild, moderate, or severe may reflect current working practice rather than a clinically proven classification. Accordingly, the cut-off gene expression levels for distinguishing active from inactive IBD identified here should be considered with some caution. Similar problems apply to commonly used markers of intestinal inflammation, such as C-reactive protein [CRP] and faecal calprotectin.[33] For example, up to 50% of patients with clinically active UC, and some patients with active CD, display normal CRP levels.[6] In addition measurement of faecal calprotectin is relevant in follow-up examinations of IBD patients, but more studies are needed to establish accurate reference values for different ages, disease subtypes, disease localization/extension, and responses to therapy.[25,34,35] The results presented here confirm

Ulcerative colitis
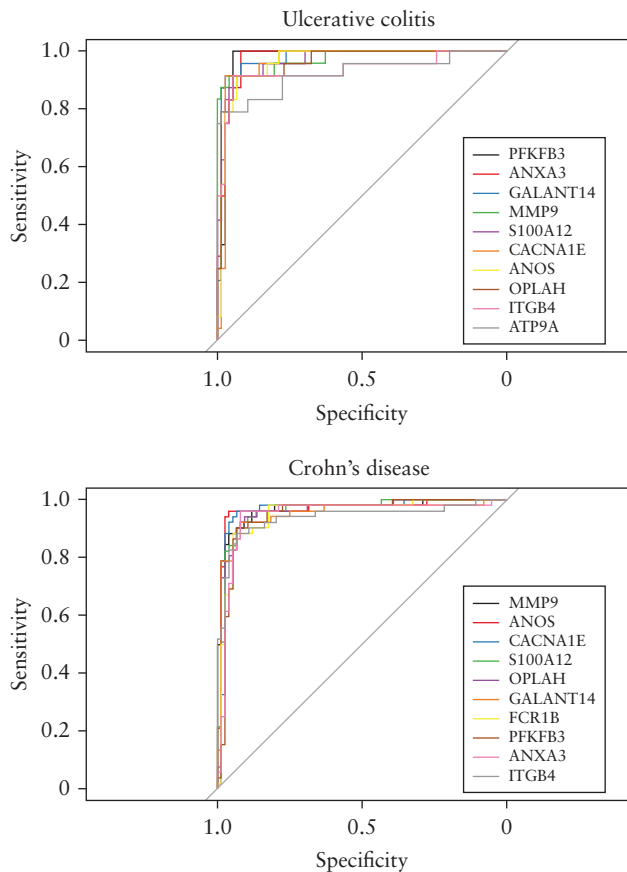

Crohn's disease

Figure 1. Diagnostic potential of candidate mRNA biomarkers. Receiver operating characteristic [ROC] curves were generated and area under the curve [AUC] values were calculated using qRT-PCR expression values obtained for the paediatric active ulcerative colitis [UC] and active Crohn's disease [CD] subgroups.

that whole-blood mRNA expression profiling may be an effective tool to monitor IBD severity in paediatric, but not adult, patients. Although this study included large groups of patients, the amounts of RNA obtained from the blood samples were only sufficient for confirmatory and replication studies on a limited number of selected transcripts. Therefore, we cannot exclude the possibility that, among the thousands of DEGs identified via high-throughput sequencing, we may have missed other potentially useful biomarkers due to methodological limitations.

In summary, this study identified and validated the use of whole-blood transcriptional biomarkers as predictors of IBD clinical activity in paediatric but not adult IBD patients. These promising results were derived from analyses of a large number of patients with a wide range of clinical activity, and were conducted simultaneously in paediatric and adult cohorts.

Funding

Conflict of Interest

The authors do not declare any conflict of interest.

Author Contributions

JO designed the study. All authors participated in generation, collection, assembly, analysis, and interpretation of data. MK and KG performed statistical analysis. JO, MK, and MM drafted the manuscript that was reviewed by all the authors.

Supplementary Data

Supplementary data are available at *ECCO-JCC* online.

References

1. Mesko B, Poliska S, Nagy L. Gene expression profiles in peripheral blood for the diagnosis of autoimmune diseases. *Trends Mol Med* 2011;**17**:223–33.
2. Gliddon HD, Herberg JA, Levin M, Kaforou M. Genome-wide host RNA signatures of infectious diseases: discovery and clinical translation. *Immunology* 2018;**153**:171–8.
3. Liew CC, Ma J, Tang HC, Zheng R, Dempsey AA. The peripheral blood transcriptome dynamically reflects system wide biology: a potential diagnostic tool. *J Lab Clin Med* 2006;**147**:126–32.
4. Wallace KL, Zheng LB, Kanazawa Y, Shih DQ. Immunopathology of inflammatory bowel disease. *World J Gastroenterol* 2014;**20**:6–21.
5. Ostrowski J, Paziewska A, Lazowska I, *et al*. Genetic architecture differences between pediatric and adult-onset inflammatory bowel diseases in the Polish population. *Sci Rep* 2016;**6**:39831.
6. Barnes EL, Liew CC, Chao S, Burakoff R. Use of blood based biomarkers in the evaluation of Crohn's disease and ulcerative colitis. *World J Gastrointest Endosc* 2015;**7**:1233–7.
7. Burakoff R, Pabby V, Onyewadume L, *et al*. Blood-based biomarkers used to predict disease activity in Crohn's disease and ulcerative colitis. *Inflamm Bowel Dis* 2015;**21**:1132–40.
8. Planell N, Masamunt MC, Leal RF, *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**:1335–46.
9. Burakoff R, Hande S, Ma J, *et al*. Differential regulation of peripheral leukocyte genes in patients with active Crohn's disease and Crohn's disease in remission. *J Clin Gastroenterol* 2010;**44**:120–6.
10. Hsu MJ, Chang YC, Hsueh HM. Biomarker selection for medical diagnosis using the partial area under the ROC curve. *BMC Res Notes* 2014;**7**:25.
11. Best WR, Becktel JM, Singleton JW, Kern F Jr. Development of a Crohn's disease activity index. National Cooperative Crohn's Disease Study. *Gastroenterology* 1976;**70**:439–44.
12. Seo M, Okada M, Yao T, Ueki M, Arima S, Okumura M. An index of disease activity in patients with ulcerative colitis. *Am J Gastroenterol* 1992;**87**:971–6.
13. Hyams JS, Ferry GD, Mandel FS, *et al*. Development and validation of a pediatric Crohn's disease activity index. *J Pediatr Gastroenterol Nutr* 1991;**12**:439–47.
14. Lichawska-Cieslar A, Pietrzycka R, Ligeza J, *et al*. RNA sequencing reveals widespread transcriptome changes in a renal carcinoma cell line. *Oncotarget* 2018;**9**:8597–613.
15. Anders S, Pyl PT, Huber W. HTSeq – a Python framework to work with high-throughput sequencing data. *Bioinformatics* 2015;**31**:166–9.
16. Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol* 2014;**15**:550.
17. Mikula M, Rubel T, Karczmarski J, Goryca K, Dadlez M, Ostrowski J. Integrating proteomic and transcriptomic high-throughput surveys for search of new biomarkers of colon tumors. *Funct Integr Genomics* 2011;**11**:215–24.
18. Zhao S, Fung-Leung W-P, Bittner A, Ngo K, Liu X. Comparison of RNA-Seq and microarray in transcriptome profiling of activated T cells. *PLoS One* 2014;**9**:e78644.
19. Zhang W, Yu Y, Hertwig F, *et al*. Comparison of RNA-seq and microarray-based models for clinical endpoint prediction. *Genome Biol* 2015;**16**:133.

20. Cabrera SM, Chen YG, Hagopian WA, Hessner MJ. Blood-based signatures in type 1 diabetes. *Diabetologia* 2016;**59**:414–25.

21. Burczynski ME, Peterson RL, Twine NC, *et al*. Molecular classification of Crohn's disease and ulcerative colitis patients using transcriptional profiles in peripheral blood mononuclear cells. *J Mol Diagn* 2006;**8**:51–61.

22. Alsobrook J, Ma T, Leighton J, *et al*. Novel genomic biomarkers that differentiate between inflammatory bowel disease and normal patients using peripheral blood specimens. *Am J Gastroenterol* 2008;**103**(s1):S439.

23. van Lierop PP, Swagemakers SM, de Bie CI, *et al*. Gene expression analysis of peripheral cells for subclassification of pediatric inflammatory bowel disease in remission. *PLoS One* 2013;**8**:e79549.

24. Heida A, Van de Vijver E, van Ravenzwaaij D, *et al*.; CACATU consortium. Predicting inflammatory bowel disease in children with abdominal pain and diarrhoea: calgranulin-C versus calprotectin stool tests. *Arch Dis Child* 2018;**103**:565–71.

25. Galgut BJ, Lemberg DA, Day AS, Leach ST. The value of fecal markers in predicting relapse in inflammatory bowel diseases. *Front Pediatr* 2017;**5**:292.

26. Heida A, Kobold ACM, Wagenmakers L, van de Belt K, van Rheenen PF. Reference values of fecal calgranulin C [S100A12] in school aged children and adolescents. *Clin Chem Lab Med* 2017;**56**:126–31.

27. Tillinger W, Jilch R, Jilma B, *et al*. Expression of the high-affinity IgG receptor FcRI [CD64] in patients with inflammatory bowel disease: a new biomarker for gastroenterologic diagnostics. *Am J Gastroenterol* 2009;**104**:102–9.

28. Minar P, Haberman Y, Jurickova I, *et al*. Utility of neutrophil Fcγ receptor I [CD64] index as a biomarker for mucosal inflammation in pediatric Crohn's disease. *Inflamm Bowel Dis* 2014;**20**:1037–48.

29. Bailey CJ, Hembry RM, Alexander A, Irving MH, Grant ME, Shuttleworth CA. Distribution of the matrix metalloproteinases stromelysin, gelatinases A and B, and collagenase in Crohn's disease and normal intestine. *J Clin Pathol* 1994;**47**:113–6.

30. Matusiewicz M, Neubauer K, Mierzchala-Pasierb M, Gamian A, Krzystek-Korpacka M. Matrix metalloproteinase-9: its interplay with angiogenic factors in inflammatory bowel diseases. *Dis Markers* 2014;**2014**:643645.

31. Nowarski R, Jackson R, Gagliani N, *et al*. Epithelial IL-18 equilibrium controls barrier function in colitis. *Cell* 2015;**163**:1444–56.

32. Kumar S. Natural killer cell cytotoxicity and its regulation by inhibitory receptors. *Immunology* 2018;**154**:383–93.

33. Peyrin-Biroulet L, Panés J, Sandborn WJ, *et al*. Defining disease severity in inflammatory bowel diseases: current and future directions. *Clin Gastroenterol Hepatol* 2016;**14**:348–54.e17.

34. Ministro P, Martins D. Fecal biomarkers in inflammatory bowel disease: how, when and why? *Expert Rev Gastroenterol Hepatol* 2017;**11**:317–28.

35. Di Ruscio M, Vernia F, Ciccone A, Frieri G, Latella G. Surrogate fecal biomarkers in inflammatory bowel disease: rivals or complementary tools of fecal calprotectin? *Inflamm Bowel Dis* 2017;**24**:78–92.