Molecular Classification of Crohn's Disease and Ulcerative Colitis Patients Using Transcriptional Profiles in Peripheral Blood Mononuclear Cells

Michael E. Burczynski,* Ron L. Peterson,[†]
Natalie C. Twine,[†] Krystyna A. Zuberek,[†]
Brendan J. Brodeur,[†] Lori Casciotti,[†]
Vasu Maganti,[‡] Padma S. Reddy,[§]
Andrew Strahs,[¶] Fred Immermann,[∥]
Walter Spinelli,** Ulrich Schwertschlag,[††]
Anna M. Slager,* Monette M. Cotreau,[††] and
Andrew J. Dorner[†]

From the Molecular Profiling and Biomarker
Discovery/Biomarker Laboratory* and Clinical Research and
Development,** Wyeth Research, Collegeville, Pennsylvania;
Molecular Profiling and Biomarker Discovery,[†] Inflammation
Research,[‡] Bioinformatics Client Services,[§] Clinical Statistics,[¶]
and Translational Research,[††] Wyeth Research, Cambridge,
Massachusetts; and Biometrics Research,[∥] Wyeth Research, Pearl
River, New York

Ulcerative colitis (UC) and Crohn's disease (CD) are common inflammatory bowel diseases producing intestinal inflammation and tissue damage. Although emerging evidence suggests these diseases are distinct, ~10% of patients remain classified as indeterminate inflammatory bowel disease even after invasive colonoscopy intended for diagnosis. A molecular diagnostic assay using a clinically accessible tissue would greatly assist in the classification of these diseases. In the present study we assessed transcriptional profiles in peripheral blood mononuclear cells from 42 healthy individuals, 59 CD patients, and 26 UC patients by hybridization to microarrays interrogating more than 22,000 sequences. Supervised analysis identified a set of 12 genes that distinguished UC and CD patient samples with high accuracy. The alterations in transcript levels observed by microarray were verified by real-time polymerase chain reaction. The results suggest that a peripheral blood mononuclear cell-based gene expression signature can provide a molecular biomarker that can complement the standard diagnosis of UC and CD. (*J Mol Diagn 2006, 8:51–61; DOI: 10.2353/jmoldx.2006.050079*)

Ulcerative colitis (UC) and Crohn's disease (CD) are two common chronic relapsing inflammatory bowel diseases (IBD) that share several demographic and clinical characteristics yet present key differences in tissue damage, suggesting distinct etiopathogenic processes. One proposed etiology of IBD is the inappropriate activation of the mucosal immune system against normal intestinal luminal bacterial flora.[1] A transmural, granulomatous inflammatory process associated with Th1-type responses is characteristic of CD, whereas inflammation in UC tends to be limited to the mucosa and contains large numbers of immunoglobulin-secreting plasma cells that appear to be associated with Th2 responses.[1] Both diseases are complex disorders in which a combination of environmental and genetic factors may determine the susceptibility of an individual to disease.[2]

The ability to quantitate the global expression profiles at the level of RNA using oligonucleotide microarrays has recently been applied to investigate transcriptional signatures present in gastrointestinal tissue obtained from CD and UC patients.[3,4] These studies identified genes involved in inflammatory responses generally up-regulated in IBD and showed that the gastrointestinal tissue transcriptomes obtained from UC and CD patients were quite distinct, with gene sets identified that appear to distinguish UC tissue from CD tissue.

In contrast to biopsies, peripheral blood is a much more accessible tissue source of cells that might be used to distinguish between UC and CD. Circulating peripheral blood mononuclear cells (PBMCs) are responsible for the comprehensive surveillance of the body for signs of infection and disease. PBMCs may therefore serve as a surrogate tissue for evaluation of disease-induced gene expression as a biomarker of disease status or severity.[5] Maas and colleagues[6] identified PBMC profiles in patients with the autoimmune diseases rheumatoid arthritis, systemic lupus erythematosus, type I diabetes, and multiple sclerosis. We have shown[7] that in the context of a nonautoimmune disease, PBMCs obtained from renal cell carcinoma patients also exhibit disease-associated transcriptomes distinct from those of healthy volunteers. Mannick and colleagues[8] recently explored expression profiles of PBMCs from seven

Table 1. Demographic Characteristics of Disease-Free Individuals, CD Patients, and Patients with UC

Type	Crohn's	UC	Normals	Normal versus IBD (*P* value)	CD versus UC (*P* value)
Number of samples	59	26	42		
Age (mean)	41.3	46.7	44.1	0.96*	0.055*
Sex	21 Males	8 Males	24 Males	0.014†	0.66†
	38 Females	18 Females	18 Females		
Race	51 Caucasian	22 Caucasian	40 Caucasian	0.09‡	0.82‡
	7 Black	3 Black	1 Asian		
	1 Hispanic	1 Hispanic	1 Indian		

*P value calculated using two-sided *t*-test with *t*-statistic based on analysis of variance error estimate.
†*P* value calculated using likelihood ratio χ^2 test comparing male to female frequencies among groups.
‡*P* value calculated using likelihood ratio χ^2 test comparing Caucasian to non-Caucasian frequencies among groups.

CD patients and five UC patients with a 2400 gene cDNA microarray and described several genes that appear differentially expressed between these diseases. In the present study, we used oligonucleotide arrays interrogating 22,000 sequences to investigate the transcriptional profiles of circulating PBMCs in a group of 42 healthy subjects and 85 IBD patients with clinical diagnoses of CD and UC. The results suggest that a molecular diagnosis of UC and CD using the transcriptional profiling of PBMC might be possible.

Materials and Methods

Patient Information and Clinical Assessments

Blood samples for pharmacogenomic analysis were collected at North American and European clinical sites from a total of 42 apparently healthy individuals, 59 CD patients, and 26 UC patients participating in three distinct clinical trials (two CD and one UC trial). Each clinical site's institutional review board or ethics committee approved this study, and no procedures were performed before obtaining informed consent from each patient. A comparison of the demographic characteristics of individuals in the present study is presented in Table 1.

CD patients had CD activity index scores (CDAI) ranging between 220 and 400 with an abdominal pain rating of ≥25 and/or a diarrhea rating of ≥25. Diagnosis of CD for at least 6 months was confirmed by radiological studies, endoscopy with histological examination, or surgical pathology; patients with a diagnosis of CD were included if the diagnosis was confirmed by a biopsy. UC patients had scores from the Physician's Global Assessment of the Mayo Ulcerative Colitis Scoring System ranging from mild to moderate (scores of 1 or 2). The diagnosis of left-sided UC was provided by endoscopy with biopsy, in addition to standard clinical criteria.

Proportions of females to males were significantly different between the healthy and IBD populations, but not distinct between the two IBD populations. Neither race (Caucasian versus non-Caucasian) nor age differed significantly between healthy and IBD populations or between the two IBD populations. Investigation of concomitant medication usage between the two IBD populations indicated that neither 5-ASA nor any of the other less-frequently used drugs reported as concomitant medications confounded the comparisons in this study.

Blood Sampling and Processing

Blood samples (8 ml) were collected into Vacutainer cell purification tubes (Becton Dickinson, Franklin Lakes, NJ) at the clinical site and shipped overnight to a central processing lab for PBMC isolation according to the manufacturer's recommendations. All PBMCs analyzed in this study were processed within 24 hours after the blood draw. Before RNA purification, complete cell counts were performed on purified PBMCs using a Pentra 60 C + hematology analyzer (ABX, Irvine, CA) to record absolute counts and percentages of neutrophils, lymphocytes, monocytes, eosinophils, and basophils. Cell counts for one PBMC sample from a UC patient were not performed, and this profile was therefore excluded from the analyses of covariance described below. Expression data from this patient were included when developing and testing prediction models. Total RNA was purified from PBMCs using the RNeasy mini column protocol (Qiagen, Valencia, CA).

Oligonucleotide Array Hybridization and Data Reduction

Total RNA (2 μg) was converted to biotinylated cRNA according to the Affymetrix protocol (Affymetrix, Santa Clara, CA). Labeled cRNA (10 μg) was fragmented and prepared for hybridization as previously described.[7] Biotinylated cRNA was hybridized to the Affymetrix HG-U133A human GeneChip array as described in the Affymetrix technical manual. Eleven biotinylated control transcripts ranging in abundance from 1:300,000 (3 ppm) to 1:100 (100 ppm) were spiked into each sample before hybridization to function as a standard curve.[9] GeneChip MAS 5.0 software was used to evaluate the specific hybridization intensity, compute signal value for each probe set, and make an absent/present call. The signal value for each probe set was then converted to a frequency value representative of the number of transcripts present in 10^6 transcripts by reference to the standard curve.[9] Each transcript was evaluated and included in the study fol-

Table 2.   Cell Composition in PBMCs from Disease-Free Individuals, CD Patients, and Patients with UC

Type	Crohn's	UC	Normals	Normal versus IBD (P value)*	CD versus UC (P value)*
Number of samples	59	25	42		
Basophil (%)	0.33	0.30	1.05	0.012	0.93
Eosinophil (%)	1.10	0.91	0.37	0.0003	0.37
Lymphocyte (%)	52.20	59.58	78.90	<0.0001	0.056
Monocyte (%)	29.41	27.63	14.65	<0.0001	0.52
Neutrophil (%)	14.96	10.58	5.00	<0.0001	0.035

*P value calculated using a two-sided *t*-test, with *t*-statistic based on analysis of variance error estimate.

lowing nonstringent criteria: called present and at or above a frequency value of 10 (10 ppm) in at least one of the samples (healthy, UC, or CD). Sequences ($n = 7908$) meeting these filtering criteria were used in the analysis.

Analysis of Covariance

Analysis of covariance methods were used to adjust for differences in PBMC cell type composition when testing for differences in mean expression among disease groups. Separate analyses of covariance were run for each transcript, using log-transformed frequency as the response measure. The analysis of covariance model included terms for disease group, gender, neutrophil percent, monocyte percent, and eosinophil percent. In the analysis of covariance, a slope describing the linear relationship between the percentage of the cell type and the expression level for a particular gene was estimated for each cell type, and a *t*-test was done to determine whether the slope was significantly different from 0 (where a slope of 0 indicates that there is no linear relationship between cell type percent and expression level).

In addition to the overall tests for treatment group differences and cell type regression effects, pairwise comparisons of disease group means adjusted for differences in cell type percentages were performed using two-sided *t*-tests, with the denominator of the *t*-statistics derived from the analysis of covariance error term. Finally, because the relative distribution of females and males was also significantly distinct among the disease groups, we included gender in the analyses of covariance. No adjustments of the raw *P* values produced by the analyses described above were done to account for the large number of statistical tests performed. A fold change filter (1.5-fold) combined with a conservative significance level of $\alpha = 0.0001$ were used to reduce the incidence of false-positive determinations.

Gene Selection and Supervised Class Prediction

Gene selection and supervised class prediction were performed using Genecluster version 2.0 (*http://www. broad/mit.edu/cancer/software/software.html*).[10] In these analyses only 4228 transcripts meeting a stringent data reduction filter (at least 50% present calls in Crohn's or UC samples and at least 50% of the Crohn's or UC samples with frequencies greater than 10 ppm) were used. Samples within each group were randomly selected for membership in a training set (75%) or a test set (25%) of profiles. Gene selection was performed using the training set of samples, and the classifier with the fewest genes that exhibited the highest overall accuracy of class assignment in the training set was identified by leave-one-out cross validation. The predictive classification model was then evaluated on samples in the test set, and the overall accuracy of class assignment for samples in the test set was reported.

For gene selection all expression data in both the training set and test set were log-transformed before analysis. In the training set of data, models containing increasing numbers of features (transcript sequences) were built using a two-sided approach (equal numbers of features in each class) with a S2N similarity metric that used median values for the class estimate. PBMC profiles from CD patients and UC patients were compared using a binary approach. Predictive gene classifiers containing between 2 and 200 genes in steps of two were evaluated by leave-one-out cross validation to identify the smallest predictive model yielding the most accurate class assignments. Prediction of class membership was performed using a weighted voting algorithm.

Ingenuity Pathway Analysis

The Ingenuity pathway analysis (IPA) tool (Ingenuity, Mountain View, CA) was used to annotate the disease-associated genes obtained from analyses of covariance. Annotations on canonical pathways and functional categories were retrieved for these gene lists from the Gene-By-Gene View and/or using the Search IPKB feature.

Real-Time Polymerase Chain Reaction (PCR) Confirmation of Microarray Results

A total of 45 ng of each PBMC RNA sample was reverse-transcribed in a 96-well plate in a 100-μl reaction using the High Capacity cDNA Archive kit (Applied Biosystems, San Diego CA). The reaction was incubated at 25°C for 10 minutes and then 37°C for 2 hours and stored at −80°C until amplification. To amplify and quantitate relatively the levels of transcripts, predesigned, gene-specific TaqMan probe and primer sets (TaqMan gene expression assays, Applied Biosystems) corresponding to the GenBank accession numbers for genes in the 12 gene classifier were used. Real-time PCR for each transcript of interest was per-

Table 3.   General Inflammatory Bowel Disease-Associated Sequences in PBMCs

Accession no.	Name	Direction in both
Hs.75716	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 2, PAI 2	↑
Hs.154654	Cytochrome P450, subfamily I (dioxin-inducible), polypeptide 1	↑
Hs.79516	Brain-abundant, membrane-attached signal protein 1	↑
Hs.177781	Unknown	↑
Hs.104624	Aquaporin 9	↑
Hs.20084	Retinoid X receptor, α	↑
Hs.2161	Complement component 5 receptor 1 (C5a ligand)	↑
Hs.865	RAP1A, member of RAS oncogene family	↑
Hs.177486	Amyloid β (A4) precursor protein (protease nexin-II, Alzheimer disease)	↑
Hs.198282	Phospholipid scramblase 1	↑
Hs.288555	ELK3, ETS-domain protein (SRF accessory protein 2)	↑
Hs.101695	NCK adaptor protein 2	↑
Hs.198282	Phospholipid scramblase 1	↑
Hs.285313	Core promoter element-binding protein	↓
Hs.151411	KIAA0916 protein	↓
Hs.20072	Myosin regulatory light chain-interacting protein	↓
Hs.81248	CUG triplet repeat, RNA-binding protein 1	↓
Hs.211610	CUG triplet repeat, RNA-binding protein 2	↓
Hs.86896	Bromodomain containing 3	↓
Hs.100555	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide 18 (Myc-regulated)	↓
Hs.143601	Hypothetical protein hCLA-iso	↓
Hs.149436	Kinesin family member 5B	↓
Hs.239483	Unknown	↓
Hs.78909	Zinc finger protein 36, C3H type-like 2	↓
Hs.85273	Retinoblastoma binding protein 6	↓
Hs.219614	F-box and leucine-rich repeat protein 11	↓
Hs.153834	Pumilio homolog 1 (Drosophila)	↓
Hs.18827	Cylindromatosis (turban tumor syndrome)	↓
Hs.127287	KIAA0794 protein	↓
Hs.73090	Nuclear factor of κ light polypeptide gene enhancer in B cells 2 (p49/p100)	↓
Hs.373557	Partial transcript encompassing THC211630 gene	↓
Hs.75243	Bromodomain containing 2	↓
Hs.118174	Tetratricopeptide repeat domain 3	↓
Hs.37096	Zinc finger protein 145 (Kruppel-like, expressed in promyelocytic leukemia)	↓
Hs.3530	FUS interacting protein (serine-arginine rich) 1	↓
Hs.83484	Meis1	↓
Hs.294014	Unknown	↓
Hs.183418/214291/355896	Cell division cycle 2-like 1 (PITSLRE proteins), cell division cycle 2-like 2	↓
Hs.77256	Enhancer of zeste homolog 2 (Drosophila)	↓
Hs.278426	PDGFA-associated protein 1	↓
Hs.10351	KIAA0308 protein	↓
Hs.18368	SR-rich protein	↓
Hs.2173	Fucosyltransferase 4 [α (1,3) fucosyltransferase, myeloid-specific]	↓
Hs.152601	UDP-glucose ceramide glucosyltransferase	↓
Hs.243901	Unknown	↓

(*Table continues*)

formed in 96-well fast block optical reaction plates in a 25-μl reaction volume (containing 1× TaqMan Fast Universal Master Mix, 1× TaqMan gene expression assay, and 2.25 ng of cDNA) using an ABI 7900HT sequence detection system (Applied Biosystems, San Francisco, CA). Default 7900 fast block cycle conditions were as follows: 95°C for 20 seconds, 40 cycles of 95°C for 1 second, and 60°C for 20 seconds. Ct values for each amplification were recorded for each target gene and the housekeeping genes *β2-micro-globulin*, *β-actin*, *18S*, and *GAPDH*. The differences between cycle thresholds for target genes and each of the four reference genes in each of the samples were calculated (ΔCt), and the average fold change in expression between UC and CD was calculated by the following formula: average fold difference = 2 raised to the power of ($\Delta Ct_{UC} - \Delta Ct_{CD}$).

Results

Cellular Composition of Purified PBMC Samples from Healthy Patients, Crohn's Patients, and UC Patients

Before the expression profiling portion of the study, the cellular compositions of the purified PBMC pellets from patients in all three groups (healthy patients, patients with CD, and patients with UC) were measured before RNA

Table 3. *Continued*

Fold difference CD versus normal	CD versus normal (analysis of covariance P value)	Fold Difference UC versus normal	UC versus normal (analysis of covariance P value)
6.93	6.84E-12	3.35	3.87E-05
3.31	6.49E-10	2.37	2.81E-05
1.94	6.81E-09	2.13	2.61E-09
1.88	6.34E-08	1.92	3.99E-07
1.88	9.92E-06	2.02	9.06E-06
1.80	1.33E-06	1.68	9.18E-05
1.74	2.65E-05	1.81	5.05E-05
1.64	5.95E-10	1.53	1.01E-06
1.63	6.11E-10	1.60	4.81E-08
1.60	3.90E-05	1.76	1.09E-05
1.59	4.25E-10	1.51	2.75E-07
1.57	2.72E-14	1.51	1.08E-10
1.56	1.02E-05	1.55	7.12E-05
2.78	8.63E-05	5.65	9.23E-09
2.47	7.92E-05	3.21	6.01E-06
2.45	3.21E-07	2.47	2.73E-06
2.44	8.74E-08	2.52	4.84E-07
2.27	5.40E-06	2.62	1.77E-06
2.24	1.22E-06	2.26	8.83E-06
2.24	3.89E-05	2.48	3.11E-05
2.20	1.20E-07	2.16	2.42E-06
2.15	2.66E-05	2.56	4.15E-06
2.10	2.47E-09	2.42	2.47E-10
2.03	1.46E-07	2.06	1.18E-06
2.01	6.23E-05	2.56	1.64E-06
1.95	9.04E-07	2.41	1.11E-08
1.92	6.49E-08	1.92	8.29E-07
1.91	7.70E-06	2.13	2.93E-06
1.83	3.31E-05	1.95	4.07E-05
1.81	3.63E-06	1.77	4.93E-05
1.81	3.43E-05	2.18	1.32E-06
1.76	1.11E-06	1.86	1.65E-06
1.76	4.63E-05	2.35	6.54E-08
1.74	8.38E-07	1.91	2.96E-07
1.72	2.20E-05	1.89	7.47E-06
1.70	2.26E-06	1.78	3.70E-06
1.68	5.43E-06	1.81	2.84E-06
1.64	1.42E-06	1.88	9.80E-07
1.63	7.72E-06	1.71	8.73E-06
1.60	6.59E-07	1.57	1.61E-05
1.60	8.99E-06	1.78	1.10E-06
1.57	1.30E-05	1.83	2.43E-07
1.54	4.08E-05	1.70	6.69E-06
1.54	3.35E-05	1.73	2.72E-06
1.51	8.84E-05	1.68	1.29E-05

isolation (Table 2). The cellular composition of PBMC samples was significantly different ($P < 0.05$) in the comparison of PBMCs from healthy patients to those from IBD patients. The overall percentages of basophils and lymphocytes were significantly lower in PBMCs from patients with IBD, whereas the percentages of eosinophils, monocytes, and neutrophils were significantly elevated in PBMCs from IBD patients. Previous studies have noted elevations in neutrophils via similar purification processes, which are attributable to changes in sedimentation density that appear related to alterations in their activation state in the peripheral blood of advanced cancer patients.[11] The selective elevation in eosinophils, monocytes, and neutrophils may be a disease-related activation event captured by the cell purification tube-based PBMC isolation process. In contrast, basophil, eosinophil, and monocyte proportions were not significantly distinct ($P < 0.05$) between CD and UC PBMC
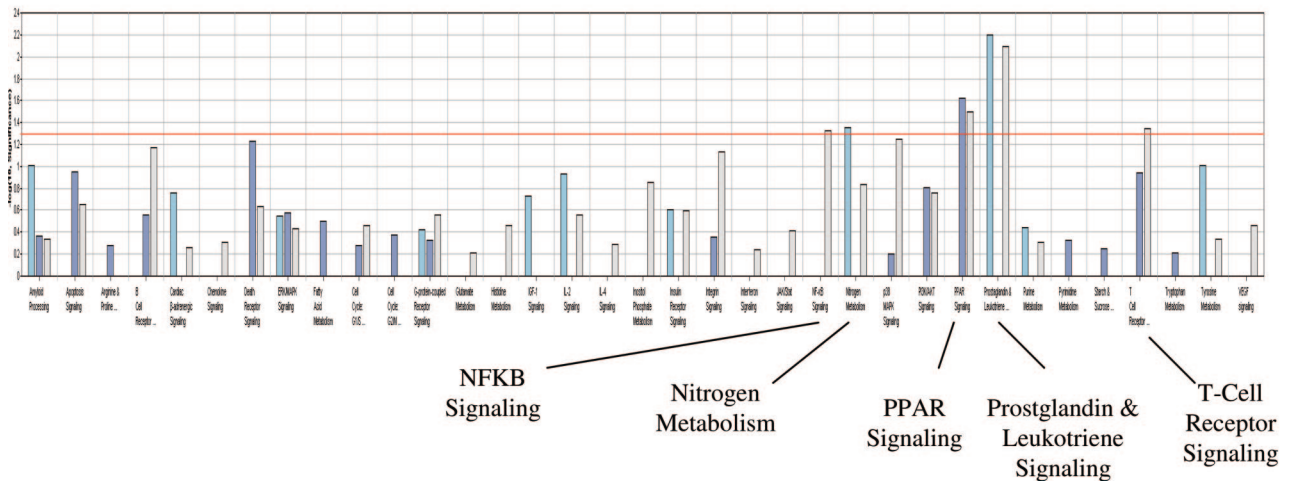
samples. In the comparison of the two IBD groups, only neutrophils were significantly different (10% versus 14%, $P = 0.035$).

Expression Level Differences in PBMCs from All IBD Patients Compared to Healthy Controls

To identify disease-associated genes that are not associated with differences in cell composition, an analysis of covariance was used to identify differentially expressed transcripts while taking into account variation in cell composition among the PBMC samples. Analyses of covariance were run for the 7908 transcripts that passed the standard expression level filter and the percentage of eosinophils, monocytes, and neutrophils were included as covariates.
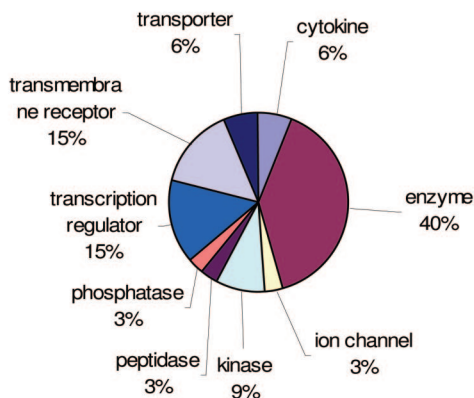
By the analysis of covariance, the levels of 220 transcripts were greater than 1.5-fold different between

A

NFKB
Signaling

Nitrogen
Metabolism

PPAR
Signaling

Prostglandin &
Leukotriene
Signaling

T-Cell
Receptor
Signaling

B

Functional Categories - CD specific transcripts

transporter 6%

cytokine 6%

transmembrane receptor 15%

transcription regulator 15%

phosphatase 3%

peptidase 3%

kinase 9%

ion channel 3%

enzyme 40%

C

Functional Categories - UC Specific Transcripts

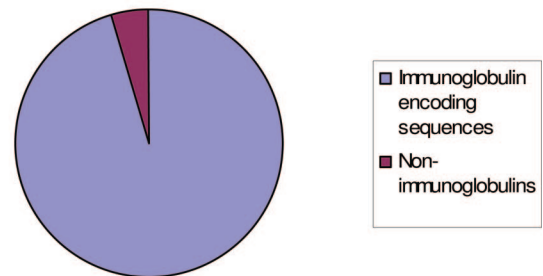- Immunoglobulin encoding sequences
- Non-immunoglobulins

Figure 1. Functional annotation and categories of transcripts identified as CD-associated, UC-associated, and differentially expressed between UC and CD. **A:** Canonical pathways overrepresented in the CD versus normal analysis of covariance comparison (light blue); the UC versus normal analysis of covariance comparison (dark blue), and the UC versus CD analysis of covariance comparison (light purple). In this panel the negative log10 of the *P* value is plotted to highlight more significant associations. The pathways interrogated (in order) were as follows: amyloid processing, apoptosis signaling, arginine and proline metabolism, B-cell receptor signaling (nonimmunoglobulin), cardiac β-adrenergic signaling, chemokine signaling, death receptor signaling, ERK/MAPK signaling, fatty acid metabolism, cell cycle regulation (G_1/S), cell cycle regulation (G_2/M), GPCR signaling, glutamate metabolism, histidine metabolism, IGF-1 signaling, IL-2 signaling, IL-4 signaling, inositol phosphate metabolism, insulin receptor signaling, integrin signaling, interferon signaling, JAK/STAT signaling, NF-κB signaling, nitrogen metabolism, p38 MAPK signaling, PI3K/AKT signaling, PPAR signaling, prostaglandin and leukotriene metabolism, purine metabolism, pyrimidine metabolism, starch and sucrose metabolism, T-cell receptor signaling, tryptophan metabolism, tyrosine metabolism, and VEGF signaling. **B:** CD-specific transcripts in PBMCs were functionally annotated using the Ingenuity pathway analysis system (Ingenuity), and the relative distribution of transcripts in each of the chosen functional categories for the CD-associated genes are presented in the pie chart. **C:** UC-specific transcripts in PBMCs were functionally annotated manually, and the relative distribution of transcripts in immunoglobulin versus nonimmunoglobulin categories is presented.

Crohns' and healthy PBMCs and possessed an unadjusted *P* value in the pairwise comparison based on the analysis of covariance of less than 0.0001, and the levels of 120 transcripts were significantly different in UC and healthy PBMCs using the same criteria as above. Forty-five of these sequences were differentially expressed in both UC and CD and these common transcripts changed in the same direction in both diseases compared to healthy levels (Table 3).

We applied an additional filter to the remaining gene sets to identify PBMC transcripts that appear differentially expressed in only one disease state. Of the 220 transcripts

Table 4.   Transcriptional Sequences Differentially Expressed in the Direct Comparison of CD Versus UC PBMC Profiles

Accession	Name	Specific to:	Fold difference	Analysis of covariance P value
Hs.90061	Progesterone receptor membrane component 1	Crohn's	2.08	2.55E-07
Hs.279843	mutL homolog 3 (*E. coli*)	Crohn's	2.00	2.75E-08
Hs.88474	Prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)	Crohn's	1.93	1.18E-05
Hs.73769	Folate receptor 1 (adult)	Crohn's	1.93	6.82E-11
Hs.89714	Chemokine (C-X-C motif) ligand 5	Crohn's	1.85	3.71E-05
Hs.83381	Guanine nucleotide binding protein (G protein), γ 11	Crohn's	1.79	8.04E-06
Hs.2359	Dual specificity phosphatase 4	Crohn's	1.76	5.30E-05
Hs.204238	Lipocalin 2 (oncogene 24p3)	Crohn's	1.75	4.35E-05
Hs.81564	Platelet factor 4 (chemokine (C-X-C motif) ligand 4)	Crohn's	1.74	4.38E-06
Hs.119257	ems1 sequence (mammary tumor and squamous cell Carcinoma-associated (p80/85 src substrate)	Crohn's	1.70	8.21E-05
Hs.26530	Serum deprivation response (phosphatidylserine binding protein)	Crohn's	1.66	5.34E-07
Hs.303023	Tubulin, β 1	Crohn's	1.65	8.39E-05
Hs.23581	Leptin receptor gene-related protein	Crohn's	1.61	7.52E-05
Hs.249216	H2B histone family, member J	Crohn's	1.61	6.91E-05
Hs.77439	Protein kinase, cAMP-dependent, regulatory, type II, β	Crohn's	1.59	4.09E-05
Hs.2178	H2B histone family, member Q	Crohn's	1.57	8.85E-06
Hs.114231	C-type lectin-like receptor-2	Crohn's	1.57	8.79E-07
Hs.12813	DKFZP434J214 protein	Crohn's	1.52	1.21E-05
Hs.2164	Pro-platelet basic protein (chemokine (C-X-C motif) ligand 7)	Crohn's	1.51	2.90E-05
Hs.300697	Immunoglobulin heavy constant γ 3 (G3m marker)	UC	3.87	9.28E-13
Hs.153261	Immunoglobulin heavy constant μ	UC	2.60	2.72E-05
n/a	Unknown EST with consensus to immunoglobulin κ orphon	UC	2.42	5.66E-05
Hs.406565	Immunoglobulin κ constant	UC	2.30	1.78E-06
n/a	28 S ribosomal RNA 5' region	UC	2.11	2.95E-07
Hs.406565	Immunoglobulin κ constant	UC	2.08	2.88E-08
Hs.406565	Immunoglobulin κ constant	UC	2.04	3.52E-07
Hs.183125	Killer cell lectin-like receptor subfamily F, member 1	UC	2.02	5.45E-05
Hs.411106	perforin 1	UC	1.98	7.07E-05
Hs.153261	Immunoglobulin heavy constant μ	UC	1.93	3.78E-05
Hs.406565	Immunoglobulin κ constant	UC	1.88	2.19E-06
Hs.406565	Immunoglobulin κ constant	UC	1.87	8.32E-09
Hs.102950	Coatomer protein complex, subunit γ, immunoglobulin λ joining 3	UC	1.74	5.47E-05
Hs.102950	Coatomer protein complex, subunit γ, immunoglobulin λ joining 3	UC	1.74	2.10E-05
Hs.406565	Immunoglobulin κ constant	UC	1.72	2.57E-07
Hs.355888	Phospholipase C, β 2	UC	1.72	2.24E-05
Hs.8272	Prostaglandin D2 synthase 21-kd (brain)	UC	1.71	5.41E-05
Hs.25338	Protease, serine, 23	UC	1.68	7.72E-05
Hs.381417	Unknown EST with consensus to immunoglobulin κ light chain variable region	UC	1.67	3.86E-05
Hs.75596	Interleukin 2 receptor, β	UC	1.65	7.19E-05
Hs.406565	Immunoglobulin κ constant	UC	1.64	2.48E-06
Hs.102950	Coatomer protein complex, subunit γ, Immunoglobulin λ joining 3	UC	1.64	4.13E-05
Hs.406565	Immunoglobulin κ constant	UC	1.63	4.95E-06
Hs.406565	Immunoglobulin κ constant	UC	1.61	3.75E-08
Hs.380156	NK-receptor, killer cell immunoglobulin-like receptor, two domains, long cytoplasmic tail, 3	UC	1.60	3.16E-08
Hs.405944	Immunoglobulin λ locus	UC	1.59	1.55E-05
Hs.348935	Immunoglobulin λ-like polypeptide 1	UC	1.58	4.69E-05
Hs.84	Interleukin 2 receptor, γ (severe combined immunodeficiency)	UC	1.53	7.84E-05
Hs.193128	Homolog of *C. elegans* smu-1	UC	1.52	5.16E-05
Hs.238944	Hypothetical protein FLJ10631	UC	1.52	3.72E-05

Table 5. 14-Sequence (12-Gene) Classifier for Assigning UC or CD on the Basis of Expression Levels in PBMCs

Classifier gene	Class	Name	Unigene ID
1	Crohns	Lipocalin 2 (oncogene 24p3)	Hs.204238
2	Crohns	mutL homolog 3 (*E. coli*)	Hs.279843
3	Crohns	Serum deprivation response (phosphatidylserine binding protein)	Hs.26530
4	Crohns	H2B histone family, member Q	Hs.2178
5	Crohns	H3 histone family, member K	Hs.70937
6	Crohns	Chemokine (C-X-C motif) ligand 5	Hs.89714
7	Crohns	Integrin, β 3 (platelet glycoprotein IIIa, antigen CD61)	Hs.87149
8	UC	Immunoglobulin heavy constant γ 3 (G3m marker)	Hs.300697
9	UC	Immunoglobulin κ constant	Hs.406565
10	UC	M27830 Human 28S ribosomal RNA gene 5' region	n/a
11	UC	Protein tyrosine phosphatase, receptor type, C-associated protein	Hs.155975
12	UC	Granzyme K (serine protease, granzyme 3; tryptase II)	Hs.3066
13	UC	Immunoglobulin κ constant	Hs.406565
14	UC	Immunoglobulin κ constant	Hs.406565

that were CD-associated, 67 sequences were not significantly altered in the UC versus healthy comparison ($P > 0.05$) and therefore appear to be CD-specific (Supplementary Table S1; *http://jmd.amjpathol.org/*). Of the 120 transcripts that were UC-associated, 22 sequences were not significantly altered in the CD versus healthy comparison ($P > 0.05$) and therefore appear to be UC-specific (Supplementary Table S2; *http://jmd.amjpathol.org/*).

The canonical gene pathways bearing the greatest likelihood of significant overrepresentation are summarized for each comparison in Figure 1A. In this analysis transcripts involved in prostaglandin metabolism were significantly overrepresented in the CD gene signature, whereas transcripts encoding proteins involved in apoptosis and B-cell signaling appear overrepresented in the UC signature. Figure 1B summarizes the diverse functional categories encompassed by the transcripts differentially expressed in CD relative to healthy controls. Major functional categories up-regulated in CD PBMCs included enzymes involved in prostaglandin metabolism, transcription regulators, and transmembrane receptors including several integrin isoforms. Finally, Figure 1C summarizes the abundant overrepresentation of immunoglobulin constant regions that was unique to the UC PBMC expression signature.

Identification of Gene Signatures Discriminating CD and UC

Because the main goal in the present study was to determine whether gene expression patterns in PBMCs of patients with CD and UC were sufficiently distinct to enable their classification on the basis of gene expression profiles in PBMCs alone, we performed a direct comparison of gene expression signatures between the two diseases. Analysis of covariance comparison of CD versus UC PBMC profiles identified 49 transcripts that were present at significantly different levels between PBMCs of CD and UC patients (1.5-fold difference, $P < 0.0001$) (Table 4).

Based on the analysis of covariance results indicating significant differences in direct comparisons of CD and

UC PBMC gene signatures, we next used a supervised class prediction approach to identify the smallest set of informative sequences capable of disease-specific classification. PBMC samples from the IBD patients were randomized into a training set composed of 44 CD and 20 UC profiles and a test set composed of 15 CD and 6 UC profiles. A 14-sequence (12-gene) classifier (Table 5) gave the highest overall accuracy (94%) in distinguishing between UC and CD PBMC profiles as evaluated by leave-one-out cross validation of the training set (Figure 2A). Increasing the size of the classifier set did not increase accuracy above this level. Figure 2B illustrates the overall expression pattern of the sequences in the classification set.

This gene classifier was used to assign class membership to the 15 CD profiles and 6 UC profiles withheld in the test set (Figure 2C). Using this predictive model all samples in the test set were correctly classified as clinically diagnosed. Only one individual in each group possessed a confidence score of less than 0.2 using this classifier, indicating the relatively high confidence with which these calls were made by the weighted voting algorithm. These results demonstrate the potential applicability of using PBMC expression profiles to aid in the molecular diagnosis of CD and UC.

Real-Time Reverse Transcriptase (RT)-PCR Confirmation of Microarray Observations

Despite the classifier's accuracy for nearest-neighbor-based class assignment in a test set of PBMC samples, the average fold changes of transcripts in the CD/UC classifiers were relatively low. We therefore performed real-time PCR to confirm the relative expression observed by Affymetrix microarray technology for CD and UC samples in this study. We used four separate housekeeping genes for normalization of the target genes (*β2-microglobulin*, *β-actin*, *GAPDH*, and *18S rRNA*). All CD and UC RNA samples in the study were converted to cDNA using the same reverse-transcription cocktail and procedure. Comparison of average fold changes calculated by microarray and real-time
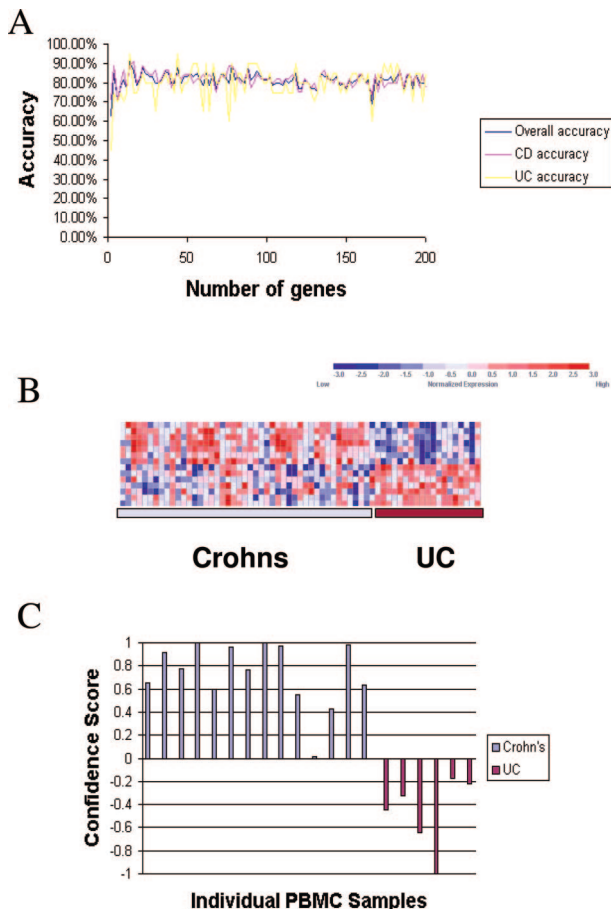
Figure 2. Supervised class prediction of CD and UC using PBMC profiles. **A:** The relative overall accuracy, accuracy of CD classification, and accuracy of UC classification for gene classifiers of increasing size are presented. The smallest predictive model with the highest accuracy (94%) on LOOCV was a classifier composed of 14 gene sequences. **B:** Expression levels for the 14 most class-correlated sequences (rows) are indicated for each of the patients in the training set (columns). Expression data for each gene across the entire data set were normalized to a mean of zero and a variance of one and relative expression is indicated according to the color scale of blue (low expression) to red (high expression) as indicated in the key. **C:** Results of weighted voting class assignment in the test set of samples. Confidence scores in favor of CD are presented as positive values, and confidence scores of the class assignments in favor of UC are presented as negative values. The overall accuracy of class assignment was 100% in the test set, in which 15 of 15 Crohn's patients were correctly classified as Crohn's, and 6 of 6 UC patients were correctly classified as UC, based solely on expression patterns in PBMCs. The actual origins of the PBMC profiles are indicated by color (CD patients, light blue; UC patients, maroon).

PCR using *β2-microglobulin* are presented in Figure 3, and relative fold changes for all 12 genes using each of the four housekeeping genes as normalizers were extremely concordant (Supplementary Table S3; *http://jmd.amjpathol.org/*). On the basis of these results, of the 12 transcripts originally identified as CD/UC discriminator genes, only the *28S rRNA* fragment appears to have been significantly overestimated by microarray hybridization.

Discussion

The focus of the present study was to determine both the commonalities and specificities of gene expression pat-
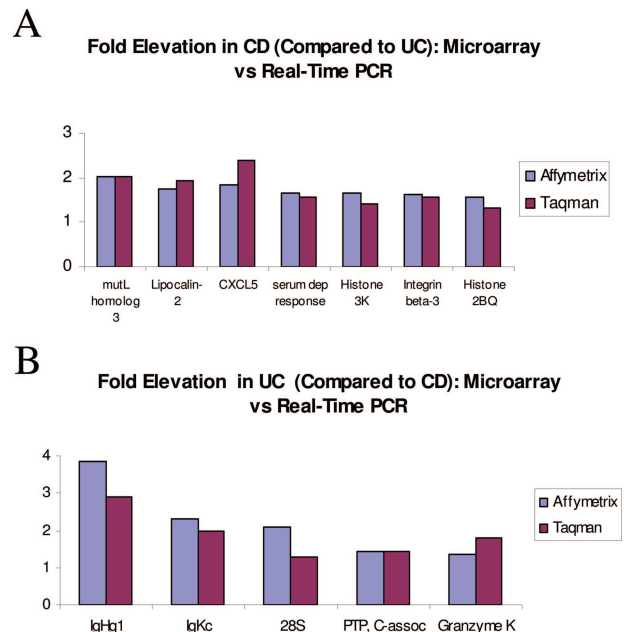

Figure 3. Real-time PCR confirmation of classifier transcript levels in CD and UC sample sets. **A:** Average fold elevation of gene classifier transcripts detected as up-regulated in CD as detected by Affymetrix microarray hybridization (blue) or real-time RT-PCR (maroon). **B:** Average fold elevation of gene classifier transcripts detected as up-regulated in UC as detected by Affymetrix microarray hybridization (blue) or real-time RT-PCR (maroon).

terns in PBMCs associated with CD and UC and whether disease-specific expression signatures could contribute to a molecular diagnosis of disease. Several dozen genes appear to be differentially expressed in both CD and UC compared with profiles for healthy patients. Many encode nuclear proteins such as transcription regulators and most are down-regulated. Examples include *NFKB2*, RNA-binding factors *CUGBP1* and *CUGBP2*, *COPEB*, *ELK3*, and *Meis 1*. Dysregulated inflammatory processes common to UC and CD may be a consequence of modulation of the activity of these transcriptional regulators.

The most highly expressed gene commonly elevated in both IBDs was the protease inhibitor *SERPINB2* (also called PAI, plasminogen activator inhibitor, type II). Increased plasminogen activator levels have been reported in mucosal lesions of IBD patients,[12] and increased PAI-1 was found in IBD patient plasma. Although distinct from PAI-1, PAI-2 shares enzyme specificity to both u-PA and to a lesser degree t-TA, and elevated PAI-2 levels are reported in rheumatoid arthritis synovial fluid.[13] These findings suggest changes in components of the fibrinolytic and coagulation system may contribute to an increased risk for thromboembolic complications and possibly to colitis and bleeding seen in IBD patients.[14] A role for PAI-2 in IBD has not been reported, but our study suggests that elevated *PAI-2* RNA levels in PBMCs are associated with disease.

Multiple functional classes of transcripts appear specifically up-regulated in PBMCs of CD patients including prostaglandin-metabolizing enzymes, chemokines, and transcriptional regulators. The CD-specific gene profile exhibited a proinflammatory gene expression profile that was not apparent in the UC PBMC profile. Prostaglandin

endoperoxide synthase 1 (*PTGS1*, cyclooxygenase 1) was significantly increased in PBMCs from CD patients, while prostaglandin D2 synthase (*PTGDS*) was decreased. These effects on the prostaglandin synthetic pathway would be expected to result in increased conversion of arachidonic acid into select prostaglandins. Although prostaglandin content is elevated in lesions of IBD patients,[15] very recent evidence suggests that levels of at least one prostaglandin (*PGE2*) are actually decreased in mononuclear cells of patients with CD,[16] and PGE2 is an important modulator of cytokine release from T lymphocytes derived from the gastrointestinal tract.[17] Several chemokines (*C-X-C ligands 4* and *7*, *platelet factor 4 variant 1*) were up-regulated in CD. Overall there was surprisingly little overlap between transcripts identified as up-regulated in the present set of CD PBMCs and those reported as up-regulated in the seven CD patients analyzed by Mannick and colleagues.[8] It is unknown whether this is attributable to the larger number of patients explored in the present study, the larger number of genes interrogated, differences in gene nomenclature, or some confounding factor between these studies. However, the most strongly up-regulated transcript in CD reported by Mannick and colleagues[8] encoded a transforming growth factor (TGF)-β-inducible transcript. In this study *TSC-22*, a distinct TGF-β-inducible transcript, was also identified as up-regulated in CD PBMCs. These observations show that up-regulation of TGF-β signal transduction appears to be evident in CD PBMCs. Constitutive elevation in this pathway could result in down-regulation of Smad-dependent pathways that may inhibit the ability of TGF-β to terminate immune responses and in turn play a causal role in the pathogenesis of CD.[8]

It is possible that a portion of the Crohn's-associated disease signature may be platelet-derived. Recent evidence has demonstrated that platelets can participate in chronic intestinal inflammation,[18] and platelets co-purified to a greater extent with the PBMCs isolated from CD patients in this study (data not shown). Thus, the detection of *platelet factor 4* and *platelet factor 4 variant 1* in the CD-associated signature could be attributable to elevated levels of co-purified platelets in isolated PBMCs. However, other transcripts among the top 10 nonmitochondrial transcripts reported in platelets[19] do not appear in the present CD-associated list of transcripts, suggesting that the levels of these anucleate cells are not the sole source of these transcripts. All of the transcripts in the CD disease signature that have been previously associated with platelets are also expressed at significant levels in purified T cells, B cells, and/or monocytes (M.E. Burczynski et al, unpublished observation), which suggests that transcripts previously associated with platelets can originate from the mononuclear cells that were isolated and profiled in this study.

The UC-specific gene set was dominated by overexpression of immunoglobulin-encoding sequences, reminiscent of the active IgG plasma cell component observed in UC patients.[20] This finding is consistent with studies on B-cell receptor gene usage that have demonstrated that infiltrating lymphocytes in UC mucosa are of peripheral rather than mucosal origin.[21,22] IgG1 and IgG4

antibodies predominate in UC, whereas IgG2 antibodies are increased in CD.[23] The prevalence of the IgG1 type has recently been explored and shown to be specific to UC and lead to greater opsonization of mucosal bacteria and a feed-forward maintenance of the polymorphonuclear leukocyte respiratory burst in UC.[24] One of the transcripts most significantly elevated in UC PBMCs in this study was annotated as immunoglobulin heavy constant gamma 3 (*IgHG3*). The region encompassed by this *IgHG3* qualifier on the Affymetrix chip actually maps (ie, shares 100% nucleotide identity by BLAST) to several sequences ascribed to immunoglobulin heavy constant gamma 1 (*G1m marker*), and has been identified as a marker of inflamed UC gastrointestinal epithelium.[3,4] These results are consistent with the previous observation that IgG1 levels in serum are significantly increased in UC patients relative to serum levels of IgG1 in patients with CD.[25]

A significant subset of patients with IBD cannot be classified by current procedures and constitute cases of indeterminate IBD.[26–28] Therefore, one of the main goals of the present study was to determine whether PBMC profiles in patients with UC and CD were sufficiently distinct to enable classification of these diseases. Results of class prediction analysis indicate that a gene signature in PBMCs can accurately discriminate UC and CD samples. Transcriptional differences are not attributable to cellular composition because cellular compositions of PBMCs from patients appear quite similar.

The disease-specific patterns, if prospectively validated in a larger population, may provide the basis for a molecular diagnosis of UC and CD and contribute to the diagnosis of patients classified as indeterminate IBD. It is quite possible that the proposed Th1 and Th2 natures of CD and UC, respectively, are mainly responsible for the differences in this study, and that other Th1- and Th2-based inflammatory diseases may bear similar signatures to those identified for CD and UC. Nonetheless the PBMC profile identified in this study appears to have clinical utility in IBD, because the gene classifier enables discrimination between these closely related disorders that are sometimes indistinguishable.

This study indicates that transcriptional profiles in the circulating monocytes, T cells, and B cells may serve as a sensitive monitor of the organism's physiological state in the context of IBD. As these cells traverse various tissues, a component of the cellular reaction to the microenvironment is a transcriptional response that can be quantitated through profiling. Expression patterns may reflect disease mechanisms that are of primary or secondary responses to disease pathophysiology. PBMCs, due to their transit through the body, may serve as an accessible surrogate monitor of tissues and systems that are not easily surveyed by common medical practices. A key challenge of expression profiling studies conducted in PBMCs will be to extend pharmacogenomic discoveries to clinical application through the development of assays incorporating gene expression for diagnostic purposes.

Acknowledgments

References

1. Podolsky DK: Inflammatory bowel disease. N Engl J Med 2002, 347:417–429

2. Bouma G, Strober W: The immunological and genetic basis of inflammatory bowel disease. Nat Rev Immunol 2003, 3:521–533

3. Warner EE, Dieckgraefe BK: Application of genome-wide gene expression profiling by high-density DNA arrays to the treatment and study of inflammatory bowel disease. Inflamm Bowel Dis 2002, 8:140–157

4. 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

5. Rockett JC, Burczynski ME, Fornace Jr AJ, Hermann PC, Krawetz SA, Dix DJ: Surrogate tissue analysis: monitoring toxicant exposure and health status of inaccessible tissues through the analysis of accessible tissues and cells. Toxicol Appl Pharmacol 2004, 194:189–199

6. Maas K, Chan S, Parker J, Slater A, Moore J, Olsen N, Aune TM: Cutting edge: molecular portrait of human autoimmune disease. J Immunol 2002, 169:5–9

7. Twine NC, Stover JA, Marshall B, Dukart G, Hidalgo M, Stadler W, Logan T, Dutcher J, Hudes G, Dorner AJ, Slonim DK, Trepicchio WL, Burczynski ME: Disease-associated expression profiles in peripheral blood mononuclear cells from patients with advanced renal cell carcinoma. Cancer Res 2003, 63:6069–6075

8. Mannick EE, Bonomolo JC, Horswell R, Lentz JJ, Serrano MS, Zapata-Velandia A, Gastanaduy M, Himel JL, Rose SL, Udall Jr JN, Hornick CA, Liu Z: Gene expression in mononuclear cells from patients with inflammatory bowel disease. Clin Immunol 2004, 112:247–257

9. Hill AA, Brown EL, Whitley MZ, Tucker-Kellogg G, Hunter CP, Slonim DK: Evaluation of normalization procedures for oligonucleotide array data based on spiked cRNA controls. Genome Biol 2001, 2:RESEARCH0055

10. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES: Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999, 286:531–537

11. Schmielau J, Finn OJ: Activated granulocytes and granulocyte-derived hydrogen peroxide are the underlying mechanism of suppression of T-cell function in advanced cancer patients. Cancer Res 2001, 61:4756–4760

12. De Bruin PAF, Crama-Bohbouth G, Verspaget HW, Verheijen JH,

Dooijewaard G, Weterman IT, Lamers CBHW: Plasminogen activators in the intestine of patients with inflammatory bowel disease. Thromb Haemost 1988, 60:262–266

13. Kruithof EKO, Baker MS, Bunn CL: Biological and clinical aspects of plasminogen activator inhibitor type 2. Blood 1995, 86:4007–4024

14. de Jong E, Porte RJ, Knot EA, Verheijen JH, Dees J: Disturbed fibrinolysis in patients with inflammatory bowel disease. A study in blood plasma, colon mucosa, and faeces. Gut 1989, 30:188–194

15. Schmidt C, Baumeister B, Kipnowski J, Schiermeyer-Dunkhase B, Vetter H: Alteration of prostaglandin E2 and leukotriene B4 synthesis in chronic inflammatory bowel disease. Hepatogastroenterology 1996, 43:1508–1512

16. Trebble TM, Arden NK, Wootton SA, Mullee MA, Calder PC, Burdge GC, Fine DR, Stroud MA: Peripheral blood mononuclear cell fatty acid composition and inflammatory mediator production in adult Crohn's disease. Clin Nutr 2004, 23:647–655

17. Barrera S, Lai J, Fiocchi C, Roche JK: Regulation by prostaglandin E2 of interleukin release by T lymphocytes in mucosa. J Cell Physiol 1996, 166:130–137

18. Danese S, de la Motte C, Fiochhi C: Platelets in inflammatory bowel disease: clinical, pathogenic, and therapeutic implications. Am J Gastroenterol 2004, 99:938–945

19. Gnatenko DV, Dunn JJ, McCorkle SR, Weissmann D, Perrotta PL, Bahou WF: Transcript profiling of human platelets using microarray and serial analysis of gene expression. Blood 2003, 101:2285–2293

20. Farrell RJ, Peppercorn MA: Ulcerative colitis. Lancet 2002, 359:331–340

21. Dunn-Walters DK, Boursier L, Hackett M, Spencer J: Biased JH usage in plasma cell immunoglobulin gene sequences from colonic mucosa in ulcerative colitis but not in Crohn's disease. Gut 1999, 44:382–386

22. Thoree VC, Golby SJ, Boursier L, Hackett M, Dunn-Walters DK, Sanderson JD, Spencer J: Related IgA1 and IgG producing cells in blood and diseased mucosa in ulcerative colitis. Gut 2002, 51:44–50

23. Kett K, Brandtzaeg P: Local IgA subclass alterations in ulcerative colitis and Crohn's disease of the colon. Gut 1987, 28:1013–1021

24. Furrie E, Macfarlane S, Cummings JH, Macfarlane GT: Systemic antibodies towards mucosal bacteria in ulcerative colitis and Crohn's disease differentially activate the innate immune response. Gut 2004, 53:91–98

25. Gouni-Berthold I, Baumeister B, Berthold HK, Schmidt C: Immunoglobulins and IgG subclasses in patients with inflammatory bowel disease. Hepatogastroenterology 1999, 46:1720–1723

26. Winther KV, Fogh P, Thomsen OO, Brynskov J: Inflammatory bowel disease (ulcerative colitis and Crohn's disease): diagnostic criteria and differential diagnosis. Drugs Today (Barc) 1998, 34:935–942

27. Bentley E, Jenkins D, Campbell F, Warren B: How could pathologists improve the initial diagnosis of colitis? Evidence from an international workshop. J Clin Pathol 2002, 55:955–960

28. Guindi M, Riddell RH: Indeterminate colitis. J Clin Pathol 2004, 57:1233–1244