Molecular Signature for Lymphatic Metastasis in Colorectal Carcinomas

Roland S. Croner, MD, Thomas Förtsch, MD,* Wolfgang M. Brückl, MD,† Franz Rödel, PhD,‡ Claus Rödel, MD,‡ Thomas Papadopoulos, MD,§ Thomas Brabletz, MD,§ Thomas Kirchner, MD,¶ Martin Sachs, PhD,‖ Jürgen Behrens, PhD,‖ Ludger Klein-Hitpass, PhD,** Michael Stürzl, PhD,†† Werner Hohenberger, MD,* and Berthold Lausen, PhD‡‡*

Purpose: TNM-staging of colorectal carcinomas (CRC) relies on the histopathologic workup of the surgically removed specimen. If valid preoperative staging methods existed, patients could be selected for adequate individual therapy before surgery. Microarray techniques provide a promising tool to identify stage-specific molecular signatures on primary tumor biopsies.

Material and Methods: Forty tumor samples of stage UICC I, II CRC, 40 samples of stage III CRC, and 25 biopsies of healthy mucosa (MC) were shock frozen in liquid nitrogen and underwent cryotomy after manual dissection for tumor tissue or MC enrichment. Isolated RNA was hybridized to GeneChips (HG-U133A, Affymetrix). Preprocessing of the microarray results was done by the robust multichip average method, and differentially expressed genes were selected by the maximum Wilcoxon statistic over 22,215 probe sets. The results were validated at an independent clinical study.

Results: Fifty differently expressed genes between stage UICC I, II versus III CRC were identified respecting the selection criteria by allowing for multiple testing. The data validation by the independent clinical study confirmed our results. In comparison to MC, the genes were over- or underexpressed. They belong to various functional groups such as cellular adhesion, transporters, signaling, metabolism, protein synthesis, gene control, and immune system.

Conclusion: Our large patient cohort and the data validation on an independent study identified 50 differentially expressed genes between CRC of different histopathologic stages. These findings indicate that molecular staging of CRC may be possible, which could help to guide individual CRC treatment before surgery.

(*Ann Surg* 2008;247: 803–810)

Lymphatic metastasis is an important predictor for tumor recurrence and survival in colorectal carcinomas (CRC).[1] Therefore, the aim of treatment regiments with curative intention is to resect the carcinoma, including all involved lymph nodes. During surgery the primary tumor is removed en bloc with all draining lymph nodes, which determine the resection area.[2] Adjuvant chemotherapy is recommended for stage UICC III colon cancer. It has been shown to reduce tumor recurrence and improve overall survival.[3] Likewise, patients with clinical stage T3 or T4 or node positive rectal cancer should receive preoperative chemoradiotherapy.[4] Primary tumor and lymph node staging currently requires the complete histopathologic workup of the surgically removed specimen. If reliable staging methods would exist before surgery, an individually tailored therapeutical approach could be designed for patients and surgery of the primary tumor could be performed depending on the tumor's biologic behavior.

Realistic estimations for imaging techniques for the detection of lymph node metastases are in the range of about 60% only.[3] The histopathologic tumor differentiation (G1–3) in preoperative endoscopically harvested snap biopsies needs further evaluation to assess its value in this regard.[3] Molecular markers would be perfectly suited for preoperative staging procedures because they can be investigated on biopsies from colonoscopy. Single molecular markers failed to achieve the level of clinical routine in this purpose because they are biased depending on the region of the snap biopsy.[5] Gene expression profiling by microarray technique allows the investigation of thousands of differentially expressed genes in parallel.

Several specific molecular signatures characterizing CRC have already been published.[6–8] The prognostic molec-

From the Departments of *Surgery, and †Internal Medicine I, University of Erlangen-Nuremberg, Erlangen; ‡Department of Radiotherapy, University of Frankfurt, Frankfurt; §Institute of Pathology, University of Erlangen-Nuremberg, Erlangen; ¶Department of Pathology, Ludwig-Maximilian-University, Munich; ‖Nikolaus-Fiebiger-Center for Molecular Medicine, University of Erlangen-Nuremberg, Erlangen; **Department of Cell Biology (Tumor Research), University of Duisburg-Essen, Medical School, Essen; and Departments of ††Molecular and Experimental Surgery, and ‡‡Medical Informatics, Biometry and Epidemiology, University of Erlangen-Nuremberg, Erlangen, Germany.

This study was sponsored by the German Federal Ministry for Education and Science (BMBF) within the National Genome Science Network (NGFN) and by the ELAN Fund, Medical Faculty of the Friedrich-Alexander-University, Erlangen-Nürnberg, Germany.

Supported by grants from the Interdisciplinary Center for Clinical Research (IZKF) of the University of Erlangen-Nuremberg.

Reprints: Roland S. Croner, MD, Department of Surgery, University of Erlangen, Krankenhausstrasse 12, D-91054 Erlangen, Germany. E-mail: Roland.Croner@uk-erlangen.de.

ISSN: 0003-4932/08/24705-0803
DOI: 10.1097/SLA.0b013e31816bcd49

ular classification of CRC regarding its metastatic behavior remains a challenge. Recently, we calculated the possibility of a molecular prediction for lymphatic metastasis based on gene expression profiling in CRC biopsies of 67%.[9] We focused on evaluating the gene expression classifier by assessing its predictive power relative to standard clinical parameters. The aim of the present study was to identify a practical handle able genomic classifier of exactly 50 genes for a possible clinical use to discriminate between lymph node negative (ie, stage UICC I/II) and lymph node positive (ie, stage UICC III) CRC.[10–12] The set of 50 genes established by our data was approved by the originally data of a similar study including less patients. This supported our findings.[13]

MATERIALS AND METHODS

Patients

All investigations were performed in accordance with the Declaration of Helsinki. After informed consent tumor biopsies from 80 patients with stage UICC I–III CRC, and biopsies of healthy mucosa (MC) from 25 independent patients that underwent surgery for CRC were harvested. Only patients who underwent surgery for the first manifestation of CRC were included in this study. No patient received preoperative radiotherapy or chemotherapy. Patients suffering from hereditary colorectal cancer or inflammatory bowel disease (Crohn's disease or ulcerative colitis) were excluded from this study. A detailed characterization of all patients included in this study is given in Table 1.

Histopathologic Quality Control

Local tumor invasion (T), lymphatic vessel (L), venous invasion (V), and tumor differentiation (G1–3) was investigated by experienced pathologists. As a routine histopathologic workup in our department, the whole mesenteric tissue of the colorectal specimens was investigated for lymph nodes. All lymph nodes were investigated by hematoxylin & eosin (HE) staining to detect lymphatic metastases. As a routine procedure in our department of pathology, they were divided and slices of both sides of the lymph node pieces were investigated by HE staining. Immunohistochemical staining was not performed. Samples of healthy MC underwent standard pathologic examination to confirm absence of disease. Detailed histologic findings are listed in Table 1.

TABLE 1. Histopathologic Carcinoma Characteristics and Tumor Localization of Patients

	UICC I	UICC II	UICC III
Number	18	22	40
Gender			
Male/female	9/9	14/8	27/13
Age*	64 ± 13	67 ± 11	67 ± 11
pT			
1	6	0	0
2	12	0	7
3a	0	6	9
3b	0	7	6
3c	0	3	5
3	0	2	1
4a	0	1	3
4b	0	2	6
4	0	1	3
pN			
0	18	22	0
1	0	0	29
2	0	0	11
pL			
0	18	19	16
1	0	3	24
pV			
0	18	22	40
1	0	0	0
M			
0	18	22	40
1	0	0	0
G			
1	0	0	0
2	18	18	29
3	0	4	10
Lymph nodes investigated†	26 (7–41)	29 (9–57)	30 (15–48)
Lymph node metastases†	0	0	3 (1–11)
Localization			
Cecum	2	3	3
Ascending cecum	2	2	4
Hepatic flexure	2	1	0
Transverse cecum	0	0	2
Splenic flexure	1	0	0
Descending cecum	0	0	4
Sigmoid colon	2	7	11
Rectum	9	9	16

*Mean and standard deviation.
†Mean and range.
pT indicates tumor invasion; pN, nodal status; pL, lymphatic vessel infiltration; pV, venous blood vessel invasion; M, distant metastasis; G, grading.

Tissue Preparation and Microarrays

Tissue samples were shock frozen in liquid nitrogen immediately after surgery to prevent bias from prolonged ischemia.[14] They were embedded in TissueTek (Zakura, Zoeterwoude, Netherlands) and stored at −80°C until further processing. The samples underwent cryotomy after manual dissection, a method of tumor tissue isolation which was described recently.[15] Total RNA was isolated using commercial kits (RNeasy-Kit, Qiagen, Hilden, Germany), following the manufacturer's protocol, including a DNAse (Qiagen) digestion. RNA quality and quantity were determined by the "Laboratory-on-a-Chip" method (Bioanalyzer 2100, Agilent Technologies, Palo Alto, CA).[16] The 3′/5′-ratios of the housekeeping genes glycerinaldehyde-3-phosphatase and β-actin supplied by the GeneChip were used as an other measures of RNA quality and to exclude partial degradation. A 3′/5′-ratio below 3 was regarded as an indicator of adequate RNA quality according to the manufacturer's protocol (Affymetrix, Santa Clara, CA).[17]

Gene expression was examined using the GeneChip technology (Affymetrix). Biotin-labeled cRNA was generated

by in vitro transcription as described previously and hybridized to the GeneChips (HG-U133A) following the manufacturer's instructions.[18]

Data Preprocessing

Gene expression measures were computed with the robust multichip average (RMA) method described in Irizarry et al[19,20] and implemented in the Bioconductor R package affy. This method includes the following successive steps: 1) background correction; 2) probe-level quantile normalization; 3) calculation of expression measures using median polish. A matrix of expression values (ME1) was computed for the 80 CEL files of tumor samples. The matrix ME1 had 22,215 rows (probe sets) and 80 columns. To allow comparison with the expression values in MC a matrix of expression values (ME2) was computed for the 80 CEL files of tumor and 25 CEL files of healthy MC samples. ME2 had 22,215 rows (probe sets) and 105 columns. ME2 was used for computing the RMA estimates of mean $\log_2$ expression values and for RMA estimates of the $\log_2$ fold change. The difference of the mean values was defined as estimates of the $\log_2$ fold change.

Selection of Differentially Expressed Genes

A set of 50 differentially expressed genes was identified by the comparison of 40 stage UICC I, II versus 40 stage III CRC samples. The R package permax was used for this purpose. The 2-sided P value using the distribution of the maximum over 22,215 single Wilcoxon statistics was computed as family wise error rate. Using ME2 the same procedure was applied for the comparison of the 25 MC samples with the 40 stage UICC I, II tumor samples, and for the comparison of the 25 MC samples with the 40 stage UICC III tumor samples. External validation of the identified set of 50 genes was done comparing the direction of the estimated fold change with the direction of the estimated fold change computed by an independent clinical study including 18 (stage UICC II) versus 18 (stage UICC III) patients.[13] The positive association of the directions was tested by Fisher exact test (1-sided). Moreover, validation was done for 2 subsets of the 50 genes: the set of the genes with 2-sided P values less than 0.01; and the set of the genes with 2-sided P values less than 0.01 and at least 1 of the 2 estimates of mean $\log_2$ expression greater than 6.

Software

The statistical analysis was performed with the open-source software R, Version 2.2.1 (http://cran.r-project.org), and Bioconductor packages (www.bioconductor.org). The following R packages were used: affy Version 1.8.1, annaffy Version 1.2.0, hgu133a Version 1.10.0, permax version 1.2.1.[21–23]

RESULTS

Patients

From 25 independent patients (male: 20, female: 5) who underwent surgery for CRC MC samples were analyzed. The median age of this patients was 62 ($\pm$10) years of age. Eighteen patients with stage UICC I, 22 patients with stage UICC II and 40 patients with stage UICC III CRC were included in the study. The median age of the patients with stage I CRC was 64 years. In stage UICC II and III the median age of patients was 67 years. No significant differences could be identified between the groups. Lymphatic vessel invasion was detected in no tumor sample of stage I but in 3 samples of stage II, and 24 samples of stage III. No venous vessel invasion was found in any carcinoma sample regardless to the UICC stages. Eighteen CRC samples of stage I and II and 29 samples of stage III were graded as G2. Four CRC samples of stage II and 10 samples of stage III were graded as G3. Nine patients of stage I, 9 patients of stage II, and 16 patients of stage III had rectal cancer. In 9 cases of stage I, 13 cases of stage II, and 24 cases of stage III the colon with various distribution from the cecum to the sigmoid colon was involved (Table 1).

Gene Expression

We identified 50 genes differentially expressed between CRC stage UICC I and II versus III with respect to our selection criteria. This set of genes was justified by multiple testing ($P < 0.015$). The expression matrices based on 80 (ME1) or 105 (ME2) biopsies resulted in slightly different fold change estimations. Forty of these genes were overexpressed, and 10 genes were underexpressed in stage UICC III versus stage UICC I and II CRC (Table 2). During the normalization of the CRC data against the MC expression levels the main part of the identified genes was detected as overexpressed in stage UICC III versus MC, and underexpressed in stage UICC I, II versus MC (Fig. 1). The identified genes could be assigned to 10 functional groups: Cellular signaling (NIBP, OR12D3, EVI1, WNT16, FSHR, MPP2, CRHR2, GH1, TBXA2R), channels, transporters and vesicle transport (SLC35D1, ATP6V0E, SLC12A4, TRPM3, COG2), cell adhesion and cytoskeleton (ADAM22, PRPH, CDH4, CLDN16), metabolism (NDUFA8, GLYAT, PSMD6, BCHE), protein synthesis (DHX15, LARS2), immune system (PGLYRP4), gene control (BAPX1, SIN3B), pseudogenes (CYP2B7P1, HBBP1, HCG4P6), miscellaneous (SPAG11, TRIM14, ZPBP), and unknown function (TNRC6B, LOC51236, RFPL3, FSD1, FLJ20259, FLJ11539, GLT25D2, UBOX5, BEAN, CFDP1, TM2D1, KIAA0427, PTHB1, YTHDC2, EFCAB1).

Subgroup Analysis

Analyzing the subgroups of rectal carcinoma patients and colon carcinoma patients separately, 6 additional genes could be identified separating between lymph node positive and negative rectal carcinomas. Four additional genes were overexpressed in rectal carcinomas stage UICC III versus stage UICC I, II within this subgroup: RIKEN cDNA 5730589L02 gene (gb: BC006309, Affymetrix ID: 211037_s_at, $P = 0.0004$), thrombospondin 3 (gb: L38969, Affymetrix ID: 209561_at, $P = 0.001$), ectropic retroviral transforming sequence b (gb: NM_004351, Affymetrix ID: 208348_s_at, $P = 0.006$) and lysophosphatidic acid acyltransferase-delta (gb: NM_020133, Affymetrix ID: 219693_at, $P = 0.006$). Two additional genes

TABLE 2. Predictive Gene Expression Profile for Colorectal Carcinomas, Which Represents Differentially Expressed Genes Between Colorectal Carcinomas Stage UICC I, II, Versus III

Affymetrix Probe Set	Gene Bank ID	Gene	Annotation	Mean I, II (log_2)	Mean III (log_2)	FC III vs. I, II (log_2)	Comparison FC Groene III vs. I, II (log_2)	Multiple *P* UICC III vs. I, II
UICC III vs. UICC I, II Up								
205433_at	NM_000055	BCHE	Butyrylcholinesterase	3.99	4.20	0.21	−0.00	0.001
211044_at	BC006333	TRIM14	Tripartite motif-containing 14	4.02	4.27	0.24	0.01	0.008
37547_at	U85995	PTHB1	Parathyroid hormone-responsive B1	4.14	4.30	0.16	0.16	0.009
215973_at	AF036973	HCG4P6	HLA complex group 4 pseudogene 6	4.36	4.60	0.25	0.06	0.013
214376_at	AI263044	EST	qz29e03.x1 NCI_CGAP_Kid11 cDNA clone	4.41	4.61	0.20	0.01	0.008
216489_at	AB046836	TRPM3	Transient receptor potential cation channel, subfamily M, member 3	4.43	4.70	0.28	0.02	0.001
211201_at	M95489	FSHR	Follicle-stimulating hormone receptor	4.48	4.70	0.22	−0.08	0.002
214068_at	AF070610	BEAN	Clone 24505	4.52	4.72	0.21	0.09	0.003
216063_at	N55205	HBBP1	Hemoglobin, beta pseudogene 1	4.48	4.77	0.29	−0.03	0.001
219791_s_at	NM_024748	FLJ11539	FLJ11539	4.55	4.81	0.26	−0.02	0.001
>209353_s_at	BC001205	SIN3B	SIN3 homolog B, transcription regulator (yeast)	4.53	4.85	0.31	−0.03	0.010
211381_x_at	AF168617	SPAG11	Sperm associated antigen 11	4.54	4.86	0.32	0.08	0.010
207021_at	NM_007009	ZPBP	Zona pellucida binding protein	4.70	4.91	0.20	−0.03	0.013
220227_at	NM_024883	CDH4	R-cadherin (retinal)	4.81	5.24	0.43	0.04	0.014
210701_at	D85939	CFDP1	Craniofacial development protein 1	4.93	5.19	0.26	−0.03	0.005
220156_at	NM_024593	EFCAB1	EF-hand calcium binding domain 1	5.03	5.33	0.30	−0.03	0.013
209883_at	AF288389	GLT25D2	Glycosyltransferase 25 domain containing 2	5.08	5.34	0.26	0.06	0.001
207031_at	NM_001189	BAPX1	Bagpipe homeobox homolog 1 (Drosophila)	5.07	5.43	0.37	0.04	0.002
206885_x_at	NM_022559	GH1	Growth hormone 1	5.24	5.54	0.29	0.00	0.005
212963_at	BF968960	TM2D1	TM2 domain containing 1	5.29	5.55	0.27	−0.01	0.008
207897_at	NM_001883	CRHR2	Corticotrophin releasing hormone receptor 2	5.39	5.75	0.35	0.02	0.004
222083_at	AW024233	GLYAT	Glycine-N-acyltransferase	5.41	5.85	0.43	0.05	0.007
214149_s_at	AI252582	ATP6V0E	ATPase, H+ transporting, lysosomal 9kDa, V0 subunit	5.41	5.91	0.50	0.02	0.005
220332_at	NM_006580	CLDN16	Claudin 16	5.55	5.80	0.26	−0.09	0.003
220944_at	NM_020393	PGLYRP4	Peptidoglycan recognition protein 4	5.43	5.97	0.54	−0.08	0.015
219170_at	NM_024333	FSD1	Fibronectin type III and SPRY domain containing 1	5.63	6.02	0.38	−0.09	0.013
221113_s_at	NM_016087	WNT16	Wingless-type MMTV integration site family, member 16	5.73	5.99	0.26	−0.13	0.014
221431_s_at	NM_030959	OR12D3	Olfactory receptor, family 12, subfamily D, member 3	5.73	6.11	0.39	−0.094	0.003
207936_x_at	NM_006604	RFPL3	Ret finger protein-like 3	5.82	6.20	0.38	−0.05	0.012
204303_s_at	NM_014772	KIAA0427	KIAA0427	5.98	6.17	0.19	0.11	0.010

(*Continued*)

TABLE 2. (*Continued*)

Affymetrix Probe Set	Gene Bank ID	Gene	Annotation	Mean I, II (log_2)	Mean III (log_2)	FC III vs. I, II (log_2)	Comparison FC Groene III vs. I, II (log_2)	Multiple *P* UICC III vs. I, II
210272_at	M29873	CYP2B7P1	Cytochrome P450, family 2, subfamily B, polypeptide 7 pseudogene 1	5.99	6.45	0.46	0.02	0.005
207984_s_at	NM_005374	MPP2	Membrane protein, palmitoylated 2 (MAGUK p55 subfamily member 2)	6.10	6.42	0.32	−0.04	0.002
208227_x_at	NM_021721	ADAM22	ADAM metallopeptidase domain 22	6.20	6.61	0.42	0.03	0.003
213847_at	NM_006262	PRPH	Peripherin	6.36	6.84	0.48	0.01	0.005
215544_s_at	AL121891	UBOX5	U-box domain containing 5	6.55	6.89	0.34	0.06	0.002
336_at	D38081	TBXA2R	Thromboxane A2 receptor	6.59	6.92	0.34	0.05	0.007
209402_s_at	AF047338	SLC12A4	Solute carrier family 12, member 4	6.79	7.10	0.31	0.05	0.001
221629_x_at	AF151022	LOC51236	Similar to brain protein 16	6.90	7.44	0.54	0.24	0.001
219071_x_at	NM_016458	C8orf30A	Chromosome 8 open reading frame 30A	7.52	8.05	0.53	0.30	0.009
56829_at	H61826	NIBP	NIK and IKK{beta} binding protein	8.04	8.38	0.34	0.13	0.002
UICC III vs. UICC I, II Down								
205835_s_at	AW975818	YTHDC2	YTH domain containing 2	4.77	4.56	−0.21	0.04	0.010
213254_at	N64803	TNRC6B	Trinucleotide repeat containing 6B	7.00	6.57	−0.43	0.00	0.008
34764_at	D21851	LARS2	Leucyl-tRNA synthetase 2, mitochondrial	7.54	6.92	−0.62	−0.10	0.014
209711_at	N80922	SLC35D1	Solute carrier family 35, member D1	8.12	7.54	−0.58	−0.33	0.004
203073_at	NM_007357	COG2	Component of oligomeric golgi complex 2	8.36	7.99	−0.36	−0.04	0.002
209174_s_at	BC000978	FLJ20259	FLJ20259	8.45	8.12	−0.33	−0.03	<0.001
221884_at	BE466525	EVI1	Ecotropic viral integration site 1	8.93	8.36	−0.56	−0.20	0.004
218160_at	NM_014222	NDUFA8	NADH dehydrogenase (ubiquinone) 1 alpha subcomplex, 8, 19kDa	9.67	9.11	−0.55	−0.20	0.002
201386_s_at	AF279891	DHX15	DEAH (Asp-Glu-Ala-His) box polypeptide 15	10.09	9.60	−0.49	−0.03	0.011
202753_at	NM_014814	PSMD6	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 6	10.38	9.89	−0.49	−0.01	0.012

Probe set, affymetrix probe ID; Accession number, gene bank ID; Annotation, gene bank description; mean I, II, mean of RMA estimate of log_2 expression value for 40 stage UICC I, II; mean III, mean of RMA estimate of log_2 expression value for 40 stage UICC III; FC III vs. I, II (log_2), RMA estimate of log_2 fold change (FC) UICC III vs. I, II; comparison FC Groene III vs. I, II (log_2), RMA estimate of log_2 fold change (FC) UICC III vs. I, II by data of the validation study Groene et al[13]; *P*, 2-sided *P*-value using the distribution of the maximum Wilcoxon statistic over 22,215 single Wilcoxon statistic. The probe IDs are arranged in 2 groups, 40 up-regulated probe IDs (left side of the plot in Fig. 1) and 10 down-regulated probe IDs (right side of the plot in Fig. 1). The probe IDs inside the 2 groups are ordered by the mean log_2 expression over the 80 cancer biopsies.

were underexpressed in stage UICC III versus, I, II rectal carcinomas: fumarate hydratase (gb: AA669797, Affymetrix ID: 214170_x_at, $P = 0.001$) and annexin A8 (gb: NM_001630, Affymetrix ID: 203074_at, $P = 0.008$). No additional genes were identified in the subgroup analysis of colon carcinomas.

External Data Validation

We validated our results on an independent clinical study that included 18 stage UICC II versus 18 stage UICC III CRC patients.[24,12] Groene et al[13] kindly computed estimates of the gene expression fold changes for our set of 50 genes on their microarray data. We obtained significant associations for the 50 genes by 32 directions in agreement ($P = 0.027$), for the 35 genes with 2-sided *P* values less than 0.01 by 25 directions in agreement ($P = 0.028$); and for the 17 genes with 2-sided *P* values less than 0.01 and at least 1 of the 2 estimates of mean log_2 expression greater than 6 by 14 directions in agreement ($P = 0.018$; cf. Table 3). Moreover, the direction of the 8 genes with the highest mean log_2 expression values are in agreement for the overexpressed group and for the underexpressed group (cf. Table 2).

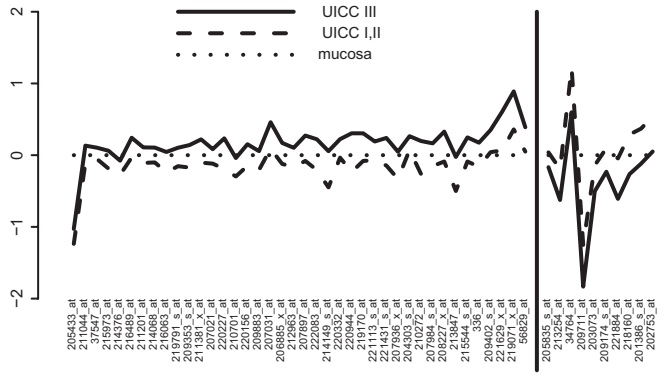

FIGURE 1. The difference of the mean $\log_2$ expression value for the 2 cancer groups relative to the mean $\log_2$ expression value for the mucosa group. The probe IDs are arranged in 2 groups, 40 up-regulated probe IDs (left side of the plot) and 10 down-regulated probe IDs (right side of the plot). The probe IDs inside the 2 groups are ordered by the mean $\log_2$ expression over the 80 cancer biopsies.

TABLE 3. Comparison of the Direction of the Estimated Fold Change With the Direction of the Estimated Fold Change Computed by the Independent Clinical Study of Groene et al[13]

	Croner\Groene	+	−	Total
A: 50 gene set ($P = 0.027$)	+	24	16	40
	−	2	8	10
		26	24	50
B: 35 gene set ($P = 0.028$)	+	20	9	29
	−	1	5	6
		21	14	35
C: 17 gene set ($P = 0.018$)	+	9	2	11
	−	1	5	6
		10	7	17

A, all 50 genes; B, 35 genes with 2-sided *P*-values less than 0.01; C, 17 genes with 2-sided *P*-values less than 0.01 and at least 1 of the 2 estimates of mean $\log_2$ expression greater than 6.

DISCUSSION

As we calculated a molecular prediction rate for lymphatic metastasis of 67% on primary tumor biopsies in CRC recently the study was expanded.[9] Using a conservative calculation based on Bonferroni's adjustment of multiple testing, a sample size of about 40 cases per group was estimated (assuming fold change 0.33, standard deviation 0.25, $\alpha = 0.05$, power 0.8). The strict control of patient selection, histopathologic CRC characterization, tissue preparation, and RNA quality guaranteed homogeneous groups as was recently suggested for microarray studies.[12] To transform classifiers to platforms that could be of broad clinical value, we calculated a practical gene set of 50 markers for lymphatic metastasis in CRC biopsies from the complete data of 22,215 probe sets.[12]

Several of our candidate genes have already been described being involved in malignant tissue transformation or tumor progression. Butyrylcholinesterase was overexpressed in healthy tissue versus carcinoma in our data. These findings correlate with descriptions in lung tumors.[25] An involvement of follicle stimulating hormone receptor and claudin 16 in the occurrence and progression of ovarian cancer was reported recently.[26,27] We evaluated both follicle stimulating hormone receptor and claudin 16 as up-regulated genes in stage UICC III CRC versus MC, and differentially expressed between stage UICC I, II versus III CRC. Another candidate of our gene list, the parathyroid hormone-responsive B1 (PTHB1) was already described in Wilms' tumors.[28] In breast cancer the thromboxane A2 receptor expression, which was up-regulated in lymph node positive CRC in our study, correlates with more aggressive tumor behavior.[29] A highly expressed member of the wnt gene family in CRC stage UICC III within our data, the WNT16, is connected to human lymphoblastoid leukemia cells.[30] All these descriptions of markers from our gene set indicate that our presented list may represent valuable markers for cancer progression in CRC. During the subgroup analysis 6 further differentially expressed genes were detected separating lymph node positive from lymph node negative rectal carcinomas. Several of these genes, for example, thrombosopondin 3 and annexin A8 have recently already been described as prognostic indicators in osteosarcoma and breast cancer.[31,32] The function during metastasis and prognostic value of these genes needs further investigation. The identification of 6 further genes associated to lymphatic metastasis in rectal carcinomas may indicate additional molecular mechanisms in rectal carcinomas during this process.

Nevertheless the differences in the analyzed gene expression values between CRC stage UICC I, II versus III are not as high as those usually found in comparisons between MC and CRC.[7] The molecular signatures possibly reflect the fact that healthy MC and tumor are histopathologically grossly different tissue types whereas the carcinomas do not vary tremendously in morphologic examinations. But gene expression profiling by microarray technology is an established method that was already used to identify the gene expression profiles in various tissue types. In our study the data were generated on a reliable commercial platform and may therefore represent a valid set of genes. The results have not been controlled by PCR or similar methods because the reproduction of microarray results by basic molecular techniques was already demonstrated successfully in many studies previously.[6,33–36] Instead we evaluated our results on an similar independent study which resulted in good concordance.[13] Groene et al[13] suggested a list of 45 genes as differentially expressed between stage UICC II versus III CRC in a smaller study including 36 patients. Groene et al published 21 down-regulated and 24 up-regulated. In our data (matrix ME1) 16 of the 21 down-regulated candidate genes of Groene's et al articles are down-regulated, and 17 of the 24 up-regulated genes are up-regulated. This comparison with an independent study supports our findings. Unfortunately, there was only a limited correlation of our gene list with other published data, as it was between these studies.[13,36–39] Reasons for this discrepancy might be the various techniques for tissue isolation, RNA preparation, kind of microarray (oligo-,

cDNA array) and hybridization. A standardization of microarray procedures for an overall comparison seems indispensable.

Many single molecular marker (eg, MMP2, MMP9, p27, E-cadherin, ICAM-1, TGF-α, survivin, p53, thymidylate synthase) have been investigated for their clinical use in CRC prognosis or response to chemotherapy and radiation.[5] Except of mismatch repair genes (eg, MLH1, MSH2) or microsatellite instability for the diagnosis of HNPCC and somatic mutations in the APC gene to identify patients with FAP, they did not reach the level of clinical routine until now.[40,41] One reason for this might be the heterogeneous gene expression pattern in CRC tissue.[42,43] A biased marker detection depending on the area of the tumor biopsy is evident. To escape this phenomenon a combination of markers which could be detected by screening for the differential expression of multiple genes seems to be more appropriate. A set of various genes might be much more resistant against this possible error. Thus, searching for disease-stage–dependent gene expression profiles could be an appropriate tool for molecular cancer staging in the future.

Evaluated findings of microarray analysis could define gene sets that might be of value for clinical diagnostics in the future. In selected cases surgical procedures and multimodal therapy regiments could be adapted to the individual tumor's aggressiveness. It could assist in decisions making for transanal resection in rectal carcinomas, especially in cases of complete tumor response after chemoradiation. In colon carcinomas stage UICC II where adjuvant chemotherapy is not recommended routinely, particularly aggressive tumors with a molecular potency for lymphatic metastasis could be selected for adjuvant therapy. The significance for clinical use of our identified gene list has to be evaluated by future investigations.

ACKNOWLEDGMENTS

The authors thank Ulrich Mansmann of the Luwdwig-Maximillians-University München, Germany, for providing us with the gene expression fold changes from his study. The authors also thank Andreas Heilmann, Medical Faculty of the Friedrich-Alexander-University, Erlangen-Nürnberg, Germany, for grouping the genes concerning their function.

REFERENCES

1. Hermanek P. Prognostic factor. Research in oncology. *J Clin Epidemiol*. 1999;52:371–374.
2. Hohenberger W, Bittorf1 B, Papadopoulos T, et al. Survival after surgical treatment of cancer of the rectum. *Langenbecks Arch Surg*. 2005;390:363–372.
3. Schmiegel W, Pox C, Adler G, et al. [S3-guideline conference "Colorectal Cancer" 2004.] *Dtsch Med Wochenschr*. 2005;130(suppl 1):S5–S53.
4. Sauer R, Becker H, Hohenberger W, et al. Preoperative versus postoperative chemoradiotherapy for rectal cancer. *N Engl J Med*. 2004;351: 1731–1740.
5. Becker HD, Hohenberger W, Jungiger T, et al. *Surgical Oncology (Chirurgische Onkologie)*. New York: Thieme-Verlag, Stuttgart; 2002.
6. Birkenkamp-Demtroder K, Christensen LL, Olesen SH, et al. Gene expression in colorectal cancer. *Cancer Res*. 2002;62:4352–4363.
7. Croner RS, Foertsch T, Brueckl WM, et al. Common denominator genes that distinguish colorectal carcinoma from normal mucosa. *Int J Colorectal Dis*. 2005;20:353–362.
8. Ghadimi BM, Grade M, Difilippantonio MJ, et al. Effectiveness of gene expression profiling for response prediction of rectal adenocarcinomas to preoperative chemoradiotherapy. *J Clin Oncol*. 2005;23:1826–1838.
9. Croner RS, Peters A, Brueckl WM, et al. Microarray versus conventional prediction of lymph node metastasis in colorectal carcinoma. *Cancer*. 2005;104:395–404.
10. Simon R. Development and evaluation of therapeutically relevant predictive classifiers using gene expression profiling. *J Natl Cancer Inst*. 2006;98:1169–1171.
11. Simon R, Radmacher MD, Dobbin K, et al. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. *J Natl Cancer Inst.* 2003;95:14–18.
12. Simon R. Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol*. 2005;23:7332–7341.
13. Groene J, Mansmann U, Meister R, et al. Transcriptional census of 36 microdissected colorectal cancers yields a gene signature to distinguish UICC II and III. *Int J Cancer*. 2006;119:1829–1836.
14. Huang J, Qi R, Quackenbush J, et al. Effects of ischemia on gene expression profiling and identification of novel genes in hepatocellular carcinomas tissue microarrays for gene amplification surveys in many different tumor types. *J Surg Res*. 2001;99:222–227.
15. Croner RS, Guenther K, Foertsch T, et al. Tissue preparation for gene expression profiling of colorectal carcinoma: three alternatives to laser microdissection with preamplification. *J Lab Clin Med*. 2004;143:344–351.
16. Nachamkin I, Panaro NJ, Li M, et al. Agilent 2100 bioanalyzer for restriction fragment length polymorphism analysis of the *Campylobacter jejuni* flagellin gene. *J Clin Microbiol*. 2001;39:754–757.
17. Hsiao LL, Dangond F, Yoshida T, et al. A compendium of gene expression in normal human tissues. *Physiol Genomics*. 2001;7:97–104.
18. Durig J, Nückel H, Hüttmann A, et al. Expression of ribosomal and translation-associated genes is correlated with a favorable clinical course in chronic lymphocytic leukemia. *Blood*. 2003;101:2748–2755.
19. Irizarry RA, Gautier L, Cope LM. In: Parmigiani G, Garrett ES, Irizarry RA, et al, eds. *The Analysis of Gene Expression Data: Methods and Software.* New York: Springer; 2003:102–119.
20. Irizarry RA, Bolstad BM, Collin F, et al. Summaries of affymetrix genechip probe level data. *Nucleic Acids Res.* 2003;31:e15.
21. Irizarry RA, Gautier L, Bolstad BM, et al. affy: *Methods for Affymetrix Oligonucleotide Arrays.* R package version 1.8.1; 2005.
22. Gray R. *Permax.* R package version 1.2.1; 2005.
23. Smith C. annaffy: *Annotation Tools for Affymetrix Biological Metadata.* R package version 1.2.0; 2005.
24. Simon RM, Dobbin K. Experimental design of DNA microarray experiments. *Biotechniques.* 2003(suppl):16–21.
25. Martinez-Moreno P, Nieto-Cerón S, Torres-Lanzas J, et al. Cholinesterase activity of human lung tumours varies according to their histological classification. *Carcinogenesis*. 2006;27:429–436.
26. Rangel LB, Sherman-Baust CA, Wernyj RP, et al. Characterization of novel human ovarian cancer-specific transcripts (HOSTs) identified by serial analysis of gene expression. *Oncogene*. 2003;22:7225–7232.
27. Yang CQ, Chan KYK, Ngan HYS, et al. Single nucleotide polymorphisms of follicle-stimulating hormone receptor are associated with ovarian cancer susceptibility. *Carcinogenesis*. 2006;27:1502–1506.
28. Vernon EG, Malik K, Reynolds P, et al. The parathyroid hormone-responsive B1 gene is interrupted by a t(1;7)(q42;p15) breakpoint associated with Wilms' tumour. *Oncogene*. 2003;22:1371–1380.
29. Watkins G, Douglas-Jones A, Mansel RE, et al. Expression of thromboxane synthase, TBXAS1 and the thromboxane A2 receptor, TBXA2R, in human breast cancer. *Int Semin Surg Oncol*. 2005;2:23.
30. Mazieres J, You L, He B, et al. Inhibition of Wnt16 in human acute lymphoblastoid leukemia cells containing the t(1;19) translocation induces apoptosis. *Oncogene*. 2005;24:5396–5400.
31. Stein T, Price KN, Morris JS. Annexin A8 is up-regulated during mouse mammary gland involution and predicts poor survival in breast cancer. *Clin Cancer Res*. 2005;11:6872–6879.
32. Dalla-Torre CA, Yoshimoto M, Lee C-H, et al. Effects of THBS3, SPARC and SPP1 expression on biological behavior and survival in patients with osteosarcoma. *BMC Cancer*. 2006;6:237.
33. Birkenkamp-Demtroder K, Olesen SH, Sørensen FB, et al. Differential gene expression in colon cancer of the caecum versus the sigmoid and rectosigmoid. *Gut*. 2005;54:374–384.

34. Bhatavdekar JM, Patel DD, Chikhlikar PR, et al. Molecular markers are predictors of recurrence and survival in patients with Dukes B and Dukes C colorectal adenocarcinoma. *Dis Colon Rectum*. 2001;44: 523–533.
35. Notterman DA, Alon U, Sierk AJ, et al. Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Res*. 2001;61:3124–3130.
36. Frederiksen CM, Knudsen S, Laurberg S, et al. Classification of Dukes' B and C colorectal cancers using expression arrays. *J Cancer Res Clin Oncol*. 2003;129:263–271.
37. Barrier A, Lemoine A, Boelle PY, et al. Colon cancer prognosis prediction by gene expression profiling. *Oncogene*. 2005;24:6155–6164.
38. Kwon HC, Kim SH, Roh MS, et al. Gene expression profiling in lymph node-positive and lymph node-negative colorectal cancer. *Dis Colon Rectum*. 2004;47:141–152.
39. Grade M, Hörmann P, Becker S, et al. Gene expression profiling reveals a massive, aneuploidy-dependent transcriptional deregulation and distinct differences between lymph node-negative and lymph node-positive colon carcinomas. *Cancer Res*. 2007;67:41–56.
40. Compton CC, et al. Prognostic factors in colorectal cancer. College of American Pathologists Consensus Statement 1999. *Arch Pathol Lab Med*. 2000;124:979–994.
41. Merg A, Lynch HT, Lynch JF, et al. Hereditary colon cancer—part I. *Curr Probl Surg*. 2005;42:195–256.
42. Baisse B, Bouzourene H, Saraga EP, et al. Intratumor genetic heterogeneity in advanced human colorectal adenocarcinoma. *Int J Cancer*. 2001;93:346–352.
43. Losi L, Baisse B, Bouzourene H, et al. Evolution of intratumoral genetic heterogeneity during colorectal cancer progression. *Carcinogenesis*. 2005;26:916–922.