

## Application of cDNA microarrays to generate a molecular taxonomy capable of distinguishing between colon cancer and normal colon

Tong-Tong Zou<sup>1,3</sup>, Florin M Selaru<sup>1,3</sup>, Yan Xu<sup>1,3</sup>, Valentina Shustova<sup>1</sup>, Jing Yin<sup>1</sup>, Yuriko Mori<sup>1</sup>, David Shibata<sup>2</sup>, Fumiaki Sato<sup>1</sup>, Suma Wang<sup>1</sup>, Andreea Olaru<sup>1</sup>, Elena Deacu<sup>1</sup>, Thomas C Liu<sup>1</sup>, John M Abraham<sup>1</sup> and Stephen J Meltzer<sup>\*,1</sup>

<sup>1</sup>Department of Medicine, Division of Gastroenterology and Greenebaum Cancer Center, University of Maryland School of Medicine and Baltimore VA Hospital, Baltimore, Maryland, MD 21201, USA; <sup>2</sup>Department of Surgery, Division of Surgical Oncology, University of Maryland School of Medicine 8-009 Bressler Research Building, 655 West Baltimore Street, Baltimore, Maryland, MD 21201, USA

**In order to discover global gene expression patterns characterizing subgroups of colon cancer, microarrays were hybridized to labeled RNAs obtained from seventeen colonic specimens (nine carcinomas and eight normal samples). Using a hierarchical agglomerative method, the samples grouped naturally into two major clusters, in perfect concordance with pathological reports (colon cancer versus normal colon). Using a variant of the unpaired *t*-test, selected genes were ordered according to an index of importance. In order to confirm microarray data, we performed quantitative, real-time reverse transcriptase–polymerase chain reaction (TaqMan RT–PCR) on RNAs from 13 colorectal tumors and 13 normal tissues (seven of which were matched normal-tumor pairs). RT–PCR was performed on the *gro1*, *B-factor*, *adican*, and *endothelin converting enzyme-1* genes and confirmed microarray findings. Two hundred and fifty genes were identified, some of which were previously reported as being involved in colon cancer. We conclude that cDNA microarraying, combined with bioinformatics tools, can accurately classify colon specimens according to current histopathological taxonomy. Moreover, this technology holds promise of providing invaluable insight into specific gene roles in the development and progression of colon cancer. Our data suggests that a large-scale approach may be undertaken with the purpose of identifying biomarkers relevant to cancer progression.**

*Oncogene* (2002) 21, 4855–4862. doi:10.1038/sj.onc.1205613

**Keywords:** colon cancer; cDNA microarrays; cluster analysis; gene filtering

### Introduction

Colorectal cancer is the second most common cause of cancer-related deaths in the United States, with an estimated 130 200 new cases and 56 300 deaths occurring in the year 2000 (American Cancer Society, 2000). Early detection of colon cancer in its premalignant stages prevents progression to invasive cancer (American Cancer Society, 2000). Therefore, it is crucial to discover lesions early in their neoplastic evolution. Furthermore, biomarkers capable of predicting progression of polyps to cancer could identify subgroups of patients most likely to develop recurrent polyps or new cancers.

cDNA microarray technologies have now made it possible to characterize tissue samples by monitoring the expression of thousands of genes simultaneously. Previous studies have been successfully performed in acute leukemia, lymphoma, breast cancer, and others (Chu *et al.*, 1998; Spellman *et al.*, 1998; Eisen and Brown 1999; Iyer *et al.*, 1999; Perou *et al.*, 1999, 2000; Alizadeh *et al.*, 2000). These breakthrough studies have begun to create a modern molecular taxonomy, which capitalizes on the statistical power of large gene expression datasets (Bassett *et al.*, 1999). This comprehensive approach permits the classification of diseases or disease lesions based on comprehensive characterizations of their phenotypes, rather than on limited histologic or other clinical features (Perou *et al.*, 1999, 2000; Alizadeh *et al.*, 2000; Selaru *et al.*, 2002a,b).

While cDNA microarray data comprises precise ratios of expression levels between each sample and a reference probe, it is difficult to interpret the enormous data flow that results from this process. In order to efficaciously extract information from this data, mathematical and bioinformatics tools have been developed. For example, the bioinformatics program Cluster (Eisen *et al.*, 1998) generates hierarchical and K-means clusters from tab-delimited text files containing expression ratios comparing two RNA species' expression levels. In addition, the program Treeview (Eisen *et al.*, 1998) generates two-dimensional graphic displays of gene expression in red and green colors (Chen *et al.*, 1998; Spellman *et al.*, 1998; Eisen and

\*Correspondence: SJ Meltzer, University of Maryland, Room N3W62, 22 S. Green Street, Baltimore, Maryland, MD 21201, USA; E-mail: smeltzer@medicine.umaryland.edu

<sup>3</sup>These authors contributed equally to this work

Received 26 February 2002; revised 18 April 2002; accepted 26 April 2002

Brown 1999; Iyer *et al.*, 1999; Perou *et al.*, 1999; 2000; Alizadeh *et al.*, 2000). Significance Analysis of Microarrays (SAM), a software program developed at Stanford University (Tusher *et al.*, 2001), was used for finding the most significantly differentially expressed genes in a cancer relative to matching normal specimens. The output consists of lists of genes ordered in decreasing order of an index of significance.

We applied SAM to identify genes whose expression was most relevant to distinguishing between normal and cancerous colonic epithelium. Our results suggest that bioinformatics analyses of microarray data can precisely classify colorectal pathological states, and that candidate genes responsible for this classification can be identified in this process.

## Results

Hybridizations were successfully performed on all specimens. Cluster software was then applied, and a dendrogram view was generated using TreeView (Figure 1).

The degree of relatedness of specimens, as well as genes, was quantitated by the closeness of branches in the dendrogram. Cluster grouped the seventeen colonic specimens into two major clusters. The first cluster contained all of the normal colon samples, while the second contained all of the colon cancer samples. This classification was derived with no previous pathological data, using only gene profiling data. Thus, normal colon tissue was differentiated from cancerous colon tissue based solely on global patterns of gene expression.

SAM was then used to identify 250 genes that were most influential in the distinction between normal and cancerous colon, with a false discovery rate (FDR) of 0.57 (Web Table, [www.microarray.umaryland.edu/manuscripts/11415/](http://www.microarray.umaryland.edu/manuscripts/11415/)). This Table contains links to further information regarding each gene (*q.v.*). Among these 250 genes, 150 were expressed at higher levels in

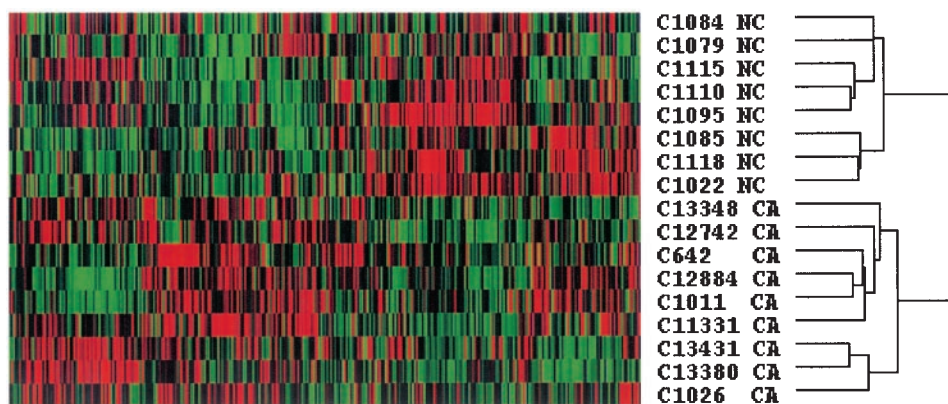
normal than in cancerous colon; the remaining 100 were expressed at higher levels in cancer than in normal tissue.

As a proof-of-principle, we verified the relative expression levels of four of these 250 genes by real-time (TaqMan) RT-PCR. The four genes tested were *adican*, *gro-1* oncogene, *properdin/B* factor, and *endothelin converting enzyme 1*. All four genes were confirmed by TaqMan RT-PCR as showing the same differential expression patterns found in the primary cDNA microarray data. The first three of these genes were overexpressed in colon cancers, while the last was overexpressed in normal colon (Figure 2).

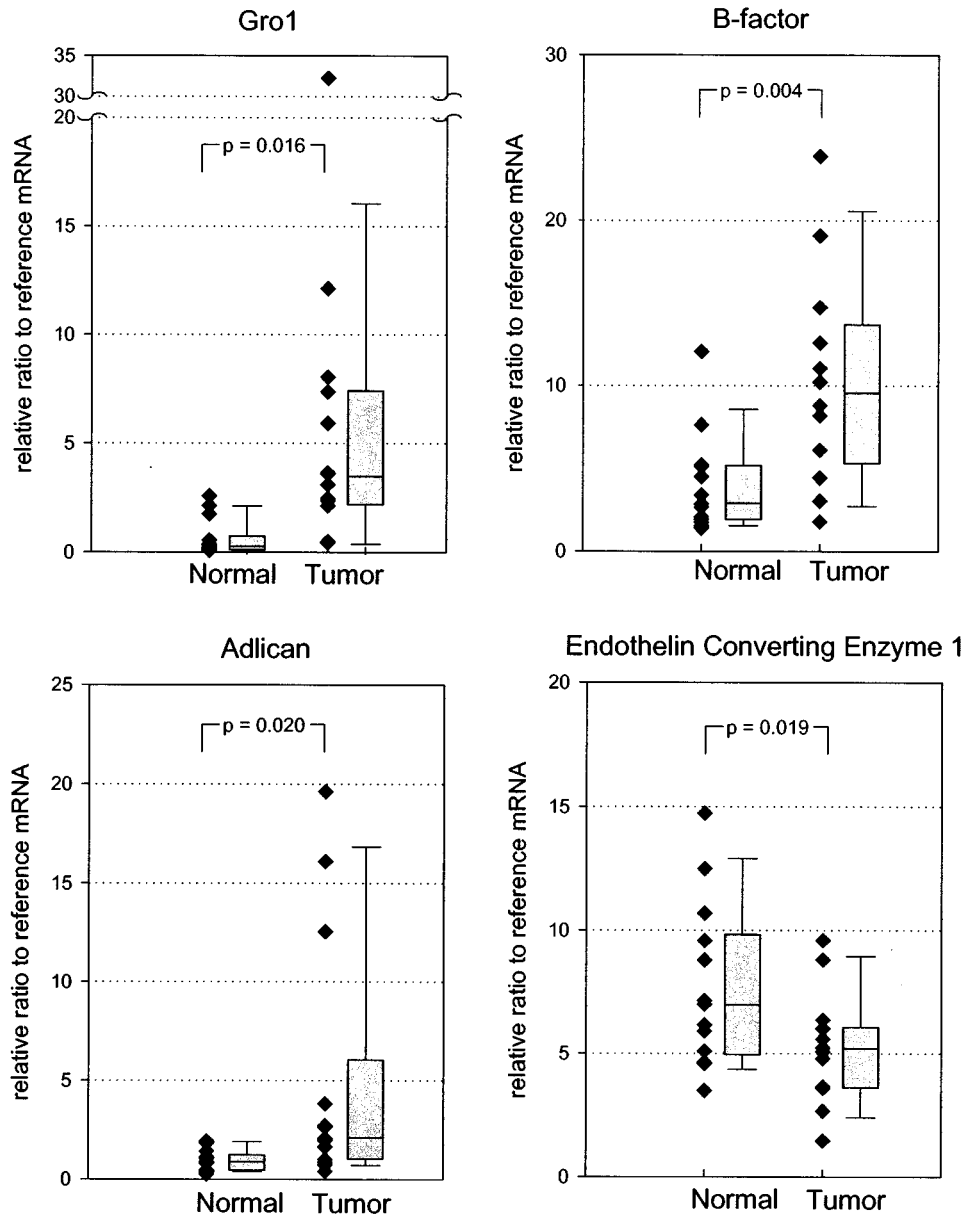
This Figure displays results of real-time PCR products performed on 13 colorectal carcinomas and 13 normal tissues, including seven paired normal and cancerous colorectal samples. Nine of the tumors and seven of the normals were analysed by both microarray and TaqMan; the remaining four cancers and six normals were analysed only by TaqMan, but were in agreement with the differential expression pattern seen in the original cDNA microarray data.

In order to obtain additional insight into global gene expression patterns and to identify individual genes distinguishing among colon cancer subgroups, we performed significance analysis of microarrays (SAM) (Tusher *et al.*, 2001) based on right versus left colon lesions and on lymph node positive (Dukes' C) versus lymph node negative (Dukes' B) tumors. Genes selected by SAM were then entered into Cluster (Eisen *et al.*, 1998; Eisen and Brown, 1999). Results of these two cluster analyses are displayed in Figure 3.

As shown in Figure 3a, clustering performed on 203 genes selected by SAM was 100% successful in correctly classifying colon cancers as arising in the right versus left colon. To our knowledge, this is the first demonstration of global gene expression differences between the right and left colon. In Figure 3b are displayed the results of clustering performed based on 224 genes selected by SAM as significantly differentially expressed between lymph node-negative versus



**Figure 1** Clustergram of colonic samples. *Vertical axis*, samples included in the study; *horizontal axis*, the 250 most significantly differentially expressed genes. For each gene and a specific sample there is a thin colored band, which visually describes the expression of that gene in the sample. *Green bands* represent underexpressed genes; *red bands*, overexpressed genes; *black bands*, genes showing approximately equal expression in tumor versus normal; and *gray bands*, missing data



**Figure 2** Real time RT-PCR verification of SAM-selected clones. The expression of the *Gro1*, *B-factor*, *adlican* and *endothelin-converting enzyme 1* genes was verified with real-time RT-PCR in colorectal tumors and normal tissues. An ABI 7700 (TaqMan) apparatus was used to generate and display real-time PCR products from 13 colorectal tumors and 13 normal tissues (including seven paired normal and tumor samples) amplified from (A) *gro1*, (B) *B-factor*, and (C) *adlican*, identified by cDNA microarray as overexpressed in colon cancers relative to normal colonic epithelia; and (D) *endothelin-converting enzyme 1*, identified by cDNA microarray as overexpressed in normal colonic epithelial versus colon cancers. Each result was expressed as the expression ratio between the sample and the reference cDNA used for making a standard curve (see Methods). Duplicates are shown for each analysis

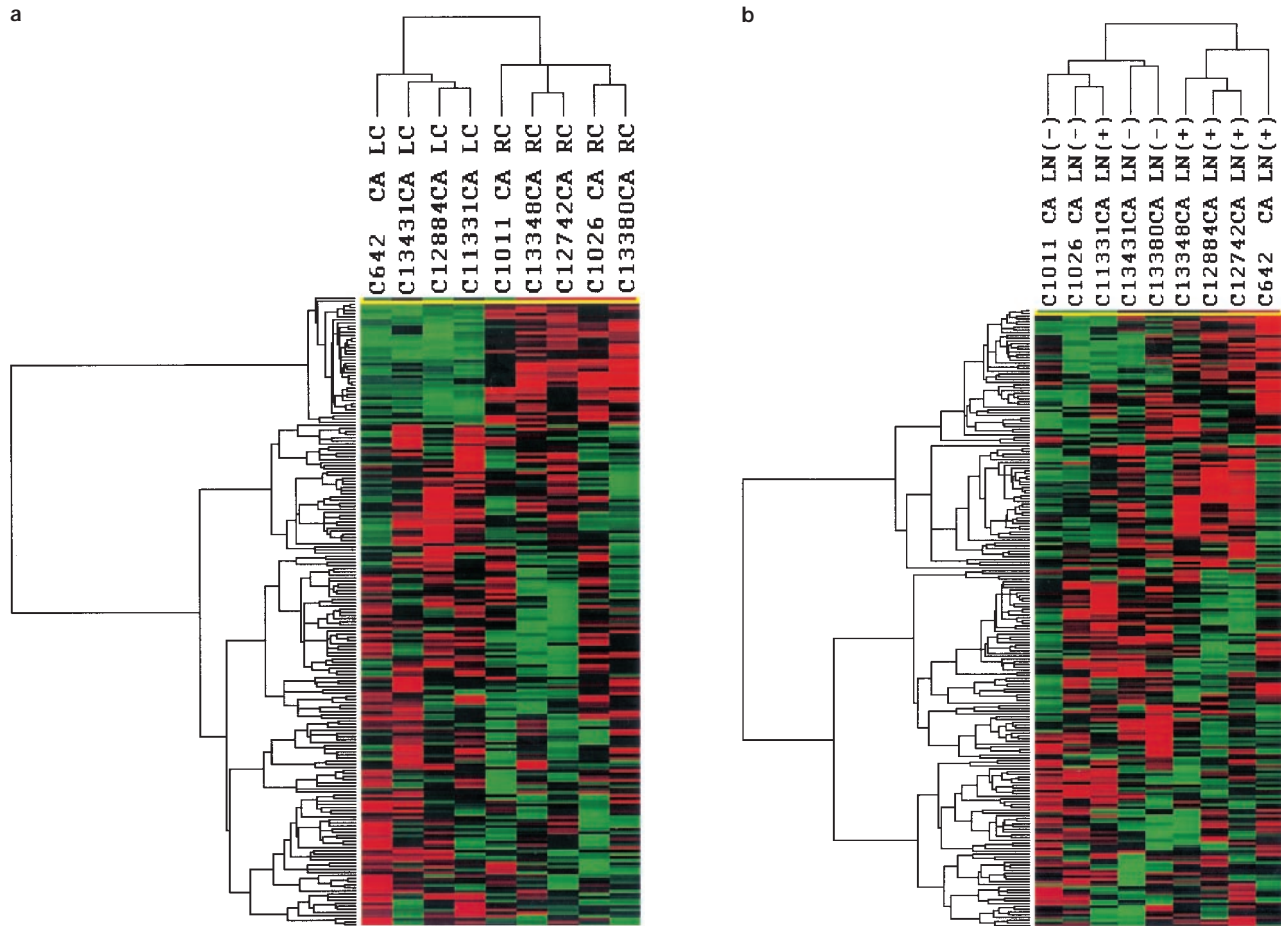
lymph node-positive tumors. Clustering was still fairly successful in this comparison, misplacing only one LN-positive specimen in the LN-negative group.

### Discussion

The above data suggest that gene filtering methods based on genomic data are capable of distinguishing normal from cancerous human colon. Furthermore,

these data show that gene comparison algorithms (such as SAM) can be used to identify genes whose differential expression is most relevant to distinguishing between two diagnoses, such as normal and cancerous colon. In addition, these results suggest the combined approaches used herein can discriminate among subgroups of colon cancer.

In our study, genes identified by SAM as being significantly differentially expressed included many with known or suspected relevance to cancer. For



**Figure 3** Clustergrams based on SAM-selected genes. (a), Clustergram based on genes significantly different between right-sided and left-sided tumors. Two-hundred and three genes were selected by SAM for their differential expression between tumors arising in the right and left colon. Clustering was then performed based on this filtered gene set, and it succeeded in accurately separating the two groups of tumors. *LC*, left colon; *RC*, right colon. Case numbers are the same as those indicated in Table 1. (b) Clustergram based on genes significantly different between LN-positive (Dukes' C) and LN-negative (Dukes' B) tumors. Two-hundred and twenty-four genes were selected by SAM based on their differential expression between LN-negative and LN-positive colon cancers, and clustering was then performed. Cluster was nearly completely accurate in its classification based on this reduced gene set, only misplacing one LN-positive case in the LN-negative colon cancer group. *LN*, lymph node. Case numbers corresponded to the patients listed in Table 1

example, among the genes whose expression was verified by quantitative RT-PCR, the *Gro1* oncogene (also known as the cytokine growth-related oncogene or melanoma growth-stimulating activity alpha) has been localized to region 4q21. This chromosomal region is frequently involved in clonal aberrations found in primary melanomas (Grammatico *et al.*, 1995). Moreover, *gro-1* has been implicated in central nervous system tumors (Robinson *et al.*, 2001). Similarly, properdin/B factor has been implicated in the etiopathogenesis of breast cancers (Perou *et al.*, 1999); and the retinoic acid receptors RAR alpha, RAR beta, and RAR gamma mRNA are overexpressed in the colon cancer cell line, LoVo (Stewart and Thomas, 1997). Conversely, many genes overexpressed in normal colon were also relevant to cancer. For example, endothelin converting enzyme 1 expression is altered in breast and

lung cancers (Ahmed *et al.*, 2000; Patel and Schrey, 1995).

Additional cancer-related genes identified by our SAM-based analysis of cDNA microarray data included members of the protein phosphatase 1 gene family, which show genetic alterations and abnormal expression in human cancers (Takakura *et al.*, 2001; Sogawa *et al.*, 1997); gastrin-releasing peptide (GRP) and its receptor (GRP-R), which are frequently expressed by cancers of the gastrointestinal tract, breast, lung, and prostate (Jensen *et al.*, 2001); members of the retinoic acid receptor gene family, which have been implicated in tumor growth inhibition (Nicke *et al.*, 1999); and certain mitogen-activated protein kinases, which have actually been reported as downregulated in colon cancers (Wang *et al.*, 2000).

The differences in gene-filtered expression patterns that we observed between right and left colon are

consistent with published observations on biological differences between the two sides of the colon (Ikeda *et al.*, 2001; Meltzer *et al.*, 1989). Moreover, the near perfect classification of LN-positive versus LN-negative tumors is not surprising, considering that this clinical parameter forms part of the basis for the traditional Dukes' classification.

The majority of published cDNA microarray studies use this technology primarily as a global gene profiling tool (Duggan *et al.*, 1999; Khan *et al.*, 1999; Alizadeh *et al.*, 2000; Elek *et al.*, 2000). Data presented in the current manuscript suggest that cDNA microarrays are also a good screening tool to identify individual genes relevant to a particular disease process, such as human cancer. Our TaqMan RT-PCR results suggest that microarrays are accurate in gauging differential expression. Moreover, our findings suggest that gene comparison algorithms, such as SAM, are highly accurate in identifying genes most influential in making distinctions between two types of study materials, such as normal and cancerous colon. Thus, they support the broad use of cDNA microarrays and bioinformatics in large-scale gene discovery studies.

## Materials and methods

### Tissues and patients

Patients were enrolled sequentially during a three-year period from the University of Maryland or Baltimore VA Hospitals. Patients with a presumed diagnosis of colonic carcinoma were enrolled in the study, with histologic confirmation after enrollment. Informed consent was obtained from all patients prior to enrollment under a protocol approved by the University of Maryland/Baltimore VA Hospital Institutional Review Board (IRB). All tissues were immediately frozen on dry ice and stored under liquid nitrogen at  $-180^{\circ}\text{C}$  until

further use. All presumptive diagnoses were confirmed by histopathologic examination: parallel samples from each anatomic location were obtained for histological analysis simultaneously with each research specimen. Clinical parameters of patients analysed in our study are included in Table 1.

### RNA extraction and amplification

Total RNA was extracted from freshly frozen tissues and cell lines with the Rneasy Mini Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. Samples were treated with Rnase-free DNase (Qiagen) on the RNeasy columns. RNA was dissolved in Rnase-free water and amplified using T7 RNA polymerase using a linear, nondistorting protocol (Van Gelder *et al.*, 1990; Luo *et al.*, 1999). Amplified RNA (aRNA) was generated from 10  $\mu\text{g}$  of total RNA using a cDNA Synthesis System (GIBCO-BRL) and the Ampliscribe T7 Transcription Kit (Epicentre Technology). Using this approach, approximately 100-fold amplification of mRNA from total RNA was obtained.

### Reference (process control) probe

A reference probe was produced from a mixture containing equimolar aliquots of RNA from the cell lines Hct116, HT29, CaCo-2, Hct15, HTB114, MCF-7, HeLa and AGS. HTB114 was derived from a patient with leukemia, HeLa from a cervical cancer, MCF-7 from a breast cancer, and AGS from a gastric cancer. The remaining four cell lines (Hct116, HT29, CaCo-2, and Hct15) were derived from colorectal cancers. These cell lines were chosen to represent a variety of cell types in order to garner a baseline (i.e., green or Cy3 expression level 1) for comparison by ratio with red (Cy5) lesion-derived signal and 2) as a process control to ensure that each clone was generating detectable signal on the array (Alizadeh *et al.*, 1998; Iyer *et al.*, 1999; Perou *et al.*, 1999, 2000).

**Table 1** Clinical characteristics of patients analysed

Sample	Age	Sex	Race	Invasive or not	Mucinous or nonmucinous	Grade	Lymph node status	Location
13348	61	F	W	invasive	mucinous	poorly differentiated	13 of 30 lymph nodes positive for metastatic adenocarcinoma	ascending colon
12742	71	M	W	invasive	nonmucinous	moderately differentiated	2 of 5 lymph nodes positive for metastatic adenocarcinoma	ascending colon
642	71	M	B	invasive	nonmucinous	moderately differentiated	*Lymph nodes positive after 2 months	rectum
12884	55	F	W	invasive	mucinous	N.A.	2 for 12 lymph nodes positive for metastatic adenocarcinoma	rectum + sigmoid colon
1011	67	M	W	invasive	nonmucinous	moderately differentiated	lymph nodes with no evidence of malignancy	cecum
11331	72	M	B	invasive	mucinous	well differentiated	3 lymph nodes positive for metastatic adenocarcinoma	sigmoid colon
13431	74	M	W	invasive	mucinous	moderately differentiated	3 lymph nodes with no evidence of malignancy	sigmoid colon
13380	79	M	W	invasive	mucinous	N.A.	10 lymph nodes with no evidence of malignancy	cecum
1026	63	M	W	invasive	nonmucinous	moderately to poorly differentiated	5 lymph nodes with no evidence of malignancy	ascending colon

W, white; B, African-American; N.A., not available

### Labeling of the aRNA probe

For each two-way comparison, 3–6  $\mu\text{g}$  of aRNA prepared from the reference cells or esophageal lesion were labeled by incorporating Cy3- or Cy5-labeled dCTP using random primers and Superscript reverse transcriptase. The resulting probes were purified with a Microcon microcentrifuge filter device and recovered in a volume of 25  $\mu\text{l}$ . These probes were used for each hybridization under a 40  $\times$  22 mm coverslip at 65°C in 2.24X SSC, 0.25% SDS in a final volume of 35  $\mu\text{l}$ .

### Preparation of micro-assay clones

We followed protocols obtained from the National Cancer Institute–Advanced Technology Center (NCI–ATC; Lance Miller, director; David Petersen, personal communication). The 95% non-redundant, sequence-verified, periodically annotation-updated cDNA library prepared by the Lawrence Livermore Laboratories was used as a source of clones (Research Genetics, Huntsville, AL, USA). One  $\mu\text{l}$  of each bacterial stock was amplified in a 108  $\mu\text{l}$  reaction containing 2.5 U Taq polymerase (Life Sciences, Gaithersburg, MD, USA). The master mix for each plate contained 9.5 ml of double-distilled water, 1 ml of 10 $\times$  buffer, 10  $\mu\text{l}$  each of 1000  $\mu\text{M}$  M13 primers (forward and reverse), 20  $\mu\text{l}$  of each of the four dNTPs (100 mM), and 50  $\mu\text{l}$  of Taq polymerase (5 U/ $\mu\text{l}$ ). PCR conditions consisted of an initial denaturation step at 96°C for 30 s followed by 30 cycles of 45 s at 94°C, 45 s at 55°C and 2 min 30 s at 72°C, then concluded by a final extension of 5 min at 72°C. The amplified inserts were then purified using a Qiagen PCR purification kit (Cat. No. 963141, QiAGEN) on a Qiagen BioRobot 9600 liquid handling robotic workstation. After purification, PCR products were desiccated in 96-well plates using a large Speed-vac apparatus, then reconstituted in 30  $\mu\text{l}$  of distilled water.

### Microarray printing

**Glass slide coating protocol** We prepared lysine-coated slides derived from the NCI–ATC and Stanford University protocols (<http://www.microarrays.org/protocols.htm>). Slides were cleaned using a solution containing 3.5N sodium hydroxide and 68% ethanol, then coated with 1% poly-L-lysine in 1% PBS. Before printing, slides were aged a minimum of 2 weeks, but not longer than 6 weeks.

**Microarray printing** The 8000 clones were printed using eight pins in a 32-pin print head (Majer Precision Engineering, Tempe, AZ, USA) on a GeneMachines Omnigridd Arrayer (GeneMachines, Oxnard, CA, USA). The printed slides were UV-crosslinked, post-treated with succinic anhydride to reduce background, and subjected to hybridization.

### Microarray hybridization and washing

Each slide was incubated in 35  $\mu\text{l}$  of hybridization solution containing Cy3- and Cy5-labeled target, 1  $\mu\text{l}$  of 50 $\times$  Denhardt's blocking solution (Sigma, St. Louis, MO, USA), 20  $\mu\text{g}$  of Human COT 1-DNA (Roche Diagnostics Corporation, Indianapolis, IA, USA), 10  $\mu\text{g}$  of yeast tRNA (Roche), 8–10  $\mu\text{g}$  of Poly-A (Roche), in 2.24 $\times$  SSC/0.25% SDS at 65°C overnight. Hybridization was performed under a 22  $\times$  40 mm cover slip. The slide was then placed in a sealed hybridization chamber (Telechem, Sunnyvale, CA, USA) containing two side wells with a total of 50  $\mu\text{l}$  of water for humidification at 65°C overnight. On the next day, the slide

was washed in 500 ml 2 $\times$  SSC, 0.1% SDS at room temperature, during which time the coverslip fell off and washing continued for 2 min. The slide was placed in 1 $\times$  SSC for 2 min at RT. It was then washed once with 0.2 $\times$  SSC at RT and once with 0.05 $\times$  SSC for 2 min and air-dried.

### Microarray scanning

The hybridized slides were scanned using a GenePix 4000A dual-laser slide scanning system (Axon) at wavelengths corresponding to each probe's unique fluorescence. The resulting GenePix report was then reformatted for importance into the Cluster software program.

### Hierarchical clustering

Data imported from GenePix were manipulated and clustered, using established algorithms implemented in the software program Cluster (<http://rana.lbl.gov/>) (Eisen *et al.*, 1998; Eisen and Brown, 1999). Prior to average linkage clustering, data was log transformed, then median centered on genes and on arrays and, finally, normalized for genes and arrays, as described (Eisen *et al.*, 1998). Average linkage clustering with centered correlation was used. TreeView software (*ibid.*) generated visual representations of the clusters.

### SAM method

SAM was used for identifying differentially expressed genes between normal colon specimens and colon cancer specimens (Tusher *et al.*, 2001). SAM, a computer program specifically designed for manipulating microarray data, reports the most statistically significant differentially expressed genes between two groups of samples. In addition SAM reports an estimate of the Median False Discovery Rate (FDR), which is the percentage of genes falsely reported as showing statistically significant differential expression. SAM uses an algorithm based on the Student's *t*-test and also performs data permutations in order to determine the FDR.

SAM compared results for tumor to those for normal colon. SAM makes this comparison by comparing tumor/reference probe ratios to normal colon/reference probe ratios (i.e., it makes direct comparisons of the primary data, which are themselves ratios of tissue RNA to reference probe RNA). Genes listed in the web link table as being overexpressed in tumor are overexpressed in tumor relative to normal colon; likewise, genes listed as overexpressed in normal colon are overexpressed in normal colon relative to tumor.

### Quantitative RT–PCR

**Samples used for verification of gene expression levels** Total RNAs from the same samples studied in cDNA microarray analysis were used for verification of gene expression levels observed on cDNA microarrays. In addition, we extracted total RNAs from seven normal and seven tumor paired colon tissues.

**Design of quantitative RT–PCR primers and probes** Quantitative RT–PCR was performed on a TaqMan real-time PCR machine (ABI 7700, Applied Biosystems, Foster City, CA, USA). For quantitative RT–PCR, the amplicon spanned an exon–exon boundary in order to exclude genomic DNA contamination. We obtained exon–exon boundary information on genes of interest from the Ensembl Genome Server of

**Table 2** TaqMan primers and probes for confirming individual gene expression (5'→3')

Gene	Sequences of primers and probes
	<i>Gro1</i>
(F)	CAAGAACATCCAAAGTGTGAACGT
(P)	FAM-CTGCGCCCAAACCCAAGTCATAGC-TAMRA
(R)	TGCAGGATTGAGGCAAGCTT
	<i>B-Factor</i>
(F)	TCCCTCCTGAAGGCTGGAA
(P)	FAM-CATCCTCATGACTGATGGATTGCACAACA-TAMRA
(R)	TGTATAGCAAGTCCCGGATCTCA
	<i>Adlican</i>
(F)	CACCTTCTCCACGTTTCACAT
(P)	FAM-CTCTCCACCATAAGGCACCTCTACTTAGCAGA-TAMRA
(R)	CTGGCAGGAAGAGTTCTAAC
	<i>Endothelin-Converting Enzyme 1</i>
(F)	TCTGGGAACACAACCAAGCA
(P)	FAM-TCGAAAACCTCCACGGCCAGCG-TAMRA
(R)	TCCTCGATCCTGGTCTCGTT

(F): forward primer, (P): dual-labeled probe, (R): reverse primer

the Sanger Centre (<http://www.ensembl.org/>). RNA sequence and exon-exon boundary information was placed into quantitative PCR primer design software (PrimerExpress version 1.5, Applied Biosystems). Primer and probes are summarized in Table 2.

Probes were labeled with the reporter dye 6-carboxyfluorescein (6'-FAM) at the 5' end and with the reporter dye 6-carboxytetramethylrodamine (TAMRA) at the 3' end.

**Acquisition and normalization of data** In each TaqMan run, serial dilutions of a single standard cDNA (derived from one colon cancer) were amplified to create a standard curve, and values of unknown samples were estimated relative to this standard curve. PCR reactions of each sample were run in duplicate; the mean value of two reactions was defined as representative of the sample. TaqMan Ribosomal RNA Primer and Probe (VIC Dye labeled: Applied Biosystems) were used to normalize quantitative data. The formula for normalization was: ratio of sample to reference cDNA = Gene(s)/Gene(r)/(Ribo(s)/Ribo(r)), where Gene(s) and Gene(r) were expression levels of each gene in the sample and reference cDNA, respectively, and Ribo(s) and Ribo(r) were ribosomal RNA expression levels in the sample and reference.

## References

- Ahmed SI, Thompson J, Coulson JM and Woll PJ. (2000). *Am. J. Respir. Cell. Mol. Biol.*, **22**, 422–431.
- Alizadeh A and Eisen M, Botstein D, Brown PO, Staudt LM. (1998). *J. Clin. Oncol. Immunol.*, **18**, 373–379.
- Alizadeh AA, Eisen MB, Davis RE and Ma C, Lossos IS, Rosenwald A, Boldrick J, Sabet H, Tran TY, Xu P, Powell JI, Yang LM, Marti G, Moore T, Hudson Jr JL, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage J, Warnke R, Staudt LM et al. (2000). *Nature*, **403**, 503–511.
- American Cancer Society. (2000). *Cancer Facts and Figures*. ([www.cancer.org](http://www.cancer.org)).
- Bassett Jr DE, Eisen MB, Boguski MS. (1999). *Nat. Genet.*, **21**, 51–55.
- Chu S, DeRisi J and Eisen M, Mulholland J, Botstein D, Brown PO, Herskowitz I. (1998). *Science*, **282**, 699–705.
- Duggan DJ, Bittner M, Chen Y, Meltzer P and Trent JM. (1999). *Nat. Genet.*, **21**, 10–14.
- Eisen MB, Brown PO. (1999). *Methods Enzymol.*, **303**, 179–205.
- Eisen MB and Spellman PT, Brown PO, Botstein D. (1998). *Proc. Natl. Acad. Sci. USA*, **95**, 14863–14868.
- Elek J, Park KH, Narayanan R. (2000). *In Vivo*, **14**, 173–182.
- Grammatico P, Roccella M, Catricala C, Roccella F, Bucher S, Mordenti C and Amantea AD, Di Rosa CD, Del Porto G. (1995). *World J. Surg.*, **19**, 350–351.
- Ikeda Y, Oda S, Abe T, Ohno S, Maehara Y and Sugimachi K. (2001). *Oncology*, **61**, 168–174.

- Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JCF, Trent JM, Staudt LM, Hudson Jr J, Boguski MS, Lashkari D, Shalon D, Botstein D and Brown PO. (1999). *Science*, **283**, 83–87.
- Jensen JA, Carroll RE and Benya RV. (2001). *Peptides*, **22**, 689–699.
- Khan J, Saal LH, Bittner ML, Chen Y, Trent JM and Meltzer PS. (1999). *Electrophoresis*, **20**, 223–229.
- Luo L, Salunga RC, Guo H, Bittner A, Joy KC, Galindo JE, Xiao H, Rogers KE, Wan JS, Jackson MR and Erlander MG. (1999). *Nat. Med.*, **5**, 117–122.
- Meltzer SJ, Zhou D and Weinstein WM. (1989). *Exp. Mol. Pathol.*, **51**, 264–274.
- Nicke B, Riecken EO and Rosewicz S. (1999). *Gut*, **45**, 51–57.
- Patel KV and Schrey MP. (1995). *Br. J. Cancer*, **71**, 442–447.
- Perou CM, Jeffrey SS, van de Rijn M, Rees CA, Eisen MB, Ross DT, Pergamenschikov A, Williams CF, Zhu SX, Lee JC, Lashkari D, Shalon D, Brown PO and Botstein D. (1999). *Proc. Natl. Acad. Sci. USA*, **96**, 9212–9217.
- Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lonning PE, Borresen-Dale AL, Brown PO and Botstein D. (2000). *Nature*, **406**, 747–752.
- Robinson S, Cohen M, Prayson R, Ransohoff RM, Tabrizi N and Miller RH. (2001). *Neurosurgery*, **48**, 864–873.
- Selaru FM, Zou T, Shustova V, Xu Y, Yin J, Mori Y, Sato F, Wang S, Oлару A, Shibata D, Greenwald BD, Krasna MJ, Abraham JM and Meltzer SJ. (2002a). *Oncogene*, **21**, 475–478.
- Selaru FM, Xu Y, Yin J, Zou T, Liu TC, Mori Y, Abraham JM, Sato F, Wang S, Twigg C, Oлару A, Shustova V, Leytin A, Shibata D, Harpaz N and Meltzer SJ. (2002b). *Gastroenterology*, **122**, 606–613.
- Sogawa K, Masaki T, Miyauchi A, Sugita A, Kito K, Ueda N, Miyamoto K, Okazaki K, Okutani K and Matsumoto K. (1997). *Cancer Lett.*, **112**, 263–268.
- Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D and Futcher B. (1998). *Mol. Biol. Cell.*, **9**, 3273–3297.
- Stewart LV and Thomas ML. (1997). *Exp. Cell Res.*, **233**, 321–329.
- Takakura S, Kohno T, Manda R, Okamoto A, Tanaka T and Yokota J. (2001). *Int. J. Oncol.*, **18**, 817–824.
- Tusher VG, Tibshirani R and Chu G. (2001). *Proc. Natl. Acad. Sci. USA*, **98**, 5116–5121.
- Van Gelder RN, von Zastrow ME, Yool A, Dement WC, Barchas JD and Eberwine JH. (1990). *Proc. Natl. Acad. Sci. USA*, **87**, 1663–1667.
- Wang Q, Ding Q, Dong Z, Ehlers RA and Evers BM. (2000). *Anticancer Res.*, **20**, 75–83.