Inflammation, adenoma and cancer: Objective classification of colon biopsy specimens with gene expression signature

Orsolya Galamb\textsuperscript{a,d,1,*}, Balázs Györfy\textsuperscript{b,1}, Ferenc Sipos\textsuperscript{a}, Sándor Spisák\textsuperscript{a}, Anna Mária Németh\textsuperscript{a}, Pál Miheller\textsuperscript{a}, Zsolt Tulassay\textsuperscript{a,d}, Elek Dinya\textsuperscript{c} and Béla Molnár\textsuperscript{a,d}

\textsuperscript{a}2nd Department of Medicine, Semmelweis University, Budapest, Hungary
\textsuperscript{b}Joint Research Laboratory of the Hungarian Academy of Sciences and the Semmelweis University for Pediatrics and Nephrology, 1st Department of Paediatrics, Semmelweis University Budapest, Hungary
\textsuperscript{c}EGIS Pharmaceuticals Ltd. Medical Department, Budapest, Hungary
\textsuperscript{d}Hungarian Academy of Science, Molecular Medicine Research Unit, Budapest, Hungary

Abstract. Gene expression analysis of colon biopsies using high-density oligonucleotide microarrays can contribute to the understanding of local pathophysiological alterations and to functional classification of adenoma (15 samples), colorectal carcinomas (CRC) (15) and inflammatory bowel diseases (IBD) (14). Total RNA was extracted, amplified and biotinylated from frozen colonic biopsies. Genome-wide gene expression profile was evaluated by HGU133plus2 microarrays and verified by RT-PCR. We applied two independent methods for data normalization and used PAM for feature selection. Leave one-out stepwise discriminant analysis was performed. Top validated genes included collagenIV\textsubscript{α1}, lipocalin-2, calumenin, aquaporin-8 genes in CRC; CD44, met proto-oncogene, chemokine ligand-12, ADAM-like decysin-1 and ATP-binding casette-A8 genes in adenoma; and lipocalin-2, ubiquitin D and IFITM2 genes in IBD. Best differentiating markers between Ulcerative colitis and Crohn's disease were cyclin-G2; tripartite motif-containing-31; TNFR shedding aminopeptidase regulator-1 and AMICA. The discriminant analysis was able to classify the samples in overall 96.2\% using 7 discriminatory genes (indoleamine-pyrrole-2,3-dioxygenase, ectodermal-neural cortex, TIMP3, fucosyltransferase-8, collectin sub-family member 12, carboxypeptidase D, and transglutaminase-2). Using routine biopsy samples we successfully performed whole genomic microarray analysis to identify discriminative signatures. Our results provide further insight into the pathophysiological background of colonic diseases. The results set up data warehouse which can be mined further.

Keywords: Gene expression signature, whole genomic oligonucleotide microarray, colon cancer, adenoma, inflammatory bowel disease

1. Introduction

Colorectal cancer (CRC) is one of the most frequent death-causing tumorous diseases in Western countries. CRC frequently follows various high-risk conditions such as adenomatous polyps and inflammatory bowel disease (IBD). The exact diagnosis of IBD types is still often difficult by conventional histology. High-density oligonucleotide microarray analysis gives an opportunity for studying the genetic and gene expression background, understanding of local pathophysiological alterations and for functional classification of diseases.

To date, microarray analyses reported in the literature were performed predominantly from surgically removed CRC samples [16], while microarray gene expression profiling of adenomas and IBDs as colorectal
diseases predisposing to CRC has been studied less. Besides the publications dealing with the comparison of gene expression profiles (GEP) of CRC and normal colonic mucosa [10,21,29], more and more scientific studies appear to focus on gene expression background of CRC progression and metastases development [1—4, 22,25,26,42,43], characterization of CRC subtypes according to mRNA expression [3,14,44], the correlation of GEP with clinicopathological parameters [3,5,7,44], and on the generation of mRNA expression based prognosis [40]. Microarray-based molecular diagnostics of malignancy in colon adenoma and CRC samples were described using 10 [1], 9 [25,26] and 4 [29] adenoma samples compared to adenocarcinoma and normal colonic tissues. Expression microarray analyses of IBD samples were performed in order to determine global GEP of mucosal samples from patients with IBD compared to normal controls [23], to identify novel candidates for ulcerative colitis (UC) and/or Crohn’s disease (CD) genetic susceptibility [24,30], to find marker genes involved in IBD-related carcinogenesis [38], to compare expression of entire chemokine family within IBD and normal patients [31], and to examine changes in GEP in peripheral blood mononuclear cells in IBD patients [27].

mRNA expression array analysis is usually performed using high volume surgically removed tissues. In the gastrointestinal tract, biopsy samples are routinely taken during the endoscopical examination with minimal intervention. The mRNA expression study of these samples could allow further insight into the development of inflammatory, preneoplastic and neoplastic diseases, and these biopsy specimens could be suitable samples for identifying early diagnostic target molecules. Colonic biopsies were applied previously for expression array analysis only in several cases [15, 23,30,31], because, even today, array technology needs significantly more RNA than can be isolated from tiny biopsy specimens. However, new techniques and commercial kits have recently become available for the reliable mRNA amplification without any effect on the original gene expression pattern [39].

Genome-wide gene expression profiling studies using microarrays have the potential to improve the diagnosis and treatment of human cancers and other disorders. However, recently appeared whole genomic oligonucleotide microarrays representing more than 47000 transcripts have not been used in any type of gastrointestinal disorders.

In the present study we aimed to find discriminatory genes between the main diagnostic groups and to develop and test validation assay system and to confirm the applicability of biopsy samples for microarray analysis-based classification. Another purpose of our work was to search for altered biological pathways for explanation of the pathomechanism of these colonic diseases based on whole genomic mRNA microarray results.

2. Materials and methods

2.1. Patients and samples

After informed consent, biopsy samples were taken from the colon during the endoscopical intervention before treatment, and stored in RNAlater Reagent (Qiagen Inc., US) at −80°C. Total RNA was extracted from biopsies of 15 patients with tubulovillous/villous adenomas, 15 with colorectal adenocarcinoma (all microsatellite negative), 9 with active ulcerative colitis (UC), 5 with active Crohn’s colitis (CD) and of 8 healthy normal controls (Table 1). Detailed patient specification is described in Supplemental Table 1.

2.2. Methods

2.2.1. Microarray analysis

Total RNA was extracted using RNeasy Mini Kit (Qiagen, US) according to the manufacturer’s instructions. Quantity and quality of the isolated RNA were tested by measuring the absorbance and agarose gel electrophoresis. Biotinylated cRNA probes were synthesized from 5–8 µg total RNA and fragmented according to the Affymetrix description using GeneChip cDNA synthesis reagents and sample cleanup module and Enzo BioArray HighYield RNA Transcript Labeling Kit (https://www.affymetrix.com/support/downloads/manuals/expressions2manual.pdf – first version). 10 µg of each fragmented cRNA sample was hybridized into HGU133 Plus2.0 array (Affymetrix Inc.) at 45°C for 16 hours. The slides were washed and stained using Fluidics Station 450 and antibody amplification staining method according to the manufacturer’s instructions. The fluorescent signals were detected by a GeneChip Scanner 3000.
Table 1

| Group                      | n  | Male/female | Age ± | Taqman available |
|----------------------------|----|-------------|-------|------------------|
| Adenoma without dysplasia  | 6  | 3/3         | 65.2  | 6                |
| Adenoma with dysplasia     | 9  | 3/6         | 73.6  | 6                |
| Ulcerative colitis         | 9  | 3/6         | 43.8  | 7                |
| Crohn’s disease            | 5  | 2/3         | 25.6  | 1                |
| CRC Dukes B                | 7  | 3/4         | 65.3  | 6                |
| CRC Dukes C-D              | 8  | 4/4         | 67.5  | 4                |
| Normal                     | 8  | 4/4         | 50.6  | 5                |

2.2.2. Statistical analysis

Pre-processing and quality control

Quality control analyses were performed according to the suggestions of The Tumor Analysis Best Practices Working Group [37]. Scanned images were inspected for artifacts, percentage of present calls (> 25%) and control of the RNA degradation were evaluated. Based on the evaluation criteria all biopsy measurements fulfilled the minimal quality requirements. According to the above recommendations we have applied two different normalization methods: MAS 5.0, and RMA [18]. MAS5.0 applies normalization on an individual chip; it has excellent specificity and good sensitivity. RMA applies cross-project normalization; it has good specificity and excellent sensitivity [37]. Further data analysis and -interpretation have been carried out with both of these pre-processing methods in order to yield the best comparison and normalization properties across all measurements.

Feature selection and cluster analysis

We have arranged the complete dataset consisting of 52 expression measurements into classes according to the histological properties of the samples. This selection procedure resulted in six new datasets (CRC/adenoma/IBD vs. normal, CRC Dukes B vs. CRC Dukes C-D, non-dysplastic adenoma vs. dysplastic adenoma, UC vs. CD), which were treated as autonomous classification problems. In order to obtain characteristic signal profiles with high predictive power we have applied the “Prediction Analysis for Microarrays” (PAM) [35]. PAM uses soft thresholding to produce a shrunken centroid, which allows the selection of genes with high predictive potential. In our experimental setup the search for a minimum number of genes with maximum predictive accuracy is not promising as we could distinguish two different groups with a very short gene list. Therefore we decided to set the PAM threshold lower for the selection of the top genes and to pick the top 100 genes for each condition. Due to the nature of PAM, at lower threshold the resulting gene list will be longer, but all genes significant at a higher threshold will be included in any selected set. Finally, the overlap of the two lists—based on two different normalizations—was taken for further analysis. The functional classification of the most differentially expressed genes were performed according to the analysis of RMA top 100 genes in each main disease groups compared to normal controls.

To visualize the discriminative patterns, hierarchical clustering was performed using the GeneSis software [34]. Annotation was performed using the Affymetrix NetAffx analysis centre (http://www.affymetrix.com/analysis/index.axf). The combined datasets for further analysis are available in the Gene Expression Omnibus databank (http://www.ncbi.nlm.nih.gov/geo/), series accession number: GSE4183.

Discriminant analysis

Further data analysis was performed in the SAS v.8.2 (SAS Institute Inc. Cary, NC, USA) program package and the discriminant analysis in SPSS v.15.0 program (SPSS Inc., USA). We have performed a stepwise discriminant analysis among the groups by forward selection of quantitative variables. The set of variables that make up each class was assumed to be multivariate normal with a common covariance matrix. Finally we used the significant discrimination model (Wilks’ Lambda was significant) and we have fixed the most important discriminatory genes. At the end of the analysis we have made the Leave-one-out classification table.

2.2.3. Taqman RT-PCR

TaqMan real-time PCR was used to measure the expression of 52 selected genes using an Applied Biosystems Micro Fluidic Card System in 36 samples, where sufficient RNA could be extracted (Table 1). The measurements were performed using an ABI PRISM® 7900HT Sequence Detection System as described in the products User Guide (http://www.appliedbiosystems.com, CA, USA). For data analysis the SDS 2.2 software was used. The extracted delta Ct values (which represent the expres-
sion normalized to the ribosomal 18S expression) were grouped according to the histological groups. Then the Student’s t-test was performed to compare the expression values between groups.

3. Results

3.1. Identification of discriminatory genes among the main diagnostic groups

CRC cases are characterized by differentially expressed genes involved in cell proliferation (pleiotrophin, insulin-like growth factor binding protein 5, REG1A, WNT1 inducible signaling pathway protein 1), adhesion (MCAM, collagens, enactin, laminin gamma 1), and transport (such as aquaporin 8, lipocalin 2 and collagen 4 A1). Adenoma cases showed altered gene expression data in transport (like ABCA8, TRPM6), adhesion (such as CXCL12, CD44, ADAM-like decysin 1), metabolism (like carbonic anhydrase I) and proliferation (such as MET protooncogene) functional groups. IBD cases are featured by the gene expression changes of immune regulation (such as IFITM3, IFITM1 and proteasome subunit, beta type, 9), cell proliferation (REG1A, tryptophanyl-tRNA synthetase, interferon stimulated gene 20kDa) and metabolism (like chitinase 3-like 1, carbonic anhydrase I, zinc finger protein 91). The top 100 genes for each condition were picked and the overlaps of the two lists, based on two different normalizations were determined. Discriminatory genes between CRC/adenoma/IBD and normal biopsy samples are listed in Table 2, and considered as the disease type specific mRNA expression markers. The list of the top 100 genes for each analysis setting including complete annotation and the complete microarray dataset are shown as supplementary information in Supplemental Table 2. The average within-PAM cross-validation misclassification error was found to be 7.3%.

To visualize the expression changes, we have clustered the selected top genes of all biopsy samples to detect similarities across the sample groups (Fig. 1). The dendrogram of the 52 colonic cases shows the discrimination potential of the selected genes. The transcripts representing the same gene are found to fit into the same cluster.

3.2. Discriminant analysis of the array results

We were able to classify the samples using 7 discriminatory genes (indoleamine-pyrrole 2,3 dioxygenase, ectodermal-neural cortex, TIMP3, fucosyltransferase 8, collectin sub-family member 12, carboxyptidase D, and transglutaminase 2) (Fig. 2). 96.2% of original grouped cases and 94.3% of cross-validated grouped cases were correctly classified (Table 3).

3.3. Discrimination of colonic disease subtypes

The subclassifying genes within the main disease groups were also identified. The metabolic and transport processes (representing genes like DnaJ homolog subfamily C member 10, coatomer protein complex subunit beta) mainly differ between CRC subgroups. In advanced stages of CRC downregulation of apoptosis (such as TIA1 cytotoxic granule-associated RNA binding protein, forkhead box O3A), and immune response (like immunoglobulin heavy constant mu, 2’-5’-oligoadenylate synthetase 2) was observed, while carbohydrate, fatty acid metabolism (like glutamine-fructose-6-phosphate transaminase 1, GDP-mannose 4,6-dehydratase, sterol-C4-methyl oxidase-like, lanosterol synthase) and energy metabolism related genes (such as ATPase inhibitory factor 1) showed higher mRNA expression levels in parallel with CRC progression. In adenoma samples upregulation of proliferation (such as interferon-gamma-inducible protein 16, aminopeptidase A, tumor protein D52-like 2) and DNA replication and transcription regulation (like IGF1, nuclear factor-like 3, zinc finger protein 452) and downregulation of immune and defense response (such as immunoglobulin heavy constant mu, T cell receptor gamma variable 9, interferon regulatory factor 4 and tryptase alpha) were found during the development of dysplastic alterations. CD cases are mainly featured by increased expression of carbohydrate metabolism genes (like galactosidase alpha, fructose-2,6-bisphosphatase 4, malte glucoseamylase), while certain cell proliferation (such as septin 10, platelet derived growth factor D, cyclin G2), apoptosis (such as BCL10, BIRC4, egl nine homolog 3), immune regulation (like decay accelerating factor, CD24, CEACAM1), transport (like dual oxidase 2, P450 (cytochrome) oxidoreductase and lipocalin 2), and ubiquitin-dependent proteasome catabolism (such as tripartite motif-containing 31, ubiquitin-conjugating enzyme E2, J1, and Ubiquitin specific protease 53) genes were found to be overexpressed in UC compared with CD cases. However, the function of many sub-
Table 2
Genes differentially expressed in CRC/adenoma/IBD compared to normal mucosa (overlap of genes identified using PAM after both RMA and MAS 5.0 normalizations)

| Probe Set ID | Gene Name                                | Gene Symbol | Gene Ontology Biological Process                              | Fold change |
|--------------|------------------------------------------|-------------|----------------------------------------------------------------|-------------|
| CRC vs normal | von Willebrand factor                     | VWF         | cell adhesion / platelet activation / response to wounding      | 3.61        |
| 211980       | collagen, type IV, alpha 1                | COL4A1      | phosphate transport                                             | 5.21        |
| 212531       | lipocalin 2 (oncogene 24p3)               | LCN2        | transport                                                       | 7.97        |
| 240157       | cavolin 1, caveloes protein, 22kDa        | CAV1        | —                                                              | —           |
| 22072        | olfactomedin 4                            | OLFM4       | —                                                              | —           |
| 22072        | ammonium homolog (mouse)                 | AMN         | —                                                              | —           |
| 203065       | aquaporin 8                               | AQP8        | water transport                                                 | —           |
| 215894       | prostaglandin D2 receptor (DP)            | PTGDR       | G-protein coupled receptor protein signaling pathway           | —           |
| 214845       | calumenin                                 | CALU        | —                                                              | 3.23        |
| 213119       | solute carrier family 36                 | SLC36A1     | amino acid transport                                            | —           |
| Adenoma vs normal | ATP-binding cassette, sub-family A (ABC1), member 8 | ABCA8 | transport                                                      | —12.78      |
| 230204       | met proto-oncogene (hepatocyte growth factor receptor) | MET | cell proliferation / cell surface receptor linked signal transduction / protein amino acid phosphorylation | —4.05       |
| 240389       | vitelliform macular dystrophy 2-like 2   | VMD2L2      | ion transport                                                   | —13.50      |
| 224412       | transient receptor potential cation channel, subfamily M, member 6 | TRPM6   | cation transport / protein amino acid phosphorylation       | —11.34      |
| 209687       | chemokine (C-X-C motif) ligand 12         | CXCL12      | G-protein coupled receptor protein signaling pathway / calcium ion homeostasis / cell adhesion / cell-cell signaling | —6.91       |
| 212014       | CD44 antigen (homing function and Indian blood group system) | CD44       | cell adhesion                                                        | 3.87        |
| 205523       | hyaluronan and proteoglycan link protein 1 | HAPLN1    | cell adhesion                                                      | —5.04       |
| 206134       | ADAM-like, decysin 1                    | ADAMDEC1    | integrin-mediated signaling pathway / negative regulation of cell adhesion / proteolysis and peptidolysis | —6.81       |
| 209835       | CD44 antigen (homing function and Indian blood group system) | CD44       | cell adhesion                                                        | 3.74        |
| 205950       | carbonic anhydrase 1                     | CA1         | one-carbon compound metabolism                                   | —9.77       |
| 228507       | Phosphodiesterase 3A, CGM-inhibited      | PDE3A       | lipid metabolism / signal transduction                           | —4.36       |
| 225275       | EGF-like repeats and discoidin 1-like domains 3 | EDIL3      | cell adhesion / development                                     | —4.38       |
| 203000       | stathmin-like 2                          | STMN2       | intracellular signaling cascade / neuron differentiation         | —5.17       |
| 236300       | Phosphodiesterase 3A, CGM-inhibited      | PDE3A       | lipid metabolism / signal transduction                           | —3.85       |
| 220724       | hypothetical protein FLJ21511            | FLJ21511    | —                                                              | —           |
| 207504       | carbonic anhydrase VII                   | CA7         | one-carbon compound metabolism                                   | —8.08       |
| 204351       | S100 calcium binding protein P           | S100P       | —                                                              | 4.38        |
| 229869       | hypothetical protein MGC45780            | MGC45780    | phosphate transport                                             | —5.04       |
| 202242       | transmembrane 4 superfamily member 2     | TM4SF2      | N-linked glycosylation                                          | —3.74       |
| 235849       | hypothetical protein MGC45780            | MGC45780    | phosphate transport                                             | —4.59       |
| 218756       | hypothetical protein MGC4172             | MGC4172     | metabolism                                                      | —3.65       |
| 215177       | integrin, alpha 6                        | ITGA6       | cell-matrix adhesion / cell-substrate junction assembly / integrin-mediated signaling pathway | —7.77       |
| 204931       | transcription factor 21                  | TCF21       | mesoderm development / regulation of transcription, DNA-dependent | —3.20       |
Table 2, continued

| Probe Set ID | Gene Name | Gene Symbol | Gene Ontology | Biological Process | Fold change |
|--------------|-----------|-------------|---------------|--------------------|-------------|
| 207003_at    | guanylate cyclase activator 2A (guanylin) | GUCA2A | signal transduction | −7.55            |
| 212203_at    | interferon induced transmembrane protein 3 (1-8U) | IFITM3 | immune response / response to biotic stimulus | 3.41          |
| 204279_at    | proteasome (prosome, macropain) subunit, beta type, 9 (large multifunctional protease 2) | PSMB9 | immune response / proteolysis and peptidolysis / ubiquitin-dependent protein catabolism | 3.77          |
| 240389_at    | interferon induced transmembrane protein 1 (9–27) | IFITM1 | cell surface receptor linked signal transduction / immune response / negative regulation of cell proliferation | −7.81         |
| 214022_at    | interferon induced transmembrane protein 1 (9–27) | IFITM1 | cell surface receptor linked signal transduction / immune response / negative regulation of cell proliferation | 2.62          |
| 201649_at    | ubiquitin-conjugating enzyme E2L 6 | UBE2L6 | ubiquitin cycle | 3.31          |
| 206059_at    | zinc finger protein 91 (HPF7, HTF10) | ZNF91 | carbohydrate metabolism / regulation of transcription, DNA-dependent | −2.22         |
| 224412_at    | transient receptor potential cation channel, subfamily M, member 6 | TRPM6 | cation transport / protein amino acid phosphorylation | −7.06         |
| 213369_at    | — | — | — | −6.27         |
| 212531_at    | lipocalin 2 (oncogene 24p3) | LCN2 | transport | 10.96         |

The functional classification of most differentially expressed genes was performed according to the analysis of RMA top 100 genes in each disease type subgroups. The list of the top 100 genes for each analysis setting including complete annotation and the complete microarray dataset are shown as supplementary information (Supplemental Table 2). Hierarchical cluster diagrams of the subgroups, based on RMA top 100 differentially expressed genes can be seen on Fig. 3.

3.4. Gene ontology of selected features

The representative gene ontology categories are shown to Tables 2 and 3. We have also mapped the selected features to chromosomes (Supplemental Fig. 1).

3.5. Taqman validation

Selection criteria for genes were the different expression in microarray analysis and the availability of validated TaqMan probes. Ten “literature” genes were also selected which were described as CRC related genes. The complete results of the TaqMan measurements are presented on Supplemental Table 3. Forty six of the 52 measured genes correlated with the results obtained using Affymetrix microarrays at a significance of \( p < 0.05 \). The expression changes of the selected genes are summarized in Table 4. The mRNA expression levels of selected discriminatory genes measured by Taqman RT-PCR are presented on Supplemental Fig. 2. Global clustering of all samples using the Taqman validated genes were also performed (Supplemental Fig. 3). Normal and UC cases belong to two distinct clusters, but clusters of CRC and adenoma cases are not clearly separated, demonstrating the expressional heterogeneity of CRC.

4. Discussion

Gene expression profiling of 52 colonic biopsy samples was done by whole genomic HGU133 Plus 2.0 microarrays in order to identify disease specific gene expression marker groups for objective classification. We aimed to develop diagnostic mRNA expression patterns for identification of adenoma and different staged CRC and of the minimal list of genes which is suitable for discrimination of different types of IBD. Examination of adenoma with/without dysplasia, CRC, IBD and normal samples in parallel can help to find condition specific gene expression alteration with a lower risk of unspecificity due to methodical reasons. Comparative microarray analysis of biopsies from all of these kinds of colonic diseases has not been reported so far in the scientific literature. Oligonucleotide whole genomic microarray analyses of biopsy samples were found to be highly standardized, reproducible and provided high quality array results regarding the array
Fig. 1. Combined clustering dendrogram of the top discriminative transcripts for IBD, adenoma and CRC. The disease samples were grouped, normalized and compared to normal controls. Samples, not included in the classification of selected disease are shaded. Repeated transcripts are marked with arrows.

Fig. 2. Discriminant analysis of colonic biopsy specimens. Note the clear separation of the single classification groups based on the discriminatory genes detailed in the results section. Group 1 = normal, Group 2 = adenoma, Group 3 = colorectal cancer, Group 4 = inflammatory bowel disease.
Table 3
Classification results\textsuperscript{b,c}

| Group | Predicted Group Membership | Total |
|-------|---------------------------|-------|
|       | 1 | 2 | 3 | 4 |     |
| Original Count | | | | | |
| 1 | 8 | 0 | 0 | 0 | 8 |
| 2 | 0 | 15 | 0 | 0 | 15 |
| 3 | 0 | 1 | 14 | 0 | 15 |
| 4 | 0 | 0 | 1 | 14 | 15 |
| % | 100.0 | 0 | 0 | 0 | 100.0 |
| 2 | 0 | 100.0 | 0 | 0 | 100.0 |
| 3 | 0 | 6.7 | 93.3 | 0 | 100.0 |
| 4 | 0 | 0 | 6.7 | 93.3 | 100.0 |

Cross-validated\textsuperscript{a} Count | | | | | |
| 1 | 8 | 0 | 0 | 0 | 8 |
| 2 | 0 | 15 | 0 | 0 | 15 |
| 3 | 1 | 1 | 13 | 0 | 15 |
| 4 | 0 | 0 | 1 | 14 | 15 |
| % | 100.0 | 0 | 0 | 0 | 100.0 |
| 2 | 0 | 100.0 | 0 | 0 | 100.0 |
| 3 | 6.7 | 6.7 | 86.7 | 0 | 100.0 |
| 4 | 0 | 0 | 6.7 | 93.3 | 100.0 |

\textsuperscript{a}Cross validation is done only for those cases in the analysis. In cross validation, each case is classified by the functions derived from all cases other than that case.

\textsuperscript{b}96.2% of original grouped cases correctly classified.

\textsuperscript{c}94.3% of cross-validated grouped cases correctly classified.

Fig. 3. Discriminating disease subtypes. Hierarchical clustering of genes associated with disease subtype-differentiation using the top 100 differentially expressed genes obtained after RMA normalization. A. CRC Dukes B vs. CRC Dukes C-D. B. Adenoma without dysplasia vs adenoma with dysplasia. C. Ulcerative colitis vs Crohn’s disease.
| Taqman ID       | Gene Symbol | Gene name                                      | Affymetrix ID | Sample groups | p value | ddCt | Fold change in microarrays |
|----------------|-------------|-----------------------------------------------|---------------|---------------|---------|------|----------------------------|
| HS00153304_m1  | CD44        | CD44 antigen                                   | 212014_s_at   | AD vs normal  | 1.82E-07 | 1.903 | 7.55                        |
| HS00157859_m1  | GUCA2A      | guanylate cyclase activator 2A                | 207003_s_at   | AD vs normal  | 9.04E-01 | -4.34 | -2.33                      |
| HS00171022_m1  | CXCL12      | chemokine (C-X-C motif) ligand 12             | 209685_s_at   | CRC vs normal | 0.00350  | -2.04 | -6.91                      |
| HS00179845_m1  | MET         | met proto-oncogene                            | 203510_at     | AD vs normal  | 1.41E-06 | 2.17  | 3.91                        |
| HS00200350_m1  | ABCA8       | ATP-binding cassette, sub-family A (ABC1), member 8 | 204719_s_at   | AD vs normal  | 0.000610 | -3.35 | -12.78                     |
| HS00205545_m1  | ADAMDEC1    | ADAM-like, decysin 1                          | 206134_at     | UC vs normal  | 8.07E-05 | -2.70 | -2.27                      |
| HS00214306_m1  | TRPM6       | transient receptor potential cation channel, subfamily M, member 6 | 224412_s_at   | AD vs normal  | 5.79E-05 | -4.73 | -11.59                     |
| HS00153408_m1  | MYC         | v-myc myelocytomatosis viral oncogene homolog (avian) | 202431_s_at   | CRC vs normal | 0.000827 | -3.09 | -2.73                      |
| HS00163669_m1  | CA2         | carbonic anhydrase II                         | 209301_at     | AD vs normal  | 9.01E-04 | 2.58  | 1.95                        |
| HS00171558_m1  | TIMP1       | tissue inhibitor of metalloproteinase 1       | 201666_at     | CRC vs normal | 0.00193  | -3.17 | -2.14                      |
| HS00236937_m1  | CXCL1       | chemokine (C-X-C motif) ligand 1              | 204470_at     | AD vs normal  | 1.66E-05 | 3.55  | 5.38                        |
| HS00236966_m1  | CXCL2       | chemokine (C-X-C motif) ligand 2              | 209777_at     | AD vs normal  | 1.24E-05 | 3.98  | 5.49                        |
| HS00266139_m1  | CA1         | carbonic anhydrase I                          | 205950_at     | CRC vs normal | 0.000219 | 2.36  | 2.03                        |
| HS00266915_m1  | ADH1B       | alcohol dehydrogenase IB (class I), beta polypeptide | 209613_at     | AD vs normal  | 0.000382 | -3.49 | -4.27                      |
| HS00154124_m1  | AQP8        | aquaporin 8                                   | 206784_at     | AD vs normal  | 9.01E-04 | -6.10 | -6.20                      |
| HS00194333_m1  | LCN2        | lipocalin 2                                   | 212531_at     | AD vs normal  | 2.67E-07 | 6.13  | 4.85                        |
| HS00197437_m1  | OLFM4       | olfactomedin 4                                | 212768_at     | CRC vs normal | 2.57E-07 | 7.04  | 4.28                        |
| HS00154230_m1  | CALU        | calumenin                                     | 214845_at     | CRC vs normal | 0.00179  | 6.20  | 3.70                        |
| HS00167995_m1  | VWF         | von Willebrand factor                         | 202112_at     | CRC vs normal | 0.000827 | 3.70  | 9.20                        |
| HS00229558_m1  | AMN         | amnionless homolog (mouse)                    | 223587_s_at   | CRC vs normal | 0.000112 | 2.44  | 2.75                        |
| HS00235003_m1  | PTGDR       | prostaglandin D2 receptor                     | 215894_at     | CRC vs normal | 0.00499  | -2.32 | -3.91                      |
| HS00266237_m1  | COL4A1      | collagen, type IV, alpha 1                    | 211980_at     | CRC vs normal | 0.0283  | 3.38  | 5.21                        |
| HS00156076_m1  | BGN         | biglycan                                      | 213905_at     | CRC vs normal | 0.120   | 1.20  | 2.24                        |
| HS00169777_m1  | PECAM1      | platelet/endothelial cell adhesion molecule   | 208983_at     | CRC vs normal | 0.764   | 1.59  | 1.59                        |
| HS00174103_m1  | IL8         | interleukin 8                                 | 202859_s_at   | CRC vs normal | 0.0283  | 7.21  | 20.20                      |
| HS00204187_m1  | DUOX2       | dual oxidase 2                                | 219727_at     | AD vs normal  | 5.77E-06 | 5.25  | 6.00                        |
| HS00160066_m1  | PI3         | protease inhibitor 3, skin-derived (SKALP)    | 203691_at     | CRC vs normal | 0.00363  | 4.91  | 9.70                        |
| HS00195812_m1  | LIPG        | lipase, endothelial                          | 219184_at     | UC vs normal  | 7.84E-05 | 6.35  | 6.43                        |
| HS00197374_m1  | UBD         | ubiquitin D                                  | 205890_s_at   | UC vs normal  | 0.000257 | 4.26  | 3.40                        |
sensitivity, present percentage and GAPDH 3′/5′ ratio. This type of analysis results in discriminant signatures, and gives an insight into the pathophysiological background of colonic diseases, and additionally, provides a data warehouse which can be further mined for in-depth pathway analyses. As recently described, joint analysis is more efficient than replication based analysis for two-stage genome-wide association studies [33]. Therefore we used a one-stage genome wide analysis to identify relevant gene expression signatures. For a classification problem comparable to our study a previous estimation suggested a required sample size of 51 subjects to detect a 2-fold change of expression level at alpha = 0.001 at the 90th percentile [41].

The main disease groups were individually compared to healthy controls. CRC samples were unequivocally distinguished according to the expression level of 13 genes. Six of them were validated by Taqman RT-PCR. Among the discriminatory genes lipocalin 2 (LCN2), collagen 4 alpha 1 (COL4A1) and aquaporin 8 (AQP8) were mentioned earlier as CRC-associated genes. LCN2 transport molecule acts as chemotactic agent and also regulates the matrix metalloproteinase-9 activity. AQP8 water channel protein is a marker for non-proliferative colonic epithelial cells, but it is not expressed by adenoma and CRC in protein level. In correlation with the findings of Fischer et al., lower AQP8 mRNA level was found in CRC and adenoma samples in our study also [13].

Adenoma cases were characterized and distinguished according to the expression changes of 27 overlapping genes. Seven of them (overexpressed CD44 and MET, and underexpressed GUCA2A, CXCL12, TRPM6, ABCA8, ADAMDEC1) were confirmed by Taqman RT-PCR. CD44 cell surface glycoprotein acts as a receptor for hyaluronic acid, which can also interact with EGFR, matrix metalloproteases and IGFBP3. The expression changes of CD44 can affect related proliferation and tumorigenesis, tumor tissue...
remodeling and immune processes. Hepatocyte growth factor receptor (MET), which was found to be overexpressed both in colon adenoma and CRC, may play an important role in colorectal tumorigenesis. Similarly to the results of Trovato et al., in our study elevated c-met mRNA level was observed both in adenoma and CRC biopsies, but CRC samples showed lower c-met expression than adenomas. Reduced expression of c-met can be associated to the progression of adenoma into carcinoma [36]. GUCA2A (guanylin) plays role in the regulation of ion transport in the colon. Expression of guanylin is downregulated in human intestinal adenomas, moreover, recent results suggest that loss of guanylin activity leads to or is a result of adenocarcinoma [8]. The chemokine ligand 12 (CXCL12) which was found underexpressed both in different carcinomas earlier [32] and adenoma samples in our study, regulates cAMP production and ion transport in intestinal epithelial cells [12]. These data can support that alterations in ion transport of the colon are involved in colorectal carcinogenesis. The exact cellular function and role in adenoma and tumor development of TRPM6, ABCA8 and ADAMDEC1 gene products have not yet been determined.

Several genes were found to show elevated mRNA level according to the adenoma-CRC sequence. TCF4 is a transcriptional factor involved in Wnt-signalling pathway which is altered in over 90% of CRCs. TCF4 participates in transcriptional regulation of genes associated with colon carcinogenesis including c-myc, cyclin D1, TCF1, PPARδ, MMP7 and MDR1. Tensin 1 (TNS1) which was also overexpressed in CRC compared to adenoma samples can also induce JNK and p38 activation leading to increased cell survival. RBMS1 is another gene which was detected as being upregulated in CRC compared to adenoma, it is also involved in the malignant transformation process. RBMS1 is a modulator of c-myc, deregulates cell cycle controls and leads cells towards transforming pathways [28]. SPARC (osteonectin) was detected as overexpressed gene in CRC in several microarray-based studies [10, 29, 42]. It is thought to play an important role in tissue remodeling, angiogenesis, and tumorigenesis. Controversial and conflicting data were published about the expression and function of insulin-like growth factor binding protein 3 and SPARCL1 in different types of cancers including CRC [19, 20].

Ten discriminatory transcripts distinguish between IBD samples and normal tissue. Overexpression of different interferon-induced genes is highly represented among the discriminatory genes. Interferon induced transmembrane protein 3 (IFITM3) was strongly expressed in severely inflamed colonic mucosa of UC both in our microarray analysis and other studies [17]. Moreover, IFITM3 showed high mRNA level in sporadic cancers, and UC-associated cancers, therefore it can be a marker for identification of high cancer-risk group within the UC. PSMB9 and UBE2L6 interferon-induced discriminatory genes are in connection with the enhanced antigen processing and presentation. LCN2 which has been mentioned above as an upregulated CRC-associated gene is also overexpressed in colonocytes and neutrophils in inflamed lesions of UC. Similarly to our findings, highly increased LCN2 mRNA levels were measured in UC samples in other microarray studies [11, 24]. Alteration of epithelial magnesium absorption was also observed in our IBD samples, as the TRPM6 (transient receptor potential cation channel, subfamily member 8) showed lower mRNA level. CD cases are mainly featured by increased expression of carbohydrate metabolism genes, while certain cell proliferation, apoptosis, immune regulation, transport, and ubiquitin-dependent protein catabolism genes were found to be overexpressed in UC compared with CD cases. However, the function of many discriminatory genes has not been identified yet. Significant overexpression of cancer-related genes (CEACAM1, -7, CD24, PDGFD) in UC is potentially important, considering reports of increased risk of developing CRC in this disease [6, 9].

For validation of the marker properties of a given gene, we should use homogeneous high case-number sample groups for microarray analysis, and further experiments, but at least RT-PCR confirmation, are needed. We focused on individual markers and independent validation of markers. Ninety-four percent of 52 selected genes which were found to be over- or underexpressed was confirmed by Taqman RT-PCR.

In conclusion, in our study we were able to distinguish not just between normal, adenoma, CRC and IBD samples, but also among the different stages of CRC using only easily-taken biopsy specimens. With a large number of samples one can establish principal gene lists that characterize distinct conditions, and if miCarray will be commercially available, the daily routine of diagnosis may be quicker and easier.
| Sample Affy ID | Sex | Age | Localization | Histology | TNM    | Grade | Dukes | Dysplasia | Taqman analysis |
|---------------|-----|-----|--------------|-----------|--------|--------|--------|------------|-----------------|
| Adenoma       |      |     |              |           |        |        |        |            |                 |
| 956*          | 956 U133 Plus 2.0.CEL | F | 84 | flexure hepatica | adenoma villous | severe |        |        |            |                 |
| 983*          | 983 U133 Plus 2.0.CEL | F | 64 | rectum | adenoma tubulovillous | severe | Yes (3.) |        |                 |
| 1115*         | 1115 U133 Plus 2.0.CEL | F | 75 | coecum | adenoma tubulovillous | severe |        |        |                 |
| 1141          | 1141 U133 Plus 2.0.CEL | M | 63 | rectum | adenoma tubulovillous | moderate | Yes (10.) |        |                 |
| 1187*         | 1187 U133 Plus 2.0.CEL | F | 80 | coecum | adenoma tubulovillous | moderate | Yes (13.) |        |                 |
| 1312*         | 1312 U133 Plus 2.0.CEL | M | 93 | rectum | adenoma villous | severe | Yes (15.) |        |                 |
| 1700*         | 1700 U133 Plus 2.0.CEL | F | 69 | coecum | adenoma tubulovillous | severe |        |        |                 |
| 1748*         | 1748 U133 Plus 2.0.CEL | F | 78 | colon desc | adenoma villous | severe | Yes (32.) |        |                 |
| 1832*         | 1832 U133 Plus 2.0.CEL | M | 56 | sigma-colon desc | adenoma villous | moderate | Yes (34.) |        |                 |
| 980           | 980 U133 Plus 2.0.CEL | F | 83 | coecum | adenoma tubulovillous | no | Yes (2.) |        |                 |
| 995*          | 995 U133 Plus 2.0.CEL | M | 58 | coecum, coecum, rectum | adenoma tubulovillous | no | Yes (4.) |        |                 |
| 1138*         | 1138 U133 Plus 2.0.CEL | M | 70 | rectum | adenoma tubulovillous | no | Yes (9.) |        |                 |
| 1154*         | 1154 U133 Plus 2.0.CEL | F | 73 | coecum | adenoma tubulovillous | no | Yes (11.) |        |                 |
| 1419*         | 1419 U133 Plus 2.0.CEL | F | 37 | rectum | adenoma villous | no | Yes (19.) |        |                 |
| 1830          | 1830 U133 Plus 2.0.CEL | M | 70 | sigma | adenoma tubulovillous | no | Yes (33.) |        |                 |
| CRC           |      |     |              |           |        |        |        |            |                 |
| 1158*         | 1158 U133 Plus 2.0.CEL | M | 56 | rectum | adenocarcinoma | T3N0M0 | G2 | B2 | Yes (12.) |                 |
| 1293*         | 1293 U133 Plus 2.0.CEL | F | 88 | ascendent colon | adenocarcinoma | T3N0M0 | G3 | B2 | Yes (14.) |                 |
| 1486*         | 1486 U133 Plus 2.0.CEL | F | 51 | hepatic flexure | adenocarcinoma | T2N0M0 | G1 | B1 | Yes (23.) |                 |
| 1708*         | 1708 U133 Plus 2.0.CEL | F | 72 | sigma | adenocarcinoma | T3N0M0 | G2 | B2 | Yes (30.) |                 |
| 1739*         | 1739 U133 Plus 2.0.CEL | F | 76 | colon desc | adenocarcinoma | T4N0M0 | G1 | B3 | Yes (31.) |                 |
| 1761          | 1761 U133 Plus 2.0.CEL | M | 58 | rectum | adenocarcinoma | T2N0M0 | G2 | B1 |        |                 |
| 1883*         | 1883 U133 Plus 2.0.CEL | M | 56 | coecum tumor | adenocarcinoma | T3N0M0 | G2 | B2 | Yes (35.) |                 |
| 1146*         | 1146 U133 Plus 2.0.CEL | M | 46 | descendent colon | adenocarcinoma | T2N1M0 | G2 | C1 |        |                 |
| 1316*         | 1316 U133 Plus 2.0.CEL | M | 85 | icenic flexure | adenocarcinoma | T3N1M1 | G2 | D | Yes (16.) |                 |
| 1377*         | 1377 U133 Plus 2.0.CEL | F | 69 | sigma | adenocarcinoma | T4N1M0 | G1-2 | C3 |        |                 |
| 1479*         | 1479 U133 Plus 2.0.CEL | F | 77 | sigma | adenocarcinoma | T4N0M1 | G1 | D |        |                 |
| 1494*         | 1494 U133 Plus 2.0.CEL | M | 73 | rectum | adenocarcinoma | T3N1M1 | G1 | D | Yes (24.) |                 |
| 1499*         | 1499 U133 Plus 2.0.CEL | M | 65 | rectum | adenocarcinoma | T3N0M1 | G1 | D |        |                 |
| 1556*         | 1556 U133 Plus 2.0.CEL | F | 66 | coecum | adenocarcinoma | T3N1M1 | G2 | D | Yes (26.) |                 |
| 1651          | 1651 U133 Plus 2.0.CEL | F | 59 | rectum | adenocarcinoma | T2N0M1 | G1 | D | Yes (28.) |                 |
| IBD           |      |     |              |           |        |        |        |            |                 |
| 939           | 939 U133 Plus 2.0.CEL | F | 34 | sigma,rectum | severe active ulcerative colitis | no |        | Yes (1.) |                 |
| 1017          | 1017 U133 Plus 2.0.CEL | F | 39 | total colon | severe active ulcerative pancolitis | no |        | Yes (5.) |                 |
| 1060          | 1060 U133 Plus 2.0.CEL | F | 73 | total colon | severe active ulcerative colitis | no |        | Yes (7.) |                 |
| 1156          | 1156 U133 Plus 2.0.CEL | M | 43 | sigma,rectum | severe active ulcerative colitis | no |        |        |                 |
| 1175          | 1175 U133 Plus 2.0.CEL | M | 43 | total colon | severe active ulcerative colitis | no |        |        |                 |
| 1368          | 1368 U133 Plus 2.0.CEL | F | 24 | total colon | severe active ulcerative colitis | mild | Yes (18.) |        |                 |
| 1533          | 1533 U133 Plus 2.0.CEL | F | 35 | rectum,sigma | severe active ulcerative colitis | no |        | Yes (25.) |                 |
| 1606          | 1606 U133 Plus 2.0.CEL | F | 35 | total colon | severe active ulcerative pancolitis | no |        | Yes (27.) |                 |
| 1665          | 1665 U133 Plus 2.0.CEL | M | 66 | total colon, mainly sigma,rectum | severe active ulcerative colitis | mild-moderate | Yes (29.) | |                 |
Supplement Table 1, continued

| Sample | Affy ID | Sex | Age | Localization | Histology           | TNM | Grade | Dukes | Dysplasia | Taqman analysis |
|--------|---------|-----|-----|--------------|---------------------|-----|-------|-------|-----------|-----------------|
| 923    | 923 U133 Plus 2.0 .CEL | F   | 37  | sigma,rectum | severe active Crohn’s disease |     |       | no    |           |                 |
| 1110   | 1110 U133 Plus 2.0 .CEL | F   | 16  | total colon  | severe active Crohn’s disease | mild | Yes (8.) |       |           |                 |
| 1118   | 1118 U133 Plus 2.0 .CEL | M   | 26  | right side   | severe active Crohn’s disease | no  |       |       |           |                 |
| 1670   | 1670 U133 Plus 2.0 .CEL | M   | 23  | total colon  | severe active Crohn’s disease | no  |       |       |           |                 |
| 1802   | 1802 U133 Plus 2.0 .CEL | F   | 26  | total colon  | severe active Crohn’s disease | no  |       |       |           |                 |
|        |         |     |     |              |                      |     |       |       |           |                 |
| Normal | 1024    | F   | 47  | normal       | normal               |     |       |       | yes (6.)  |                 |
| 1081   | 1081 U133 Plus 2.0 .CEL | F   | 57  | normal       | normal               |     |       |       |           |                 |
| 1114   | 1114 U133 Plus 2.0 .CEL | F   | 60  | normal       | normal               |     |       |       |           |                 |
| 1122   | 1122 U133 Plus 2.0 .CEL | M   | 45  | normal       | normal               |     |       |       |           |                 |
| 1357   | 1357 U133 Plus 2.0 .CEL | M   | 47  | normal       | normal               | yes (17.) |       |       |           |                 |
| 1431   | 1431 U133 Plus 2.0 .CEL | M   | 50  | normal       | normal               | yes (20.) |       |       |           |                 |
| 1440   | 1440 U133 Plus 2.0 .CEL | F   | 44  | normal       | normal               | yes (21.) |       |       |           |                 |
| 1456   | 1456 U133 Plus 2.0 .CEL | M   | 55  | normal       | normal               | yes (22.) |       |       |           |                 |

Microsatellite instability analysis was done from samples marked with *. GEO Accession numbers: “normal: GSM95473 – GSM95480”, “adenoma: GSM95481 – GSM95495”, “CRC: GSM95496 – GSM95510”, “IBD: GSM95511 – GSM95525”.

Supplement Figure 1. Chromosomal localization of discriminatory genes.

The Supplemental Table 2 and the Supplemental Table 3 are available on the http://3dhistech.com/en/article/DMA-511 website.
Supplement Figure 2. mRNA expression levels of selected discriminatory genes measured by Taqman RT-PCR. dCt is the expression value normalized to the ribosomal 18S protein.

Supplement Figure 3. Global clustering of all samples measured by Taqman RT-PCR according to the Taqman validated genes. Probably due the heterogeneity of CRC samples they do not cluster together during global clustering using Taqman-generated results.
Acknowledgements

We thank hereby the support of Professor Zoltán Szentirmay for the microsatellite instability analysis, of László Nagy MD, PhD and Beáta Scholtz PhD for Taqman real-time PCR analysis and of Zoltán Szallási MD, PhD for manuscript review. The study was supported in part by the National Office for Research and Technology, Hungary.

References

[1] D Agrawal, T Chen, R Iby, J Quackenbush, A.F. Chambers, M. Szabo, A. Cantor, D. Coppola and T.J. Yeatman, Osteopontin identified as colon cancer tumor progression marker, C R Biol 326 (2003), 1041–1043.

[2] E. Bandres, V. Catalan, I. Sola, B. Honorato, E. Cubedo, [3] F. Bertucci, S. Salas, S. Eysteries, V. Nasser, P. Finetti, C. Ginestier, E. Charafe-Jauffret, B. Loriod, L. Bachelart, J. D. Birnbaum and R. Houlgatte, Gene expression profiling of dysregulated apoptosis is a major mechanism in the lymph node involvement in colorectal carcinoma, J Oncol Rep 12 (2004), 287–292.

[4] K. Birkenkamp-Demtroder, S.H. Olesen, F.B. Sorensen, S. Laurberg, P. Laiho, L.A. Aaltonen, S. Laurberg, P. Laiho, L.A. Aaltonen and T.F. Orntoft, Differential gene expression in colon cancer DNA microarrays and correlation with pathological parameters, Oncogene 23 (2004), 1377–1391.

[5] K. Birkenkamp-Demtroder, S.H. Olesen, F.B. Sorensen, S. Laurberg, P. Laiho, L.A. Aaltonen and T.F. Orntoft, Differential gene expression in colon cancer of the caecum versus the sigmoid and rectosigmoid, Gut 54 (2005), 374–384.

[6] O. Brostrom, R. Lofberg, B. Nordenvall, A. Ost and G. Hellers, Colon cancer and novel susceptibility candidate genes, Hepatology 38, 374–384.

[7] S.T. Chiu, F.J. Hsieh, S.W. Chen, C.L. Chen, H.F. Shu and H. Li, Clinical-pathologic correlation of up-regulated genes identified using cDNA microarray and real-time reverse transcription-PCR in human colorectal cancer, Cancer Epidemiol Biomarkers Prev 14 (2005), 437–443.

[8] M.B. Cohen, J.A. Hawkins and D.P. Witte, Guanylin mRNA expression in human intestine and colorectal adenocarcinoma, Lab Invest 78 (1998), 101–108.

[9] C.M. Costello, N. Mah, R. Hasler, P. Rosenstiell, G.H. Waetzig, A. Hahn, T. Lu, Y. Gurbuzz, S. Nikolaus, M. Albrecht, J. Hampe, R. Lucius, G. Kroppe, H. Eickhoff, H. Lehrach, T. Lengauer and S. Schreiber, Dissection of the inflammatory bowel disease transcriptome using genome-wide cDNA microarrays, PLoS Med 2 (2005), e199.

[10] R.S. Croner, T. Foertsch, W.M. Brueckl, K. Guenther, R. Siebenhaar, C. Stremmel, K.E. Matzel, T. Papadopoulos, T. Kirchner, J. Behrens, L. Klein-Hitpass, M. Stuerzl, W. Hohenberger and B. Reingruber, Common denominator genes that distinguish colorectal carcinoma from normal mucosa, Int J Colorectal Dis 20 (2005), 353–362.

[11] T.P. Dooley, E.V. Curto, S.P. Reddy, R.L. Davis, G.W. Lambert, T.W. Wilborn and C.O. Elson, Regulation of gene expression in inflammatory bowel disease and correlation with IBD drugs: screening by DNA microarrays, Inflamm Bowel Dis 10 (2004), 1–14.

[12] M.B. Dwinnell, H. Ogawa, K.E. Barrett and M.F. Kagnoff, SDF-1/CXCR1 regulates cAMP production and ion transport in intestinal epithelial cells via CXCR4, Am J Physiol Gastrointest Liver Physiol 286 (2004), G844–G850.

[13] H. Fischer, R. Stenling, C. Rubio and A. Lindblom, Differential expression of aquaporin 8 in human colonic epithelial cells and colorectal tumors, BMC Physiol 1 (2001), 1.

[14] C.M. Frederiksen, S. Knudsen, S. Laurberg and T.F. Orntoft, Classification of Dukes’ B and C colorectal cancers using expression arrays, J Cancer Res Clin Oncol 129 (2003), 263–271.

[15] O. Galamb, F. Sipos, E. Dinya, S. Spisak, Z. Tulassay and B. Molnar, mRNA expression, functional profiling and multivariate classification of colon biopsy specimen by cDNA overall glass microarray, World J Gastroenterol 12 (2006), 6998–7006.

[16] O. Galamb, F. Sipos, K. Fischer, Z. Tulassay and B. Molnar, The results of the expression array studies correlate and enhance the known genetic basis of gastric and colorectal cancer, Cytotherapy 8 Clin Cytom 68 (2005), 1–17.

[17] T. Hisamatsu, M. Watanabe, H. Ogata, T. Ezaki, S. Hozawa, H. Ishii, T. Kanai and T. Hibi, Interferon-inducible gene family I-8U expression in colitis-associated colon cancer and severely inflamed mucosa in ulcerative colitis, Cancer Res 59 (1999), 5927–5931.

[18] R.A. Irizarry, L. Gautier and L.M. Cope, The Analysis of Gene Expression Data: Methods and Software, Springer Verlag chapter 4, 2003.

[19] S. Kansra, D.Z. Ewton, J. Wang and E. Friedman, IGFBP-3 mediates TGF beta 1 proliferative response in colon cancer cells, Int J Cancer 87 (2000), 373–379.

[20] I. Kirman, N. Poltoratskaia, P. Sylla and R.L. Whelan, Insulin-like growth factor-binding protein 3 inhibits growth of experimentally colon carcinoma, Surgery 136 (2004), 205–209.

[21] O. Kitahara, Y. Furukawa, T. Tanaka, C. Kihara, K. Ono, R. Yanagawa, M.E. Nita, T. Takagi, Y. Nakamura and T. Tsunoda, Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia, Cancer Res 61 (2001), 3544–3549.

[22] H.C. Kwon, S.H. Kim, M.S. Roh, J.S. Kim, H.S. Lee, H.J. Choi, J.S. Jeong, H.J. Kim and T.H. Hwang, Gene expression profiling in lymph node-positive and lymph node-negative colorectal cancer, Dis Col Rect 47 (2004), 141–152.

[23] T. Langmann, C. Moehle, R. Maurer, M. Scharl, G. Liebsch, A. Zahn, W. Strehmel and G. Schmitz, Loss of detoxification receptor target genes, Gastroenterology 127 (2004), 26–40.

[24] I.C. Lawrance, C. Fiocchi and S. Chakravarti, Ulcerative colitis and Crohn's disease: distinctive gene expression profiles in inflamed mucosa in inflammatory bowel disease: dysregulation of pregnane X receptor target genes, Gastroenterology 127 (2004), 26–40.

[25] M. Li, Y.M. Lin, S. Hasegawa, T. Shimokawa, K. Murata, M. Kameyama, O. Ishikawa, T. Katagiri, T. Tsunoda, Y. Nakamura and Y. Furukawa, Genes associated with liver metastasis of colon cancer, identified by genome-wide cDNA microarray, Int J Oncol 24 (2004), 305–312.

[26] Y.M. Lin, Y. Furukawa, T. Tsunoda, C.T. Yue, K.C. Yang and Y. Nakamura, Molecular diagnosis of colorectal tumors
by expression profiles of 50 genes expressed differentially in adenomas and carcinomas, Oncogene 21 (2002), 4120–4128.

[27] E.E. Mannick, J.C. Bonomolo, R. Horswell, J.J. Lentz, M.S. Serrano, A. Zapata-Velardia, M. Gastanaduy, J.L. Himel, S.L. Rose, J.N. Udall Jr, C.A. Hornick and Z. Liu, Gene expression in mononuclear cells from patients with inflammatory bowel disease, Clin Immunol 112 (2004), 247–257.

[28] T. Niki, S. Izumi, Y. Saegusa, T. Takai, S.M. Iguchi-Ariga and H. Ariga, MSSP promotes ras/myc cooperative cell transforming activity by binding to c-Myc, Genes Cells 5 (2000), 127–141.

[29] D.A. Notterman, U. Alon, A.J. Sierk and A.J. Levine, Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays, Cancer Res 61 (2001), 3124–3130.

[30] S. Okahara, Y. Arimura, T. Yabana, K. Kobayashi, A. Gotoh, S. Motoya, A. Imamura, T. Endo and K. Imai, Inflammatory gene signature in ulcerative colitis with cDNA macroarray analysis, Aliment Pharmacol Ther 21 (2005), 109–120.

[31] K. Shibuta, N.A. Begum, M. Mori, K. Shimoda, T. Akiyoshi and M. Grosso, S. Alesci, F. Trimarchi and G. Barresi, Immunohistochemical expression of HGF, c-MET and transcription factor STAT3 in colorectal tumors, Eur J Histochem 48 (2004), 291–297.

[32] R. Yanagawa, Y. Furukawa, T. Tsunoda, O. Kitahara, M. Kameyama, K. Murata, O. Ishikawa and Y. Nakamura, Genomic screening of genes showing altered expression in liver metastases of human colorectal cancers by cDNA microarray, Neoplasia 3 (2001), 395–401.