Multigene Analysis Can Discriminate Between Ulcerative Colitis, Crohn's Disease, and Irritable Bowel Syndrome

PETRA VON STEIN,* ROBERT LOFBERG,‡ NIKOLAI V. KUZNETSOV,* ALEXANDER W. GIELEN,* JAN–OLOV PERSSON,§ ROLF SUNDBERG,§ KARIN HELLSTROM,* ANDERS ERIKSSON,‖ RAGNAR BEFRITS,¶ AKE OST,# and OLIVER D. VON STEIN*

*InDex Diagnostics AB, Stockholm, Sweden; ‡Department of Medicine, Karolinska Institute and IBD Unit at Sophiahemmet, Stockholm, Sweden; §Statistical Consultancy, Mathematical Statistic, Stockholm University, Stockholm, Sweden; ‖Department of Internal Medicine, Sahlgrenska University Hospital, Gothenburg, Sweden; ¶Gastroenterology and Hepatology Clinic, Karolinska University Hospital, Solna, Sweden; and #Gastrointestinal Pathology, Aleris-Medilab AB, Täby and Karolinska Institute, Stockholm, Sweden

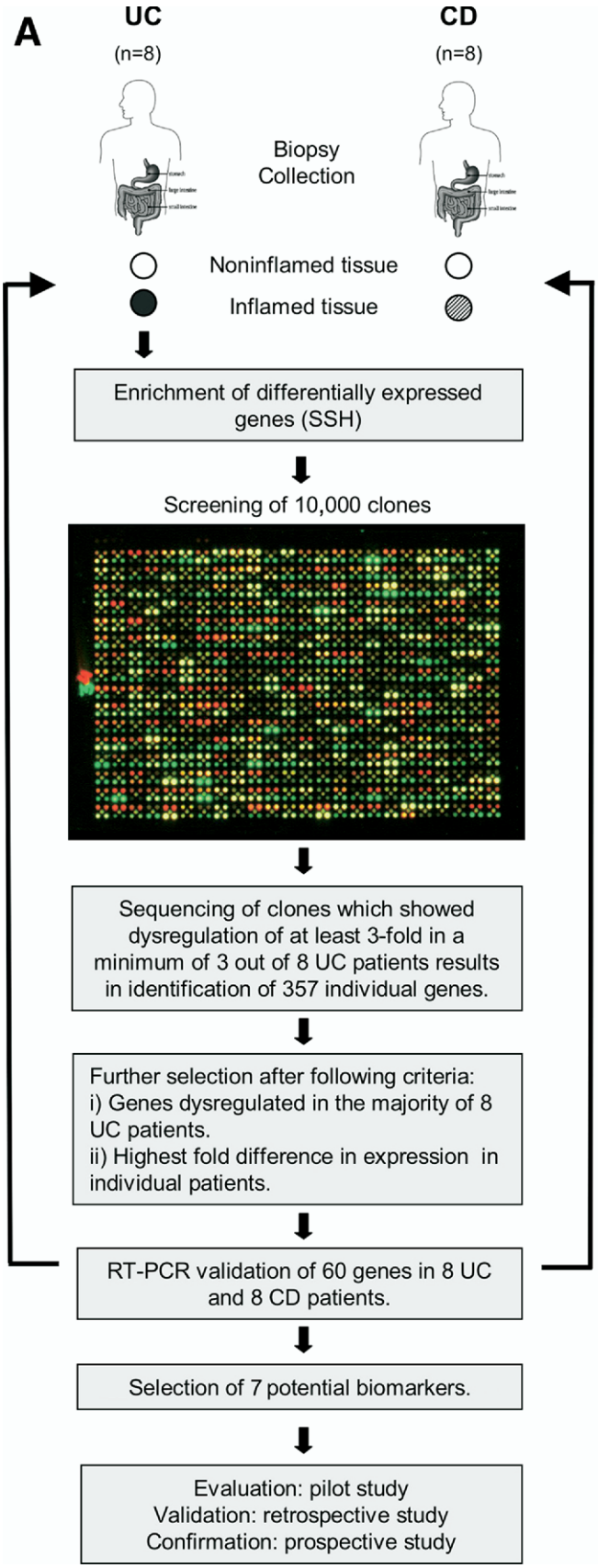
See CME quiz on page 2153; see editorial on page 2164.

Background & Aims: **Inflammatory bowel diseases (IBDs) and the irritable bowel syndrome (IBS) are heterogeneous disorders of the gastrointestinal tract and can profoundly affect the quality of life. Because many of the symptoms of IBD are similar to those of IBS, the former may be misdiagnosed. In addition, the 2 major forms of IBD, ulcerative colitis (UC) and Crohn's disease (CD), have overlapping nonspecific, pathologic features leading to difficulties in assessing colonic inflammation and hence the term *IBD unclassified* has been proposed. The aim of this study was to identify and assess the utility of a certain set of marker genes that could help to distinguish IBS from IBD, and further to discriminate between UC and CD.** ***Methods:*** **Subtractive suppression hybridization was used to identify IBD-specific genes in colonic mucosal biopsy specimens. In quantitative polymerase chain reaction experiments, the differential expressions of identified genes then were analyzed using a classification algorithm and the possible clinical value of these marker genes was evaluated in a total of 301 patients in 3 stepwise studies.** ***Results:*** **Seven marker genes were identified as differentially expressed in IBD, making it possible to discriminate between patients suffering from UC, CD, or IBS with area under the receiver-operating characteristic curves ranging from 0.915 to 0.999 ($P < .0001$) using the clinical diagnosis as gold standard.** ***Conclusions:*** **Expression profiling of relevant marker genes in colonic biopsy specimens from patients with IBD/IBS-like symptoms may enable swift and reliable determination of diagnosis, ultimately improving disease management.**

Inflammatory bowel disease (IBD) is a term encompassing several conditions involving chronic inflammation in the gastrointestinal tract with 2 major forms: ulcerative colitis (UC) and Crohn's disease (CD). Environmental factors, infectious microbes, genetic susceptibility, ethnic origin, and a dysregulated immune system are shown to be critical factors in the pathogenesis of IBD.[1] In 20%–30% of IBD cases, patients do not meet the diagnostic criteria of UC or CD and are classified as *chronic IBD* or *IBD unclassified* (IBDU).[2–4] The symptoms of IBD are relatively unspecific and only a few if any endoscopic or radiologic findings may unequivocally distinguish between UC and CD. The picture is complicated further by the fact that other gastrointestinal disorders such as *irritable bowel syndrome* (IBS) can present symptoms very similar to IBD, making the differentiation of IBS and IBD difficult. Moreover, some IBS patients may develop IBD or mild IBD and yet retain predominantly IBS-like symptoms.[5] Critical for making the distinction is a careful history assessment, various laboratory tests for general signs of inflammation (eg, C-reactive protein, erythrocyte sedimentation rate, platelet count), and endoscopic, histologic, and radiologic investigations. IBD serologic tests and fecal markers have been investigated intensively to enhance the diagnosis of IBD. Fecal markers such as lactoferrin and calprotectin are indicative for any inflammation in the gastrointestinal tract. They can discriminate between inflammatory diseases and functional disorders without inflammation but not between UC and CD. Serologic markers include autoantibodies to neutrophils (antineutrophil antibodies [ANCA], perinuclear ANCA [pANCA]) and antimicrobial antibodies (immunoglobulin [Ig]G and IgA anti-*Saccharomyces cerevisiae* antibodies [ASCA], anti-outer membrane porin C, anti-I2,

Abbreviations used in this paper: ASCA, anti-*Saccharomyces cerevisiae* antibodies; AUC, area under the receiver-operating characteristic curve; GRO-α, growth-related oncogene α; IBDU, unclassified inflammatory bowel disease; IBS, irritable bowel syndrome; MMP-7, matrix metalloproteinase 7; pANCA, perinuclear antineutrophil antibodies; qPCR, quantitative polymerase chain reaction; RegIV, regenerating protein IV; ROC, receiver-operating characteristic; SLC, solute carrier; SPAP, small protein associated with PDZ domain–containing protein 1.

0016-5085/08/$34.00
doi:10.1053/j.gastro.2008.02.083



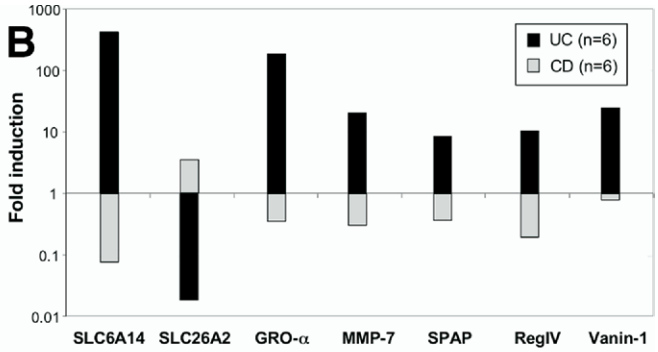


Figure 1. Screening and identification of 7 IBD-specific genes. (*A*) From 8 UC and 8 CD patients, mucosal biopsy specimens were collected from inflamed and noninflamed regions of the colon. RNA material derived from the UC patients was used to enrich dysregulated genes using the subtractive suppression hybridization (SSH) procedure. After the screening process, genes were selected for validation in 8 UC and CD patients resulting in the selection of 7 genes. (*B*) Seven potential IBD discriminating genes were found to be oppositely regulated in inflamed and noninflamed colon biopsy specimens. Plotted are the mean ΔCt value of fold induction/repression of SLC6A14, SLC26A2, GRO-α, MMP-7, SPAP, RegIV, and Vanin-1 in UC (■) and CD (□) patients.

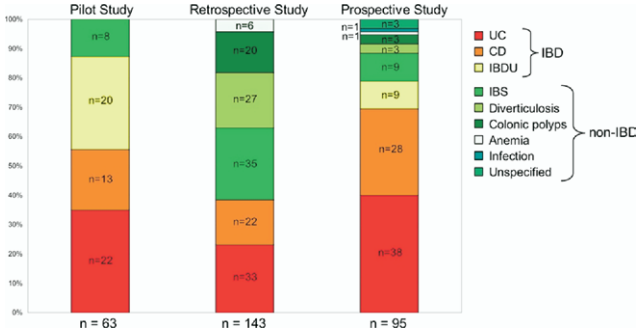


Figure 2. Overview of the included patients in the 3 performed studies.

and anti-CBir1).[6] UC commonly shows a pANCA+/ASCA- pattern whereas CD typically shows a pANCA-/ASCA+ pattern. The former combination has a reported sensitivity for UC of 44%–58% and a specificity of 81%. Similarly, the latter pattern has a sensitivity for CD of 30%–64% and a specificity of 92%–97%.[6,7] The newer generation of serologic markers such as outer membrane porin C, anti-I2, and anti-CBir1 are specific to subgroups of CD and it was shown that, for example, anti-CBir reactivity is associated significantly with some subforms of CD but is associated negatively with UC-like CD.[8] Nevertheless, the combination of different serologic markers seems to only slightly improve the performance and an increase in specificity often is observed with a decrease in sensitivity.[7,9] Currently, the overall assessment of the clinician based on endoscopy, histopathology, patient history, and radiology still remains the gold standard.[9]

A reliable method for discriminating between patients suffering from UC, CD, and IBS would allow more appropriate treatment regimens to be made on a case-by-case basis. By using the subtractive suppression hybridization technology[10,12] we could identify 7 IBD-specific genes: solute carrier (SLC)6A14, SLC26A2, small protein associated with PDZ domain–containing protein 1 (SPAP), regenerating protein IV (RegIV), Vanin-1, matrix metalloproteinase 7 (MMP-7), and growth-related oncogene α (GRO-α). By using the differential expression profiles of these genes, we developed a rapid, sensitive, and reproducible method that combines fluorescence-based real-time polymerase chain reaction (quantitative PCR [qPCR]) with a specially constructed algorithm for quantification and analysis of the 7 marker genes. By using this multigene diagnosis, not only was it possible to distinguish between IBS and IBD, but also between UC and CD with a high sensitivity and specificity as shown in 3 different studies.

Materials and Methods

Study Subjects

Pilot study. Biopsy samples were collected in a prospective manner at the IBD-unit at Sophiahemmet (Stockholm, Sweden). Fifty-five IBD patients (23 women and 32 men) with a mean age of 44.6 years (range, 19–76 y), and a mean disease duration of 4.5 years (range, 1 mo to 18 y) and 8 IBS patients were included.

Retrospective study. Complementary DNA (cDNA) samples derived from mucosal biopsy samples of 55 IBD and 88 non-IBD patients were used (65 women and 78 men), collected at the Sahlgrenska University Hospital (Gothenburg, Sweden) as described previously.[11] The mean age was 53 years (range, 21–88 y), and the mean disease duration was 12.4 years.

Prospective study. A 6-center study in Sweden was performed in accordance with good clinical practice (monitored by Trial Form Support AB) and comprised 95 patients (47 women and 48 men) with a mean age of 41.5 years (range, 20–71 y).

Biopsy Specimen Collection: Pilot and Prospective Studies

Colorectal mucosal biopsy samples were taken during routine colonoscopy from patients selected on the basis of clinical and pathologic evidence of having one of the following conditions: UC, CD, IBDU in an active state of the disease (mild, moderate, or severe), or a gastrointestinal disorder such as IBS, diverticulosis, or colorectal polyps. Up to 3 biopsy specimens were collected from an inflamed site in the colon and at standardized locations in case of disorders without signs of inflammation. For subtractive suppression hybridization and screening purposes, additional biopsy specimens from noninflamed areas of the colon were collected.

RNA Extraction and cDNA Synthesis: Pilot and Prospective Studies

Biopsy samples were homogenized using a Pellet Pestle Motor Homogenizer (Kontes) before total RNA was isolated using a Qiagen RNeasy Kit according to the manufacturer's guidelines. Two micrograms of each RNA sample were used for first-strand cDNA synthesis using 10 pmol/L of the oligo-dT-primer (5′-t_{20}nv-3′). Buffer, deoxynucleotide triphosphates, and reverse transcriptase (Superscript II) were supplied by Invitrogen and the reactions were performed according to the manufacturer's guidelines. The reaction mixture for first-strand synthesis was pre-incubated for 10 minutes at 65°C in a PCR sprint cycler (Hybaid) and chilled on ice before the enzyme Superscript II was added and incubated further for 1 hour at 42°C in a PCR cycler.

Isolation of Dysregulated Genes

Starting material was 2 μg of total RNA derived from biopsy specimens resected from inflamed and noninflamed sites of the colon from 8 UC patients. The subtractive suppression hybridization was performed as described.[10,12] Approximately 10,000 clones from the pooled subtractions were picked using BioPick (BioRo-

Table 1. Summary of All Statistical Data Analyses

A. Pilot study

Data set					
Disease	Control	AUC	Correct prediction	Sensitivity	Specificity
UC (n = 22)	CD (n = 13)	**0.962**	94%	96%	92%

B. Retrospective study

Data set					
Disease	Control	AUC	Confirmed diagnosis	Sensitivity	Specificity
UC (n = 33)	non-UC (n = 110)	**0.979**	97%	88%	99%
CD (n = 22)	non-CD (n = 121)	**0.959**	93%	70%	98%
IBD (n = 55)	non-IBD (n = 88)	**0.954**	91%	85%	94%

C. Prospective study

Data set					
Disease	Control	AUC	Correct prediction	Sensitivity	Specificity
UC (n = 38)	non-UC (n = 48)	**0.937**	85%	90%	86%
CD (n = 28)	non-CD (n = 58)	**0.915**	85%	68%	92%
IBD (n = 75)	non-IBD (n = 20)	**0.999**	98%	95%	95%

NOTE. Discriminating potential was estimated through ROC analysis of a disease vs a control group compared to the clinical diagnosis. Depicted are the areas under the ROC curves (AUC) and the correct prediction, sensitivity, and specificity values obtained with the conventional cut-off value of 0.5 (50%). All analyses had a P value of $< .0001$.

botics) and amplified as described.[13,14] PCR products were spotted on Hybond N^+ membrane (Amersham) using MicroGrid TAS (BioRobotics), and filters were screened as described with 16 subtractive probes derived from single subtractions of 8 UC patients.[13,14] The expression data were analyzed using ArrayVision software 6.0 (Imaging Research Inc). Selected genes were validated using reverse-transcriptase PCR of the original UC patients and 8 additional CD patients. Accession numbers were as follows: NM_007231 (SLC6A14), NM_000112 (SLC26A2), NM_001511 (GRO-α), NM_004666 (Vanin-1), BC003635 (MMP-7), BC012303 (SPAP), and BC017089 (RegIV).

Real-Time PCR (qPCR)

The analyses were performed on a 7000 (pilot study) or a 7500 real-time PCR system (Applied Biosystems) using SYBR Green PCR Kits (Eurogentec; Applied Biosystems) according to the manufacturer's guidelines. Reactions were performed in triplicates using 1 μL 1:10 diluted cDNA per reaction. The data of the real-time analysis were analyzed using SDS 1.3 software (Applied Biosystems). γ-actin (5′-gtgcagggtattaacgtgtcaggg–3′/5′-ccaactcaaagcaagtaacagcccacgg–3′); SLC6A14 (5′-gagcaaagaggtggatattctggc–3′/5′-ctccccagtcagggtatggaattg–3′); SLC26A2 (5′-cacctaaagctattatgcaggagg–3′/5′-ctcctcaattcatgacctgtgggc–3′); GRO-α (5′-gccaatgagatcattgtgaaggca–3′/5′-caacatgagaaatgttgaccacac–3′); MMP-7 (5′-cactgttcttccactccatttagc–3′/5′-gacatctacccactgcaagtatag–3′); SPAP (5′-gttcctggtcctcgttgcaatcgc–3′/5′-ccatcgaagagtaccttccatctg–3′); RegIV (5′-ggtgatatcatcatgagacccagc–3′/5′-ctttaaactcaggatagatgccag–3′); Vanin-1 (5′-ccaactgactgatagactctgagc–3′/5′-ggcatagatcactactgcaagtgc–3′); lactoferrin (5′-gacccttgatggtggtttcatata-3′/5′-cataatagtgagttcgtggctgtctt-3′).

Data Analysis and Statistics

STATA (Stata Corp LP) was used for the statistical calculations, leading up to algorithms and class membership probabilities. Receiver-operating characteristic (ROC) plots were constructed using Prism 4 (GraphPad Inc) and box plots were constructed using DeltaGraph 5 (Redrock).

Pilot study. To classify a patient as UC or CD a linear discriminant function was computed using qPCR data (ΔCt values) of all 7 markers (d1–d7) under the assumption of normal distribution of the data (common covariance matrix) and an equal proportion of UC and CD in the population.[15] Based on the first principal components of the same data set and assumptions, a classification algorithm was constructed to calculate the probability of IBDU and IBS patients to have UC.

Retrospective and prospective studies. qPCR data (ΔCt values) were used to classify the patients into 3 diagnoses (UC, CD, and non-IBD). The classification algorithm used was based on a quadratic discriminant function under the assumption of normally distributed data and equal population proportions but without a common covariance matrix.[15]

Table 2. Multigene Diagnosis of IBDU Patients From the Pilot Study

Patient #	Patient code	d1	d2	d3	d4	d5	d6	d7	Probability of having UC	Multigene diagnosis	Clinical diagnosis (after 4–12 months)
44	IBDU-1	8.82	4.44	7.88	7.73	4.94	4.2	9	0.250	CD	CD
45	IBDU-2	7.5	4.33	5.07	6	3.96	2.16	6.56	0.758	UC	IBDU
46	IBDU-3	5.07	7.17	4.39	6.17	3.42	2.41	6.37	0.928	**UC**	**CD**
47	IBDU-4	20.8	7.86	3.08	9.82	9.24	12.74	7.81	0.006	CD	CD
48	IBDU-5	4.44	3.8	3.99	6.16	2.14	1.57	5.15	0.931	UC	UC
49	IBDU-6	7.54	4.73	7.43	7.4	4.17	1.72	7.94	0.543	UC	UC
50	IBDU-7	6.86	6.27	5.58	9.23	3.86	4.68	9.02	0.474	CD	CD
51	IBDU-8	7.14	3.3	5.48	7.49	4.27	4.31	8.79	0.439	CD	IBDU
52	IBDU-9	6.09	7.39	5.07	7.55	3.35	1.78	7.06	0.857	UC	UC
53	IBDU-10	8.87	3.1	8.85	9.28	6.01	2.41	10.44	0.108	CD	IBDU
54	IBDU-11	6.28	5.77	0.67	4.78	4.15	2.65	6.14	0.950	UC	IBDU
55	IBDU-12	4.7	7.31	2.89	7.07	1.9	2.42	6.47	0.947	UC	UC
56	IBDU-13	5.14	5.54	3.96	4.16	1.26	1.32	5.36	0.969	UC	IBDU
57	IBDU-14	5.15	5.98	3.99	7.56	2.99	1.66	7.29	0.879	UC	IBDU
58	IBDU-15	3.7	5.52	1.51	4.27	1.92	0.52	6.2	0.985	UC	IBDU
59	IBDU-16	12.46	1.31	7.41	10.9	5.06	2.98	10.76	0.024	CD	IBDU
60	IBDU-17	4.86	7.28	1.9	5.06	2.92	1.43	6.13	0.976	UC	UC
61	IBDU-18	4.59	5.94	3.55	4.26	2.98	−1.3	6.05	0.979	UC	IBDU
62	IBDU-19	4.74	7.23	2.93	3.54	3.8	1.79	8.1	0.969	UC	UC
63	IBDU-20	5.36	5.24	3.52	5.24	4	3.02	6.38	0.911	UC	IBDU

NOTE. The classification algorithm derived from firm diagnosed CD and UC cases from the pilot study were used to determine the probability of having UC from raw data values (d1–d7) of IBDU cases. If probability >0.5 the diagnosis was UC, otherwise CD. After 4–12 months 10 of 20 patients were given a firm clinical diagnosis whereby in 9/10 cases the multigene analysis predicted this diagnosis correctly (90% correct prediction). **Bold** denotes misclassification.

Calculated probabilities were used as values for ROC plots. The diagnosis, based on the overall assessment of the treating clinician, served as the gold standard and was based on the patient's disease history, colonoscopy, and histopathology data. All ROC plots were performed using a 95% confidence interval.

Ethical Considerations

The pilot and prospective studies were approved by the Regional Ethical Committee at the Karolinska Institutet in Stockholm. The retrospective analysis was approved by the Regional Ethical Committee in Gothenburg.

Results

Screening for IBD-Specific Genes

To identify IBD-specific genes we used a screening procedure with the subtractive suppression hybridization method, which enables the enrichment and isolation of dysregulated genes by comparing 2 cDNA populations,[10,12] in this case inflamed and noninflamed biopsy samples of 8 UC patients. For each patient, 2 subtractions were performed, enabling the isolation of both up-regulated and down-regulated genes specific to the inflamed biopsy specimens (Figure 1*A*). This resulted in the identification of 357 dysregulated genes. The expression levels of 60 genes, selected on the basis of their strong up-regulation or down-regulation in at least 4 of the 8 UC patients, were validated by reverse-transcription PCR analysis in the 8 original UC patient samples. In addition, these 60 genes also were analyzed by reverse-transcription PCR in 8 CD patient samples to further specify their expression pattern. Interestingly, 7 of the genes showed an opposite expression pattern in UC patients compared with CD patients in at least 6 of 8 UC and CD patients (Figure 1*B*). Four of these genes (GRO-α, MMP-7, RegIV, and Vanin-1) are known to be involved in inflammation, 2 other genes (SLC6A14 and SLC26A2) belong to the large family of solute carrier energy-dependent transport molecules, and, last, SPAP was involved, with a probable function in cross-membrane transport.

Pilot Study

This explorative study was performed to evaluate the potential of the 7 genes as a tool for a diagnostic method. It included in total 63 patients (Figure 2) and was performed in 2 parts. In the first part, 1–3 biopsy specimens collected from 22 active UC and 13 active CD cases were analyzed via qPCR. The ΔCt values of the 7 marker genes were used to design a classification algorithm that collectively considered the expression status of all 7 markers. With a conventional cut-off value at class membership probability of 0.5 (UC, >0.5; CD, <0.5), the diagnosis of UC or CD was predicted correctly in 33 of all 35 IBD patients (94%), with a sensitivity for UC of 96% (21 of 22) and a specificity of 92% (12 of 13), meaning that in 96% of UC patients the method predicted correctly that the patient was afflicted with UC and in 92% of CD patients the method predicted correctly that the patient was afflicted with CD (Table 1). To validate a

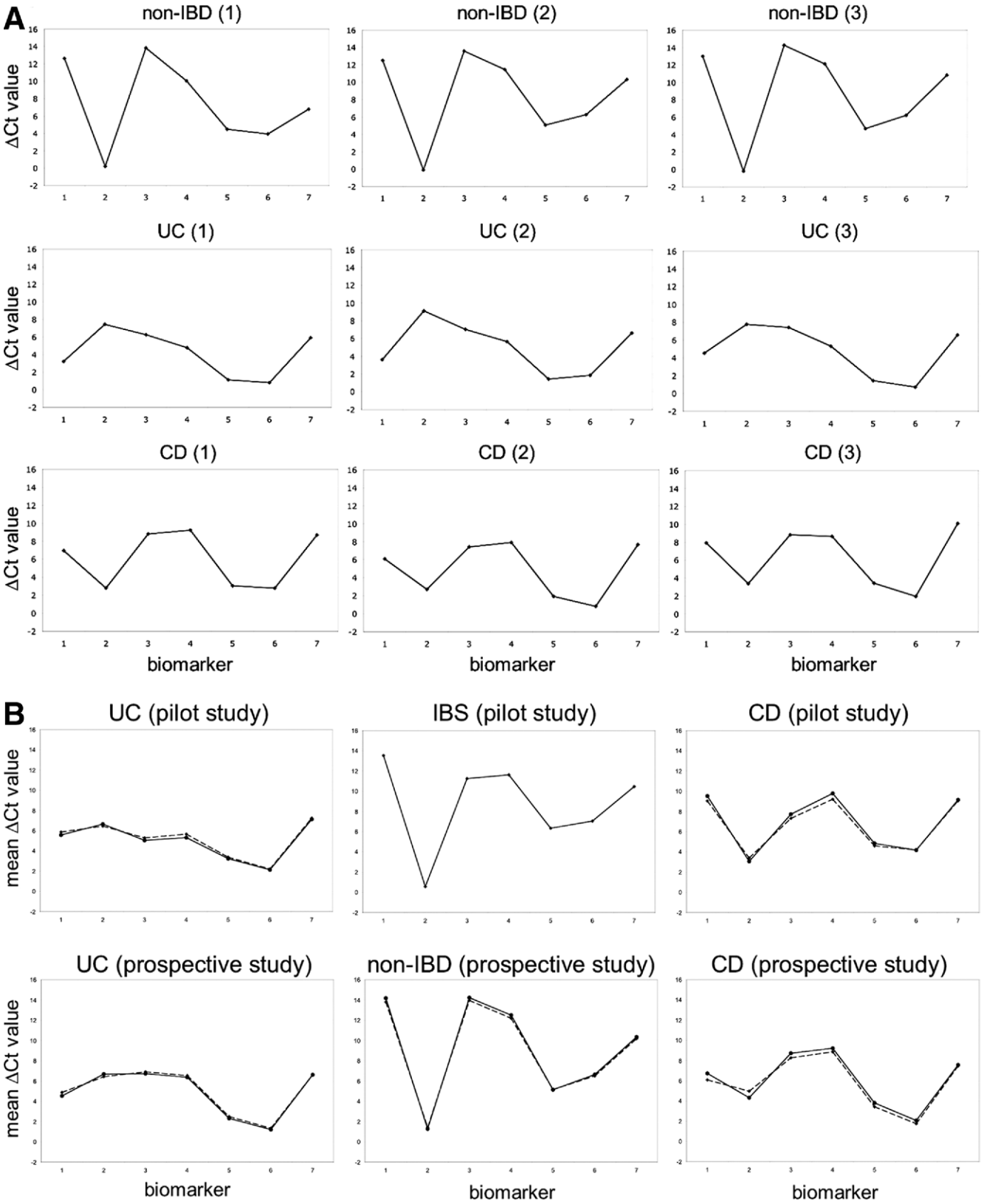


Figure 3. Each disease has a typical expression signature. Depicted are ΔCt values for each of the marker genes plotted horizontally to obtain a specific profile pattern. (*A*) Expression signatures of 3 UC, 3 CD, and 3 IBS patients. (*B*) Expression signatures derived by plotting the mean ΔCt values (*y-axis*) against the individual markers (*x-axis)* obtained from the pilot and prospective study. The qPCR derived ΔCt values were grouped according to the clinical diagnosis (*dashed line*) or according to the multigene algorithm outcome (*solid line*) and the mean values of each group for all 7 markers were plotted. (*C*) Expression signatures of inflamed tissue of other gastrointestinal diseases. Shown are 4 cases of collagenous colitis, 2 cases of diverticulitis, and 1 case each of infectious colitis and solitary rectal ulcer (*red line/full circle*). As references, the mean of 5 healthy volunteers (*green line/open square*) were used as well as the mean of all UC cases of prospective study (*dashed line*), the mean of all CD cases of prospective study (*dotted line*) and a noninflamed control of solitary rectal ulcer (*blue line/open square*). [1]SLC6A14, [2]SLC26A2, [3]GRO-α, [4]MMP-7, [5]SPAP, [6]RegIV, [7]Vanin-1.

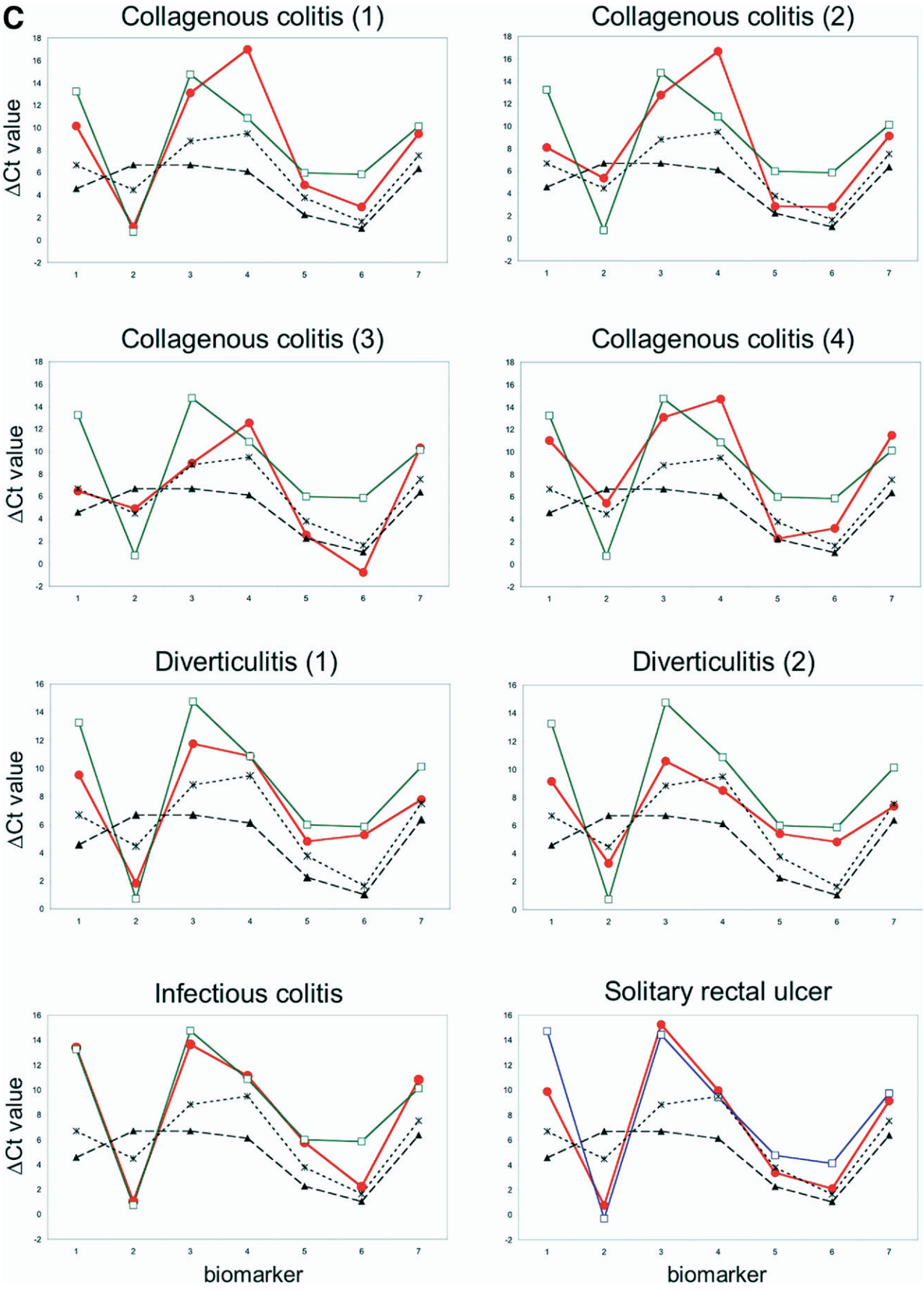


Figure 3. Continued.

Table 3. Multigene Diagnosis of Selected Examples From the Retrospective Study

#	Classification	Diagnosis	d1	d2	d3	d4	d5	d6	d7	Probability of having P(CD)	Probability of having P(non-IBD)	Probability of having P(UC)	Multigene diagnosis
1	non-IBD	Anemie D649	9.822	−0.243	15.846	11.867	3.456	2.712	10.177	0.01	0.99*	0	non-IBD
2	non-IBD	Anemie D649	9.883	0.442	14.162	13.363	3.84	3.356	8.703	0	0.98	0.02	non-IBD
3	non-IBD	Diverticulosis K579	9.156	0.324	12.559	12.339	3.349	1.765	9.995	0.13	0.87	0	non-IBD
4	non-IBD	Diverticulosis K579	10.346	1.235	16.324	13.85	3.966	2.124	7.513	0	0.99	0.01	non-IBD
5	non-IBD	Polyp colon K635	9.333	0.083	13.104	13.368	4.277	4.72	7.722	0	0.99	0.01	non-IBD
6	non-IBD	Polyp colon K635	9.952	−0.346	13.126	12.963	4.158	4.016	9.29	0	1	0	non-IBD
7	non-IBD	IBS K599	13.085	3.395	17.848	15.811	6.731	4.465	11.592	0	1	0	non-IBD
8	non-IBD	IBS K599	11.417	0.392	13.787	10.541	0.119	2.995	8.775	0.1	0.89	0.01	non-IBD
1	active CD	CD colon K501	5.311	1.857	10.641	11.153	1.692	0.37	8.895	0.96	0.01	0.04	CD
2	active CD	CD colon K501	13.13	0.955	13.676	9.978	3.674	3.183	7.852	0.61	0.39	0	CD
3	active CD	CD combined K508	7.886	2.472	9.139	11.876	2.189	−1.987	5.436	1	0	0	CD
4	active CD	CD combined K508	7.344	1.682	10.871	12.008	4.451	1.874	10.049	0.81	0.19	0	CD
1	active UC	Pancolitis K510	5.04	2.748	10.654	11.139	0.406	−1.591	8.860	0.35	0	0.65	UC
2	active UC	Pancolitis K510	3.15	4.086	7.257	10.25	0.4	−0.318	4.967	0.04	0	0.96	UC
3	active UC	Proctitis K512	5.575	6.142	5.564	8.355	0.885	−0.171	6.46	0	0	1	UC
4	active UC	Proctitis K512	4.683	6.697	8.512	9.882	1.708	0.394	8.366	0	0	1	UC
5	active UC	Proctosigmoiditis K513	5.849	5.178	12.491	14.205	1.982	−0.781	7.953	0.03	0	0.97	UC
6	active UC	Proctosigmoiditis K513	5.617	7.094	9.868	10.56	2.536	−0.64	8.1	0	0	1	UC
1	inflamed control	Diverticulitis K578	9.927	0.337	12.934	13.254	4.22	5.728	7.354	0.03	0.96	0.1	non-IBD
2	inflamed control	rectal ulcer K626	9.877	0.739	15.26	9.957	3.36	2.091	9.128	0.15	>0.84	0.01	non-IBD
3	inflamed control	infectious colitis A047	13.407	1.032	13.646	11.121	5.785	2.209	10.822	0.01	0.99	0	non-IBD
4	inflamed control	collagen colitis K528	8.101	5.367	12.781	16.667	2.857	2.798	9.123	0.28	0.71	0.01	non-IBD

NOTE. The classification algorithm derived from raw data values (d1–d7) of the retrospective study was used to determine the probability of having UC, CD, and non-IDB. Examples are depicted for each disease class (non-IBD, CD, UC, inflamed controls). The probability of patient #1 of the non-IBD class is 0.99 (*) indicating a 99% chance of not having IBD.

diagnostic method, it should be compared with the commonly used gold standard, which usually is performed by ROC analysis.[16] The area under the ROC curve (AUC) is a measure for the value of the diagnostic test. For example, the best possible prediction method would yield an AUC of 1.0, representing 100% sensitivity and 100% specificity. Thus, the probability values derived from the classification model were compared with the clinical diagnosis in a ROC plot. This resulted in an AUC of 0.962 with an indicated *P* value of less than .0001, illustrating high significance.

In the second part, qPCR data of the 7 markers derived from 20 active IBDU patients were analyzed. In this case a classification algorithm based on the first principal components of the qPCR data from the first part were implemented into a Microsoft Excel format and the probabilities of having UC were calculated. By using the cut-off value of 0.5 (50%), a multigene diagnosis could be given (Table 2). This diagnosis was compared against the clinical diagnosis concluded by the blinded, treating physician some 4–12 months later. In this way, 9 of 10 cases were predicted correctly after later clinical examinations in which the treating physician could provide a firm diagnosis. Despite later clinical evaluations, no firm diagnosis could be made for the remaining 10 IBDU patients and, consequently, the multigene diagnosis for these patients could not yet be confirmed. Moreover, the same criterion was used on biopsy specimens from 8 IBS patients, and they all were classified correctly as certain non-UC cases (probability for UC, <.02). The difference in the expression profiles between UC, CD, and IBS can be visualized by plotting the actual ΔCt values for the 7 markers per patient (Figure 3*A*). Illustrated are profiles derived from 3 individual patients per indication and each disease appears to have a distinct expression signature that is remarkably similar for patients of the same disease. This finding is supported in Figure 3*B* where the expression signatures of the disease groups derived from the mean ΔCt values of the pilot and prospective study are plotted. Also shown is the comparison of the progression of the signatures when the ΔCt raw data were grouped according to the multigene algorithm or the clinical diagnosis. The close resemblance of the curves indicates again that these expression signatures seem to be characteristic for each patient group. Interestingly, other rare gastrointestinal inflammatory diseases seem to have expression signatures not typical for either UC or CD, suggesting that they could constitute their own specific signature (Figure 3*C*). The multigene diagnosis appears to confirm this possibility because such cases were classified as non-IBD (Table 3). However, further studies are required addressing larger cohorts of these rare gastrointestinal diseases to enable a proper evaluation.

Retrospective Study

In this study, the utility of all markers as a multigene diagnostic tool was validated further. A total of 143 cDNA samples were received from a previously established biobank derived from biopsy specimens from 55 active IBD patients (33 UC and 22 CD patients) and 88 non-IBD patients diagnosed with IBS, diverticulosis, colonic polyps, or anemia (Figure 2). The material was analyzed by qPCR of the marker genes, followed by computing a new classification algorithm that could now calculate the probability of having UC, CD, or neither of them (non-IBD), instead of only the probability of having UC as in the pilot study. In Table 3 examples for each patient group are depicted. Patient 1 of the non-IBD patients, for example, has a probability of 99% to not be afflicted with UC or CD. To measure the effectiveness of the method different ROC plots were performed analyzing UC against non-UC, CD against non-CD, and IBD against non-IBD (Table 1). This resulted in AUC values of 0.979, 0.959, and 0.954, respectively. In Table 1 the correct prediction and sensitivity/specificity at a cut-off value of 50% are summarized. In all cases the specificity was very high and the sensitivity was somewhat lower, but by changing the cut-off value of the UC/non-UC discrimination to 0.2 the sensitivity and specificity are both 94%. Taken together, the results of the pilot study could be confirmed in a larger cohort retrospective study, and the possibility to effectively discriminate IBD from non-IBD patients was validated.

Prospective Study

To confirm data derived from the pilot and the retrospective studies, a controlled and monitored prospective study was performed in a blinded fashion. A total of 95 patients were included in the study and the respective numbers of patients in each disease indication is shown in Figure 2. Biopsy specimens were collected from patients diagnosed with either UC or CD in an active state of the disease for re-evaluation of the diagnosis, or patients with IBD-like symptoms without any former diagnosis to evaluate the disease. These could be cases of IBD (UC, CD, and IBDU) as well as non-IBD cases such as IBS, diverticulosis, or colonic polyps. Patient history, endoscopic examinations, and results of histopathologic assessments formed the basis for the clinical diagnosis and served as the gold standard. In detail, 30% of the included patients were not previously diagnosed and another 30% had a change in their diagnosis from the time point of inclusion to the final assessment. The biopsy specimens were analyzed for their expression profile by using the described marker genes in qPCR analysis and the probabilities for having UC, CD, and non-IBD were calculated using the same classification algorithm as in the retrospective study. The probability values were depicted in ROC plots (Figure 4) to show the effectiveness of the method. UC could be distinguished from non-UC with an AUC of 0.937, CD from non-CD with an AUC of 0.915, and IBD from non-IBD with an AUC of 0.999, confirming the high discriminating potential of the method.

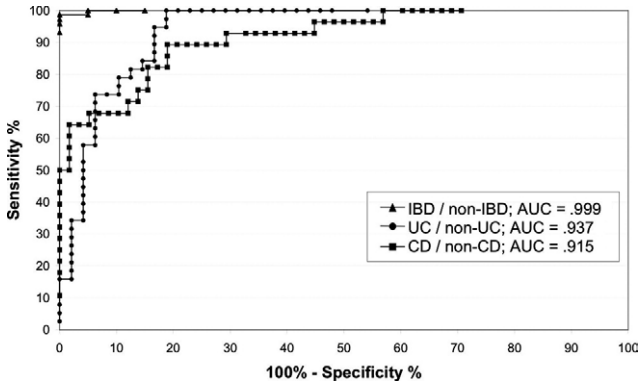


Figure 4. ROC plots of the prospective study. The calculated probabilities of having UC or CD or non-IBD derived from the multigene analysis of the 7 marker genes were compared with the overall clinical assessment as gold standard. Each ROC plot had a *P* value of less than .0001.

Discriminating Potential of the Individual Markers

In order to gain insight regarding the discriminating potential of each marker on a singular basis, ΔCt values were plotted in a box plot format for the 3 disease indications for the pilot and the prospective study (Figure 5). For comparison, the expression levels of lactoferrin, a fecal marker for inflammation, was analyzed. In the pilot study, the expression levels differ substantially enough to allow the discrimination between UC and IBS. However, lactoferrin mRNA expression levels show considerable overlap when considering UC versus CD and a slight overlap with CD versus IBS, indicating little likelihood of discriminating between these indications at the mRNA level. The analysis of the 7 markers revealed a good separation with only minor overlapping for several of the markers (eg, SLC6A14, GRO-α, SPAP, RegIV), suggesting that a number of these markers, when considered singularly, had already significant potential to discriminate between these 3 disease groups. In thorough statistical analysis, all possible combinations of the markers were tested for their diagnostic potential. The combination of SLC6A14 and RegIV for example, gave a correct prediction of 91% with a sensitivity of 95% and a specificity of 85% (cutoff 0.5). However, the use of all 7 markers gave, as expected, the best results. Box plot analyses on the prospective study data indicated a good UC/CD discriminating potential for SLC6A14, SLC26A2, GRO-α, and SPAP and a good IBD/non-IBD discriminating potential for SLC6A14, GRO-α, and RegIV. The lactoferrin control had in this study only a minor IBD/non-IBD discriminating potential.

Discussion

Gastrointestinal disorders such as IBD can dramatically affect quality of life,[17] and the diagnosis can be a

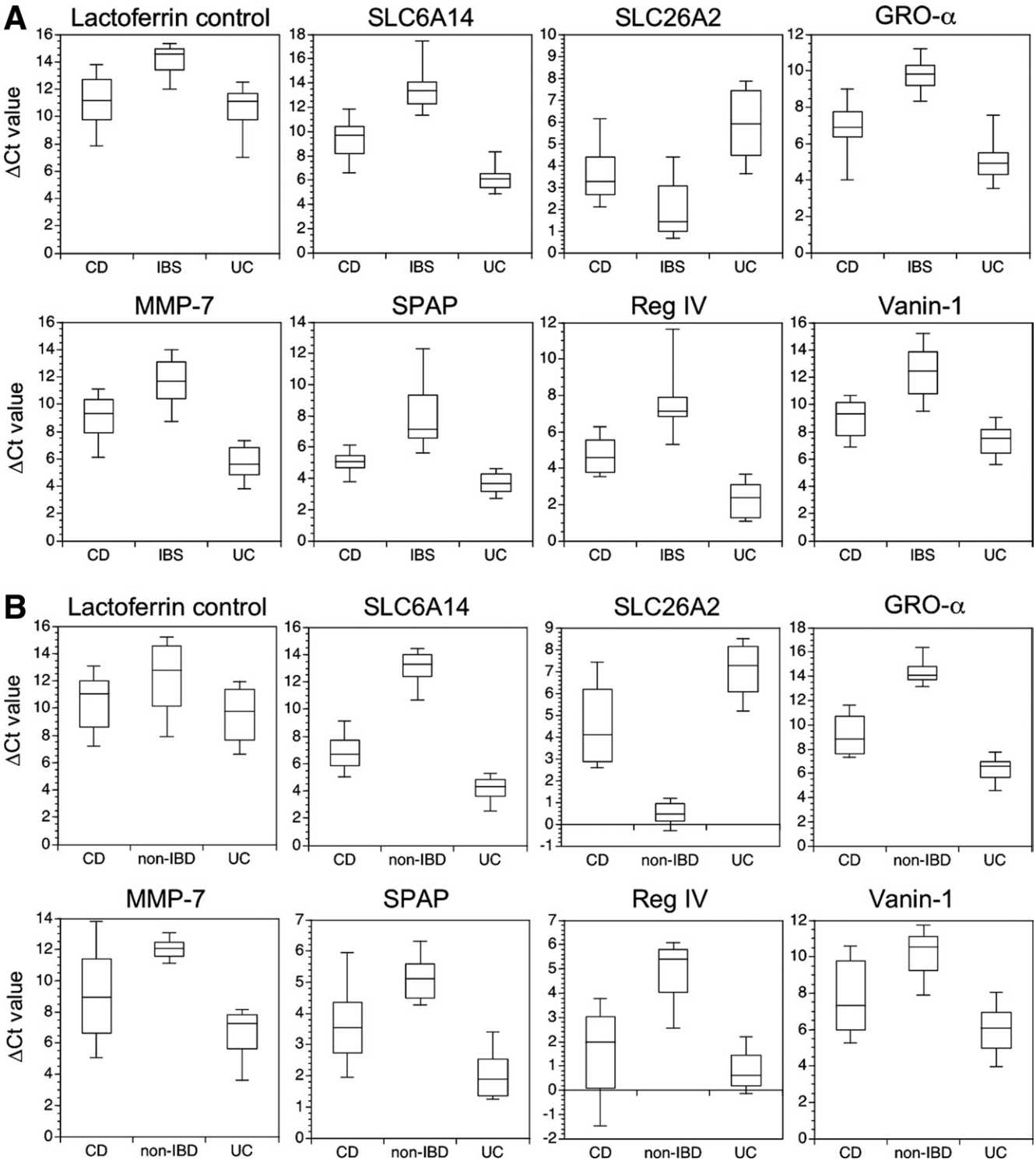


Figure 5. Individual box plot analysis of all 7 markers in comparison to the control lactoferrin. Patient cDNA was analyzed by qPCR and the resulting ΔCt values plotted per disease (CD, IBS/non-IBD, and UC). Shown are the median, the upper and lower quartiles, and the whole distribution of the values of all included patients per group (*A*) Pilot study CD (n=13), IBS (n=8), UC (n=22). (*B*) Prospective study CD (n=22), non-IBD (n=88), UC (n=55).

challenging process. In up to 30% of patients, only a temporary diagnosis of IBDU can be made. An early and accurate assessment of the diagnosis could have important clinical implications in the choice of medical treatment and for surgery considerations (eg, when colectomy, and later construction of a pelvic pouch, will be contemplated). Genetic and serologic studies have shown heterogeneity of UC and CD. With new arising techniques, it is important to improve

the accuracy in diagnosis and prognosis of the disease to enable the most appropriate choice of treatment.[18]

In this article, we introduce a method that is capable of differentiating IBD from non-IBD patients and, in addition, distinguishing with a high probability if the patient is suffering from UC or CD. The first step was the identification of 7 genes as IBD-specific genes by a special isolation and screening approach. SLC14A6 and SLC26A2 belong to a large family of solute carrier energy-dependent transport molecules that are responsible for amino acid or sulfate transport. Solute carriers are known to be associated with IBD. Gene variants of SLC22A4 (OCTN1) and SLC22A5 (OCTN2) are located in the IBD5 locus with evidence for an association with CD[19,20] and UC.[21] During our studies other groups also could show that several solute carriers are dysregulated in IBD.[11,22] Interestingly, solute carrier transport molecules such as SLC26A6 and OCTNs are regulated by PDZ domain–containing protein 1,[23,24] an adaptor protein known to interact in vivo with another marker gene, SPAP. The connection of SPAP to PDZ domain proteins and membrane-bound location suggest a function in intestinal ion transport.[25] Vanin-1 is a GPI-anchored pantetheinase, highly expressed in gut and liver.[26] It is involved in the regulation of inflammation and tissue repair and it has been shown that Vanin-1–deficient mice are protected against colitis.[26,27] RegIV is a member of the regenerating gene family that are involved mainly in liver, pancreatic, gastric, and intestinal cell proliferation and differentiation.[28] They play a role in tissue injury, inflammation, and carcinogenesis. RegIV originally was identified in a high-throughput sequence analysis of a large IBD library.[29] MMP-7 is a metalloproteinase belonging to a family of extracellular matrix–degrading endopeptidases that play a role in extracellular matrix turnover and are recognized as important players in inflammation and carcinogenesis[30,31] and recently, reports suggesting their involvement in the pathogenesis of IBD have been published.[32] GRO-α is a CXC chemokine with pleiotropic effects in immunity and is made at the site of inflammation to recruit and activate neutrophils.[33] It is one of many proinflammatory chemokines produced by human colon epithelial cells during the acute mucosal inflammatory response.[34] Recent research has shown that serum levels of GRO-α are increased in patients with IBD.[35] Collectively, the marker genes of the presented IBD diagnostic method seem to play different roles either in the pathogenesis or as effectors in IBD. Their differential expression pattern in UC and CD could be derived from different immunophenotypes.[36–38] Three of the genes (Vanin-1, RegIV, and MMP-7) also are involved in mucosal turnover and could serve as markers for mucosal healing. It has been suggested recently to use mucosal healing as primary outcome measure in future treatment studies.[39]

After having identified this variety of genes, the intention was to find out if a multigene analysis with these markers could serve as a diagnostic tool for patients presenting with IBD-like symptoms. In a pilot study, qPCR data derived from multigene analysis of firm diagnosed UC and CD cases were taken to calculate the probabilities of having UC using an implemented classification algorithm. These probabilities then were plotted in ROC curves and compared with the clinical diagnosis to illustrate the effectiveness of the method. The calculated AUC of 0.962 is, to the best of our knowledge, the highest reported AUC value for the discrimination between UC and CD. Recently, a new set of CD-specific serologic markers were evaluated (antilaminaribioside carbohydrate antibody, antichitobioside carbohydrate antibody, antimannobioside carbohydrate antibody, and outer membrane porin) and compared with gASCA (IgG ASCA) and pANCA. The addition of a further serologic marker, antilaminaribioside carbohydrate antibody, to gASCA/pANCA slightly improved the AUC to 0.809 from a value of 0.795 derived from just gASCA/pANCA alone.[9] This shows clearly the discriminative potential of this proposed method as a diagnostic tool. But, is the method also able to predict a diagnosis in IBDU patients? To answer this question, the qPCR data were analyzed using the implemented classification algorithm. In 90% of the later firm diagnosed cases, the multigene diagnosis could correctly predict the clinical diagnosis. By contrast, using pANCA/ASCA in IBDU patients gave a sensitivity/specificity for CD of 67%/78% and for UC of 78%/67%. Analysis of pANCA/ASCA in IBDU patients may be of limited use because half of the IBDU patients were seronegative.[40] The proposed method could provide a more robust tool in diagnosing IBDU patients, but this has to be confirmed in larger cohort studies over a longer period of time. Nevertheless, the IBDU data derived from the pilot and the prospective studies interestingly gave no separate expression signature specific for IBDU patients. They match the signatures of either UC or CD, sometimes with a slight reduction of the probability for UC or CD (Table 2). We could observe that cases diagnosed with UC for several years with sudden appearance of giant cells (CD phenotype) seem to have a specific expression signature not typical for CD. So far, the best predictors for such cases of changed diagnosis are nonbloody diarrhea at initial presentation and more than 10% weight loss.[41] Expression profiling of a much bigger cohort of patients could provide additional data because subgrouping of patients seems to be important to ultimately identify the primal therapeutic targets in appropriate subgroups of patients.[42]

Although the discrimination between UC and CD by the multigene analysis in both the pilot and retrospective studies indicated high sensitivity/specificity, true diagnostic potential of the described genes can be shown only when a variety of non-IBD patients also are considered.[43] To address this issue, IBS, diverticulosis, colonic polyps, and anemia patients were analyzed in addition to the

well-defined IBD patients and the calculations were performed for each disease group. The separation of IBD from non-IBD with an AUC of 0.954 indicates also a high specificity of these markers for IBD. By comparison, a study analyzing gASCA/pANCA/antilaminaribioside carbohydrate antibody achieved an AUC of 0.849 for the discrimination between IBD and non-IBD.[9]

In the prospective study, biopsy specimens were collected during the disease evaluation process and analyzed as previously described. The multigene diagnosis then was compared with the overall assessment of the clinicians by ROC plots. The discrimination between IBD and non-IBD patients with an AUC of nearly 1 (0.999) exemplified the high specificity but also the high sensitivity for IBD. Regarding the IBD patients, 85% of all cases were predicted correctly. Of interest was the observation of an apparent discordance of 31% between the overall clinical assessment and the histopathologic diagnosis. The multigene diagnosis had a concordance of 70% with the histopathology assessment, indicating that the presented method is capable of specifying a diagnosis before any specific features in the histopathology are detectable.

In summary, the 3 described studies have indicated an obvious utility of applying a multigene approach to distinguish between disease sets of similar symptoms. Although the identification of the described marker genes derived on the basis of their dysregulated expression in inflamed biopsy samples from UC patients, the studies presented here clearly indicate their usefulness at discriminating IBD from non-inflammatory conditions with high sensitivity and specificity. Moreover, in cases of suspected IBD, the described marker genes differed sufficiently in their expression between UC and CD that the 2 diseases could be segregated effectively with high sensitivity and specificity. Lastly, and perhaps of greater clinical importance, are cases of IBDU in which the need to reach a firm diagnosis as early as possible is crucial. The method illustrated here, demonstrated in the pilot study that a correct proposal of diagnosis could be reached in 90% of cases in the absence of sufficient clinical evidence. However, we acknowledge that the conclusions presented here are drawn from a limited number of patient samples and additional studies are required addressing also inflammatory conditions of the bowel other than IBD.

Nevertheless, the diagnostic method presented here could prove to be a useful addition to the current repertoire of clinical measures routinely employed in the classification of true IBD cases, and also for early discrimination from the more prevalent non-IBD cases. It would complement existing practices and could be adopted easily on a routine basis.

References

1. Baumgart DC, Carding SR. Inflammatory bowel disease: cause and immunobiology. Lancet 2007;369:1627–1640.
2. Satsangi J, Silverberg MS, Vermeire S, et al. The Montreal classification of inflammatory bowel disease: controversies, consensus, and implications. Gut 2006;55:749–753.
3. Martland GT, Shepherd NA. Indeterminate colitis: definition, diagnosis, implications and a plea for nosological sanity. Histopathology 2007;50:83–96.
4. Mitchell PJ, Rabau MJ, Haboubi NY. Indeterminate colitis. Tech Coloprocol 2007;11:91–96.
5. Talley NJ. Irritable bowel syndrome. Intern Med J 2006;36:724–728.
6. Bruining DH, Loftus EV. Evolving diagnostic strategies for inflammatory bowel disease. Curr Gastroenterol Rep 2006;8:478–485.
7. Bossuyt X. Serologic markers in inflammatory bowel disease. Clin Chem 2006;52:171–181.
8. Papadakis KA, Yang H, Ippoliti A, et al. Anti-flagellin (CBir1) phenotypic and genetic Crohn's disease associations. Inflamm Bowel Dis 2007;13:524–530.
9. Ferrante M, Henckaerts L, Joossens M, et al. New serological markers in inflammatory bowel disease are associated with complicated disease behaviour. Gut 2007;56:1394–1403.
10. von Stein OD. Isolation of differentially expressed genes through subtractive suppression hybridization. Methods Mol Biol 2001;175:263–278.
11. Flach CF, Eriksson A, Jennische E, et al. Detection of elafin as candidate biomarker for ulcerative colitis by whole genome microarray screening. Inflamm Bowel Dis 2006;12:837–842.
12. Diatchenko L, Lau YC, Campbell AP, et al. Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. Proc Natl Acad Sci U S A 1996;93:6025–6030.
13. von Stein OD, Thiess WG, Hofman M. A high throughput screening for rarely transcribed differentially expressed genes. Nucleic Acids Res 1997;25:2598–2602.
14. Wiklund I-M, Kuznetsov N, Lofberg R, et al. Identification of WASP and FKBP-like (WAFL) protein as a potential candidate gene in inflammatory bowel disease by subtractive suppression hybridization approach. Gut (submitted).
15. Naes T, Isaksson T, Feran T, et al. A user friendly guide to multivariate calibration and classification. NIR Publications, Chichester, UK, 2002.
16. Zweig MH, Campbel G. Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. Clin Chem 1993;39:561–577.
17. Oxelmark L, Magnusson A, Löfberg R, et al. Group-based intervention program in inflammatory bowel disease patients: effects on quality of life. Inflamm Bowel Dis 2007;13:182–190.
18. Sands BE. Inflammatory bowel disease: past, present, and future. J Gastroenterol 2007;42:16–25.
19. Peltekova VD, Wintle RF, Rubin LA, et al. Functional variants of OCTN cation transporter genes are associated with Crohn's disease. Nat Genet 2004;36:471–475.
20. Törkvist L, Noble CL, Lördal M, et al. Contribution of the IBD5 locus to Crohn's disease in the Swedish population. Scand J Gastroenterol 2007;42:200–206.
21. Waller S, Tremelling M, Bredin F, et al. Evidence for association of OCTN genes and IBD5 with ulcerative colitis. Gut 2006;55:809–814.
22. Lawrance IC, Fiocchi C, Chakravarti S. Ulcerative colitis and Crohn's disease: distinctive gene expression profiles and novel susceptibility candidate genes. Hum Mol Genet 2001;10:445–456.
23. Aronson PS. Role of SLC26-mediated Cl-/base exchange in proximal tubule NaCl transport. Novartis Found Symp 2006;273:148–158.
24. Sugiura T, Kato Y, Kubo Y, et al. Mutation in an adaptor protein PDZK1 affects transport activity of organic cation transporter

OCTNs and oligopeptide transporter PEPT2. Drug Metab Pharmacokinet 2006;21:375–383.
25. Silver DL, Wang N, Vogel S. Identification of small PDZK1-associated protein, DD96/MAP17, as a regulator of PDZK1 and plasma high density lipoprotein levels. J Biol Chem 2003;278: 28528–28532.
26. Galland F, Malergue F, Bazein H, et al. Two human genes related to murine vanin-1 are located on the long arm of human chromosome 6. Genomics 1998;53:203–213.
27. Berruyer C, Pouyet L, Millet V, et al. Vanin-1 licences inflammatory mediator production by gut epithelial cells and controls colitis by antagonizing peroxisome proliferator-activated receptor gamma activity. J Exp Med 2006;203:2817–2827.
28. Zhang YW, Ding LS, Lai MD. Reg gene family and human diseases. World J Gastroenterol 2003;9:2635–2641.
29. Hartupee JC, Zhang H, Bonaldo MF, et al. Isolation and characterization of a cDNA encoding a novel member of the human regenerating protein family: RegIV. Biochim Biophys Acta 2001; 1518:287–293.
30. Page-McCaw A, Ewald AJ, Werb Z. Matrix metalloproteinases and the regulation of tissue remodelling. Nat Rev 2007;8:221–233.
31. Verma RP, Hansch C. Matrix metalloproteinases (MMPs): chemical-biological functions and (Q)SARs. Bioorg Med Chem 2007; 15:2223–2268.
32. Ravi A, Garg P, Sitaraman SV. Matrix metalloproteinases in inflammatory bowel disease: boon or a bane? Inflamm Bowel Dis 2007;13:97–107.
33. Persson-Dajotoy T, Andersson P, Bjartell A, et al. Expression and production of the chemokine growth-related oncogene-alpha by human eosinophils. J Immunol 2003;170:5309–5316.
34. Yang SK, Eckmann L, Panja A, et al. Differential and regulated expression of C-X-C, C-C, and C-chemokines by human colon epithelial cells. Gastroenterology 1997;113:1214–1223.
35. Mitsuyama K, Tsuruta O, Tomiyasu N, et al. Increased circulating concentrations of growth-related oncogene (GRO)-α in patients with inflammatory bowel disease. Dig Dis Sci 2006;51:173–177.
36. Marks DJ, Harbord MW, MacAllister R, et al. Defective acute inflammation in Crohn's disease: a clinical investigation. Lancet 2006;367:668–678.
37. Schmidt C, Giese T, Ludwig B, et al. Expression of interleukin-12-related cytokine transcripts in inflammatory bowel disease: elevated interleukin-23p19 and interleukin-27p28 in Crohn's disease but not ulcerative colitis. Inflamm Bowel Dis 2005; 11:16–23.
38. Fuss IJ, Becker C, Yang Z, et al. Both IL-12p70 and IL-23 are synthesized during active Crohn's disease and are down-regulated by treatment with anti-IL12 p40 monoclonal antibody. Inflamm Bowel Dis 2006;12:9–15.
39. Rutgeerts P, Vermeire S, Van Assche G. Mucosal healing in inflammatory bowel disease: impossible ideal or therapeutic target? Gut 2007;56:453–455.
40. Jossens S, Reinisch W, Vermeire S, et al. The value of serological markers in indeterminate colitis: a prospective follow-up study. Gastroenterology 2002;122:1242–1247.
41. Melmed GY, Elashoff R, Chen GC, et al. Predicting a change in diagnosis from ulcerative colitis to Crohn's disease: a nested, case-control study. Clin Gastroenterol Hepatol 2007;5:602–608.
42. Targan SR, Karp LC. Inflammatory bowel disease diagnosis, evaluation and classification: state-of-the-art. Curr Opin Gastroenterol 2007;23:390–394.
43. Sands BE. From symptom to diagnosis: clinical distinctions among various forms of intestinal inflammation. Gastroenterology 2004;126:1518–1532.

Received September 5, 2007. Accepted February 28, 2008.

Address requests for reprints to: Petra von Stein, InDex Diagnostics AB, Retziuslaboriet, Karolinska Institutet, Scheeles väg 1, 17 177 Stockholm, Sweden. e-mail: petra.stein@indexdiag.com; fax: (46) 8-508-847-39.

Supported by a grant from the foundation for clinical research at Sophiahemmet, Stockholm, Sweden.