Stage-Specific Alterations of the Genome, Transcriptome, and Proteome During Colorectal Carcinogenesis

Jens K. Habermann,[1,2*] Ulrike Paulsen,[1,3] Uwe J. Roblick,[2,3] Madhvi B. Upender,[1] Lisa M. McShane,[4] Edward L. Korn,[4] Danny Wangsa,[1] Stefan Krüger,[5] Michael Duchrow,[3] Hans-Peter Bruch,[3] Gert Auer,[2] and Thomas Ried[1]

[1]Genetics Branch, National Cancer Institute, NIH, Bethesda, MD
[2]Unit of Cancer Proteomics, Department of Oncology and Pathology, Karolinska Institutet, Stockholm, Sweden
[3]Department of Surgery, University of Schleswig-Holstein, Campus Lübeck, Germany
[4]Biometric Research Branch, National Cancer Institute, NIH, Bethesda, MD
[5]Institute of Pathology, University of Schleswig-Holstein, Campus Lübeck, Germany

To identify sequential alterations of the genome, transcriptome, and proteome during colorectal cancer progression, we have analyzed tissue samples from 36 patients, including the complete mucosa-adenoma-carcinoma sequence from 8 patients. Comparative genomic hybridization (CGH) revealed patterns of stage specific, recurrent genomic imbalances. Gene expression analysis on 9K cDNA arrays identified 58 genes differentially expressed between normal mucosa and adenoma, 116 genes between adenoma and carcinoma, and 158 genes between primary carcinoma and liver metastasis ($P < 0.001$). Parallel analysis of our samples by CGH and expression profiling revealed a direct correlation of chromosomal copy number changes with chromosome-specific average gene expression levels. Protein expression was analyzed by two-dimensional gel electrophoresis and subsequent mass spectrometry. Although there was no direct match of differentially expressed proteins and genes, the majority of them belonged to identical pathways or networks. In conclusion, increasing genomic instability and a recurrent pattern of chromosomal imbalances as well as specific gene and protein expression changes correlate with distinct stages of colorectal cancer progression. Chromosomal aneuploidies directly affect average resident gene expression levels, thereby contributing to a massive deregulation of the cellular transcriptome. The identification of novel genes and proteins might deliver molecular targets for diagnostic and therapeutic interventions. This article contains Supplementary Material available at http://www.interscience.wiley.com/jpages/1045-2257/suppmat.

INTRODUCTION

The progression of colorectal cancer is defined by the sequential acquisition of genetic alterations (Fearon and Vogelstein, 1990). On the cytogenetic level, many of these aberrations can be visualized as specific chromosomal gains and losses. These aneuploidies result in a recurrent pattern of genomic imbalances, which is specific for these tumors and strictly conserved (Ried et al., 1996). For instance, one of the earliest acquired genetic abnormalities during colorectal tumorigenesis are copy number gains of chromosome 7 (Bomme et al., 1994). These trisomies can already be observed in benign polyps and can emerge in otherwise stable, diploid genomes. At later stages, e.g., in adenomas with high-grade dysplasias or in invasive carcinomas, additional specific cytogenetic abnormalities become common, such as gains of chromosome and chromosome arms 8q, 13, and 20q, and losses that map to 8p, 17p, and 18q (http://cgap.nci.nih.gov/Chromosomes/Mitelman). These chromosomal aneuploidies are accompanied by specific mutations in oncogenes and tumor suppressor genes, including *KRAS*, *APC*, and *TP53* (Vogelstein and Kinzler, 2004). It is therefore likely that both chromosomal aneuploidies and specific gene mutations contribute to tumorigenesis. However, it remains yet to be explored which systematic, global alterations of the cellular transcriptome are required for the sequential steps of cellular immortalization and transformation. The immediate consequences of chromosomal aneuploidies remain elusive vis-a-vis the deregulation of the cellular transcriptome during the genesis of colorectal cancer. Available methodology for parallel gene

Supported by: Swedish Cancer Society (Cancerfonden); Cancer Society (Cancerföreningen), Sweden; Intramural Research Program of the NIH, National Cancer Institute.

*Correspondence to: Jens K. Habermann, Genetics Branch, Center for Cancer Research, NCI, NIH, 50 South Drive, Rm 1408, Bethesda, MD 20892-8010, USA. E-mail: habermaj@mail.nih.gov

Received 9 May 2006; Accepted 29 August 2006

DOI 10.1002/gcc.20382

Published online 16 October 2006 in Wiley InterScience (www.interscience.wiley.com).

expression profiling now allows one to address these questions (Schena et al., 1996). Similarly, proteomic techniques have rapidly evolved and have the potential to be used to monitor disease specific alterations (Misek et al., 2004). For instance, two-dimensional gel electrophoresis (2DE) can be used to identify differentially expressed peptides and subsequent mass spectrometry can then be used for protein identification (Oppermann et al., 2000). Such approaches enable discernment of tumor specific quantitative or qualitative changes of protein patterns. In an attempt to, first, identify gene expression patterns and genetic pathways associated with malignant transformation, second, analyze how precisely chromosomal aneuploidies affect resident gene expression levels, and third, discover novel diagnostically and therapeutically relevant biomarkers of colon cancer progression, we have mapped genomic imbalances and associated gene and protein expression changes in a series of well-characterized, microdissected surgical specimens from colon cancer patients. In several cases, multiple tissue samples were collected at various disease stages from the same patients.

MATERIALS AND METHODS

Clinical Samples

The study is based on colorectal adenocarcinoma specimens from 36 patients (20 male, 16 female). None of these patients had signs of hereditary predisposition for colorectal cancer (e.g., FAP, HNPCC). The tumors were diagnosed at the University Hospital Schleswig-Holstein, Campus Lübeck, Germany. Tissue samples were collected during surgery from different stages of the adenoma-carcinoma-sequence comprising 16 normal mucosa specimen, 17 adenomas, 20 primary sporadic colorectal carcinomas, and 13 liver metastases adhering to the guidelines of the local ethical review board. None of the patients received neoadjuvant therapy prior to operation. We were successful in collecting the complete sequence of normal mucosa-adenoma-carcinoma for eight individual patients, paired normal mucosa and carcinoma samples from an additional eight patients, and paired carcinoma and metastasis from two patients. All samples were stained with hematoxylin & eosin and re-evaluated by one experienced pathologist (S.K.). Tumor staging was done according to the UICC system and the histological differentiation was assessed according to the WHO classification. Before further processing, all normal mucosa samples and adenomas were microdissected. Normal mucosa was taken from the colectomy specimen in secure distance to the present carcinoma by the pathologist. According to the assessment of the pathologist, the obtained normal mucosa did not harbor adenomatous or cancer tissue. Each sample was divided into two parts: one for simultaneous DNA and RNA extraction, and one for protein extraction. For protein extraction, a subfraction of the obtained samples was checked for the amount of representative tumor or epithelial cells, respectively, by Giemsa staining. Samples were used only when containing at least 80% tumor cells. The main fraction of cells was then used for protein extraction and the protein pellets were frozen at -80° until further processing for 2DE. The clinical data are summarized in Table 1.

Image Cytometry

Image cytometry was performed on Feulgen-stained histological sections. The staining procedure, internal standardization, tumor cell selection, and analysis were based on methods described previously (Auer et al., 1980). In brief, all DNA-values were expressed in relation to the corresponding staining controls, which were given the value 2c, denoting the normal diploid DNA-content. All specimens were divided into two main groups: (i) diploid cases with a distinct peak in the normal 2c region and no cells exceeding 5c, (ii) aneuploid cases with a main peak around the 4c region and varying numbers of cells (>5%) exceeding 5c.

Comparative Genomic Hybridization

DNA was extracted from fresh frozen tissue subsequent to RNA extraction using Trizol (Invitrogen, Carlsbad, CA) according to the manufacturer's protocol. Quality of DNA and RNA was assessed by gel electrophoresis and spectrophotometric analysis prior to comparative genomic hybridization (CGH) or expression array analysis.

The DNA was then labeled by nick-translation using biotin-11-dUTP (Roche Diagnostic Corporation, Indianapolis, IN) as described previously (Ried et al., 1996). Genomic DNA from cytogenetically normal individuals was labeled with digoxigenin-12-dUTP (Roche Diagnostic Corporation, Indianapolis, IN) and was used as control. Simultaneous hybridization of the test and control DNA was performed on karyotypically normal metaphase chromosomes using an excess of Cot1-DNA (Invitrogen). The biotin-labeled sequences were visualized with avidin-FITC (Vector Laboratories, Burlingame, CA) and the digoxigenin la-

TABLE 1. Summary of Clinical and Experimental Parameters of 36 Patients with Sporadic Colorectal Carcinoma

Case	Age	Sex	Surgery	Histology (TNM, Grading)	UICC stage	Location	Lymph. Met.	Survival	Died	CGH	Array	2-DE
1	61	M	R0	Normal mucosa				50 month	Alive	3695	No	Yes
				Adenoma						3708	Yes	Yes
				Adenoma (D3, aneuploid)						No	Yes	Yes
				Carcinoma (T3N0M0, G3)	II	Rectum	0/18			No	No	Yes
2	75	F	R0	Adenoma		Ascendence		40 month	Alive	3712	Yes	No
				Adenoma		Ascendence				3713	Yes	No
				Carcinoma (T3N0M0, G2)	II	Caecum	0/42			3729	Yes	No
3	78	M	R0	Normal mucosa				17 month	Yes	3692	No	Yes
				Adenoma (D3, aneuploid)						3706	No	Yes
				Carcinoma (TisN0M1, in adenoma)	IV	Sigma	1/27			3722	Yes	Yes
				Carcinoma (T4N1M1, G3)	IV	Transversum				3721	Yes	Yes
4	60	M	R1	Normal mucosa				4 month	Yes	No	No	Yes
				Adenoma						3707	Yes	Yes
				Carcinoma (T3N0M0, G1)	II	Ascendence	0/16			3723	No	Yes
5	74	M	R0	Normal mucosa				41 month	Alive	3696	No	Yes
				Adenoma						3709	No	Yes
				Carcinoma (T3N1M0, G3)	III	Right Flexur	1/27			3727	Yes	Yes
6	71	F	R0	Normal mucosa				48 month	Alive	3690	Yes	Yes
				Adenoma (tubular, D1)						No	No	Yes
				Carcinoma (T3N0M0, G2)	II	Right Flexur	0/21			3717	Yes	Yes
7	77	M	R0	Normal mucosa				42 month	Alive	No	No	Yes
				Adenoma						No	No	Yes
				Carcinoma (T2N1M0, G3)	III	Rectum	1/16			3720	Yes	Yes
8	76	M	R0	Normal mucosa				21 month	Alive	3699	No	Yes
				Adenoma						3711	Yes	Yes
				Carcinoma (T3N1M1, G3)	IV	Rectum	2/19			No	No	Yes
9	58	M	R0	Normal mucosa				30 month	Alive	3700	No	No
				Adenoma						No	Yes	No
				Carcinoma (T3N0M0, G2)	II	Sigma	0/13			3730	Yes	No
10	78	F	R0	Normal mucosa				45 month	Yes	3688	No	Yes
				Carcinoma (T3N1M0, G2)	III	Rectum	3/19			3716	Yes	Yes
11	65	F	R0	Normal mucosa				7 month	Yes	3691	No	Yes
				Carcinoma (T3N2M1, G3)	IV	Ascendence	11/15			3718	Yes	Yes
12	59	M	R0	Normal mucosa				5 month	Yes	No	No	Yes
				Carcinoma (T3N2M1, G2)	IV	Sigma	5/14			3719	Yes	Yes
13	71	F	R0	Normal mucosa				52 month	Alive	No	No	Yes
				Carcinoma (T2N0M0, G2)	I	Caecum	0/12			No	No	Yes
14	76	M	R0	Normal mucosa				30 month	Yes	3693	Yes	Yes
				Carcinoma (T3N1M1, G2)	IV	Rectum	3/19			3725	Yes	Yes
15	75	F	R0	Normal mucosa				12 month	Yes	3694	No	Yes
				Carcinoma (T4N1M1, G2)	IV	Rectum	2/18			3726	Yes	Yes

(Continued)

TABLE 1. Summary of Clinical and Experimental Parameters of 36 Patients with Sporadic Colorectal Carcinoma (Continued)

Case	Age	Sex	Surgery	Histology (TNM, Grading)	UICC stage	Location	Lymph. Met.	Survival	Died	CGH	Array	2-DE
16	64	M	R0	Normal mucosa				17 month	Yes	No	Yes	No
				Carcinoma (T3N2M1, G3)	IV	Sigma	4/10			No	Yes	No
17	81	M	R1	Normal mucosa				1 month	Yes	3687	Yes	Yes
				Carcinoma (T3N1M1, G3)	IV	Ascendence	1/15			3715	No	Yes
				Metastasis	IV	Liver				3737	No	Yes
18	76	M	R0	Carcinoma (T3N2M1, G3)	IV	Sigma	4/14	42 month	Yes	3714	Yes	Yes
				Metastasis	IV	Liver				3732	Yes	Yes
19	60	M		Adenoma (tubular)						3701	No	No
20	71	F		Adenoma (tubular-villous)						3702	No	No
21	55	F		Adenoma (tubular)						3703	No	No
22	68	M		Adenoma (tubular-villous)						3704	No	No
23	64	F		Adenoma (tubular-villous)						3705	No	No
24	43	M		Adenoma						3710	Yes	No
25	71	M	R0	Carcinoma (T3N1M0, G2)	III	Rectum	1/24	52 month	Alive	3724	No	No
26	72	F	R1	Metastasis (T0N0M1, G3)	IV	Liver		66 month	Alive	3731	No	No
27	62	F	R0	Metastasis (T0N0M1, G3)	IV	Liver		61 month	Alive	No	Yes	No
28	63	F	R1	Metastasis (T0NXM1, G2)	IV	Liver		43 month	Yes	3734	No	No
29	63	M	R0	Metastasis (T0N0M1, G2)	IV	Liver		27 month	Yes	3735	No	No
30	84	M	R0	Metastasis (T0N0M1, G3)	IV	Liver		48 month	Alive	3736	Yes	No
31	64	F	R0	Metastasis (T0N0M1, G3)	IV	Liver		40 month	Alive	3738	No	No
32	63	M	R0	Metastasis (T0N0M1, G2)	IV	Liver		10 month	Yes	3739	Yes	No
33	65	F	R2	Metastasis (T3N2M1, G3)	IV	Liver		7 month	Yes	3740	No	No
34	38	F	R0	Metastasis (T0N0M1, G2)	IV	Liver		36month	Alive	3741	Yes	No
35	54	M	R2	Metastasis (T0N0M1, G2)	IV	Liver		28 month	Alive	3742	Yes	No
36	58	F	R0	Metastasis (T0N0M1, GX)	IV	Liver		34 month	Alive	3743	Yes	No

Abbreviations used: F, female; M, male; Lymph. Met., lymph node metastasis; R, surgical resection; R0, total resection of tumor (curative resection); R1, incomplete tumor resection; CGH, if CGH could be performed the case annotation of the SKY-M-FISH-CGH database is given (http://www.ncbi.nlm.nih.gov/sky/).

beled sequences were detected with a mouse-derived antibody against digoxigenin followed by a secondary rhodamine conjugated anti-mouse antibody (Sigma–Aldrich, Milwaukee, WI). Detailed protocols can be retrieved from http://riedlab.nci.nih.gov. Quantitative fluorescence imaging and CGH analysis was performed using Leica CW4000 Karyo V1.0 software (Leica Imaging Systems, Cambridge, UK). Interpretation of changes at chromosome regions 1pter and chromosomes 16, 19, and 22 required careful examination because these loci are prone to artifacts due to the high proportion of repetitive sequences. CGH profiles of all cases as well as CGH comparison tools can be found at http://www.ncbi.nlm.nih.gov/sky/skyweb.cgi.

Microarray Analysis

Total RNA was extracted using Trizol (Invitrogen, Carlsbad, CA) followed by Qiagen RNeasy column purification (Qiagen, Valencia, CA). All samples were hybridized against human reference RNA of the same lot (Stratagene, La Jolla, CA; lot no. 0810006) using a slightly modified protocol from Hedge et al. (2000). Extraction and hybridization protocols can be viewed in detail at http://www.riedlab.nci.nih.gov. In brief, 20 μg of total RNA was reverse transcribed using random primers and converted into cDNA using reverse transcriptase. After incorporation of aminoallyl-conjugated nucleotides, the RNA was indirectly labeled with Cy3 (tumor RNA) and Cy5 (reference RNA) (Amersham, Piscataway, NJ). Each sample was hybridized against the reference RNA in a humid chamber (Arraylt^{TM} Hybridization Cassette, TeleChem Intl., Sunnyvale, CA) for 16 hr at 42°C, washed, and scanned using an Axon GenePix 4000B Scanner (Axon Instruments, Union City, CA). We used custom-prepared arrays obtained from the National Cancer Institute's Advanced Technology Center (ATC). All arrays contained 9,128 cDNAs denatured and immobilized on a poly-L-lysine-coated glass surface, covering on average 28.8% of the total number of genes localized on any particular chromosome arm. All hybridizations were performed on one print batch. The gene annotation file for these arrays is Hs-UniGEM2-v2px-32Bx18Cx18R.gal. GenePix software version 4.0.1.17 was used to apply the GAL file through an interactive gridding process: all images of the scanned microarray slides were meticulously inspected for artifacts. Empty spots and aberrant spots and slide regions were flagged for exclusion from analyses (Korn et al., 2004). Based on RNA quality assessment and hybridization quality controls, arrays of 3 normal mucosa samples, 8 adenomas, 15 carcinomas, and 7 liver metastases were used to identify differentially expressed genes. Log ratios for each spot were calculated as follows: in each channel, the signal was calculated as foreground mean minus background median. If the signal was less than 100 in any single channel, the signal value in that channel was set to 100. If the signal was less than 100 in both channels, the spot was flagged as unreliable and not used in any further analyses. For all remaining (nonflagged) spots, a log ratio was calculated as $\log_2$[(green signal)/(red signal)]. M versus A plots were constructed with vertical (y) axis corresponding to the log ratio and average of the log red and green signals on the horizontal (x) axis. An intensity-based normalization curve was estimated by applying a locally weighted regression (lowess) smoother to the M versus A plot. The lowess curve was calculated using the statistical software Splus (version 6.0, Release 2, Insightful Corporation). Normalization was accomplished by subtracting the lowess curve-determined correction factor from each log ratio. To avoid unduly influencing results by a few very extreme expression ratios, any normalized log ratios >5 or <-5 (i.e., ratios >32 or $<1/32$) were truncated at 5 or -5 prior to subsequent statistical analyses. Comparisons of gene expression between stage groups (normal mucosa, adenoma, carcinoma, metastasis) were performed using two-sample pooled-variance t-statistics. Unless otherwise noted, a gene was considered differentially expressed between groups if the two-sample t-statistic calculated for that gene reached statistical significance with $P < 0.001$. This conservative significance level was used to control the average number of false positive findings for each two-group comparison to no more than one false positive per thousand genes tested.

To examine the association between gene expression and gene copy number changes (measured by CGH), CGH and gene expression data were used only from those tissue samples for which both analyses could be performed in parallel (Table 1). For each array and particular chromosome arm, a chromosome arm-level average log expression ratio was calculated as the average of the log expression ratios observed for that array for all genes known to reside on that chromosome arm. For gene copy number changes, the 500 ratio measurements per chromosome of tumor/reference calculated by Leica CW4000 imaging and analysis software (Leica, Cambridge, UK) were equally distributed over the total number of bp for each chromosome and plotted using Excel (Microsoft, Redmond, WA). The 500 ratio values per chromosome were log trans-

formed before the average genomic copy number of each chromosome arm was calculated using the data points corresponding to each arm, excluding values that mapped to the centromeric and pericentromeric heterochromatic regions. Average ratio values for the p-arm of the acrocentric chromosomes 13, 14, 15, 21, and 22 and the entire Y chromosome were not determined. A linear mixed model was applied for each chromosome arm separately. The dependent variable in each linear mixed model was the chromosome arm level average log expression ratio. Independent variables in each model were the continuous CGH values (fixed effect), disease stage (normal mucosa, adenoma, carcinoma, metastasis) as categorical fixed effects, and random subject effects. A correlation was considered to be statistically significant with $P < 0.05$.

Biological Pathway Analysis

As an exploratory approach, we used Ingenuity Pathways Analysis (IPA) software (Ingenuity, Mountain View, CA) to assess the involvement of significantly differentially expressed genes in known pathways and networks. The term "pathway" is used to refer to canonical pathways that annotate genes of interest to known metabolic or signaling cascades as described in the Kyoto Encyclopedia of Genes and Genomes (KEGG, www.genome.jp/kegg). Networks, however, are determined by IPA as groups of genes that interact with a defined gene of interest, i.e., in our analysis a gene significantly differentially expressed among certain stages of colorectal tumorigenesis. In the here presented analysis, "focus genes" are defined as genes that were differentially expressed ($P < 0.001$) in our group comparisons and that additionally were part of IPA database genes. IPA networks can include several genes or proteins and also allow identification of indirect interactions between focus genes. Thus, focus genes are used to map molecular networks that indicate how these genes of interest may influence each other above and beyond already known interaction maps as derived from, e.g., KEGG pathways. The IPA generated networks are listed in a certain order, with the top networks having a lower likelihood, that the generation of the networks was serendipitous.

Confirmation of Gene Expression Changes by Quantitative RT-PCR (Taqman)

cDNAs were prepared by reverse transcribing 25 ng of total cellular RNA using the Single-Strand cDNA Synthesis Kit (Diagnostic Corporation, Indianapolis, IN) according to manufacturer's protocol. Quantitative-PCR (Q-PCR) analysis was performed using TaqMan probes (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions, in 10 μl final volumes, in 384-well microtiter plates (Haslett et al., 2002). Thermocycling conditions using an Applied Biosystems ABI-7900 SDS were as follows: 50°C for 2 min, 95°C for 10 min, 40 cycles of 95°C for 10 sec, and 60°C for 1 min. Specific primers for Q-PCR of *GAPDH*, *GUS*, and *18S* as housekeeping genes and differentially expressed genes (*APC*, *MYC*, *EAF1*, *EGR1*, *EPHA7*, *HRG*, *MFAP4*, *SERPINA1*, *OPCML*, and *CLCA1*) were designed using Applied Biosystems Assay-by-Design primer design software, and their sequences are proprietary. The differentially expressed genes were selected randomly from those genes that were significantly differentially expressed in more than two-stage comparisons. Accurate quantification of each mRNA was achieved using the normalization of the sample ΔC_T values to one reference. This value, is referred to as the $\Delta\Delta C_{T\text{-Sample}}$ value ($\Delta\Delta C_{T\text{-Sample}} = \Delta C_{T\text{-Sample}} - \Delta C_{T\text{-Reference}}$). The target mRNA expression was normalized separately to *GAPDH*, *GUS*, and *18S* expression, and the relative expression was calculated back to the controls for each cell type. The confirmation of the cDNA expression analysis by TaqMan was done by comparing the average expression levels between stage groups, e.g., if the trend of increasing or decreasing cDNA expression over the sequence of malignant transformation was the same as assessed by TaqMan.

Proteomic Analysis

A subset of the samples processed for CGH and gene expression was also analyzed by two-dimensional gel electrophoresis and subsequent mass spectrometry (Table 1). A detailed description of proteomic sample processing has been published by Roblick et al. (2004).

RESULTS

Here we describe a comprehensive evaluation of sequential alterations of the genome, transcriptome, and proteome during colorectal tumorigenesis in individual patients. We were interested in identifying dynamic changes during early steps of malignant transformation, tumor progression, and metastasis, and in deciphering how chromosomal aneuploidies affect the transcriptional equilibrium of cancer cells.

Aneuploidy

As a first approximation to genomic alterations during tumorigenesis, we measured the nuclear DNA content using static image cytometry. All normal mucosa samples were diploid. Aneuploid cell

populations were observed in two of nine adenomas and all of the primary colon carcinomas and liver metastases (data not shown). These results are consistent with previously reported analyses and confirm the sequential genomic destabilization during tumorigenesis (Steinbeck et al., 1994; Ried et al., 1996). We then analyzed the samples using CGH to gain insight into the chromosomal distribution of genomic imbalances. Eleven normal mucosa samples, 13 colorectal adenomas, 16 colorectal adenocarcinomas, and 12 liver metastases were successfully analyzed. We did not detect genomic copy number changes in the normal mucosa samples. Specific gains of chromosomes or chromosome arms 7 and 20q (both in 23% of cases), and 13 (15%) were mapped in the adenomas. The carcinoma samples showed gains on 20q (88%), 7p (50%), 7q, 8q, 13, 20p, and X (all 44%), and a loss of 18 (56%), 4q (44%), and 8p (31%). Gains of chromosomes or chromosome arms 20q (92%), 8q (67%), 7p, and 13 (58%), 7q and 20p (50%), and X (42%) and losses of 18q (83%) and 8p (67%) were observed in the metastases. By dividing the frequency of observed copy number imbalances with the respective case numbers, we calculated an average number of copy number alteration (ANCA) index. The ANCA was 0.92 in the adenomas, 6.5 in the carcinomas, and 6.75 in the metastases, indicating increasing genomic instability during tumor progression. The increasing ANCA index was significant for the progression from normal mucosa to adenomas ($P < 0.05$) and carcinomas ($P < 0.001$) and metastasis ($P < 0.001$). Localized high-level copy number increases (high-level gain defined here as more than three copies of that particular chromosome present) were detected more frequently in the metastases (1.25) than in the primary carcinomas (0.44) and adenomas (0.08). This increase of amplifications was significant between each stage comparison ($P < 0.02$). The summary of the results and a comparison between the three stages is presented in Figure 1A. Patterns of genomic imbalances in individual cases can be retrieved from http://www.ncbi.nlm.nih.gov/sky/skyweb.cgi according to the case annotation listed in Table 1.

Gene Expression by cDNA Arrays

We next wished to identify how global gene expression levels were affected during tumorigenesis. Total RNA of each sample was hybridized on 9K cDNA arrays against a human reference RNA pool (Stratagene). After RNA quality assessment and hybridization quality controls, 3 normal mucosa samples, 8 adenomas, 15 carcinomas, and 7 metastases were included in the analysis. Median-normalized, truncated log ratios were used for *t*-test group comparison analysis. Here, we report only three stage comparisons: normal mucosa versus adenoma samples, adenoma versus carcinoma samples, and carcinoma versus metastasis samples. A total of 368 genes were differentially expressed between the four stages of colorectal cancer progression ($P < 0.001$). Of those, 58 genes were differentially expressed between normal mucosa and adenoma (20 genes were upregulated, 38 genes had lower expression levels); 116 genes were differentially expressed between adenoma and carcinoma (80 elevated, 36 decreased), and 158 genes were differentially expressed between primary carcinoma and liver metastasis (138 elevated, 20 decreased). Complete lists of these genes can be found as supplementary information (Supplementary Tables 1a–1d; Supplementary material for this article can be found at http://www.interscience.wiley.com/jpages/1045-2257/suppmat). A list of significantly differentially expressed genes that occur in more than one stage comparison, together with genes that constitute focus genes as defined by IPA (see below), is provided in Table 2. As shown in the principal component analysis (PCA)-plot in Figure 1B, differences in gene expression profiles of unsupervised data allow separation of adenomas, carcinomas, and metastases.

In an attempt to decipher the involvement of deregulated genes in genetic networks, we used IPA software (Ingenuity, Mountain View, CA). The application of IPA-based algorithms to our data set identified 50 individual pathways that were affected by gene expression changes of at least one focus gene (see Materials and Methods for definition of IPA terminologies). The number of affected pathways increased with stage progression: six pathways were affected by four focus genes in the transition from normal mucosa to adenoma; 14 pathways by nine focus genes from adenoma to carcinoma; and 17 pathways by 13 focus genes in the progression from carcinoma to metastasis (data not shown).

We then selected further for pathways that were affected in more than one stage comparison and identified 20 pathways (Table 3). Of these 20 pathways, 14 were affected by four focus genes, reflecting the multiple functions of any one of those focus genes. Ten pathways were affected in all three-stage comparisons, namely apoptosis signaling, B-cell receptor signaling, G1/S checkpoint regulation, G2/M damage checkpoint regulation, ERK/MAPK signaling, fructose and mannose metabolism, integrin signaling, NF-kB signaling, PI3/AKT signaling, and PPAR signaling.

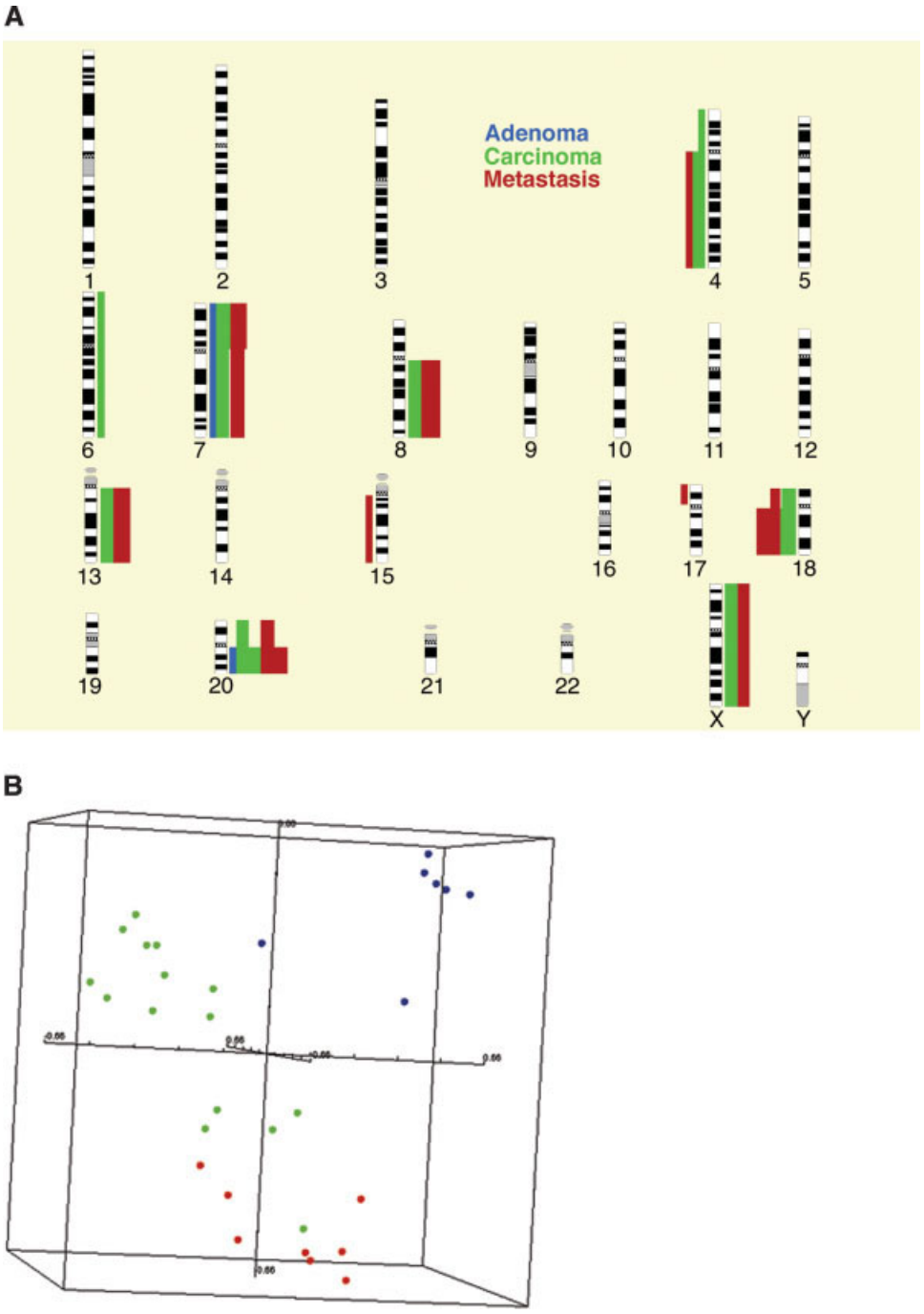

Figure 1. A: Comparison of genomic imbalances in sporadic adenomas (A, $n = 14$), sporadic colorectal carcinomas (SCC, $n = 15$), and liver metastases of SCC (M, $n = 12$). Bars on the left side of the chromosome ideogram denote a loss of sequence in the tumor genome, bars on the right side a gain. The number of alterations per chromosome is normalized to 10 cases for each disease stage. B: Principal component analysis (PCA; centered correlation (Pearson) as distant metric) of gene expression for adenomas (red), primary carcinomas (green), and liver metastases (blue).

In addition to canonical pathways that were related to genes differentially expressed during sequential tumorigenesis, we also analyzed which of the identified genes were selected as focus genes in high-ranked networks as identified by IPA. In the progression from normal mucosa to adenoma, the focus genes *EPHB2*, *TKT*, *SLC27A2*, and *YWHAE* were found to be involved in Network 1. *EPHB2* is known to have an increased expression in colon cancer and was expressed 7.1-fold higher

TABLE 2. Summary of Significantly Differentially Expressed Genes that Were Chosen as Focus Genes and/or Were Differentially Expressed in More than One Stage Progression[a]

Group	Identifier no.	Symbol	NW	Function	MAP	Ratio A/N	*P* N-A	Ratio Ca/A	*P* A-Ca	Ratio Ca/N	*P* N-Ca	Ratio M/Ca	P Ca-M
A	Hs.380030			Homo sapiens cDNA FLJ30193 fis	18q	0.06	**0.0004**	0.87	0.1091	0.05	**0.0000**	1.23	0.1421
	Incyte 604856			Nicotinamide N-methyltransferase		0.96	0.8712	2.06	**0.0007**	1.97	0.0525	3.30	**0.0000**
	Incyte 2192092			Homo sapiens clone 24627 mRNA	3	0.81	0.1058	0.72	**0.0009**	0.58	**0.0004**	0.87	0.1812
	Hs.119065			ESTs	1q	0.58	**0.0007**	1.05	0.5534	0.61	**0.0005**	1.18	0.2662
	Hs.69771	BF		Immuneresponse, high in colon cancer	6p	2.98	0.1708	1.23	0.5141	3.67	**0.0007**	5.00	**0.0000**
	Hs.1498	HRG		Histidine-rich glycoprotein	3q	0.33	**0.0001**	1.02	0.9010	0.33	0.0374	13.61	**0.0000**
	Hs.1327	BCHE		Butyrylcholinesterase	3q	0.27	0.0251	0.69	0.0237	0.18	**0.0005**	2.25	**0.0009**
	Hs.63131	EAF1		ELL associated factor 1	3p	9.49	0.0014	0.27	**0.0003**	2.58	0.0493	18.93	**0.0000**
	Hs.21356	ETL1		Actin-dependent regulator of chromatin	4q	6.02	0.0019	0.34	**0.0007**	2.07	0.0752	19.69	**0.0000**
	Hs.77719	GGCX		Gamma-glutamyl carboxylase	2p	1.78	0.0049	1.12	0.1698	2.01	**0.0000**	1.43	**0.0009**
	Hs.155090	GNB5		Signal transduction pathways	15q	0.41	**0.0009**	0.73	0.0428	0.30	**0.0005**	1.24	0.3507
	Hs.111301	MMP2		Matrix metalloproteinase 2	16q	0.37	0.0604	3.93	**0.0009**	1.47	0.5010	0.21	**0.0003**
	Hs.143482	PPID		Expressed in bilary tract cancer/melanoma	4q	1.80	**0.0003**	0.95	0.4570	1.71	**0.0000**	1.08	0.3755
	Hs.73962	EPHA7		Tumor suppressor gene?	6q	0.11	**0.0001**	1.15	0.3497	0.12	**0.0001**	0.80	0.2320
	Hs.125124	EPHB2		Expressed in colon cancer	1p	7.14	**0.0001**	0.60	0.0410	4.31	**0.0005**	0.86	0.6119
	Hs.296049	MFAP4		Involved in cell adhesion	17p	0.13	**0.0000**	2.64	**0.0004**	0.33	0.0081	0.55	0.0420
	Hs.406684	SCN7A		Sodium channel, voltage-gated, type VII	2q	0.45	0.0095	1.09	0.3618	0.49	**0.0008**	0.70	**0.0007**
	Hs.286084	CENTG3		Centaurin, gamma 3	7q	1.15	0.2154	1.47	**0.0003**	1.70	**0.0010**	1.06	0.5892
	Hs.102991	GEMIN7		Pre-mRNA splicing factor activity	19q	0.89	0.6668	0.54	**0.0000**	0.48	**0.0009**	0.83	0.1917
	Hs.110736	SLC12A2		Up-regulated by mitogenic factor 3	5q	5.32	**0.0005**	0.45	**0.0003**	2.37	0.0079	0.52	0.0380
	Hs.297681	SERPINA1		Serine proteinase inhibitor, clade A	14q	5.74	0.0053	0.28	**0.0007**	1.63	0.3024	17.15	**0.0000**
B	Hs.75514	NP	1	Immunodificiency, high in ovarian cancer	14q	2.56	0.0443	1.26	0.1597	3.22	**0.0003**	0.55	**0.0004**
	Hs.326035	EGR1	2	Transcription, tumor-suppressor-gene?	5q	0.27	**0.0004**	2.52	**0.0009**	0.67	0.2942	1.54	0.1065
	Hs.878	SORD	4	Activated by MYCN, expressed in melanoma	15q	3.09	**0.0002**	0.49	**0.0000**	1.53	0.0812	1.65	0.0060
	Hs.79474	YWHAE	1	Cell division, carcionogenesis	17p	2.21	**0.0007**	0.63	**0.0006**	1.40	0.0484	0.88	0.3038
C	Hs.11729	SLC27A2	3	Solute carrier family 27, member 2	15q	4.11	**0.0006**	0.59	0.0044	2.43	0.0141	1.87	0.0100
	Hs.89643	TKT	1	Expressed in ALM with trisomy 8	3p	2.37	**0.0007**	0.90	0.4096	2.13	0.0017	1.04	0.7586
	Hs.73090	NFKB2	1	Expressed in CLL B cells	10q	0.84	0.3147	1.64	**0.0005**	1.38	0.0748	1.06	0.6522
	Hs.99029	CEBPB	1	Expressed in breast ca	20q	0.94	0.7720	1.75	**0.0004**	1.65	0.0350	0.93	0.6106
	Hs.83583	ARPC2	1	Actin related protein 2/3 complex, subunit 2	2q	0.91	0.7429	1.59	**0.0006**	1.45	0.0981	0.94	0.6141
	Hs.77890	GUCY1B3	2	Guanylate cyclase 1, soluble, beta 3	4q	0.53	0.0084	1.78	**0.0008**	0.94	0.8038	0.78	0.1511
	Hs.2488	LCP2	2	Cell movement, stimulates NFKB	5q	0.69	0.1153	2.12	**0.0005**	1.47	0.1870	1.04	0.8496
	Hs.82568	CYP27A1	1	Cytochrome P450, family 27	2q	0.60	0.0904	0.90	0.7254	1.51	0.0477	2.98	**0.0001**
	Hs.1872	PCK1	2	Phosphoenolpyruvate carboxykinase 1	20q	0.23	0.0193	0.45	0.2329	1.91	0.1392	6.28	**0.0003**
	Hs.2890	PRKCG	1	Expressed in HCT116 cell line	19q	0.42	0.0883	0.84	0.7598	2.00	0.0601	9.89	**0.0000**
	Hs.75183	CYP2E1	6	Role in electron transport, cancer?	10q	0.86	0.1928	0.98	0.9390	1.14	0.4648	19.27	**0.0000**
	Hs.22003	SLC6A1	9	Solute carrier family 6 (GABA)	3p	0.81	0.1849	1.00	0.9844	1.24	0.0778	3.19	**0.0000**
	Hs.22003	SLC6A1	9	Solute carrier family 6 (GABA)	3p	1.08	0.7720	1.15	0.3852	1.06	0.6789	3.45	**0.0002**
	Hs.378199	GSTA2	4	Glutathione S-transferase A2	6p	0.77	0.3231	1.02	0.9099	1.33	0.0789	5.88	**0.0002**
	Hs.75615	APOC2	3	Mutation causes hyperlipoproteinemia type IB	19q	0.67	0.2371	0.79	0.5401	1.19	0.4447	30.06	**0.0000**
	Hs.83834	CYB5	2	Expressed in rat hepatoma cells	18q	1.24	0.3480	1.13	0.5576	0.91	0.5567	2.55	**0.0003**

(Continued)

TABLE 2. Summary of Significantly Differentially Expressed Genes that Were Chosen as Focus Genes and/or Were Differentially Expressed in More than One Stage Progression[a] (Continued)

Group	Identifier no.	Symbol	NW	Function	MAP	Ratio A/N	*P* N-A	Ratio Ca/A	*P* A-Ca	Ratio Ca/N	*P* N-Ca	Ratio M/Ca	P Ca-M
	Hs.2899	HPD	1	Expressed through HINF1A, bind cancer cells	12q	0.85	0.3172	0.77	0.4379	0.92	0.6566	19.97	**0.0000**
	Hs.161640	TAT	2	Tyrosine aminotransferase	16q	0.72	0.1768	0.60	0.0623	0.83	0.2622	13.84	**0.0000**
	Hs.178295	GUCY1A2	2	Increases expression of FOS	11q	0.84	0.4210	0.95	0.8284	1.14	0.4379	2.02	**0.0009**
	Hs.183671	TDO2	1	Tryptophan 2,3-dioxygenase	4q	1.04	0.9171	1.44	0.1929	1.39	0.0244	5.55	**0.0001**
D	P04792	HSPB1	1	Apoptosis, role in carcinogenesis	7q			Significantly higher protein expression in tumor and metastasis					
	P02570	ACTB	2	Function in cell- motility, structure, and integrity	7p			Significantly higher protein expression in tumor and metastasis					
	P02571	ACTG1	2	Function for cytoskeleton, role in cell motility	17q			Significantly higher protein expression in tumor and metastasis					

[a]Significant *P*-values are highlighted in bold ($P < 0.001$). Group A describes genes that were differentially expressed exclusively in more than one stage progression. Group B lists genes that were differentially expressed in more than one stage progression and that were simultaneously also identified as focus genes of networks by Ingenuity Pathways Analysis (IPA). Group C describes genes and Group D proteins that were differentially expressed in only one stage comparison and that were also identified as focus genes/proteins by IPA.

in the adenomas and 4.3-fold higher in the carcinomas compared with normal mucosa (both $P < 0.001$) (Liu et al., 2002). *TKT* is expressed in acute myeloid leukemia blast cells with a trisomy 8 and was 2.3-fold higher expressed in the adenomas when compared with normal mucosa ($P < 0.001$) (Virtaneva et al., 2001). *SLC12A2* has been shown in mouse models to be downregulated by *CTNNB1* (β-catenin) or upregulated by *MYOD1* (myogenic factor 3) (Bergstrom et al., 2002; Gounari et al., 2002). Hypermethylation of the *MYOD1* gene in patients with colorectal cancer has also been described as a prognostic factor for shorter survival times (Hiranuma et al., 2004). In our analysis, *SLC12A2* was initially highly expressed in the adenoma versus normal mucosa samples (5.3-fold, $P < 0.001$) and then significantly lower expressed in the carcinomas compared with the adenomas (0.4-fold, $P < 0.001$), however, still higher than the normal mucosa (2.3-fold, not significant). *YWHAE* is involved in cell division and its homozygous deletion triggers carcinogenesis in small cell lung carcinoma (Konishi et al., 2002). It was expressed 2.21-fold higher in the adenomas ($P < 0.001$) compared to normal mucosa and 0.63-fold lower expressed in the carcinoma compared to the adenomas ($P < 0.001$). During the progression from adenoma to carcinoma, 16 focus genes mapped to Network 1, and during progression from carcinoma to metastasis 17 focus genes were found to be involved in Network 1. A complete list with associated fold expression changes is provided as supplementary data. We also found that mean gene expression levels of 616 genes increased constantly (yet not necessarily significantly by individual *t*-test comparisons) from normal mucosa to adenoma, carcinoma, and metastasis, while they decreased for 1,100 genes. We were also interested in correlating the results of our expression profiling with known canonical genetic pathways involved in colorectal tumorigenesis. We therefore extracted the expression levels of 22 genes involved in such pathways from our expression data, including *MYC*, *EGFR*, *CTNNB1*, *APC*, *WAF1*, and others. The results are presented in Figure 2A and show that although the majority of these genes are altered consistent with data in the literature, most of them did not reach the threshold of significance that we applied to our dataset.

Confirmation of Gene Expression Changes by Quantitative RT-PCR

Expression levels of 14 genes were analyzed by quantitative reverse transcriptase-PCR for all different stages using three different housekeeping

TABLE 3. Canonical Pathways Involved in Colorectal Carcinogenesis as they Relate to Networks (NW) and Focus Genes/Proteins Identified by Ingenuity Pathways Analysis (IPA)[a]

	N versus A		A versus Ca		Ca versus M		Benign vs malignant	
Canonical pathways	NW	Focus genes	NW	Focus genes	NW	Focus genes	NW	Focus protein
Apoptosis signaling	1		1	NFKB2	1/2/4		1	
B cell receptor signaling	7		1/2	NFKB2, EGR1	2		1	
Cell cycle: G1/S checkpoint regulation	1		1		1/2/4		2	
Cell cycle: G2/M DNA damage checkpoint	1	YWHAE	1	YWHAE	1/4			
Death receptor signaling			1	NFKB2	2		1/2	HSPB1
ERK/MAPK signaling	4		1/2/4		1/2/4	PRKCG	1/2	HSPB1
Estrogen receptor signaling	1/8				2	PCK1	1/6	
Fatty acid metabolism	3	SLC27A2			6	CYP2E1		
Fructose and mannose metabolism	4	SORD	4	SORD	2			
G-protein-coupled receptor signaling			1	NFKB2	1/2	PRKCG	1	
IL-6 signaling			1	NFKB2, CEBPB	2/4		1/2	HSPB1
Integrin signaling	1		1/2	ARPC2	1/4		2	ACTB, ACTG1
NF-kB signaling	1		1/2	NFKB2	2/14		1	
PI3/AKT signaling	1	YWHAE	1	YWHAE	1/2/4			
PPAR signaling	1		1	NFKB2	2/4		1/6	
Purine metabolism			2	GUCY1B3	1/2/7	NP, GUCY1A2		
SAPK/JNK signaling	1				1/4		1	
Starch and sucrose metabolism			1		1/2	NP		
T cell receptor signaling			1/2	NFKB2, LCP2	2/4		1	
VEGF signaling	1	YWHAE	1	YWHAE			2	ACTB, ACTG1

[a]The selected focus genes and proteins were significantly differentially expressed in between stage progressions from normal mucosa (N) via adenoma (A) to invasive carcinoma (Ca) and metastasis (M).

genes (*GAPDH*, *GUS*, *18S*) as internal controls. These experiments confirmed the array expression values for 11 of 14 genes (*APC*, *MYC*, *EGR1*, *EPHA7*, *HRG*, *MFAP4*, *SERPINA1*, *CLCA1*, *EPHB2*, *TKT*, and *YWHAE*), thus attesting to the reliability of our array expression data. For the three remaining genes (*OPCML*, *EAF1*, and *SLC27A2*), the array derived expression values could be confirmed only for certain stage comparisons (data not shown).

Correlation of Chromosomal Aberrations and Gene Expression Changes

Specific chromosomal aneuploidies are recurrent genomic alterations in epithelial cancers. It is therefore of fundamental biological significance to understand how precisely these aneuploidies deregulate the cellular transcriptome. We conducted a statistical analysis using linear mixed models (see Material and Methods, Microarray analysis subsection) to examine the association between chromosomal arm copy number changes as detected by CGH and changes of arm-specific average gene expression levels. For instance, chromosome arm 7p was gained in ten of the samples analyzed, and the average gene expression levels on this chromosome arm were significantly upregulated ($P < 0.0001$) (Table 4). A similar result was obtained for several other chromosome arms that were commonly subject to copy number changes in colorectal carcinomas, such as 7q, 8p and 8q, 13q, 18p and 18q, and 20p and 20q. The results for these chromosome arms with a significance of $P < 0.05$ are illustrated in Figure 3A. Figure 3B shows the average gene expression for those samples without any copy number alteration (left) and the average gene expression for those samples with a loss or gain of that arm (right). The majority of chromosomal copy number changes (69%) exerted direct effects on resident gene expression levels. This association was, however, not uniform; for instance, we could not detect a correlation of genomic copy number and average expression levels for chromosome arm 17p, which is frequently lost. The results are summarized in Table 4.

Correlation of Gene and Protein Expression Changes

A total of 14 normal mucosa specimens, 9 adenomas, 14 carcinomas, and 2 metastases included in this study were part of a larger tissue collection that was analyzed also for protein expression profiling using two-dimensional gel electrophoresis (Table 1). These results were presented separately by Roblick et al. (2004). A total of 112 polypeptide spots were

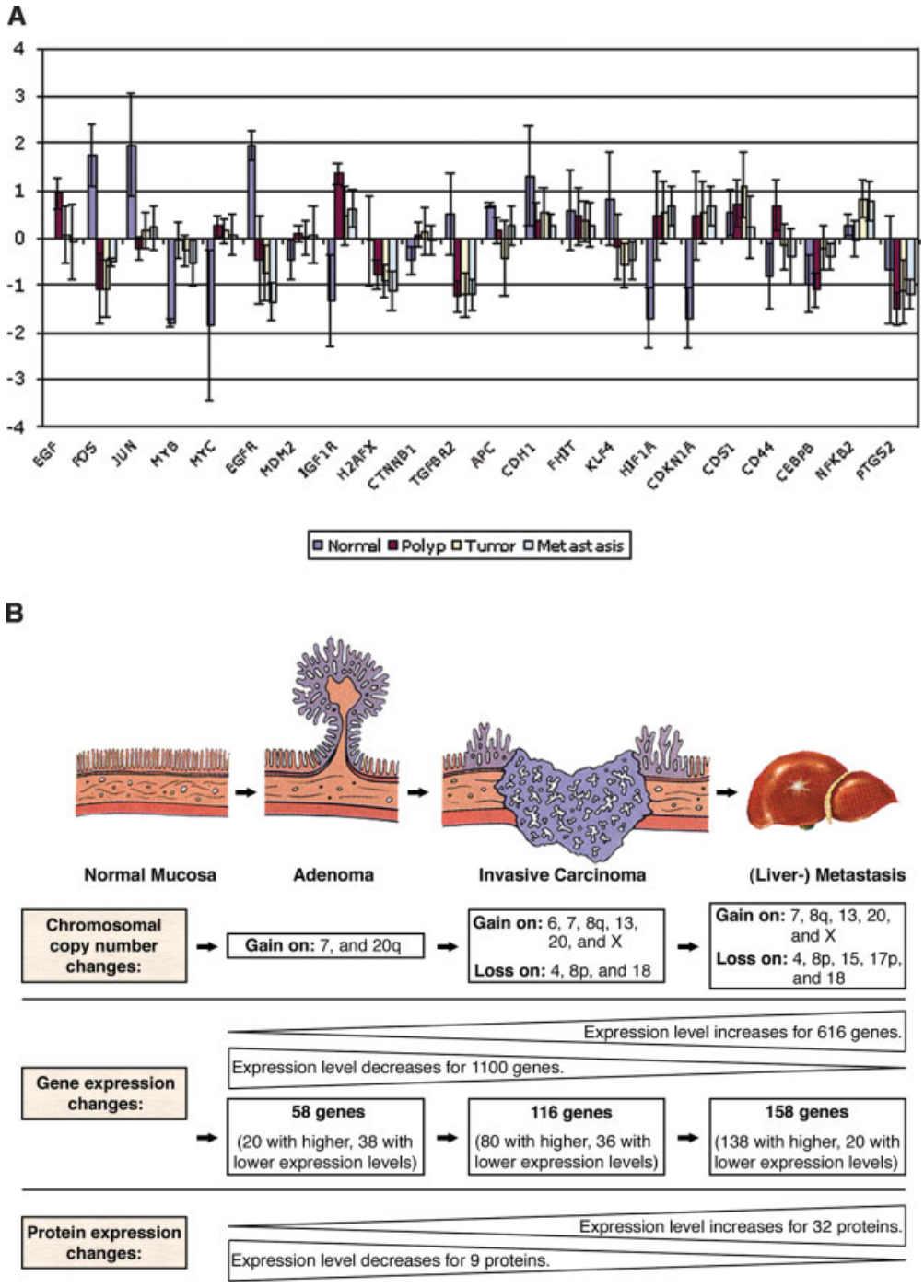

Figure 2. A: Expression profiles of genes known to be important in colorectal carcinogenesis that were also present on the used array platform (plotted is the expression ratio of these genes as compared to the reference RNA as obtained by the cDNA expression arrays). B: Summary figure of genomic, transcriptional, and proteomic changes along the "adenoma-carcinoma-sequence."

selected using Boolean analysis and the Mann–Whitney test ($P < 0.05$) and showed at least a two-fold differential expression between the four stages of carcinogenesis. Of those polypeptides, 72 could be characterized by mass spectrometry. Interestingly, these 72 polypeptides belong to 42 individual proteins, suggesting a high frequency of post-translational modifications, such as phosphorylation. For 27 of these 42 proteins, the corresponding cDNA was included on our microarray platform.

TABLE 4. Regression Analysis of Chromosomal Aberrations and Average Gene Expression per Chromosomal Arm[a]

Map	*P*-value	Loss	Partial loss	No aberration	Partial gain	Gain	Amplicon	Number of cases
1p	**0.002**	0	0	30	0	0	0	30
1q	**<0.0001**	0	0	27	0	3	0	30
2p	**0.001**	0	0	29	0	1	0	30
2q	0.275	0	0	29	0	1	0	30
3p	**<0.0001**	1	0	29	0	0	0	30
3q	0.375	0	0	30	0	0	0	30
4p	**0.005**	3	0	27	0	0	0	30
4q	0.458	6	0	24	0	0	0	30
5p	**0.003**	0	0	29	0	1	0	30
5q	0.084	2	1	27	0	0	0	30
6p	**<0.0001**	0	0	29	0	1	0	30
6q	0.886	0	0	26	0	3	1	30
7p	**<0.0001**	0	0	19	0	10	1	30
7q	**0.009**	0	0	21	1	8	0	30
8p	**<0.001**	8	0	22	0	0	0	30
8q	**0.024**	0	0	20	1	8	1	30
9p	**<0.001**	3	0	27	0	0	0	30
9q	**<0.0001**	3	0	27	0	0	0	30
10p	**0.003**	1	0	28	0	1	0	30
10q	**0.044**	1	0	28	0	1	0	30
11p	**<0.0001**	1	0	29	0	0	0	30
11q	**0.009**	1	0	29	0	0	0	30
12p	**0.018**	0	0	26	1	3	0	30
12q	**<0.001**	0	1	27	0	2	0	30
13q	**<0.001**	0	0	19	0	8	3	30
14q	**<0.0001**	1	0	28	0	1	0	30
15q	**<0.001**	4	0	26	0	0	0	30
16p	0.468	0	0	29	0	1	0	30
16q	0.628	0	0	29	0	1	0	30
17p	0.294	4	0	26	0	0	0	30
17q	0.309	1	0	26	0	3	0	30
18p	**<0.0001**	8	0	21	0	1	0	30
18q	**0.022**	12	2	15	0	1	0	30
19p	0.353	1	0	29	0	0	0	30
19q	0.358	0	0	29	0	1	0	30
20p	**<0.0001**	3	0	18	0	9	0	30
20q	**0.036**	0	0	10	0	16	4	30
21q	0.565	4	0	26	0	0	0	30
22q	0.363	2	0	28	0	0	0	30
Xp	**<0.0001**	0	0	21	0	8	1	30
Xq	**0.040**	1	0	20	0	8	1	30

[a]The "*P*-value" column gives the *P*-values for the association of arm-level expression with CGH results after adjusting for stage type (e.g. normal mucosa, adenoma, carcinoma, or metastasis). Significant *P*-values and correlations are highlighted in bold.

Out of those 27 differentially expressed proteins, 12 corresponding genes showed a similar trend in transcriptional expression as observed for translational changes; however, it did not reach significance (Supplementary Table 2).

Our analysis of genetic changes during sequential cellular transformation of colon epithelium in individual patients revealed disturbances of multiple genetic pathways. In addition, chromosomal aneuploidies directly impact on resident gene expression levels, therefore enhancing the complexity of transcriptional deregulation in colorectal carcinomas and associated metastases beyond mere expression changes of specific oncogenes or tumor suppressor genes.

DISCUSSION

Here we report a comprehensive characterization of sequential genomic, transcriptional and proteomic alterations during colorectal tumorigenesis (Fig. 2B). We determined the degree of genomic instability during carcinogenesis by DNA content measurements and CGH and analyzed global gene expression profiles using cDNA arrays. Microarray-based gene expression profiling was complemented by identification of protein expression levels.

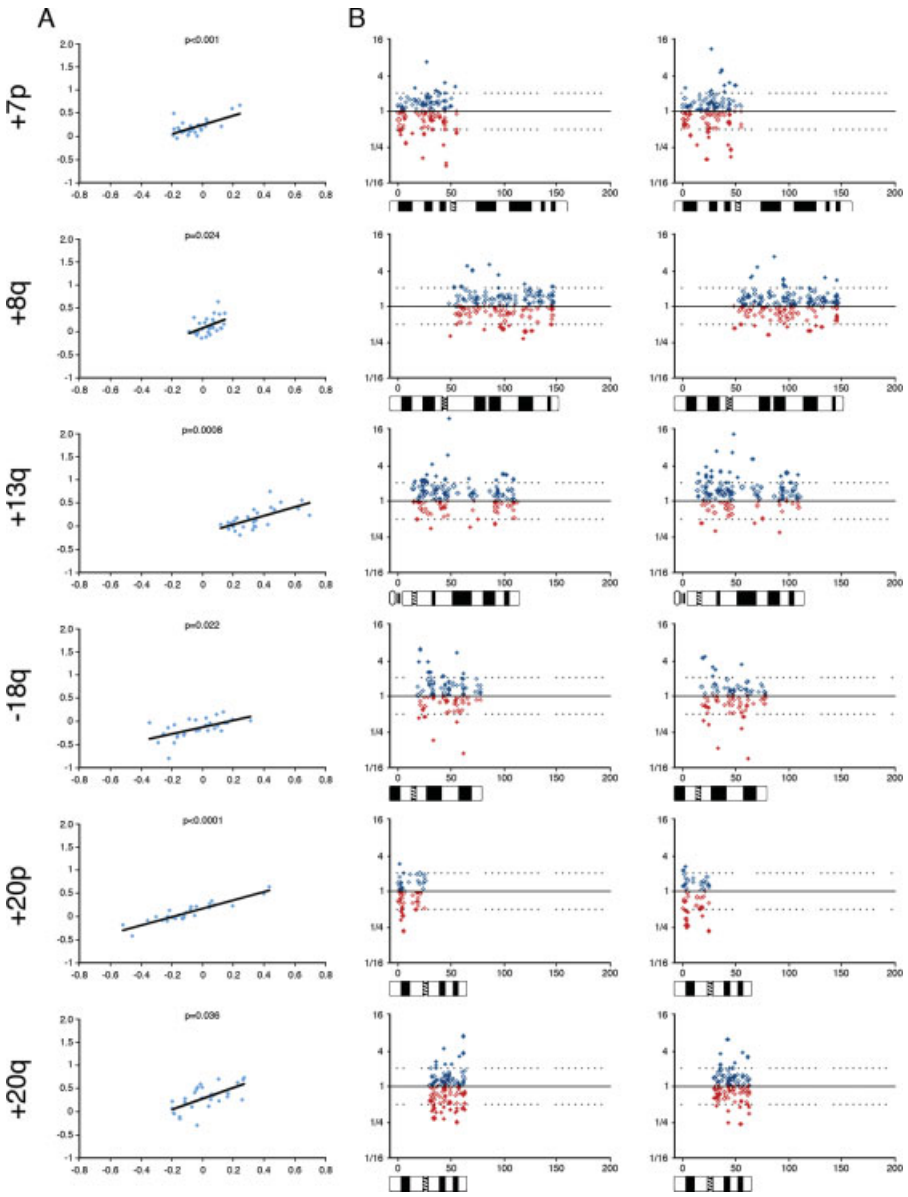

Figure 3. Correlation between gene expression and alterations of chromosome copy number. A: The average gene expression value (*y*-axis) is plotted against the average CGH ratio value (*x*-axis) for each of the 30 samples (including normal mucosa, adenomas, carcinomas, and metastasis samples) for which both the analyses had been performed. The *P*-value is indicated. The directionality of the copy number change most commonly observed is indicated 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 samples without (left) copy number alteration, and to the right with a full loss (18q) or full gain (7p, 8q, 13q, 18q, 20p, and 20q) as indicated in Table 4. These plots correspond to the graphs in A. Information about the number of samples in each category are included in Table 4. Blue, genes with increased expression relative to the reference RNA; red, genes with decreased expression relative to the reference RNA.

Our CGH analyses confirmed the importance of specific chromosomal aneuploidies during colorectal tumorigenesis. As reported previously by us and others, gains of chromosomes and chromosome arms 7, 8q, 13q, and 20q, as well as losses that map to 8p, 17p, and 18q are common (Ried et al., 1996; Bardi et al., 1997). Despite gross variation in the nuclear DNA content (genomic aneuploidy) specific chromosomal imbalances are constantly selected for. We therefore consider it important to evaluate how these aneuploidies affect global gene expression profiles. It is for this reason that the results from expression profiling were analyzed not only for the purpose of gene discovery and identification of potential biomarkers for diagnostic purposes, but also to address transcriptional deregulation as a consequence of aneuploidy.

In a first step, we identified genes differentially expressed between normal colon mucosa and adeno-

mas ($n = 58$), adenomas and carcinomas ($n = 116$), and carcinomas and metastases ($n = 158$). While the number of deregulated genes increased sequentially during progression from one stage to another, we observed redundancy in terms of involvement of gene networks and affected pathways. For instance, for all three comparisons, the genes that were sequentially up- or downregulated during malignant transformation belonged to networks that influence signaling pathways involved in cell cycle control, apoptosis, angiogenesis, and proliferation. Our results are therefore consistent with the interpretation that during the transition from normal colorectal mucosa to adenomas, carcinomas, and metastases several genetic pathways are altered in specific ways. Regarding the expression levels of genes known to be involved in canonical signaling pathways, we noted that not all were altered in expected directions. For instance, we observed downregulation of *EGFR* and the oncogenes *FOS* and *JUN*, which is anti-intuitive. However, we detected upregulation of *MYC* and *HIF1*, while expression of the tumor suppressor genes *TGFBR2*, *APC*, and *CDKN1A* decreased, as expected. Of the significantly differentially expressed genes, several were affected in more than one stage comparison, such as, e.g., *HRG*, *PPID*, *EPHA7*, and *EPHB2*. These genes also showed a low intragroup SD but a large intergroup mean expression difference and could therefore serve as biomarkers for stage-specific diagnosis or as novel therapeutic targets. Interestingly, the Wnt pathway target genes *EPHB2*, *CD44*, *MYB*, and *SLC12A2* showed a significant elevated expression level in the adenomas, when compared with normal mucosa, indicating a strong involvement of the Wnt pathway in early malignant transformation (see also the Wnt homepage of Roel Nusse at http://www.stanford.edu/~rnusse/wntwindow.html) (van de Wetering et al., 2002). While the *CTNNB1* expression itself increased over the sequence of malignant transformation (Fig. 2A), this change was not significant according to our conservative threshold level.

We also identified 42 differentially expressed proteins that underwent frequent posttranslational modification. Only 5% of these proteins are membrane proteins, whereas 84% of these are localized in the cytoplasm. This distribution could simply reflect that differentially expressed proteins are less frequent cytoskeletal proteins. Alternatively, it could well be that the pH-range as well as the resolution capacity of the 2-DE technique preselects for cytoplasmatic proteins. This preselection could account for the fact that no direct match of differentially expressed genes and proteins could be observed. On our array platform presenting ~10,000 gene transcripts (which covers around a third of the entire human genome), only 27 transcripts corresponding to 42 differentially expressed proteins could be directly compared for their expression. Out of those, 12 genes showed a similar trend but did not reach our conservative significance level (Supplementary Table 2). Table 3 shows that differentially expressed genes and proteins, however, did interact within the same canonical pathways. From yeast and mammalian model systems, it is well known that gene and protein expression can differ from each other. Especially protein pairs encoded by coexpressed genes interact with each other more frequently than with random proteins. Furthermore, the mean similarity of expression profiles is significantly higher for respective interacting protein pairs than for random ones (Grigoriev, 2001).

Comparison with published studies proved to be difficult (Shih et al., 2005). This was mainly due to the fact that most published comparisons focused on differential gene expression comparing normal epithelium and cancer, rather than specific stage comparisons in individual patients. Comparison was further complicated by the fact that virtually all studies used different array platforms. Despite these impediments several consistent changes became obvious (see also Supplementary Tables 3a–3d); as in at least three independently conducted studies (Notterman et al., 2001; Takemasa et al., 2001; Bertucci et al., 2004; Kim et al., 2004; Shih et al., 2005), we observed the specific upregulation of *SLC12A2* (solute carrier family 12, member 2) and *SORD* (sorbitol dehydrogenase) and the pronounced downregulation of *SEPP1* (selenoprotein P, plasma, 1), providing supporting evidence that all three genes might play a yet underestimated role in the progression of colorectal cancers. We also extracted from our dataset expression levels of genes known to be involved in colorectal tumorigenesis. Very few reached the significance levels required in our analysis; however, most changed in expected directions.

The genesis of solid tumors is accompanied, if not caused, by the acquisition and maintenance of specific chromosomal aneuploidies. These aneuploidies can emerge in premalignant adenomas, where they can be the first detectable genomic alteration (Bomme et al., 1994). In many instances, mutations of common tumor suppressor genes, such as *TP53*, are not required for the acquisition

of these abnormalities (Steinbeck et al., 1994). Once present, these specific chromosomal aneuploidies are faithfully maintained, even in highly advanced primary tumors and in their local and distant metastases (and after prolonged tissue culture ex vivo as well) (Ried et al., 1999; Ghadimi et al., 2000; Platzer et al., 2002). It is therefore reasonable to postulate a prominent selective role for genomic imbalances during carcinogenesis. The collection of primary adenoma and tumor samples that we have used here is no exception. The CGH analysis mirrored results from our own group and others as previously described (Bomme et al., 1994; Ried et al., 1996). This strict conservation of genomic imbalances that so clearly defines colorectal carcinomas prompted us to identify the immediate consequences of chromosomal aneuploidies on resident gene expression levels. This question has been addressed in several tumor entities and in experimental systems, and the balance of evidence suggests that genomic imbalances result in massive and global deregulation of cellular transcriptomes that affect average expression levels of genes on aneuploid chromosomes (Phillips et al., 2001; Hyman et al., 2002; Matzke et al., 2003; Upender et al., 2004; Heidenblad et al., 2005; Grade et al., 2006; Tsafrir et al., 2006). We were then eager to explore the relationship of genomic imbalances and gene expression levels during colorectal tumorigenesis. Our results indicate that there is in general a significant correlation of chromosomal copy number and average transcriptional activity of resident genes. This correlation applies to almost all chromosomes that are frequently affected in colorectal tumors; however, a few exceptions exist. For instance, gene expression levels on the short arm of chromosome 17 seem to be counter intuitive if compared to the observed chromosome arm copy number decrease. It remains to be established whether this particular effect on chromosome arm 17p is a reflection of local transcriptional upregulation by a compensatory pathway or, e.g., through chromatin remodeling events, or whether upregulation of expression levels of 17p genes is a result of gene expression changes that occur elsewhere in the genome.

In conclusion, increasing genomic instability, aneuploidy and a recurrent pattern of chromosomal aberrations are accompanied by distinct gene- and protein-expression patterns that correlate with certain stages of colorectal cancer progression. It became also evident that specific and recurrent chromosomal aberrations exert a strong and direct influence on average gene expression levels of all resident genes on the affected chromosomes. We confirmed low and high expression levels of canonical genes involved in colorectal tumorigenesis and identified novel cancer-associated transcripts. In addition, differentially expressed proteins undergo extensive posttranslational modifications. Selected genes and proteins now warrant further validation in prospective studies for use in diagnostics and potentially as therapeutic targets.

ACKNOWLEDGMENTS

The authors express their gratitude to Buddy Chen and Joseph Cheng for manuscript editing and IT support, Susanne Becker and Birgitta Sundelin for excellent technical assistance, Claudia Killaitis for clinical data logistics, Annemarie Aumüller, Elke Gheribi, Vera Grobleben, Gisela Grosser-Pape, and Regina Kaatz for clinical sample collection, and Vasuki Gobu and Klaus Moderegger for extracting chromosome CGH values from the Leica software. Ulrike Paulsen received a travel allowance from the Boehringer Ingelheim Fonds, Germany.

REFERENCES

Auer GU, Caspersson TO, Wallgren AS. 1980. DNA content and survival in mammary carcinoma. Anal Quant Cytol 2:161–165.

Bardi G, Parada LA, Bomme L, Pandis N, Willen R, Johansson B, Jeppsson B, Beroukas K, Heim S, Mitelman F. 1997. Cytogenetic comparisons of synchronous carcinomas and polyps in patients with colorectal cancer. Br J Cancer 76:765–769.

Bergstrom DA, Penn BH, Strand A, Perry RL, Rudnicki MA, Tapscott SJ. 2002. Promoter-specific regulation of MyoD binding and signal transduction cooperate to pattern gene expression. Mol Cell 9:587–600.

Bertucci F, Salas S, Eysteries S, Nasser V, Finetti P, Ginestier C, Charafe-Jauffret E, Loriod B, Bachelart L, Montfort J, Victorero G, Viret F, Ollendorff V, Fert V, Giovaninni M, Delpero JR, Nguyen C, Viens P, Monges G, Birnbaum D, Houlgatte R. 2004. Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. Oncogene 23:1377–1391.

Bomme L, Bardi G, Pandis N, Fenger C, Kronborg O, Heim S. 1994. Clonal karyotypic abnormalities in colorectal adenomas: Clues to the early genetic events in the adenoma–carcinoma sequence. Genes Chromosomes Cancer 10:190–196.

Fearon ER, Vogelstein B. 1990. A genetic model for colorectal tumorigenesis. Cell 61:759–767.

Ghadimi BM, Sackett DL, Difilippantonio MJ, Schrock E, Neumann T, Jauho A, Auer G, Ried T. 2000. 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 27:183–190.

Gounari F, Signoretti S, Bronson R, Klein L, Sellers WR, Kum J, Siermann A, Taketo MM, von Boehmer H, Khazaie K. 2002. Stabilization of β-catenin induces lesions reminiscent of prostatic intraepithelial neoplasia, but terminal squamous transdifferentiation of other secretory epithelia. Oncogene 21:4099–4107.

Grade M, Ghadimi BM, Varma S, Simon R, Wangsa D, Barenboim-Stapleton L, Liersch T, Becker H, Ried T, Difilippantonio MJ. 2006. Aneuploidy-dependent massive deregulation of the cellular transcriptome and apparent divergence of the Wnt/β-catenin signaling pathway in human rectal carcinomas. Cancer Res 66:267–282.

Grigoriev A. 2001. A relationship between gene expression and protein interactions on the proteome scale: Analysis of the bacteriophage T7 and the yeast *Saccharomyces cerevisiae*. Nucleic Acids Res 29:3513–3519.

Haslett JN, Sanoudou D, Kho AT, Bennett RR, Greenberg SA, Kohane IS, Beggs AH, Kunkel LM. 2002. Gene expression comparison of biopsies from Duchenne muscular dystrophy (DMD) and normal skeletal muscle. Proc Natl Acad Sci USA 99:15000–15005.

Hegde P, Qi R, Abernathy K, Gay C, Dharap S, Gaspard R, Hughes JE, Snesrud E, Lee N, Quackenbush J. 2000. A concise guide to cDNA microarray analysis. Biotechniques 29:548–550; 552–554; 556 passim.

Heidenblad M, Lindgren D, Veltman JA, Jonson T, Mahlamaki EH, Gorunova L, van Kessel AG, Schoenmakers EF, Hoglund M. 2005. Microarray analyses reveal strong influence of DNA copy number alterations on the transcriptional patterns in pancreatic cancer: Implications for the interpretation of genomic amplifications. Oncogene 24:1794–1801.

Hiranuma C, Kawakami K, Oyama K, Ota N, Omura K, Watanabe G. 2004. Hypermethylation of the MYOD1 gene is a novel prognostic factor in patients with colorectal cancer. Int J Mol Med 13:413–417.

Hyman E, Kauraniemi P, Hautaniemi S, Wolf M, Mousses S, Rozenblum E, Ringner M, Sauter G, Monni O, Elkahloun A, Kallioniemi OP, Kallioniemi A. 2002. Impact of DNA amplification on gene expression patterns in breast cancer. Cancer Res 62:6240–6245.

Kim H, Nam SW, Rhee H, Shan Li L, Ju Kang H, Hye Koh K, Kyu Kim N, Song J, Tak-Bun Liu E. 2004. Different gene expression profiles between microsatellite instability-high and microsatellite stable colorectal carcinomas. Oncogene 23:6218–6225.

Konishi H, Nakagawa T, Harano T, Mizuno K, Saito H, Masuda A, Matsuda H, Osada H, Takahashi T. 2002. Identification of frequent G(2) checkpoint impairment and a homozygous deletion of 14-3-3 ε at 17p13.3 in small cell lung cancers. Cancer Res 62:271–276.

Korn EL, Habermann JK, Upender MB, Ried T, McShane LM. 2004. Objective method of comparing DNA microarray image analysis systems. Biotechniques 36:960–967.

Liu W, Ahmad SA, Jung YD, Reinmuth N, Fan F, Bucana CD, Ellis LM. 2002. Coexpression of ephrin-Bs and their receptors in colon carcinoma. Cancer 94:934–939.

Matzke MA, Mette MF, Kanno T, Matzke AJ. 2003. Does the intrinsic instability of aneuploid genomes have a causal role in cancer? Trends Genet 19:253–256.

Misek DE, Imafuku Y, Hanash SM. 2004. Application of proteomic technologies to tumor analysis. Pharmacogenomics 5:1129–1137.

Notterman DA, Alon U, Sierk AJ, Levine AJ. 2001. Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. Cancer Res 61:3124–3130.

Oppermann M, Cols N, Nyman T, Helin J, Saarinen J, Byman I, Toran N, Alaiya AA, Bergman T, Kalkkinen N, Gonzalez-Duarte R, Jornvall H. 2000. Identification of foetal brain proteins by two-dimensional gel electrophoresis and mass spectrometry comparison of samples from individuals with or without chromosome 21 trisomy. Eur J Biochem 267:4713–4719.

Phillips JL, Hayward SW, Wang Y, Vasselli J, Pavlovich C, Padilla-Nash H, Pezullo JR, Ghadimi BM, Grossfeld GD, Rivera A, Linehan WM, Cunha GR, Ried T. 2001. The consequences of chromosomal aneuploidy on gene expression profiles in a cell line model for prostate carcinogenesis. Cancer Res 61:8143–8149.

Platzer P, Upender MB, Wilson K, Willis J, Lutterbaugh J, Nosrati A, Willson JK, Mack D, Ried T, Markowitz S. 2002. Silence of chromosomal amplifications in colon cancer. Cancer Res 62:1134–1138.

Ried T, Heselmeyer-Haddad K, Blegen H, Schrock E, Auer G. 1999. Genomic changes defining the genesis, progression, and malignancy potential in solid human tumors: A phenotype/genotype correlation. Genes Chromosomes Cancer 25:195–204.

Ried T, Knutzen R, Steinbeck R, Blegen H, Schrock E, Heselmeyer K, du Manoir S, Auer G. 1996. Comparative genomic hybridization reveals a specific pattern of chromosomal gains and losses during the genesis of colorectal tumors. Genes Chromosomes Cancer 15:234–245.

Roblick UJ, Hirschberg D, Habermann JK, Palmberg C, Becker S, Kruger S, Gustafsson M, Bruch HP, Franzen B, Ried T, Bergmann T, Auer G, Jornvall H. 2004. Sequential proteome alterations during genesis and progression of colon cancer. Cell Mol Life Sci 61:1246–1255.

Schena M, Shalon D, Heller R, Chai A, Brown PO, Davis RW. 1996. Parallel human genome analysis: Microarray-based expression monitoring of 1000 genes. Proc Natl Acad Sci USA 93:10614–10619.

Shih W, Chetty R, Tsao MS. 2005. Expression profiling by microarrays in colorectal cancer (review). Oncol Rep 13:517–524.

Steinbeck RG, Heselmeyer KM, Auer GU. 1994. DNA ploidy in human colorectal adenomas. Am J Epidemiol 16:196–202.

Takemasa I, Higuchi H, Yamamoto H, Sekimoto M, Tomita N, Nakamori S, Matoba R, Monden M, Matsubara K. 2001. Construction of preferential cDNA microarray specialized for human colorectal carcinoma: Molecular sketch of colorectal cancer. Biochem Biophys Res Commun 285:1244–1249.

Tsafrir D, Bacolod M, Selvanayagam Z, Tsafrir I, Shia J, Zeng Z, Liu H, Krier C, Stengel RF, Barany F, Gerald WL, Paty PB, Domany E, Notterman DA. 2006. Relationship of gene expression and chromosomal abnormalities in colorectal cancer. Cancer Res 66:2129–2137.

Upender MB, Habermann JK, McShane LM, Korn EL, Barrett JC, Difilippantonio MJ, Ried T. 2004. Chromosome transfer induced aneuploidy results in complex dysregulation of the cellular transcriptome in immortalized and cancer cells. Cancer Res 64:6941–6949.

van de Wetering M, Sancho E, Verweij C, de Lau W, Oving I, Hurlstone A, van der Horn K, Batlle E, Coudreuse D, Haramis AP, Tjon-Pon-Fong M, Moerer P, van den Born M, Soete G, Pals S, Eilers M, Medema R, Clevers H. 2002. The β-catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. Cell 111:241–250.

Virtaneva K, Wright FA, Tanner SM, Yuan B, Lemon WJ, Caligiuri MA, Bloomfield CD, de La Chapelle A, Krahe R. 2001. Expression profiling reveals fundamental biological differences in acute myeloid leukemia with isolated trisomy 8 and normal cytogenetics. Proc Natl Acad Sci USA 98:1124–1129.

Vogelstein B, Kinzler KW. 2004. Cancer genes and the pathways they control. Nat Med 10:789–799.