Gene Expression Profiling Reveals a Massive, Aneuploidy-Dependent Transcriptional Deregulation and Distinct Differences between Lymph Node–Negative and Lymph Node–Positive Colon Carcinomas

Marian Grade,[1] Patrick Hörmann,[2] Sandra Becker,[2] Amanda B. Hummon,[2] Danny Wangsa,[2] Sudhir Varma,[3] Richard Simon,[3] Torsten Liersch,[1] Heinz Becker,[1] Michael J. Difilippantonio,[2] B. Michael Ghadimi,[1] and Thomas Ried[2]

[1]Department of General Surgery, University Medical Center, Göttingen, Germany and [2]Genetics Branch and [3]Biometrics Research Branch, National Cancer Institute, NIH, Bethesda, Maryland

Abstract

To characterize patterns of global transcriptional deregulation in primary colon carcinomas, we did gene expression profiling of 73 tumors [Unio Internationale Contra Cancrum stage II (n = 33) and stage III (n = 40)] using oligonucleotide microarrays. For 30 of the tumors, expression profiles were compared with those from matched normal mucosa samples. We identified a set of 1,950 genes with highly significant deregulation between tumors and mucosa samples (P < 1e−7). A significant proportion of these genes mapped to chromosome 20 (P = 0.01). Seventeen genes had a >5-fold average expression difference between normal colon mucosa and carcinomas, including up-regulation of *MYC* and of *HMGA1*, a putative oncogene. Furthermore, we identified 68 genes that were significantly differentially expressed between lymph node–negative and lymph node–positive tumors (P < 0.001), the functional annotation of which revealed a preponderance of genes that play a role in cellular immune response and surveillance. The microarray-derived gene expression levels of 20 deregulated genes were validated using quantitative real-time reverse transcription-PCR in >40 tumor and normal mucosa samples with good concordance between the techniques. Finally, we established a relationship between specific genomic imbalances, which were mapped for 32 of the analyzed colon tumors by comparative genomic hybridization, and alterations of global transcriptional activity. Previously, we had conducted a similar analysis of primary rectal carcinomas. The systematic comparison of colon and rectal carcinomas revealed a significant overlap of genomic imbalances and transcriptional deregulation, including activation of the Wnt/β-catenin signaling cascade, suggesting similar pathogenic pathways. [Cancer Res 2007;67(1):41–56]

Note: Supplementary data for this article are available at Cancer Research Online (http://cancerres.aacrjournals.org/).

Requests for reprints: Thomas Ried, Genetics Branch, Center for Cancer Research, National Cancer Institute, NIH, Room 1408, Building 50, 50 South Drive, Bethesda, MD 20892-8010. Phone: 301-594-3118; Fax: 301-435-4428; E-mail: riedt@mail.nih.gov or B. Michael Ghadimi, Department of General Surgery, University Medical Center, Robert-Koch-Str. 40, 37075 Göttingen, Germany. Phone: 49-551-39-6162; Fax: 49-551-39-6106; E-mail: mghadim@uni-goettingen.de.

doi:10.1158/0008-5472.CAN-06-1514

Introduction

The advent and maturation of methodologies for parallel gene expression profiling allows the systematic interrogation of modifications of cellular transcriptomes, and global gene expression profiles have been described for a plethora of human diseases, including cancer (1). These studies now go beyond the mere description of discerning genetic changes in cancer samples and noncancerous tissue but extend to specific questions that are relevant for the clinical management of this disease. Examples include improved cancer classification, the development of prognostic profiles, and the prediction of individual responses to therapeutic interventions (2, 3).

Colorectal carcinomas, with an incidence of some 150,000 cases in the United States alone, were among the first cancers systematically analyzed by global gene expression profiling (4). The well-established linear progression from normal epithelium to dysplastic lesions of increasing morphologic abnormality and finally to locally invasive and metastatic disease also allowed the exploration of sequential transcriptional changes that occur during tumorigenesis (5–8). In addition, specific signatures associated with tumor stage and lymph node and liver metastases were described (9–15), and aneuploidy-dependent transcriptional deregulation was the focus of more recent reports (16, 17). Primary tumors and derived cell lines were used to establish profiles of response to chemotherapy and combined modality therapy (18, 19) and to analyze drug resistance (20, 21) and clinical recurrence (22, 23). The literature has been recently reviewed (24–26).

We have now focused our analysis on four specific aspects of colon tumorigenesis: (*a*) delineation of gene expression differences of primary colon cancers and adjacent normal mucosa, (*b*) identification of gene expression changes that distinguish colon tumors with and without lymph node metastases, (*c*) deciphering the consequences of chromosomal aneuploidies on resident gene expression levels, and (*d*) a systematic comparison of colon and rectal carcinomas, tumors that emerge in an anatomically and physiologically closely related environment. This comparison has become possible because we have previously applied analogous techniques to the analysis of primary rectal carcinomas (16).

Materials and Methods

Patients and sample collection. For this study, we collected tumor specimens from 73 patients with primary adenocarcinomas of the colon who were treated at the Department of General Surgery, University Medical Center, Göttingen, Germany. All tumors were located at least 16 cm above

Table 1. Clinical data of 73 patients and experimental setup

Colon cancer patient	Age (y)	Sex	Histopathology	Tumor expression	Mucosa expression	CGH
UICC stage II						
1	58	M	pT_3 pN_0 (0/17) M_0 R_0 G_2	X	X	
2	74	M	pT_3 pN_0 (0/19) M_0 R_0 G_2	X		
3	72	M	pT_3 pN_0 (0/29) M_0 R_0 G_2	X		
4	25	M	pT_3 pN_0 (0/31) M_0 R_0 G_3	X		X
5	68	F	pT_3 pN_0 (0/16) M_0 R_0 G_2	X		
6	73	F	pT_4 pN_0 (0/17) M_0 R_0 G_3	X	X	X
7	50	M	pT_3 pN_0 (0/25) M_0 R_0 G_2	X		
8	74	M	pT_3 pN_0 (0/44) M_0 R_0 G_2	X		X
9	34	F	pT_3 pN_0 (0/31) M_0 R_0 G_1-G_2	X	X	
10	77	M	pT_3 pN_0 (0/20) M_0 R_0 G_2	X	X	
11	85	M	pT_3 pN_0 (0/21) M_0 R_0 G_2	X	X	
12	39	F	pT_3 pN_0 (0/27) M_0 R_0 G_2	X		X
13	78	M	pT_3 pN_0 (0/39) M_0 R_0 G_2	X		
14	70	M	pT_3 pN_0 (0/23) M_0 R_0 G_2	X	X	
15	71	F	pT_3 pN_0 (0/31) M_0 R_0 G_3	X		X
16	60	M	pT_3 pN_0 (0/15) M_0 R_0 G_2	X		X
17	68	M	pT_3 pN_0 (0/18) M_0 R_0 G_2	X		
18	70	M	pT_3 pN_0 (0/27) M_0 R_0 G_2	X	X	
19	74	F	pT_4 pN_0 (0/57) M_0 R_0 G_2	X		X
20	74	M	pT_3 pN_0 (0/28) M_0 R_0 G_2	X	X	
21	65	M	pT_3 pN_0 (0/24) M_0 R_0 G_2	X		X
22	81	F	pT_3 pN_0 (0/15) M_0 R_0 G_2	X	X	
23	84	M	pT_3 pN_0 (0/21) M_0 R_0 G_3	X	X	X
24	75	M	pT_3 pN_0 (0/17) M_0 R_0 G_2	X	X	X
25	63	M	pT_3 pN_0 (0/29) M_0 R_0 G_2	X	X	
26	72	F	pT_3 pN_0 (0/20) M_0 R_0 G_2	X		X
27	66	M	pT_3 pN_0 (0/26) M_0 R_0 G_2	X		X
28	63	M	pT_3 pN_0 (0/20) M_0 R_0 G_2	X	X	
29	85	M	pT_3 pN_0 (0/12) M_0 R_0 G_1-G_2	X	X	
30	67	F	pT_3 pN_0 (0/35) M_0 R_0 G_2	X		X
31	54	F	pT_3 pN_0 (0/28) M_0 R_0 G_3	X	X	
32	65	M	pT_3 pN_0 (0/23) M_0 R_0 G_2	X		X
33	72	M	pT_3 pN_0 (0/17) M_0 R_0 G_2	X	X	
UICC stage III						
34	93	F	pT_3 pN_1 (2/17) M_0 R_0 G_2	X		X
35	42	M	pT_4 pN_1 (2/51) M_0 R_0 G_2	X		
36	41	M	pT_3 pN_2 (15/42) M_0 R_0 G_2	X		
37	79	M	pT_3 pN_1 (1/25) M_0 R_0 G_2	X		X
38	74	M	pT_2 pN_1 (1/23) M_0 R_0 G_2-G_3	X	X	
39	52	M	pT_3 pN_1 (1/28) M_0 R_0 G_2	X	X	X
40	68	M	pT_3 pN_1 (1/21) M_0 R_0 G_2	X	X	X
41	50	M	pT_3 pN_1 (1/33) M_0 R_0 G_2	X		
42	79	M	pT_3 pN_1 (1/2) M_0 R_0 G_2	X	X	X
43	43	M	pT_3 pN_2 (10/28) M_0 R_0 G_2	X		
44	79	F	pT_3 pN_1 (2/26) M_0 R_0 G_2	X	X	
45	66	F	pT_4 pN_2 (4/36) M_0 R_0 G_2	X		X
46	77	M	pT_3 pN_2 (8/16) M_0 R_0 G_3	X	X	
47	62	M	pT_3 pN_2 (12/13) M_0 R_0 G_2	X		X
48	56	F	pT_3 pN_2 (5/23) M_0 R_0 G_2	X		X
49	68	F	pT_4 pN_2 (9/21) M_0 R_0 G_2	X		X
50	36	M	pT_3 pN_1 (2/39) M_0 R_0 G_1	X		
51	78	F	pT_3 pN_2 (4/23) M_0 R_0 G_2	X	X	
52	65	F	pT_3 pN_2 (5/21) M_0 R_0 G_3	X		X
53	68	M	pT_4 pN_2 (11/26) M_0 R_0 G_2	X	X	
54	73	F	pT_3 pN_1 (3/22) M_0 R_0 G_2	X		X
55	66	M	pT_3 pN_1 (1/20) M_0 R_0 G_2	X		
56	76	M	pT_3 pN_1 (2/20) M_0 R_0 G_2	X	X	

(*Continued on the following page*)

Table 1. Clinical data of 73 patients and experimental setup (Con'td)

Colon cancer patient	Age (y)	Sex	Histopathology	Tumor expression	Mucosa expression	CGH
57	70	M	pT_3 pN_2 (21/55) M_0 R_0 G_3	X		
58	66	M	pT_3 pN_2 (1/32) M_0 R_0 G_2	X		X
59	75	F	pT_2 pN_1 (1/16) M_0 R_0 G_2	X		X
60	50	M	pT_3 pN_1 (2/24) M_0 R_0 G_2	X		X
61	57	M	pT_3 pN_2 (12/25) M_0 R_0 G_2	X		
62	72	M	pT_3 pN_1 (2/22) M_0 R_0 G_3	X	X	
63	72	M	pT_3 pN_1 (1/32) M_0 R_0 G_2	X		X
64	76	M	pT_4 pN_2 (4/39) M_0 R_0 G_2	X		X
65	56	M	pT_3 pN_1 (2/22) M_0 R_0 G_2-G_3	X	X	
66	70	M	pT_2 pN_2 (4/20) M_0 R_0 G_2	X		X
67	61	M	pT_4 pN_1 (3/24) M_0 R_0 G_2	X	X	
68	76	M	pT_3 pN_2 (12/22) M_0 R_0 G_3	X	X	
69	69	M	pT_2 pN_1 (2/16) M_0 R_0 G_2	X		
70	81	M	pT_3 pN_2 (12/21) M_0 R_0 G_2	X		X
71	63	M	pT_3 pN_1 (1/18) M_0 R_0 G_3	X		
72	57	M	pT_2 pN_1 (2/18) M_0 R_0 G_3	X		
73	75	M	pT_3 pN_2 (4/45) M_0 R_0 G_2	X	X	

Abbreviations: M, male; F, female.

the anocutaneous verge and were classified based on the WHO histopathologic typing of colorectal cancers (27). The specimen collection includes 33 lymph node–negative tumors [T_3-$T_4N_0M_0$; Unio Internationale Contra Cancrum (UICC) stage II] and 40 lymph node–positive tumors (T_2-T_4N_1-N_2M_0; UICC stage III). After surgery, tumor resections were immediately stored on ice and then inspected by an experienced pathologist. Consistent with standard procedures, samples were only considered when the tissue contained at least 70% of tumor cells. Representative sections were macrodissected from the tumors and snap-frozen in liquid nitrogen. When possible, a representative biopsy of normal colonic mucosa was also obtained (n = 30). The clinical data and experimental setup are summarized in Table 1.

RNA isolation and labeling. The amount of material was in the range of 24 to 370 mg. RNA extraction was done using TRIZOL (Invitrogen, Carlsbad, CA) according to standard procedures[4] and resulted in RNA amounts that averaged 207 μg; 20 μg of total RNA were reverse transcribed into cDNA using random primers and reverse transcriptase. After incorporation of aminoallyl-dUTP followed by chemical coupling of Cy3 (Amersham, Piscataway, NJ), cDNA quantification and labeling efficiency was determined using the Nanodrop quantification device (Nanodrop, Rockland, DE). Control cDNA was generated by labeling of a reference mRNA pool (Stratagene, La Jolla, CA) as mentioned above using Cy5 (Amersham).

Expression profiling. Expression profiling was carried out on National Cancer Institute oligonucleotide arrays (21,543 features) as previously described (16) using the Operon V2 oligo set. Briefly, 20 μg of Cy3-labeled test cDNA and 20 μg of Cy5-labeled reference cDNA were hybridized at 42°C overnight in specifically designed hybridization cassettes (TeleChem International, Sunnyvale, CA). After hybridization, slides were washed and scanned on an Axon scanner using GenePixPro (3.0) software (Axon Instruments, Union City, CA). Spot quality was assessed according to criteria in GenePixPro (3.0) software. Background subtraction and normalization was done upon data extraction from the CIT/NIH microarray database, mAdb.[5] Spots with a size of <10 μm or an intensity of <100 in both the red and green channels were eliminated, followed by removal of features that were uninformative in >50% of available arrays.

Quantitative real-time PCR. Gene expression levels were validated by quantitative reverse transcription-PCR (RT-PCR) with Power SYBR Green technology (Applied Biosystems, Inc., Foster City, CA). For each RT-PCR reaction, 300 ng cDNA was used. PCR was done with the default variables of the Applied Biosystems' Prism 7000 sequence detector, except for a total reaction volume of 25 μL. Primers were obtained from Operon Technologies, Inc. (Huntsville, AL). The sequences for the primers used here are provided in Supplementary Table S1 and correspond to the same region of the genes interrogated by the microarray. Each sample was analyzed in triplicate, and each data point was calculated as the median of the three measured C_T values.

DNA isolation and comparative genomic hybridization. After successful RNA extraction, DNA was isolated using sodium citrate/ethanol (details of the experimental procedures are provided at the following web site: http://www.riedlab.nci.nih.gov/protocols.asp). On average, DNA amounts of 205 μg were obtained. Comparative genomic hybridization (CGH) was done for 32 tumors as previously reported (28).

Statistical analysis. Statistical analyses were done using the *BRBArray-Tools* package (version 3.1.0) for microarray analysis developed at the Biometrics Research Branch of the National Cancer Institute[6] and MATLAB (version 6.5) from The Mathworks (Natick, MA).

A class comparison analysis was done using the expression data of 73 primary colon carcinomas and 30 matched mucosa samples. The two-sample t statistic with randomized variance (29) was used to measure the difference in gene expression between the two classes. The randomized variance model assumes that the variance of the expression of each gene is randomly drawn from an inverse-γ distribution and enables sharing of variance information among genes without assuming all genes have the same variance. Class prediction analysis was done using the Diagonal Linear Discriminant classifier (30, 31). We then used leave-one-out cross-validation (LOOCV) to estimate the extent to which tumor samples could be discerned from normal mucosa (32, 33).

We also did a class comparison analysis using the expression data of the 33 lymph node–negative tumors (UICC stage II) and the 40 lymph node–positive tumors (UICC stage III) following the procedures described above for the analysis of tumor versus normal mucosa. Genes were considered as

[4] http://www.riedlab.nci.nih.gov/protocols.asp.
[5] http://nciarray.nci.nih.gov/.

[6] http://linus.nci.nih.gov/BRB-ArrayTools.html.

being differentially expressed at a significance of $P < 0.001$. We then did class prediction analysis, again using a significance threshold of $P < 0.001$.

To assess the consequences of chromosomal aneuploidies on global gene expression levels, we established genomic copy number changes for 32 of the 73 colon carcinomas using chromosomal CGH. We then plotted the tumor/reference ratio measurements per chromosome arm against the expression values of its resident genes, excluding values that mapped to the centromeric and pericentromeric heterochromatic regions. For the comparison here, we considered only those copy number alterations that affected entire chromosome arms.

Biological pathway analysis. Gene lists both for the discernment of tumor versus mucosa and lymph node–positive versus lymph node–negative tumors were assessed for known biological interactions and involvement in canonical pathways using Ingenuity Pathway Analysis (IPA; Ingenuity, Mountain View, CA).

Results

Comparison of primary colon carcinomas and normal colon mucosa: differentially expressed genes. Here, we have used global gene expression profiling of 73 primary colon carcinomas and 30 matched normal mucosa samples on oligonucleotide microarrays to generate signatures of malignant transformation in these common tumors. The clinical data of the patients are summarized in Table 1. After normalization and filtering of the array data, 16,037 of the 22,543 printed features were available for further analyses. Based on a group comparison, we identified a set of 4,371 genes that was differentially expressed between colon cancers and normal mucosa at a significance of $P < 0.0001$ (using the two-sample t statistic). To increase the confidence that the detected genes point to relevant biological pathways in colonic carcinogenesis, we applied the same additional stringent selection criteria that had already been applied to a set of rectal carcinomas in a previous study (16): (*a*) at a significance value of $P < 1e-7$, 2,074 features (which correspond to 1,950 annotated genes) were differentially expressed between carcinomas and normal mucosa; 1,582 of these genes were up-regulated in the carcinoma samples, whereas 368 showed decreased expression (Supplementary Table S2). An exclusive comparison of the 30 matched tumors and mucosa revealed 1,102 differentially expressed genes at a $P < 1e-7$ (Supplementary Table S3). (*b*) Seventeen genes were at least 5-fold deregulated between the average of the normal mucosa and the carcinomas. Twelve of these genes were up-regulated (including *MYC*), and five showed reduced expression in the tumor samples (Table 2A). (*c*) In contrast to the results from the rectal carcinomas, only one gene (*HMGA1*) was always >2-fold higher expressed between any given tumor and its matched mucosa. This gene also satisfied the criteria defined in (*a*) and (*b*). Interestingly, *HMGA1*, a putative oncogene, is involved in the *MYC*-signaling pathway, in chromatin remodeling, and inhibits the function of *TP53* family members in cancer (34). Furthermore, we identified two genes (*SLC12A2* and *RPL13*) for which the minimum expression levels in the tumors were always higher than the maximum in the normal mucosa samples.

Comparison of primary colon carcinomas and normal colon mucosa: class prediction using gene expression profiles. We then wished to explore whether the gene expression levels of the carcinomas and normal mucosa samples were sufficiently different to discern the two groups. This was done using an established LOOCV with five different classifiers and a significance level of $P = 0.0001$. Two of these classifiers (1 nearest neighbor and 3 nearest neighbor) resulted in a correct prediction rate of 100%, whereas the remaining three (support vector machine, diagonal linear discriminant, and compound covariate predictor) showed accuracies ranging from 96% to 98%. This is equivalent to the results generated from the analysis of the rectal carcinomas (16).

Functional annotation of differentially expressed genes. The functional annotation of the differentially expressed genes and their affiliation with specific genetic pathways was interrogated using the IPA software. In the analysis summarized here, we targeted our search to the 17 genes with a >5-fold difference in average expression between tumor and mucosa, of which 10 genes were represented in the manually curated knowledge bank of IPA. Interestingly, the 10 focus genes clustered tightly into one network (as defined by IPA), with *MYC* in a central location regulated by and regulating the neighboring genes (Fig. 1). As expected, *MYC* shows on average a >5-fold expression increase in the tumors ($P < 1e-7$). This expression increase is accompanied by transcriptional activation of *HMGA1*, whose expression is on average up-regulated >5-fold as well ($P < 1e-7$). Other genes whose levels of transcriptional deregulation are intuitive include *RPL36A, LCN2, S100A11, RRM2*, and *FABP1*. The top cellular categories in this network are cancer, cell cycle, and cell assembly and organization.

We then asked to which extent genes reported to be involved in colorectal carcinogenesis and/or in the Wnt/β-catenin signaling pathway were affected in our colon data set. Most of these genes were previously studied in rectal carcinomas (16). At a significant threshold of $P < 1e-7$, 2,074 of 16,037 features on the arrays were differentially expressed between tumor and mucosa samples (13%). However, 28% (26 of 94) of the genes involved in colorectal cancer and 25% (8 of 32) of those that are part of the Wnt/β-catenin signaling pathway were significantly deregulated at $P < 1e-7$ in colon cancers. Using the binomial distribution, we calculated the probability of these percentages occurring by chance as $P < 0.001$ and $P = 0.046$, respectively. These data support the interpretation that these pathways are involved in the genesis of colon and rectal cancers.

Genomic clustering of differentially expressed genes. We had previously shown that differentially expressed genes in rectal carcinomas revealed a predilection for certain chromosomes. For instance, genes on chromosomes 13 and 20 were more frequently differentially expressed compared with genes on other chromosomes (16). The genomic clustering of the 1,950 genes differentially expressed ($P < 1e-7$) in the colon cancer samples analyzed here was not as obvious as in the rectal cancers, and the number of deregulated genes was in general a reflection of the number of genes on the array (Fig. 2*A*). However, when analyzing the proportion of deregulated genes that were increased in expression, we could show that chromosomes 13 and 20 again contained more genes that were overexpressed in the tumor samples. The results, however, were only significant for chromosome 20 ($P = 0.01$, using the binomial distribution). Although chromosome 13 also exhibits a higher proportion of overexpressed genes, the statistical value ($P = 0.08$) was lower (Fig. 2*B*).

Comparison of lymph node–positive and lymph node–negative cancers: differentially expressed genes. We then asked whether gene expression patterns in primary colon cancers that, at time of diagnosis, are negative or positive for lymph node metastases (UICC stages II and III, respectively), are different. This could point to biological differences between these groups and could potentially allow conclusions as to whether the propensity for the development of lymph node metastases is an inherent feature of the primary tumor. Of the 16,037 features that passed the filtering criteria, we identified 74 features (68 genes) that were significantly differentially expressed ($P < 0.001$) between the lymph node–negative and lymph node–positive carcinomas (Table 2C).

Table 2. Expression ratio for genes of interest

Unigene	Gene name	Description	Map	Av. tumor/av. mucosa	*P*
A. Genes with 5-fold differential expression between colon cancer and mucosa					
Hs.546343	*CLCA4*	Chloride channel, calcium-activated, family member 4	1p31-p22	0.1454	6.00e−7
Hs.380135	*FABP1*	Fatty acid binding protein 1, liver	2p11	0.1974	<1e−7
Hs.406691	*HIST1H2AJ*	Histone 1, H2aj	6p22-p21.3	6.2340	<1e−7
Hs.182432	*HIST1H2BM*	Histone 1, H2bm	6p22-p21.3	5.0377	<1e−7
Hs.518805	*HMGA1*	High mobility group AT-hook 1, transcript variant 5	6p21	5.3492	<1e−7
Hs.458414	*IFITM1*	IFN-induced transmembrane protein 1 (9-27)	11p15.5	5.7868	<1e−7
Hs.525648	*IGHG1*	Immunoglobulin heavy constant γ 1 (G1m marker)	14q32.33	5.0495	<1e−7
Hs.204238	*LCN2*	Lipocalin 2 (oncogene 24p3)	9q34	5.7985	2.48e−5
Hs.272789	*MS4A12*	Membrane-spanning 4-domains, subfamily A, member 12	11q12	0.1534	<1e−7
Hs.202453	*MYC*	v-*myc* myelocytomatosis viral oncogene homologue (avian)	8q24.12-q24.13	5.4948	<1e−7
Hs.232165	*PRV1*	Polycythemia rubra vera 1	19q13.2	0.1337	<1e−7
Hs.432485	*RPL36A*	Ribosomal protein L36a	Xq22.1	5.8333	<1e−7
Hs.368304	*RPS2*	Ribosomal protein S2	16p13.3	5.9091	<1e−7
Hs.226390	*RRM2*	Ribonucleotide reductase M2 polypeptide	2p25-p24	5.1410	<1e−7
Hs.417004	*S100A11*	S100 calcium binding protein A11 (calgizzarin)	1q21	5.3681	<1e−7
Hs.162585	*SLC12A2*	Solute carrier family 12 (sodium/potassium/ chloride transporters), member 2	5q23.3	13.2698	<1e−7
Hs.302738	*SLC26A2*	Solute carrier family 26 (sulfate transporter), member 2	5q31-q34	0.1748	<1e−7
B. Genes reported to be effected in human colorectal cancer					
Hs.12341	*ADAR**	Adenosine deaminase, RNA-specific, transcript variant ADAR-c	1q21.1-q21.2	1.9791	<1e−7
Hs.158932	*APC* †	Adenomatosis polyposis coli	5q21-q22	0.7384	0.0413
Hs.512765	*AXIN1**	Axin 1, transcript variant 2	16p13.3	1.7782	<1e−7
Hs.156527	*AXIN2**	Axin 2 (conductin, axil)	17q23-q24	2.0671	2.0e−7
Hs.485139	*BAK1* †	BCL2-antagonist/killer 1	6p21.3	0.9225	0.2576
Hs.150749	*BCL2* ‡	B-cell CLL/lymphoma 2	18q21.33	0.9042	0.2731
Hs.514527	*BIRC5*§	Baculoviral IAP repeat-containing 5, transcript variant 1	17q25	1.0325	0.8202
Hs.500812	*BTRC* ‡	β-Transducin repeat containing, transcript variant 2	10q24.32	0.5603	0.0012
Hs.418533	*BUB3*‖	BUB3 budding uninhibited by benzimidazoles 3 homologue (yeast), transcript variant 1	10q26	0.6673	0.0123
Hs.141125	*CASP3* ‡	Caspase-3, apoptosis-related cysteine protease, transcript variant β	4q34	1.2255	0.0026
Hs.523852	*CCND1**	Cyclin D1 (PRAD1: parathyroid adenomatosis 1)	11q13	2.2010	4.0e−7
Hs.502328	*CD44*¶	CD44 antigen (homing function and Indian blood group system)	11p13	2.5796	<1e−7
Hs.461086	*CDH1* ‡	Cadherin 1, type 1, E-cadherin (epithelial)	16q22.1	1.1212	0.3919
Hs.370771	*CDKN1A* ‡	Cyclin-dependent kinase inhibitor 1A (p21, Cip1), transcript variant 1	6p21.2	0.8853	0.2620
Hs.238990	*CDKN1B* ‡	Cyclin-dependent kinase inhibitor 1B (p27, Kip1)	12p13.1-p12	1.0632	0.4335
Hs.446747	*CLN3*‖	Ceroid-lipofuscinosis, neuronal 3, juvenile (Batten, Spielmeyer-Vogt disease)	16p12.1	1.2865	0.0967
Hs.442592	*CSNK1A1* ‡	Casein kinase 1, α1	5q32	0.9459	0.5741
Hs.446484	*CSNK2A1**	Casein kinase 2, α 1 polypeptide, transcript variant 3	20p13	0.4632	<1e−7
Hs.208597	*CTBP1*‖	COOH-terminal binding protein 1, transcript variant 1	4p16	1.4123	1.0e−7
Hs.476018	*CTNNB1**	Catenin (cadherin-associated protein), β 1, 88 kDa	3p21	1.8724	<1e−7
Hs.82407	*CXCL16* ‡	Chemokine (C-X-C motif) ligand 16	17p13	0.7647	0.0007
Hs.156316	*DCN*‖	Decorin, transcript variant C	12q13.2	0.6550	0.0444
Hs.172648	*DLX4*‖	Distal-less homeobox 4, transcript variant 2	17q21.33	1.4522	0.0001
Hs.335034	*DPYD**	Dihydropyrimidine dehydrogenase	1p22	0.2588	1.4e−5
Hs.488293	*EGFR* ‡	Epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homologue, avian)	7p12	0.7427	0.0147
Hs.488293	*EGFR* ‡	Epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homologue, avian)	7p12	1.1724	0.0933
Hs.488293	*EGFR* ‡	Epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homologue, avian)	7p12	1.0370	0.5818

(*Continued on the following page*)

Table 2. Expression ratio for genes of interest (Cont'd)

Unigene	Gene name	Description	Map	Av. tumor/av. mucosa	*P*
Hs.488293	*EGFR* ‡	Epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homologue, avian)	7p12	1.0497	0.6724
Hs.488293	*EGFR* ‡	Epidermal growth factor receptor (erythroblastic leukemia viral (v-erb-b) oncogene homologue, avian)	7p12	1.0297	0.7796
Hs.249718	*EIF4E* ‡	Eukaryotic translation initiation factor 4E	4q21-q25	1.0917	0.5827
Hs.76753	*ENG*¶	Endoglin (Osler-Rendu-Weber syndrome 1)	9q33-q34.1	1.7069	<1e−7
Hs.523329	*EPHB2*§	EPH receptor B2, transcript variant 1	1p36.1-p35	2.0082	0.0001
Hs.437008	*EPHB4*∥	EPH receptor B4	7q22	0.8902	0.2954
Hs.446352	*ERBB2* ‡	v-erb-b2 erythroblastic leukemia viral oncogene homologue 2, neuro/glioblastoma-derived oncogene homologue (avian), transcript variant 1	17q21.1	1.1790	0.1872
Hs.434059	*ETV4*¶	Ets variant gene 4 (E1A enhancer binding protein, E1AF)	17q21	2.0244	<1e−7
Hs.26770	*FABP7*¶	Fatty acid binding protein 7, brain	6q22-q23	0.5682	<1e−7
Hs.444552	*FLJ12529*¶	Pre-mRNA cleavage factor I, 59 kDa subunit	11q12.2	1.6352	<1e−7
Hs.126057	*FRAT1* ‡	Frequently rearranged in advanced T-cell lymphomas, transcript variant 1	10q24.1	1.0079	0.9435
Hs.94234	*FZD1* ‡	Frizzled homologue 1 (*Drosophila*)	7q21	0.8509	0.0884
Hs.292493	*G22P1*∥	Thyroid autoantigen 70 kDa (Ku antigen)	22q13.2-q13.31	0.7030	4.9e−5
Hs.292493	*G22P1**	Thyroid autoantigen 70 kDa (Ku antigen)	22q13.2-q13.31	1.8593	<1e−7
Hs.234896	*GMNN* ‡	Geminin, DNA replication inhibitor	6p22.2	1.5305	0.0003
Hs.58561	*GPR87*¶	G protein–coupled receptor 87	3q24	0.5108	<1e−7
Hs.445733	*GSK3B*¶	Glycogen synthase kinase 3 β	3q13.3	1.9103	<1e−7
Hs.116462	*HNF4A* †	Hepatocyte nuclear factor 4, α, transcript variant 2	20q12-q13.1	1.7249	0.0002
Hs.530227	*HSF1* ‡	Heat shock transcription factor 1	8q24.3	0.8966	0.3072
Hs.487062	*IGF2R* ‡	Insulin-like growth factor 2 receptor	6q26	0.9967	0.9816
Hs.522818	*L1CAM* ‡	L1 cell adhesion molecule, transcript variant 2	Xq28	1.1216	0.0601
Hs.125132	*LEF1* ‡	Lymphoid enhancer-binding factor 1	4q23-q25	1.1640	0.4959
Hs.102267	*LOX**	Lysyl oxidase	5q23.2	1.6019	8.3e−6
Hs.485968	*MAP3K7* ‡	Mitogen-activated protein kinase kinase kinase 7, transcript variant D	6q16.1-q16.3	1.1370	0.0983
Hs.507681	*MAP3K7IP1* †	Mitogen-activated protein kinase kinase kinase 7 interacting protein 1	22q13.1	1.4271	0.0002
Hs.549053	*MICA* ‡	MHC class I polypeptide-related sequence A	6p21.3	1.2301	0.0131
Hs.80976	*MKI67* ‡	Antigen identified by monoclonal antibody Ki-67	10q25-qter	1.2428	0.0904
Hs.195364	*MLH1*§	MutL homologue 1, colon cancer, nonpolyposis type 2 (*Escherichia coli*)	3p21.3	0.8917	0.2921
Hs.83169	*MMP1*∥	Matrix metalloproteinase 1 (interstitial collagenase)	11q22.3	0.9556	0.7422
Hs.156519	*MSH2* ‡	MutS homologue 2, colon cancer, nonpolyposis type 1 (*E. coli*)	2p22-p21	1.2519	0.1718
Hs.445052	*MSH6* ‡	MutS homologue 6 (*E. coli*)	2p16	0.7931	0.0074
Hs.202453	*MYC*¶	v-*myc* myelocytomatosis viral oncogene homologue (avian)	8q24.12-q24.13	5.4948	<1e−7
Hs.25960	*MYCN*§	v-*myc* myelocytomatosis viral-related oncogene, neuroblastoma derived (avian)	2p24.1	0.5318	0.0001
Hs.208759	*NLK* ‡	Nemo-like kinase	17q11.2	1.3011	0.0070
Hs.519452	*NPM1*¶	Nucleophosmin (nucleolar phosphoprotein B23, numatrin)	5q35	3.7391	<1e−7
Hs.515524	*NUCB1*∥	Nucleobindin 1	19q13.2-q13.4	1.9420	<1e−7
Hs.147433	*PCNA*¶	Proliferating cell nuclear antigen, transcript variant 2	20pter-p12	3.6092	<1e−7
Hs.549112	*PHLDA1*¶	Pleckstrin homology-like domain, family A, member 1	12q15	2.1962	0.0001
Hs.77274	*PLAU*¶	Plasminogen activator, urokinase	10q24	2.0827	0.0001
Hs.485196	*PPARD* ‡	Peroxisome proliferative activated receptor, delta, transcript variant 1	6p21.2-p21.1	0.8934	0.0761
Hs.500466	*PTEN*∥	Phosphatase and tensin homologue (mutated in multiple advanced cancers 1)	10q23.3	0.6022	1.0e−7
Hs.196384	*PTGS2*∥	Prostaglandin-endoperoxide synthase 2 (prostaglandin G/H synthase and cyclooxygenase)	1q25.2-q25.3	0.7627	0.2710
Hs.43666	*PTP4A3*∥	Protein tyrosine phosphatase type IVA, member 3, transcript variant 2	8q24.3	2.4590	<1e−7
Hs.413812	*RAC1* ‡	Ras-related C3 botulinum toxin substrate 1 (rho family, small GTP binding protein Rac1), transcript variant Rac1	7p22	0.7153	0.0026
Hs.502875	*RELA* ‡	v-rel reticuloendotheliosis viral oncogene homologue A, nuclear factor of κ light polypeptide gene enhancer in B cells 3, p65 (avian)	11q13	1.0933	0.1077

(*Continued on the following page*)

Table 2. Expression ratio for genes of interest (Cont'd)

Unigene	Gene name	Description	Map	Av. tumor/av. mucosa	*P*
Hs.170019	*RUNX3* ‡	Runt-related transcription factor 3	1p36	0.9194	0.4741
Hs.514913	*SERPINB2* ‡	Serine (or cysteine) proteinase inhibitor, clade B (ovalbumin), member 2	18q21.3	1.0481	0.8699
Hs.414795	*SERPINE1* ‡	Serine (or cysteine) proteinase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	7q21.3-q22	1.1902	0.2752
Hs.213424	*SFRP1* ‡	Secreted frizzled-related protein 1	8p12-p11.1	0.7990	0.1974
Hs.936	*SLC34A1* ∥	Solute carrier family 34 (sodium phosphate), member 1	5q35	0.9409	0.5700
Hs.75862	*SMAD4* †	SMAD, mothers against DPP homologue 4 (*Drosophila*)	18q21.1	0.7643	0.0104
Hs.48029	*SNAI1* ∥	Snail homologue 1 (*Drosophila*)	20q13.1-q13.2	0.9292	0.1323
Hs.360174	*SNAI2* ‡	Snail homologue 2 (*Drosophila*)	8q11	0.7397	0.0165
Hs.2316	*SOX9* ¶	SRY (sex determining region Y)-box 9 (campomelic dysplasia, autosomal sex-reversal)	17q24.3-q25.1	1.6483	1.0e−7
Hs.524461	*SP1* †	Sp1 transcription factor	12q13.1	1.3991	0.0132
Hs.195659	*SRC* †	v-src sarcoma (Schmidt-Ruppin A-2) viral oncogene homologue (avian), transcript variant 2	20q12-q13	1.1934	0.0793
Hs.437058	*STAT5A* †	Signal transducer and activator of transcription 5A	17q11.2	0.3030	<1e−7
Hs.23582	*TACSTD2* §	Tumor-associated calcium signal transducer 2	1p32-p31	1.5260	0.0779
Hs.552578	*TCF1* ‡	Transcription factor 1, hepatic; LF-B1, hepatic nuclear factor (HNF1), albumin proximal factor	12q24.2	1.2816	0.0002
Hs.1103	*TGFB1* ‡	Transforming growth factor, β 1 (Camurati-Engelmann disease)	19q13.1	1.1986	0.0859
Hs.369397	*TGFBI* ¶	Transforming growth factor, β-induced, 68 kDa	5q31	4.8397	<1e−7
Hs.494622	*TGFBR1* ∥	Transforming growth factor, β receptor I (activin A receptor type II-like kinase, 53 kDa)	9q22	0.9566	0.7910
Hs.82028	*TGFBR2* ‡	Transforming growth factor, β receptor II (70/80 kDa)	3p22	0.8568	0.2352
Hs.104839	*TIMP2* ‡	Tissue inhibitor of metalloproteinase 2	17q25	0.7570	0.1109
Hs.104839	*TIMP2* ∥	Tissue inhibitor of metalloproteinase 2	17q25	1.7871	<1e−7
Hs.297324	*TIMP3* *	Tissue inhibitor of metalloproteinase 3 (Sorsby fundus dystrophy, pseudoinflammatory)	22q12.3	0.7265	8.7e−6
Hs.197320	*TLE1* †	Transducin-like enhancer of split 1 [E(sp1) homologue, *Drosophila*]	9q21.32	1.5402	2.0e−7
Hs.197320	*TLE1* †	Transducin-like enhancer of split 1 [E(sp1) homologue, *Drosophila*]	9q21.32	0.4857	5.0e−7
Hs.408312	*TP53* ‡	Tumor protein p53 (Li-Fraumeni syndrome)	17p13.1	0.7269	0.0474
Hs.369762	*TYMS* ∥	Thymidylate synthetase	18p11.32	1.3628	0.0038
Hs.73793	*VEGF* ¶	Vascular endothelial growth factor	6p12	1.9404	<1e−7
Hs.284122	*WIF1* †	WNT inhibitory factor 1	12q14.3	0.7679	0.0056
Hs.492974	*WISP1* §	WNT1 inducible signaling pathway protein 1, transcript variant 1	8q24.1-q24.3	1.0663	0.4602
Hs.388739	*XRCC5* ‡	X-ray repair complementing defective repair in Chinese hamster cells 5 (double-strand break rejoining; Ku autoantigen, 80 kDa)	2q35	1.0716	0.5546

C. Genes with differential expression between lymph node–positive (UICC stage III) and lymph node–negative (UICC stage II) tumors

Unigene	Gene name	Description	Map	Av. UICC stage III/av. UICC stage II	*P*
Hs.12341	*ADAR*	Adenosine deaminase, RNA-specific, transcript variant ADAR-c	1q21.1-q21.2	0.767	4.9e−5
Hs.119591	*AP2S1*	Adaptor-related protein complex 2, σ 1 subunit, transcript variant AP17	19q13.2-q13.3	0.723	8.8e−6
Hs.433291	*ARD1*	ARD1 homologue, *N*-acetyltransferase (*Saccharomyces cerevisiae*)	Xq28	0.791	0.0003
Hs.521056	*ATP5J2*	ATP synthase, H^+ transporting, mitochondrial F0 complex, subunit f, isoform 2, nuclear gene encoding mitochondrial protein, transcript variant 1	7q22.1	0.654	0.0007
Hs.324521	*C14orf156*	Chromosome 14 open reading frame 156	14q24.3	0.624	4.7e−6
Hs.368149	*CCT7*	Chaperonin containing TCP1, subunit 7 (eta), transcript variant 1	2p13.2	0.698	0.0002
Hs.524216	*CDCA3*	Cell division cycle associated 3	12p13	0.711	0.0005
Hs.414565	*CLIC1*	Chloride intracellular channel 1	6p22.1-p21.2	0.705	0.0004
Hs.176615	*COLEC10*	Collectin sub-family member 10 (C-type lectin)	8q23-q24.1	0.579	0.0001
Hs.5120	*DNCL1*	Dynein, cytoplasmic, light polypeptide 1	12q24.23	0.749	0.0002
Hs.131431	*EIF2AK2*	Eukaryotic translation initiation factor 2-α kinase 2	2p22-p21	0.717	0.0006

(*Continued on the following page*)

Table 2. Expression ratio for genes of interest (Cont'd)

Unigene	Gene name	Description	Map	Av. UICC stage III/av. UICC stage II	*P*
Hs.530096	*EIF3S2*	Eukaryotic translation initiation factor 3, subunit 2 β, 36 kDa	1p34.1	0.749	0.0008
Hs.415846	*FTCD*	Formiminotransferase cyclodeaminase, transcript variant B	21q22.3	0.803	0.0009
Hs.301961	*GSTM1*	Glutathione *S*-transferase M1	1p13.3	0.613	0.0005
Hs.119192	*H2AFZ*	H2A histone family, member Z	4q24	0.722	0.0002
Hs.20521	*HRMT1L2*	HMT1 hnRNP methyltransferase-like 2 (*S. cerevisiae*), transcript variant 3	19q13.3	0.772	0.0005
	HSPC003	HSPC003 protein		0.709	0.0001
Hs.515126	*ICAM1*	Intercellular adhesion molecule 1 (CD54), human rhinovirus receptor	19p13.3-p13.2	0.529	7.7e−6
Hs.624	*IL8*	Interleukin 8	4q13-q21	0.431	0.0006
Hs.91142	*KHSRP*	KH-type splicing regulatory protein (FUSE binding protein 2)	19p13.3	0.773	0.0010
Hs.182507	*KRTHB5*	Keratin, hair, basic, 5	12q13	1.363	0.0008
Hs.549159	*LOC51326*	ADP-ribosylation factor-like	17q21.31	1.246	0.0009
Hs.432453	*MAP3K8*	Mitogen-activated protein kinase kinase kinase 8	10p11.23	0.711	0.0003
Hs.355867	*MARS*	Methionine-tRNA synthetase	12q13.2	0.750	0.0008
Hs.532833	*MC4R*	Melanocortin 4 receptor	18q22	1.532	0.0001
Hs.471918	*MCPIP*	MCP-1 treatment-induced protein	1p34.3	0.668	0.0001
Hs.367842	*MKI67IP*	MKI67 (FHA domain) interacting nucleolar phosphoprotein	2q14.3	0.748	0.0007
Hs.190086	*MRCL3*	Myosin regulatory light chain	18p11.31	0.663	0.0002
Hs.476706	*MRPL37*	Mitochondrial ribosomal protein L37, nuclear gene encoding mitochondrial protein	1p32.1	0.789	0.0008
Hs.144941	*MUF1*	MUF1 protein	1p34.1	0.694	0.0002
Hs.551508	*NEB*	Nebulin	2q22	1.317	0.0002
Hs.155396	*NFE2L2*	Nuclear factor (erythroid-derived 2)-like 2	2q31	0.770	0.0006
Hs.209113	*NOMO1*	NODAL modulator 1	16p13.11	0.770	0.0009
Hs.515876	*NRBP*	Nuclear receptor binding protein	2p23	0.740	0.0002
Hs.446427	*OAZ1*	Ornithine decarboxylase antizyme 1	19p13.3	0.753	0.0003
Hs.239499	*PDCD11*	Programmed cell death 11	10q24.33	0.728	1.6e−5
Hs.443831	*PDCD5*	Programmed cell death 5	19q12-q13.1	0.656	3.8e−5
Hs.530479	*PMF1*	Polyamine-modulated factor 1	1q12	0.810	0.0009
Hs.47062	*POLR2I*	Polymerase (RNA) II (DNA directed) polypeptide I, 14.5 kDa	19q12	0.725	3.0e−5
Hs.434937	*PPIB*	Peptidylprolyl isomerase B (cyclophilin B)	15q21-q22	0.719	0.0009
Hs.17883	*PPM1G*	Protein phosphatase 1G (formerly 2C), magnesium-dependent, γ isoform, transcript variant 2	2p23.3	0.705	2.2e−5
Hs.467192	*PPP2R1A*	Protein phosphatase 2 (formerly 2A), regulatory subunit A (PR 65), α isoform (PPP2R1A), mRNA.	19q13.41	0.827	0.0009
Hs.432121	*PRDX2*	Peroxiredoxin 2	19p13.2	0.671	0.0001
Hs.446260	*PSMA6*	Proteasome (prosome, macropain) subunit, α type, 6	14q13	0.719	0.0006
Hs.89545	*PSMB4*	Proteasome (prosome, macropain) subunit, β type, 4	1q21	0.748	0.0004
Hs.211594	*PSMC4*	Proteasome (prosome, macropain) 26S subunit, ATPase, 4, transcript variant 2	19q13.11-q13.13	0.677	5.3e−6
Hs.369125	*PSMD14*	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 14	2q24.2	0.696	2.0e−5
Hs.78466	*PSMD8*	Proteasome (prosome, macropain) 26S subunit, non-ATPase, 8	19q13.2	0.821	0.0003
Hs.75348	*PSME1*	Proteasome (prosome, macropain) activator subunit 1 (PA28 α), transcript variant 1	14q11.2	0.732	0.0002
Hs.434081	*PSME2*	Proteasome (prosome, macropain) activator subunit 2 (PA28 β)	14q11.2	0.582	1.0e−7
Hs.157351	*PTD004*	GTP-binding protein PTD004, transcript variant 1	2q31.1	0.755	0.0006
Hs.279529	*PX19*	Px19-like protein	5q35.3	0.735	0.0007
Hs.77510	*RFWD3*	Ring finger and WD repeat domain 3	16q22.3	0.696	0.0008
Hs.269004	*SLC36A1*	Solute carrier family 36 (proton/amino acid symporter), member 1	5q33.1	1.500	0.0009
Hs.515472	*SNRPD2*	Small nuclear ribonucleoprotein D2 polypeptide 16.5 kDa, transcript variant 1	19q13.2	0.680	0.0006
Hs.490394	*SSBP1*	Single-stranded DNA binding protein 1	7q34	0.723	0.0001
Hs.194385	*STAP2*	Signal-transducing adaptor protein-2	19p13.3	0.591	0.0002
Hs.172772	*TCEB2*	Transcription elongation factor B (SIII), polypeptide 2 (18 kDa, elongin B), transcript variant 1	16p12.3	0.704	0.0002

(*Continued on the following page*)

Table 2. Expression ratio for genes of interest (Cont'd)

Unigene	Gene name	Description	Map	Av. UICC stage III/av. UICC stage II	*P*
Hs.371282	*TCF3*	Transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47) (TCF3), mRNA	19p13.3	0.754	0.0007
Hs.371282	*TCF3*	Transcription factor 3 (E2A immunoglobulin enhancer binding factors E12/E47)	19p13.3	0.733	0.0007
Hs.518123	*TFG*	TRK-fused gene, transcript variant 1	3q12.2	0.739	0.0001
Hs.197320	*TLE1*	Transducin-like enhancer of split 1 [E(sp1) homologue, *Drosophila*]	9q21.32	0.762	0.0007
Hs.146070	*TPM3*	Tropomyosin 3	1q21.2	0.746	0.0009
Hs.44532	*UBD*	Ubiquitin D	6p21.3	0.325	1.3e−5
Hs.119251	*UQCRC1*	Ubiquinol-cytochrome *c* reductase core protein I	3p21.3	0.710	0.0004
Hs.77578	*USP9X*	Ubiquitin-specific protease 9, X-linked (fat facets-like, *Drosophila*), transcript variant 2	Xp11.4	0.695	0.0003
Hs.520974	*YWHAG*	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, γ polypeptide	7q11.23	0.734	0.0004
Hs.292575	*ZNF37A*	Zinc finger protein 37a (KOX 21)	10p11.2	0.702	0.0008
		Human DNA sequence from PAC 30P20 on chromosome Xq21.1-Xq21.3. Contains set pseudogene, ESTs and STS		0.697	0.0001
		Human DNA sequence from clone 38C16 on chromosome 6q22.33-24.1. Contains GAPD (glyceraldehyde 3-phos		0.696	0.0003
		Human DNA sequence from clone 522P13 on chromosome 6p21.31-22.3. Contains a 60S ribosomal protein L2		0.724	0.0004
Hs.532392		CDNA FLJ13112 fis, clone NT2RP3002587	1	1.346	0.0006
		Human DNA sequence from PAC 124O9 on chromosome 6q21. Contains DNAJ2 (HDJ1) like pseudogene, ESTs, S		0.731	0.0006
		Human DNA sequence from clone RP1-9E21 on chromosome 1q24-25. Contains a pseudogene similar to calpon		0.764	0.0008

*Significant in colon, same direction in rectum: 81% (25 of 31).
† Only values for the colon data set: 10 genes.
‡ Not significant in either data set: 37 genes.
§Significant in rectum, same direction in colon: 65% (20 of 31).
‖Significant (in either or both data sets) but opposite direction: 32% (16 of 50).
¶Significant in both data sets AND same direction: 93% (14 of 15).

Under the given experimental conditions, the likelihood of finding 74 features by chance has a probability of $P = 0.013$. Of these features, 68 were down-regulated in the lymph node–positive tumors, and six were up-regulated compared with the lymph node–negative tumors. The separation of the two groups based on these 74 features is displayed as a multidimensional scaling analysis (MDA) in Fig. 3*A*. We then did a class comparison with this set of genes, which achieved a sensitivity for the detection of lymph node–positive tumors that does not exceed 75%, and a specificity of 63%.

Functional annotation of genes differentially expressed in node-negative and node-positive tumors. In analogy to the analysis of genes involved in colon carcinogenesis, we employed IPA for the functional annotation of genes that were differentially expressed when comparing lymph node–negative with lymph node–positive tumors. Here, the genes were assigned to eight different networks (four with more than one focus gene), of which the two highest scoring networks were connected through *IL15RA* (see Supplementary Figure for further detail). Interestingly, the highest ranked network contained a plethora of genes involved in cell movement and immune response and genes involved in the development and function of the hematologic system. *IL8* is a prominent node in this network. The second network contained genes involved in cell-cell signaling and interactions and, just as network one, in immune response and the development and function of the hematologic system. The central connecting genes in this network are *IFNG* (IFN-γ) and *TNF*.

Validation of gene expression levels using quantitative RT-PCR. To validate the gene expression levels derived from expression profiling on the arrays, we did quantitative RT-PCR of 20 genes. These genes were selected because they were among those that were highly differentially expressed between the colon cancer and normal mucosa (*FABP1, HMGA1, MYC, RPL36A*, and *SLC12A2*), because of their involvement in the Wnt/β-catenin signaling pathway (*CD44, CTNNB1, GSK3B, PCNA, PLAU, SOX9, TGFB1, TWIST1*, and *VEGF*), or because of their differential expression between the lymph node–positive and lymph node–negative tumors (*IL8, UBD, ICAM1, PPM1G*, and *PSMC4*). *IL15RA* was validated because it connected the two highest scoring networks in the functional annotation analysis using IPA. The ratios of gene expression levels between tumor samples and matched normal mucosa were compared for up to 15 patients.

The RT-PCR analyses confirmed the expression levels for four highly differentially expressed genes (*HMGA1, MYC, RPL36A*, and *SLC12A2*), whereas *FABP1* showed down-regulation in the tumors

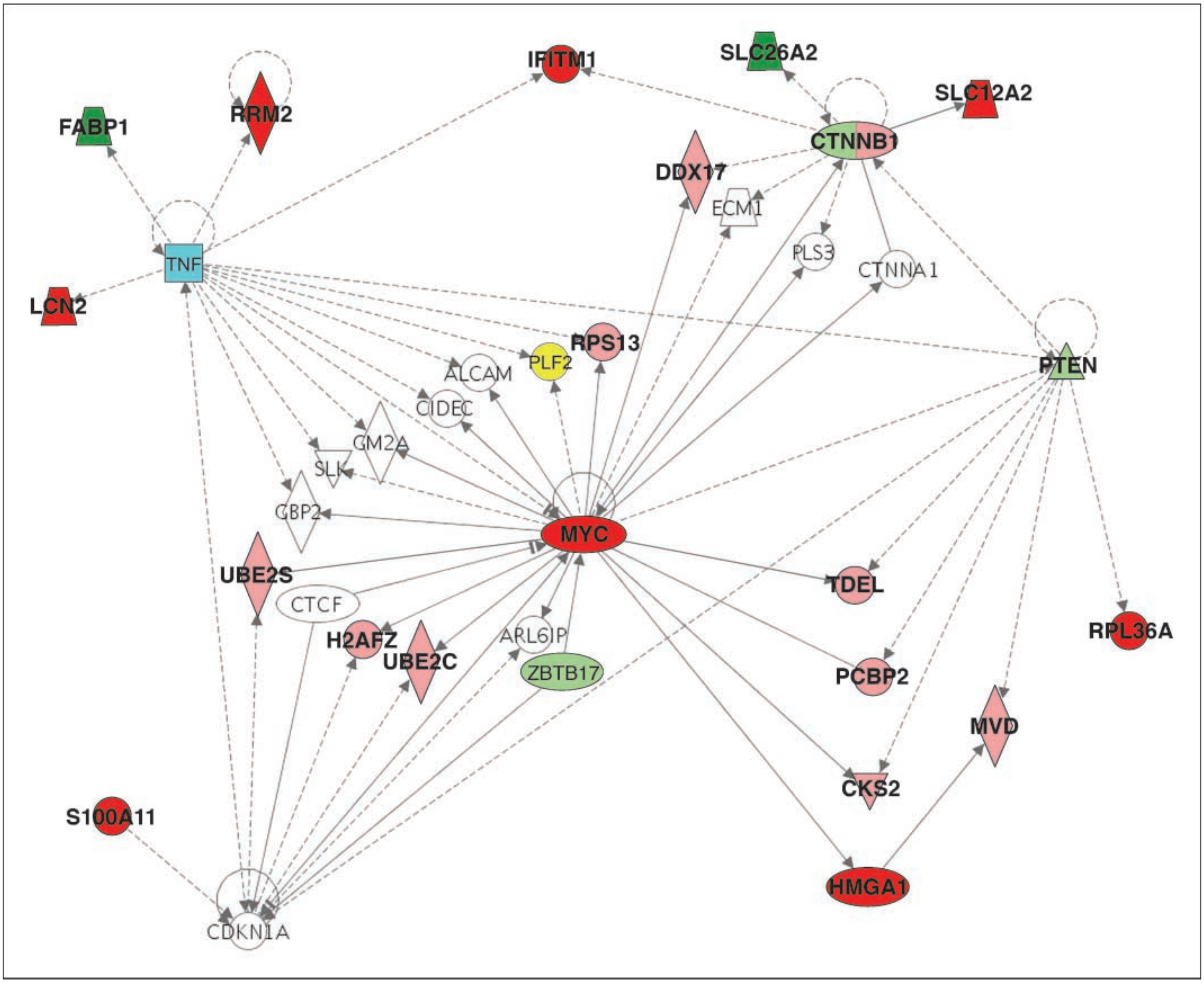


Figure 1. Network annotation of genes with greater than 5-fold expression change in colon tumors relative to normal colonic epithelium using IPA. *Red,* genes up-regulated in colon tumors; *green,* genes down-regulated in colon tumors. *Dark shade,* genes with >5-fold differential expression; *light shade,* genes with lower difference in expression. *PLF2* (*yellow*) was not spotted on the array, and *TNF* (*blue*) did not meet the filtering criteria. Genes whose names are printed in bold font were deregulated significantly ($P < 0.0001$). *CTNNB1* was present twice on the arrays and revealed conflicting results. *MYC* assumes a central position in the network.

by microarray analysis (0.64) but was up-regulated when analyzed by RT-PCR (1.86). The expression levels of all nine genes involved in the Wnt/β-catenin signaling pathways were confirmed.

The differences in gene expression levels in the node-positive versus node-negative tumors were confirmed for at least six patients in each group. The gene expression differences for *IL8, UBD, ICAM1*, and *PPM1G* were consistent between the platforms, but we could not confirm the directionality of the difference for *IL15RA* and *PSMC4*. In summary, 17 of 20 genes tested (85%) revealed concordant results between the microarray and the RT-PCR–based measurements.

Effects of chromosomal copy number changes and aneuploidy on average gene expression levels. Chromosomal aneuploidies are arguably the most common genetic aberrations in epithelial cancers. To assess their consequence on global gene expression levels, we mapped genomic imbalances from 32 of the colon carcinomas analyzed here using chromosome CGH (UICC stage II, $n = 14$ and UICC stage III, $n = 18$). The most frequent genomic gains occurred on chromosome arms 7p (66%), 8q (31%), 13q (66%), 20p (37%), and 20q (62%), whereas frequent losses mapped to chromosome arms 17p (43%) and 18q (47%). For a detailed case summary, see http://www.ncbi.nlm.nih.gov/sky/skyweb.cgi. Only two of the lymph node–negative tumors showed gains of chromosome arm 8q, whereas eight lymph node–positive tumors revealed copy number increases. This confirmed the results of previous analyses from our own laboratories (16, 35, 36) and from the literature (for a recent review, see ref. 37).

After having established the patterns and percentages of chromosomal copy number changes, we were in the position to query how precisely these imbalances affect the transcriptional activity of the resident genes. Towards this end, we measured average chromosome arm–specific gene expression levels (relative to the Stratagene reference RNA) for the 32 tumors for which we had done CGH analysis. These values were then compared with and plotted against the CGH ratio values for the respective chromosome arms in analogy to our previous analysis of rectal carcinomas (16). To calculate statistical correlations, we determined the percent correlation and the R^2 values between the average arm expression values and the average CGH measurements. In general, there was a strong positive correlation between the chromosome arm copy number and the average expression of its resident genes. Figure 4*A* shows the results for the commonly aneuploid chromosome arms (7p, 8q, 13q, 18q, 20p, and 20q). The correlation coefficients and significance values for all of the chromosome arms are presented in Supplementary Table S4. The median (52%) and average (49%) of the correlation coefficients are consistent with our observations in the rectal tumors (55% and 51%, respectively). As previously described (16, 38), we also plotted the average expression of each gene along the length of the chromosome arm for those chromosomes with frequent copy number changes and compared it with those cases in which these particular chromosomes were not subject to copy number alterations (Fig. 4*B*). The association of chromosome arm average gene expression levels and chromosomal copy numbers is depicted as a positively correlated general shift in the expression profiles.

Biological pathway concordance between primary colon and rectal carcinomas. In a previous study, we focused on the gene expression profiles of rectal carcinomas (16). In an attempt to determine the degree to which rectal and colon cancers are similar, we compared the genes that were deregulated at $P < 0.0001$ in the two data sets and represented on both oligonucleotide array platforms. After removing duplicate genes and probes that do not correspond to known genes, we obtained 2,978 genes (out of the original 4,371) with $P < 0.0001$ from the colon data set. When the

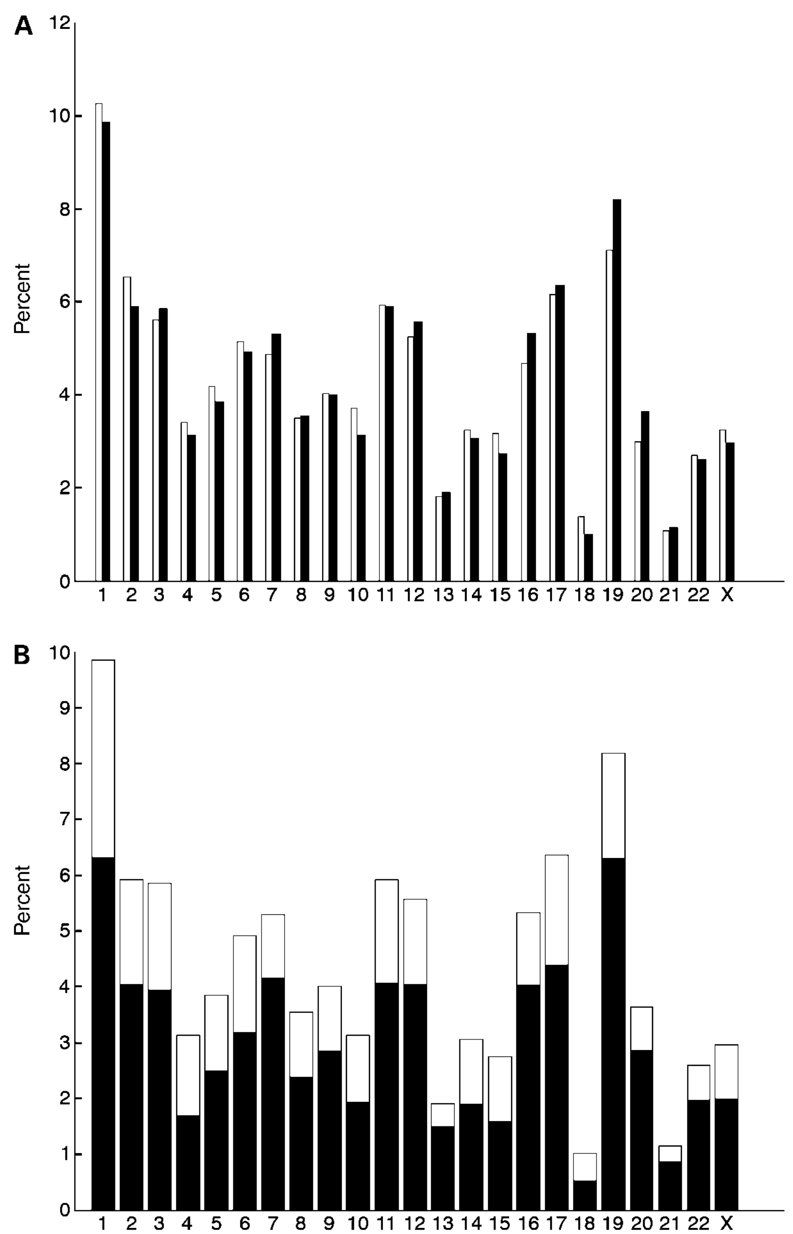


Figure 2. Chromosomal localization of genes with significant expression changes. *A,* 94% of the 16,037 genes that passed the filtering criteria had chromosome mapping locations. *White columns,* percentages of these genes that map to each chromosome; 94% of the 4,371 genes differentially expressed in the tumors with $P < 0.0001$ had known chromosome locations. *Black columns,* percentages of these genes, which map to each chromosome. *B,* percentage of genes indicated as black columns in (*A*) that were up-regulated (*black*) or down-regulated (*white*) in the tumors relative to the normal colon mucosa.

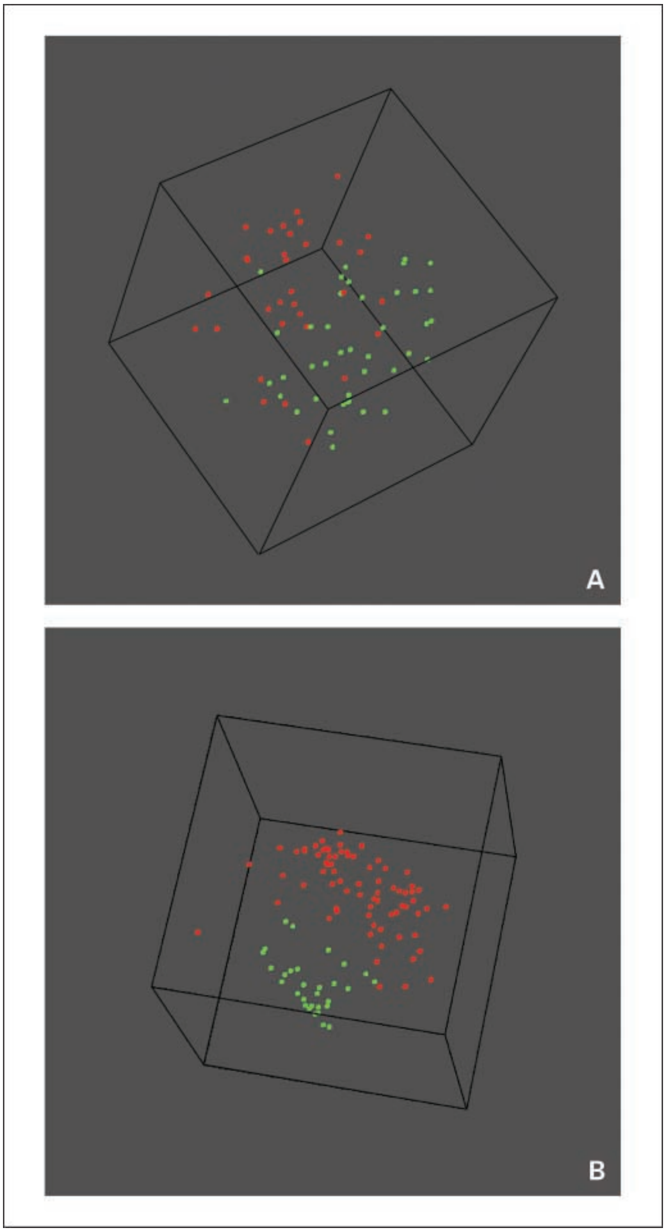


Figure 3. *A,* MDA of lymph node–negative (*red*) and lymph node–positive (*green*) colon cancers based on the set of 74 differentially expressed features. *B,* MDA of 73 colon cancers (*red*) and 30 matched normal mucosa samples (*green*). Note the stringent separation of the two groups.

same criteria were applied to the genes significantly deregulated between rectal tumors and normal rectal mucosa, we were left with 1,374 genes (out of the original 1,722) from the rectal data set. There was a considerable overlap of 490 genes between the two data sets (Supplementary Table S5). The probability that this overlap was due to chance is significantly small ($P < 0.001$). For 96 of these genes, the regulation is divergent. However, 394 genes were deregulated in analogous directions (i.e., up-regulated in the tumor samples of colon and rectum or vice versa). $P < 0.001$ was also obtained for the probability that this level of correlated deregulation is due to chance. Both *P*s were calculated by repeatedly taking random sets of 2,978 and 1,371 genes from the colon and rectal data set, respectively, and finding the number of times the overlap was equal to or above the overlap we observed.

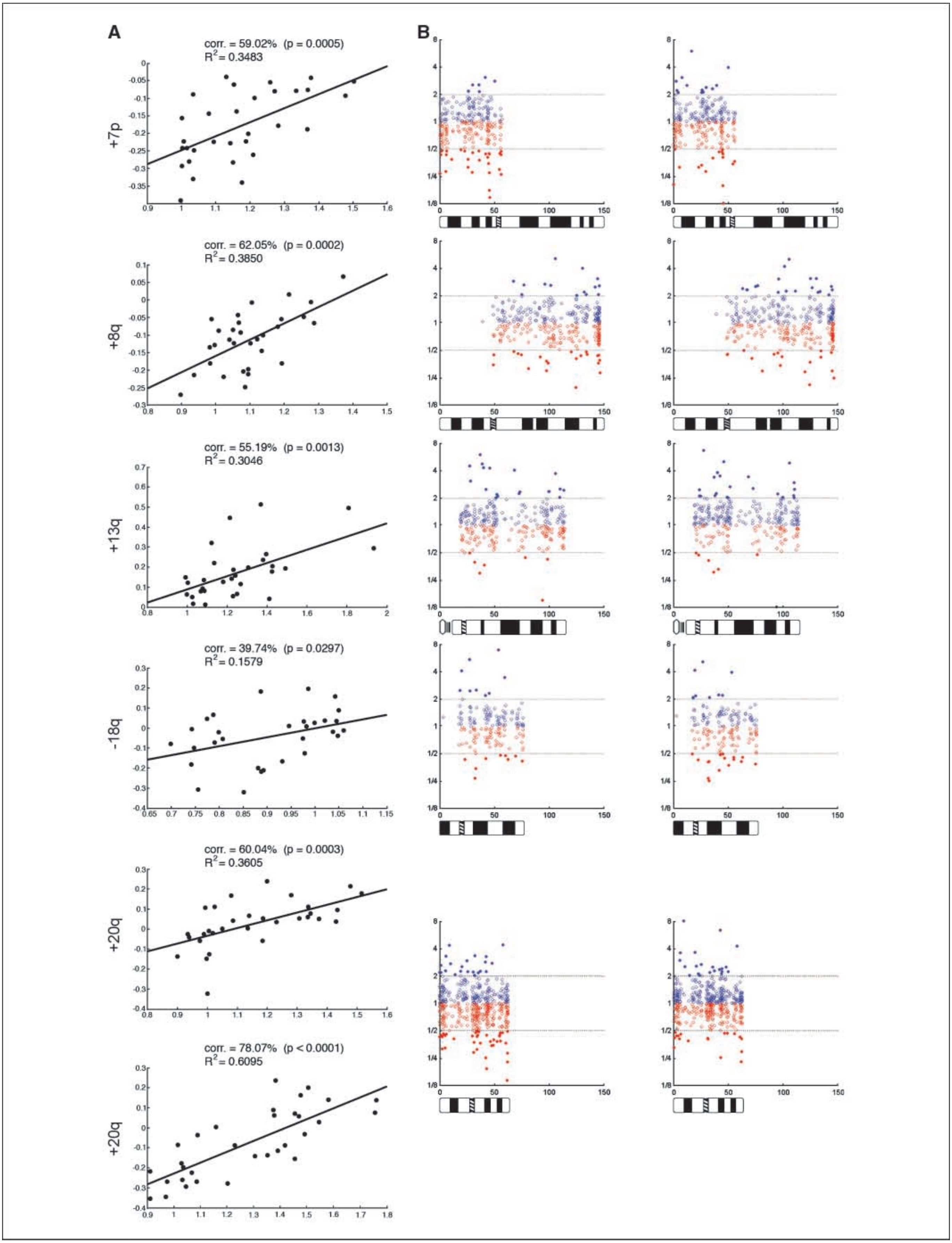
We were then curious to establish the similarity of gene expression changes between the rectum and colon for those genes previously reported to be involved in colorectal carcinogenesis (Table 2B), which included members of the canonical Wnt/β-catenin signaling pathway. We found that 81% (25 of 31) of the genes significantly deregulated in colon cancers ($P < 0.0001$) have a change in expression in the same direction in rectal carcinomas. Sixty-five percent (20 of 31) of those significantly deregulated in cancers of the rectum ($P < 0.0001$) had a change in the same direction in colon cancers. A subset of 15 genes was significantly deregulated in both the colon and rectal carcinomas, only one of which (*FLJ12529*) changed expression in opposite directions in the two data sets. In addition, 16 genes had significantly altered gene expression in one data set but showed either no change or an insignificant change in the opposite direction in the other data set, and 37 genes were not differentially expressed in either data set.

Discussion

We have focused here on the establishment of gene expression profiles of locally advanced colon carcinomas (UICC stages II and III). Our analyses were directed towards addressing four major clinically and biologically relevant questions: first, could we identify relevant genes that distinguish primary colon cancers and adjacent normal mucosa; second, could we detect differences in the gene expression profiles of primary tumors with and without lymph node metastases; third, how was the tumor transcriptome affected by specific chromosomal aneuploidies present in virtually all sporadic colon cancers; and fourth, how closely are primary colon cancers related to carcinomas of the rectum?

Comparison of colon cancer versus mucosa. The systematic comparison of gene expression profiles in primary colon cancer and normal colon mucosa revealed 4,371 differentially expressed genes at a significance of $P < 0.0001$. Using a significance value of $P < 1e-7$ as previously described for rectal carcinomas (16), 1,950 annotated genes were deregulated. When we applied as a selection criterion that the average expression levels had to be at least 5-fold different between cancer and mucosa, 17 genes were identified (Table 2A). Looking at this small subset of genes, we were reassured to find highly overexpressed genes that are known to be associated with malignant transformation, including *MYC*, whose role in colon carcinogenesis is well established. Interestingly, we also identified *HMGA1* as one of the highly up-regulated genes. *HMGA1* was the only gene that showed a 5-fold average difference between tumors and mucosa while always being >2-fold higher expressed in any given tumor compared with its associated mucosa. Of note, it also fulfilled the Bonferroni correction (i.e., it was significantly

Figure 4. Correlation of chromosomal copy numbers and resident gene expression levels. *A,* average CGH ratio value (*x-axis*) is plotted against the average gene expression value (*y-axis)* for each of 32 patients for whom we had done both analyses. The percentage correlation, its *P*s, and R^2 are indicated in each plot. The directionality of the copy number change is represented as a + (gain) or − (loss) preceding the chromosome number. *B,* the average expression of each gene along the length of the chromosome is plotted for those carcinomas without (*left*) and with (*right*) a copy number alteration. These plots correspond to the graphs in (*A*). *Blue,* genes with increased expression; *red,* genes with decreased expression relative to the reference RNA.


A
B
corr. = 59.02% (p = 0.0005)
R² = 0.3483
+7p
corr. = 62.05% (p = 0.0002)
R² = 0.3850
+8q
corr. = 55.19% (p = 0.0013)
R² = 0.3046
+13q
corr. = 39.74% (p = 0.0297)
R² = 0.1579
-18q
corr. = 60.04% (p = 0.0003)
R² = 0.3605
+20q
corr. = 78.07% (p < 0.0001)
R² = 0.6095
+20q

deregulated at a significance of $P < 1e-7$). As a member of the high mobility group family, *HMGA1* promotes tumor progression and metastasis and was previously reported to be highly expressed in colorectal cancers (39). Just recently, Frasca et al. showed that down-regulation of *HMGA1* via siRNA enhances the apoptotic pathway through reactivation of inactivated tumor suppressor genes, including *p53* (34).

We then systematically compared our list of 17 genes with published gene lists that were specifically derived from microarray experiments (4–6, 15, 16, 40–44), not including supplementary data sets. Only four genes were previously reported to be differentially expressed between colorectal cancers and normal epithelium. These genes include *MYC, IFITM1*, and *LCN2*, all of which were up-regulated in the tumors, consistent with our findings. Of those genes with higher expression in the tumors whose involvement in colon tumorigenesis was hitherto not known, we identified 13 genes involved in the maintenance of nucleosome structure (*HIST1H2BM* and *HIST1H2AJ*), cell division or proliferation (*RPL36A* and *RPS2*), and other cellular pathways. These genes can be considered potential novel diagnostic and therapeutic molecular targets. The independent validation of the expression levels of four of these genes (*SLC12A2, RPL36A, MYC*, and *HMGA1*) in a subset of 15 patients revealed concordant results between the array platform and RT-PCR–based analyses. The validation of other relevant genes, such as those involved in the Wnt/β-catenin signaling pathway, showed concordance for all nine genes.

When we did a functional annotation of our 17 genes using the IPA software, we were intrigued by the degree of coherence of the deregulated genes. In contrast to our previous analysis of rectal carcinomas (16), the 10 genes with >5-fold average deregulation present in the IPA knowledge bank were all included in one network (Fig. 1). *MYC* appears at a central integrating position, which attests once more to the dominant role of this gene in colon carcinogenesis. Of the 25 additional genes that were part of this network, yet not included in our list of 17 genes, 12 were significantly deregulated ($P < 0.0001$), whereas the remaining 11 were not. One of these genes (*PLF2*) was not spotted on the array, and *TNF* did not fulfill our filtering criteria. As expected, *CTNNB1* (β-catenin) showed significant up-regulation in the tumors.

The analysis of the normal rectal mucosa samples that we recently described (16) revealed a clear separation of the normal mucosa samples into two distinct classes, for which we could not identify an obvious explanation, such as, for example, proximity of the sampling area to the primary tumor. We were therefore curious as to whether this phenomenon also surfaced in our normal colon mucosa samples. Although the separation of colon cancer samples and normal mucosa was very stringent, the normal mucosa samples were all grouped in one MDA cluster (Fig. 3*B*). Therefore, only one class of normal mucosa samples was observed, and the clustering into two distinct groups might just reflect an idiosyncrasy of the rectal mucosa samples.

Comparison of lymph node–negative and lymph node–positive colon cancers. Here, we aimed to determine whether the propensity for the development of lymph node metastases is an inherent feature of the primary tumors and could therefore be unveiled using gene expression profiling. It was for this reason that we deliberately focused on UICC stage II and UICC stage III carcinomas, whose main discerning feature is the lymph node status (82% of all 73 tumors included here belonged to T category 3). We enriched this sample selection for T category 3 tumors because we surmised that this selection would result in the highest probability of identifying the gene expression signature of lymphatic metastases, should there be any. The analysis was therefore not confounded by potential gene expression differences attributable to different T categories.

Seventy-four spotted array features showed significantly different expression values ($P < 0.001$) between the lymph node–positive and lymph node–negative tumors, of which 68 represented annotated genes. All but five of these genes were down-regulated in the node-positive tumors. Functional annotation of these genes using IPA suggested a preponderance of genes involved in immune response based on IL-8 signaling, cell motility, and posttranslational modifications. Taken together, these results would be in general compatible with the interpretation that pathways of immune surveillance, cell motility, and apoptosis are differentially regulated in UICC stage II and UICC stage III tumors. Activation of the immune response, accompanied by an enhancement of the apoptotic machinery, could therefore synergize to reduce the likelihood for lymphatic metastasis. Although we believe this to be an intriguing and reasonable interpretation of our results, we realize that further functional investigations of the involvement of these genes remain to be done. However, we validated the gene expression differences for *IL8, UBD, ICAM1*, and *PPM1G*, which were consistent between the platforms. *IL15RA* and *PSMC4* showed inverse expression levels when comparing the microarray and RT-PCR results. When we surveyed the literature for studies aimed at identifying differentially expressed genes between node-negative and node-positive colorectal carcinomas (9, 13, 14), none of our identified 68 genes was previously reported. This lack of overlap is likely due to different patient selection criteria or differences in array platforms and analytic strategies.

The pretherapeutic assessment of the lymph node status of colon carcinomas is of little clinical effect because it would not influence the treatment choice. From a tumor biological point of view, however, it still remains a matter of debate if the genetic make up of a solid tumor determines its metastatic potential. Based on a class prediction analysis of the primary tumors, we were not able to reliably distinguish between those tumors from which cells had infiltrated to the lymph node and those that had not. We are thus reluctant in extrapolating that one can readily determine by looking solely at the primary tumor whether or not it is associated with synchronous lymph node metastases. Thus, it is possible that any subsequent change to a cell in the primary tumor allowing it to metastasize is carried away with the cell as it migrates into the periphery and is therefore not observed. Not mutually exclusive is the possibility that only a small pool, if any, of cells with metastatic potential persist in the primary tumor. It therefore remains to be determined whether the capability of a primary tumor to metastasize would require additional mutations, or whether this capability would be engrained in its specific gene expression profiles. This hypothesis has been put forward based on studies using gene expression profiles in different solid tumors (45). However, this supposition, which would contradict the more generally accepted dogma that metastasis requires additional mutations followed by clonal selection and expansion, did not remain uncontested (46, 47).

Consequences of chromosomal aneuploidy on resident gene expression levels. Specific chromosomal aneuploidies are the defining feature of epithelial cancers (37, 48, 49). Numerous studies, using cytogenetic and molecular cytogenetic techniques, have established that colorectal tumorigenesis is invariably accompanied (or in fact caused) by the acquisition and maintenance of chromosomes and chromosome arms 7, 8q, 13q, and 20 and losses that map to 4q, 8p, 17p, and 18q (35, 37, 50). The tumors included in our sample collection here are no exception. Only with the advent of methods for parallel gene expression profiling has it become possible to identify the consequences of these dominant genetic aberrations on the global tumor transcriptome. Understanding the direct role of genomic imbalances on resident gene expression levels is, of course, paramount to the understanding of basic characteristics of tumor biology: the effects of these aneuploidies could range from the deregulation of just a few candidate genes on these chromosomes to global changes of the transcriptional equilibrium of most or all of the resident genes, which would result in an enormous amount of altered message.

We and others have therefore conducted analyses that allowed for the simultaneous mapping of genomic copy number changes and average gene expression levels in colorectal cancers and model systems thereof (11, 16, 17, 38). These studies now suggest that chromosomal aneuploidies, and thus genomic imbalances, result in an alteration of transcriptional activity that is correlated to the variation in genomic copy number. For instance, the introduction of extra copies of chromosome 7 in the karyotypically stable colon cancer cell line DLD1 significantly increased the average gene expression levels of genes on that chromosome by a factor of 1.25. A similar picture emerged in rectal carcinomas (16) and, recently, in colorectal cancer (17). The results from our analysis revealed a statistically significant positive correlation between gene expression levels and chromosomal copy numbers (Fig. 4*A* and *B*), thereby supporting these previous interpretations. We can thus conclude that genomic imbalances contribute to a massive, aneuploidy-dependent deregulation of global transcriptional activities in colon and rectal carcinomas. The relative role of these changes for the acquisition and maintenance of the malignant phenotype vis-à-vis the activation or inactivation of specific oncogenes and tumor suppressor genes remains to be established.

Comparison of rectal and colon cancers. We conducted a comprehensive comparison of gene expression changes in rectal and colon carcinomas. First, we matched the lists of differentially expressed genes and discovered a significant resemblance of transcriptional deregulation. After removing duplicate genes and probes that do not correspond to known genes, 1,374 genes were deregulated in rectal cancers, whereas 2,978 genes were altered in expression in colon cancers compared with their respective normal epithelium ($P < 0.0001$). Of the 490 genes common among these lists (Supplementary Table S5), 80% ($n = 394$) were deregulated in the same direction, which is significant ($P = 0.001$). Second, we identified a high overlap for those genes previously reported to be involved in colorectal carcinogenesis or the canonical Wnt/β-catenin signaling pathway (Table 2B). Of note, 81% (25 of 31) of the genes significantly deregulated in the colon ($P < 0.0001$) were deregulated in the same direction in the rectum, whereas 65% (20 of 31) of those significantly deregulated in the rectum ($P < 0.0001$) were deregulated in the same direction in the colon. Fourteen genes were significantly deregulated in the same direction in both the colon and rectal carcinomas. In conclusion, this suggests a marked similarity of transcriptional deregulation in tumors arising in these distinct anatomic locations. These results corroborate our interpretation that rectal and colonic carcinogenesis requires, in general, deregulation of similar genetic pathways.

However, there seem to be some discrepancies in the details of how these pathways are altered. The foremost is that the expression level of *PTGS2* (*COX-2*), which was significantly up-regulated in the rectal carcinomas, did not seem to be affected at the transcriptional level in the colon tumors. This could possibly be due to alternative functional up-regulation in colonic cancer induced by, for example, posttranslational modification. Three downstream genes affected by *PTGS2* through prostaglandin E_2 synthesis, however, were similarly affected in the colon and the rectum. That is, neither *BCL2* nor *EGFR* was increased, but *VEGF* was ($P < 1e-7$ in both data sets), again implying that signaling through this pathway in colorectal tumors is geared towards increased vascularization and not cell survival or increased proliferation. Very interestingly, we again observed a highly significant up-regulation ($P < 1e-7$) of *GSK3B* in the colon cancers. This was verified by RT-PCR and is now the second example that the gene encoding this β-catenin degradation complex member is strongly overexpressed (16). Other members of this complex (*Axin1* and *Axin2*) also had increased expression levels, again at odds with the idea that the Wnt/β-catenin signaling pathway is activated.

That being said, an examination of β-catenin expression itself revealed a significant increase (1.87, $P < 1e-7$). Additional supporting evidence that the Wnt/β-catenin signaling pathway is activated in these colon carcinomas comes from an analysis of downstream targets, such as *Sox9* (1.65, $P < 1e-7$), *CCND1* (2.20, $P < 4e-7$), *CD44* (2.58, $P < 1e-7$), *EPHB2* (2.01, $P = 0.0001$), *VEGF* (1.94, $P < 1e-7$), *CTBP1* (1.41, $P = 1e-7$), and *MYC* (5.49, $P < 1e-7$). The central role of *MYC* activation was confirmed. Not only did this gene assume an integrating position in the IPA networks, but its deadly message was also significantly up-regulated in both rectal and colon cancers ($P < 0.0001$ and $P < 1e-7$, respectively). It would therefore be intriguing to explore novel avenues to target *MYC* therapeutically.

Acknowledgments

Received 5/1/2006; revised 10/2/2006; accepted 10/26/2006.

Grant support: Intramural research program of the NIH, National Cancer Institute.

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank Buddy Chen for IT support and assistance with figures and tables and Dr. L. Füzesi for pathology reports and sample requisition.

References

1. Ladanyi M, Gerald WL. Expression profiling of human tumors: diagnostic and research applications. Totowa (NJ): Humana Press; 2003.
2. Chung CH, Bernard PS, Perou CM. Molecular portraits and the family tree of cancer. Nat Genet 2002; 32:533–40.
3. Ludwig JA, Weinstein JN. Biomarkers in cancer staging, prognosis and treatment selection. Nat Rev Cancer 2005;5:845–56.
4. Alon U, Barkai N, Notterman DA, et al. Broad patterns of gene expression revealed by clustering analysis of tumor and normal colon tissues probed by oligonucleotide arrays. Proc Natl Acad Sci U S A 1999; 96:6745–50.
5. Notterman DA, Alon U, Sierk AJ, Levine AJ. Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. Cancer Res 2001;61:3124–30.
6. Kitahara O, Furukawa Y, Tanaka T, et al. Alterations of gene expression during colorectal carcinogenesis revealed by cDNA microarrays after laser-capture microdissection of tumor tissues and normal epithelia. Cancer Res 2001;61:3544–9.
7. Lechner S, Muller-Ladner U, Renke B, Scholmerich J, Ruschoff J, Kullmann F. Gene expression pattern of laser

microdissected colonic crypts of adenomas with low grade dysplasia. Gut 2003;52:1148–53.
8. Nosho K, Yamamoto H, Adachi Y, Endo T, Hinoda Y, Imai K. Gene expression profiling of colorectal adenomas and early invasive carcinomas by cDNA array analysis. Br J Cancer 2005;92:1193–200.
9. 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;47:141–52.
10. Li M, Lin YM, Hasegawa S, et al. Genes associated with liver metastasis of colon cancer, identified by genome-wide cDNA microarray. Int J Oncol 2004;24:305–12.
11. Platzer P, Upender MB, Wilson K, et al. Silence of chromosomal amplifications in colon cancer. Cancer Res 2002;62:1134–8.
12. D'Arrigo A, Belluco C, Ambrosi A, et al. Metastatic transcriptional pattern revealed by gene expression profiling in primary colorectal carcinoma. Int J Cancer 2005;115:256–62.
13. Croner RS, Peters A, Brueckl WM, et al. Microarray versus conventional prediction of lymph node metastasis in colorectal carcinoma. Cancer 2005;104:395–404.
14. Frederiksen CM, Knudsen S, Laurberg S, Orntoft TF. Classification of Dukes' B and C colorectal cancers using expression arrays. J Cancer Res Clin Oncol 2003;129:263–71.
15. Koehler A, Bataille F, Schmid C, et al. Gene expression profiling of colorectal cancer and metastases divides tumours according to their clinicopathological stage. J Pathol 2004;204:65–74.
16. Grade M, Ghadimi BM, Varma S, et al. Aneuploidy-dependent massive deregulation of the cellular transcriptome and apparent divergence of the Wnt/β-catenin signaling pathway in human rectal carcinomas. Cancer Res 2006;66:267–82.
17. Tsafrir D, Bacolod M, Selvanayagam Z, et al. Relationship of gene expression and chromosomal abnormalities in colorectal cancer. Cancer Res 2006;66:2129–37.
18. Mariadason JM, Arango D, Shi Q, et al. Gene expression profiling-based prediction of response of colon carcinoma cells to 5-fluorouracil and camptothecin. Cancer Res 2003;63:8791–812.
19. 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;23:1826–38.
20. Schmidt WM, Kalipciyan M, Dornstauder E, et al. Gene expression profiling of colon cancer reveals a broad molecular repertoire in 5-fluorouracil resistance. Int J Clin Pharmacol Ther 2003;41:624–5.
21. de Angelis PM, Fjell B, Kravik KL, et al. Molecular characterizations of derivatives of HCT116 colorectal cancer cells that are resistant to the chemotherapeutic agent 5-fluorouracil. Int J Oncol 2004;24:1279–88.
22. Wang Y, Jatkoe T, Zhang Y, et al. Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. J Clin Oncol 2004;22:1564–71.
23. Arango D, Laiho P, Kokko A, et al. Gene-expression profiling predicts recurrence in Dukes' C colorectal cancer. Gastroenterology 2005;129:874–84.
24. Shih W, Chetty R, Tsao MS. Expression profiling by microarrays in colorectal cancer (Review). Oncol Rep 2005;13:517–24.
25. Bustin SA, Dorudi S. Gene expression profiling for molecular staging and prognosis prediction in colorectal cancer. Expert Rev Mol Diagn 2004;4:599–607.
26. Augenlicht LH, Velcich A, Klampfer L, et al. Application of gene expression profiling to colon cell maturation, transformation and chemoprevention. J Nutr 2003;133:2410–6S.
27. Sobin LH, Wittekind C. UICC: TNM classification of malignant tumors. 5th ed. New York: John Wiley & Sons; 1997.
28. Ghadimi BM, Sackett DL, Difilippantonio MJ, et al. Centrosome amplification and instability occurs exclusively in aneuploid, but not in diploid colorectal cancer cell lines, and correlates with numerical chromosomal aberrations. Genes Chromosomes Cancer 2000;27:183–90.
29. Wright GW, Simon RM. A random variance model for detection of differential gene expression in small microarray experiments. Bioinformatics 2003;19:2448–55.
30. Dudoit S, Fridlyand J. A prediction-based resampling method for estimating the number of clusters in a dataset. Genome Biol 2002;3:RESEARCH0036.
31. Dudoit S, Fridlyand J, Speed TP. Comparison of discrimination methods for the classification of tumors using gene expression data. J Am Stat Assoc 2002;97:77–87.
32. Ntzani EE, Ioannidis JP. Predictive ability of DNA microarrays for cancer outcomes and correlates: an empirical assessment. Lancet 2003;362:1439–44.
33. Simon R, Radmacher MD, Dobbin K, McShane LM. Pitfalls in the use of DNA microarray data for diagnostic and prognostic classification. J Natl Cancer Inst 2003;95:14–8.
34. Frasca F, Rustighi A, Malaguarnera R, et al. HMGA1 inhibits the function of p53 family members in thyroid cancer cells. Cancer Res 2006;66:2980–9.
35. Ried T, Knutzen R, Steinbeck R, et al. Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. Genes Chromosomes Cancer 1996;15:234–45.
36. Ghadimi BM, Grade M, Liersch T, et al. Gain of chromosome 8q23–24 is a predictive marker for lymph node positivity in colorectal cancer. Clin Cancer Res 2003;9:1808–14.
37. Grade M, Becker H, Liersch T, Ried T, Ghadimi BM. Molecular cytogenetics: genomic imbalances in colorectal cancer and their clinical impact. Cell Oncol 2006;28:71–84.
38. Upender MB, Habermann JK, McShane LM, et al. Chromosome transfer induced aneuploidy results in complex dysregulation of the cellular transcriptome in immortalized and cancer cells. Cancer Res 2004;64:6941–9.
39. Fedele M, Bandiera A, Chiappetta G, et al. Human colorectal carcinomas express high levels of high mobility group HMGI(Y) proteins. Cancer Res 1996;56:1896–901.
40. Clarke PA, George ML, Easdale S, et al. Molecular pharmacology of cancer therapy in human colorectal cancer by gene expression profiling. Cancer Res 2003;63:6855–63.
41. Birkenkamp-Demtroder K, Christensen LL, Olesen SH, et al. Gene expression in colorectal cancer. Cancer Res 2002;62:4352–63.
42. Zou TT, Selaru FM, Xu Y, et al. Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon. Oncogene 2002;21:4855–62.
43. Zhang L, Zhou W, Velculescu VE, et al. Gene expression profiles in normal and cancer cells. Science 1997;276:1268–72.
44. Agrawal D, Chen T, Irby R, et al. Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling. J Natl Cancer Inst 2002;94:513–21.
45. Ramaswamy S, Ross KN, Lander ES, Golub TR. A molecular signature of metastasis in primary solid tumors. Nat Genet 2003;33:49–54.
46. Fidler IJ, Kripke ML. Genomic analysis of primary tumors does not address the prevalence of metastatic cells in the population. Nat Genet 2003;34:49–54.
47. Hunter K, Welch DR, Liu ET. Genetic background is an important determinant of metastatic potential. Nat Genet 2003;34:23–54.
48. Heim S, Mitelman F. Cancer cytogenetics. 2nd ed. New York: Wiley-Liss; 1995.
49. Ried T, Heselmeyer-Haddad K, Blegen H, Schrock E, Auer G. Genomic changes defining the genesis, progression, and malignancy potential in solid human tumors: a phenotype/genotype correlation. Genes Chromosomes Cancer 1999;25:195–204.
50. Bardi G, Sukhikh T, Pandis N, Fenger C, Kronborg O, Heim S. Karyotypic characterization of colorectal adenocarcinomas. Genes Chromosomes Cancer 1995;12:97–109.