GASTROENTEROLOGY 2005;129:874–884

Gene-Expression Profiling Predicts Recurrence in Dukes' C Colorectal Cancer

DIEGO ARANGO,*,† PAIVI LAIHO,* ANTTI KOKKO,* PIA ALHOPURO,* HELI SAMMALKORPI,* REIJO SALOVAARA,*,§ DANIEL NICORICI,¶ SAMPSA HAUTANIEMI,¶ HAFID ALAZZOUZI,† JUKKA–PEKKA MECKLIN,‖ HEIKKI JÄRVINEN,§ AKSELI HEMMINKI,# JAAKKO ASTOLA,¶ SIMO SCHWARTZ Jr,† and LAURI A. AALTONEN*

*Department of Medical Genetics, Biomedicum Helsinki, University of Helsinki, Helsinki, Finland; †Molecular Biology and Biochemistry Research Center (CIBBIM), Valle Hebron Research Institute, Barcelona, Spain; §Department of Pathology, Haartman Institute, University of Helsinki, Helsinki, Finland; ¶Institute of Signal Processing, Tampere University of Technology, Tampere, Finland; ‖Department of Surgery, Jyväskylä Central Hospital, Jyväskylä, Finland; and #Cancer Gene Therapy Group, University of Helsinki, and Department of Oncology, Helsinki University Central Hospital, Helsinki, Finland

Background & Aims: **Although approximately 50% of Dukes' C colorectal cancer patients are surgically cured, it is currently not possible to distinguish these patients from those at high risk of recurrence. The recent advent of routine adjuvant chemotherapy for these patients has greatly complicated the identification of new markers predicting the response to surgery, which is now reliant on archived materials. Microarray analysis allows fine tumor classification but cannot be used with paraffin-embedded archival samples.** ***Methods:*** **We used microarray analysis of a unique set of fresh-frozen tumor samples from Dukes' C patients who had surgery as the only form of treatment to identify molecular signatures that characterize tumors from patients with good and bad prognosis.** ***Results:*** **Unsupervised hierarchical clustering and a K-nearest neighbors–based classifier identified groups of patients with significantly different survival ($P = .019$ and $P = .0001$). Expression profiling outperformed previously reported genetic markers of prognosis such as *TP53* and *K-RAS* mutational status and allelic imbalance in chromosome 18q, which were of limited prognostic power in this study. Functional categories significantly enriched in gene-expression differences included protein transport and folding. The prognostic potential of the *RAS* homologue *RHOA*, one of the most differentially expressed genes, was further investigated using immunohistochemistry and a tissue microarray containing 137 independent Dukes' C tumor samples. Reduced RHOA expression was associated with significantly shorter survival ($P = .01$).** ***Conclusions:*** **This study shows that gene-expression profiling of surgical tumor samples can predict recurrence in Dukes' C patients. Therefore, this approach could be used to guide decisions concerning the clinical management of these patients.**

Colorectal cancer is one of the leading causes of cancer-related death in the Western world, and a large proportion of these patients are diagnosed with locally advanced disease with regional lymph node metastasis (Dukes' C stage). Despite the high incidence of this disease, the clinical management of Dukes' C colorectal cancer patients is far from optimal. Surgical resection prevents recurrence in approximately 50% of Dukes' C patients. Because of the relatively high risk of recurrence in these patients, several large multi-institutional clinical trials have been performed in the last 2 decades to investigate the benefits of adjuvant chemotherapy. 5-Fluorouracil (5-FU)–based therapy was found to prevent recurrence in 10%–20% of patients.[1,2] However, because it is currently not possible to accurately distinguish surgically cured patients from those at high risk of recurrence, the great majority of Dukes' C patients are administered adjuvant chemotherapy; this results in a large number of surgically cured patients undergoing chemotherapeutic treatment without benefit. In addition, routine administration of adjuvant chemotherapy to these patients greatly complicates the identification of surgically cured patients, because it is not possible to distinguish patients cured by surgery from those with a good response to the adjuvant treatment. Therefore, experiments addressing this question heavily rely on archived materials collected before adjuvant treatment was incorporated into routine clinical practice.

Abbreviations used in this paper: 5-FU, 5-fluorouracil; PCR, polymerase chain reaction; tRNA, transfer RNA.

0016-5085/05/$30.00
doi:10.1053/j.gastro.2005.06.066

The potential of several markers, such as *TP53* or *K-RAS* mutational status and loss of heterozygosity in chromosome 18q, to identify patients who vary in their probability of recurrence after surgery has been previously investigated.[3–12] However, the accuracy of these markers is limited, and conflicting reports can be found in the literature.[3–12] As a result, these markers have not become routinely used in the clinical management of these patients. Identification of additional means of distinguishing patients with good and poor prognosis after surgery would allow targeting aggressive chemotherapeutic treatment to patients who can potentially benefit from it and would spare surgically cured patients the side effects, toxicity, and cost associated with such treatment.

Microarray analysis allows assessment of the level of expression of thousands of genes simultaneously,[13–16] and the potential of expression profiling as a tool to predict the prognosis of cancer patients has been previously realized for different types of cancer, including lymph node–negative Dukes' B colorectal cancer.[17–20] In addition, microarray analysis can provide significant insight into biological differences between tumors from patients with good and bad prognosis and can be used as a screening tool to find individual markers that could identify groups of patients who differ in their prognosis. However, it is currently not possible to use microarray analysis with RNA extracted from archived paraffin-embedded tumor samples.

In this study, we used high-density oligonucleotide microarray analysis and a unique set of fresh-frozen tumor samples from Dukes' C colorectal cancer patients who had surgery as the only form of treatment, to identify patterns of gene expression that characterize tumors from patients with good and bad prognosis. Moreover, we compared the accuracy of these profiles of expression with that of other genetic markers of prognosis and found that expression profiling outperformed other markers tested. In addition, the level of protein expression of the *RAS* homologue *RHOA*, one of the genes with the most significantly different expression at the messenger RNA level in tumors from patients with good and bad prognosis, was identified as a useful immunohistochemical marker capable of predicting the prognosis of an independent set of 137 tumor samples from Dukes' C colorectal cancer patients.

Materials and Methods

Patients, Samples, and RNA Extraction

Among the 1042-fresh frozen colorectal cancer samples available in our tumor bank (collected from 1994 to 1998), we selected samples from Dukes' C patients (281 patients) who had surgery as the only form of treatment (91 patients). Patients who died of causes unrelated to colorectal cancer were excluded. These samples were collected before adjuvant chemotherapy became standard practice for these patients and constitute, therefore, a unique set of samples to investigate response to surgery. Moreover, complete follow-up was available for all these patients for at least 6 years (mean follow-up, 8.9 years), thus allowing analysis of long-term survival. Informed consent for genetic analysis of the tumor sample was obtained from each patient according to the Human Investigations and Ethical Committee–approved research proposal. Frozen sections from OCT-embedded samples (Tissue-Tek, Zoeterwoude, The Netherlands) of these 91 tumors were cut, and after toluidine blue staining and histological verification, 47 of them were found to contain at least 70% tumor cells. Frozen tissue samples were macrodissected from the selected areas of the frozen OCT blocks and homogenized by using a tissue homogenizer (Ultra-Turrax T8; IKA Labortechnik, Staufen, Germany) in 1 mL of TRIzol reagent (Invitrogen, Carlsbad, CA). Total RNA was extracted with TRIzol as recommended by the manufacturer, followed by an RNeasy cleanup step (Qiagen, Germantown, MD). The RNA quality of these 47 samples was then assessed with a 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA), and 25 of them showed optimal quality for oligonucleotide microarray analysis, as shown the integrity of the observed 18S and 28S ribosomal RNA bands. Of these, 10 had disease recurrence within 5 years of surgery (poor prognosis), and 15 had at least 5 years of disease-free survival after surgical treatment (good prognosis) (Table 1).

Microarray Analysis

Target preparation and hybridization. Eight micrograms of total RNA was reverse-transcribed, and the resulting complementary DNA template was used for the in vitro transcription reaction. This resulted in biotin-labeled chromosomal RNA as described in the Supplementary Materials (http://research.med.helsinki.fi/cancerbio/gastro). Microarrays were scanned with a GeneChip Scanner GA2500, and MAS 5.0 software (both from Affymetrix, Santa Clara, CA) was used to obtain quantitative expression information for all 22,283 probe sets in the HG-U133A microarrays. Detection call cutoff values used for present and marginal genes were $\alpha_1 = 0.05$ and $\alpha_2 = 0.065$, respectively. The percentage of present and marginal calls ranged from 37% to 52%, and the 3′/5′ ratio of β-actin and glyceraldehyde phosphate dehydrogenase control genes was <2.8. All chips were then scaled to the average gene expression of all 25 chips, and the expression of each probe set was divided by the mean expression of that probe across all the samples so that the resulting mean expression for every probe was 1.

Identification of differentially expressed genes. Only genes with detectable expression levels in at least 50% of the samples (10,035 genes) were included in this analysis. Gene expression for samples with absent calls was substituted with the gene average expression across all samples to allow

Table 1. Clinical Features of Patients in the Study and Results of the K-Nearest Neighbors Classifier

Patient no.	MSI status	Grade[a]	Age (*y*)	Dukes'	DFS (*y*)	True prognosis[b]	Predicted prognosis[c]
367	MSS	3	88	C	0.5	Bad	Bad
176	MSS	2	63	C	0.5	Bad	Good
146	MSS	2	61	C	0.6	Bad	Bad
616	MSS	2	55	C	0.7	Bad	Bad
981	MSS	2	56	C	0.9	Bad	Bad
707	MSS	2	86	C	1.1	Bad	Good
416	MSS	2	63	C	1.2	Bad	Bad
931	MSS	2	63	C	2.2	Bad	Bad
986	MSS	2	83	C	2.5	Bad	Bad
463	MSS	2	58	C	2.6	Bad	Bad
2000	MSS	2	68	C	5.5	Good	Good
988	MSS	2	55	C	6.0	Good	Good
917	MSS	2	72	C	6.2	Good	Good
939	MSS	1	64	C	6.4	Good	Good
711	MSS	2	72	C	7.0	Good	Good
634	MSS	1	86	C	7.0	Good	Good
613	MSS	2	74	C	7.1	Good	Good
482	MSS	1	84	C	7.2	Good	Bad
476	MSS	3	59	C	7.5	Good	Good
466	MSS	1	70	C	7.7	Good	Good
381	MSS	2	83	C	8.0	Good	Good
330	MSS	2	81	C	8.1	Good	Good
280	MSS	1	66	C	8.2	Good	Good
259	MSS	2	70	C	8.2	Good	Good
117	MSS	2	84	C	9.1	Good	Good

MSI, microsatellite instability; MSS, microsatellite stable; DFS, disease-free survival.
[a]Histological grade.
[b]Observed prognosis.
[c]Prognosis predicted by the 5-nearest neighbors (NN) classifier using 17 genes. Bad, recurrence within 5 years of surgery; good, no recurrence within 5 years of surgery.

class permutation analysis (see below) as previously described.[21] Genes differentially expressed in patients with good and bad prognosis were identified by using a Student *t* test. The class labels were then randomly permuted, and a new *t* test was run with the new groupings for every gene. This process was repeated 1 million times, and the percentage of times was scored in which a *P* value was lower than the true *P* value initially calculated. We selected as differentially expressed those genes that were found to have a *P* value lower than the initial *P* value in $<1\%$ of the permutations.

Functional group enrichment analysis. Gene ontology terms (http://www.geneontology.org) were used to annotate all genes with present calls in at least 50% of samples (10,035 genes). A total of 1384 partially overlapping functionally related categories of genes were identified. Categories with <5 genes or >500 genes were removed from this analysis, because they were considered to be too specific or too general to be useful. Using the list of 218 unique genes differentially expressed, we identified categories that were significantly enriched in the number of genes with different expression in tumors from patients with good and bad prognosis ($P < .01$). These analyses were implemented using GoMiner software,[22] and a Fisher exact test was used to identify significantly enriched categories.

Unsupervised hierarchical clustering. The 5102 genes that were expressed at detectable levels (present and marginal calls) in all 25 samples were selected, and then all data were $\log_2$-transformed and the intensity of both the arrays and the genes were centered on the median. These data were used to hierarchically cluster all 25 samples by using Cluster software and were visualized by using TreeView software.[23]

Classification of tumors from patients with good and bad prognosis. The 5102 genes with present or marginal calls in all 25 samples were used for these analyses. Each gene profile was discretized independently to binary values across the samples by applying the Lloyd algorithm.[24] The Lloyd algorithm minimizes the average discretization error, which represents the distance between the data and the discrete representation. Also, the Lloyd algorithm can be understood as a particular case for 1-dimensional data of the K-means clustering method. These data were then used to generate a K-nearest neighbors classifier capable of using the expression profile of these 25 samples to distinguish patients with good and bad prognosis. These analyses were implemented as previously described by using GeneCluster II software.[17] The number of genes used in this classifier varied from 1 to 100, and the number of neighbors varied from 1 to 10. The predictions of the results of every classifier tested were then validated using a leave-one-out cross-validation approach as previously described.[17,25,26] Briefly, 1 sample is initially removed from the analysis, and the remaining 24 samples are used to build a classifier. The generated classifier is then

applied to the sample initially left out, and the class call is recorded. This cross-validation loop is iteratively repeated 25 times each time leaving out a different sample, and the final result is 25 true class/predicted class pairs. The accuracy of the predictor can thus be assessed by the number of correct calls made by the classifier.[17,25,26] As an additional control, the sample class labels were randomly permuted 200 times, and the average prediction accuracy of the classifier for these random groupings was computed. Using a leave-one-out approach minimizes the bias that could occur when trying to validate a classifier with the sample set used for training of the classifier.

Quantitative Real-Time Reverse-Transcription Polymerase Chain Reaction

Aliquots of RNA (100 ng) were reverse-transcribed using Superscript II according to the manufacturer's recommendations (Invitrogen). Five microliters of the undiluted reverse transcriptase reaction were used to polymerase chain reaction (PCR)-amplify RHOA, CLTC, MTMR1, ARCN1, and IDH3G by using Assays-on-Demand TaqMan primers and probes and TaqMan Universal PCR Master Mix in a GeneAmp 5700 Sequence Detection System (all from Applied Biosystems, Branchburg, NJ). Relative gene-expression levels were quantified with the $\Delta\Delta C_T$ method with β-actin as a housekeeping standard control, as previously described.[26–28]

TP53, *K-RAS*, and *RHOA* Sequencing and Assessment of 18q Allelic Imbalance

An extended set of 46 fresh-frozen colorectal cancer samples were screened for *K-RAS* and *TP53* mutations and were assessed for allelic imbalance in chromosome 18q. The mutation hotspots of *K-RAS* (codons 12, 13, and 61) were PCR-amplified as described previously.[29,30] Exons 2–11 of TP53 were PCR-amplified as previously described.[25,31] The complete coding sequences of *RHOA* (exons 2–5) were sequenced in 5 of the tumor samples that showed the highest RHOA protein levels and in 5 of the tumors with the lowest expression, as described at http://research.med.helsinki.fi/cancerbio/gastro. Two polymorphic microsatellite markers in 18q21 (D18S1110 and D18S1156) were used to assess the allelic imbalance in this region, as previously reported.[28,32] The size of the amplified PCR products was assessed by using an Applied Biosystems ABI3730 Automatic DNA sequencer. Allelic imbalance was scored if there was a difference >40% in the abundance of an allele between normal and tumor samples.[32,33]

Tissue Microarray and Immunohistochemistry

An independent set of formalin-fixed, paraffin-embedded samples from 137 Dukes' C colorectal tumors and 16 lymph node metastases was used for immunohistochemical assessment of RHOA expression levels using a tissue microarray. All 137 tumor samples were from patients with Dukes' C colorectal cancer: 61 of them had surgery as the only form of treatment, and 76 received also 5-FU–based adjuvant chemotherapy. After histological examination of H&E-stained tumor sections, areas containing a high proportion of tumor cells were selected. Triplicate 0.6-mm cores from every sample were arrayed in a fresh paraffin block using a Beecher Instruments tissue arrayer (Silver Spring, MD).

Unstained 4-μm sections from the tissue microarray were mounted on slides coated with 3-aminopropyl-triethoxy-silane (Sigma, St Louis, MO). The sections were treated with 3% H_2O_2 solution for 10 minutes to block endogenous peroxidase activity. Sections were then treated in a microwave oven in 10 mmol/L sodium citrate buffer (pH 6) for 5 minutes at 800 W and 10 minutes at 450 W. For immunohistochemical analysis, the commercial PowerVision Poly-HRP IHC detection kit was used (ImmunoVision Technologies, Brisbane, CA), as previously reported.[28,32] An anti-RHOA monoclonal antibody raised against a peptide corresponding to amino acids 120–150 of RHOA of human origin was used at a 1:1000 dilution for 1 hour at room temperature (RHOA clone 26C4; Santa Cruz Biotechnology Inc, Santa Cruz, CA). The specificity of this antibody has been previously tested in formalin-fixed paraffin-embedded samples.[34] RHOA expression was evaluated in the 137 triplicate tumor samples and 16 lymph node metastases with the investigator blinded to the clinical data. A semiquantitative scale from 0 to 4 was used to measure the intensity of the staining.

To investigate how survival differences in the high- and low-RHOA groups change as a function of the staining cutoff level selected, we systematically calculated the mean survival in both groups, as well as the hazard ratio and the log-rank *P* value for every possible grouping resulting from increasing the number of patients allocated to the low-RHOA group from 1 to 137, starting with the patient with the lowest RHOA tumor level (see Supplementary Materials, Table 4; available at http://research.med.helsinki.fi/cancerbio/gastro).

Results

Genes Differentially Expressed in Tumors From Patients With Good and Bad Prognosis

A total of 236 sequences (2.3% of the total), representing 218 unique genes, were identified as differentially expressed between tumors from patients with or without recurrence within 5 years of initial surgery (poor and good prognosis, respectively) among the 10,035 sequences with present calls in at least 50% of the samples (Supplementary Materials, Table 1; available at http://research.med.helsinki.fi/cancerbio/gastro). Of these, 160 had lower expression and 58 had higher expression in tumors from patients with poor prognosis compared with tumors from patients with good prognosis. The relative expression of 5 genes showing significant expression differences and used in all 25 iterations of the K-nearest neighbors–based classifier (see below; *RHOA*, *CLTC*, *MTMR1*, *ARCN1*, and *IDH3G*) was inde-

pendently determined using quantitative real-time reverse-transcription PCR in the 22 tumors for which sufficient RNA was available and could be optimally PCR-amplified. A highly significant correlation was observed between the 110 determinations made by real-time reverse-transcription PCR and oligonucleotide microarray analysis (Pearson's $r = .4$, $P < .0001$; Supplementary Materials, Figure 1; available at http://research.med.helsinki.fi/cancerbio/gastro).

Functional Group Enrichment Analysis

Gene ontology terms were used to classify all 10,035 genes with a present call in at least 50% of the tumors into 1385 biological processes, cellular components, and molecular functions. Functional group enrichment analysis[22,35] was used to identify categories with a significant enrichment in the number of genes differentially expressed in tumors from patients with good and bad prognosis (see Materials and Methods). Using a Fisher exact test, we identified 24 partially overlapping subcategories with a P value $<.01$ among the 1385 gene categories tested (Supplementary Materials, Table 2; available at http://research.med.helsinki.fi/cancerbio/gastro). These overlapping subcategories can be grouped into 6 main functional groups: protein transport, protein folding, transfer RNA (tRNA) ligase activity, chemotaxis, muscle contraction, and negative regulation of enzyme activity (Supplementary Materials, Table 2; available at http://research.med.helsinki.fi/cancerbio/gastro).

Use of Microarray Analysis to Identify Patients With Good and Bad Prognosis

We next used gene-expression profiling of the tumor samples obtained at the time of surgery to distinguish patients with good and bad long-term prognosis. Unsupervised hierarchical clustering using the expression levels of the 5102 sequences that were expressed at detectable levels in all 25 samples in the study showed that most (7 of 10) tumors from patients who had recurrence within 5 years after surgery (bad prognosis) clustered together in 1 of the 3 main sub-branches generated and that 12 of the 15 tumors from patients with good prognosis clustered in the other 2 sub-branches (Figure 1*A*). Moreover, a significant difference ($P < .019$) was observed in the disease-free survival of patients in the sub-branch dominated by patients with bad prognosis and the 2 sub-branches with mainly patients with good prognosis (Figure 1*B*).

The expression of these 5102 genes was then used to build a K-nearest neighbors–based supervised classifier capable of discriminating with higher accuracy tumors from patients with good and bad prognosis (see Materials and Methods). We used a leave-one-out approach to minimize the possible bias created when trying to validate a classifier by using the sample set used to train it. With this supervised approach, 1 sample is initially removed from the analysis, and the remaining 24 samples are used to identify the N genes that are most

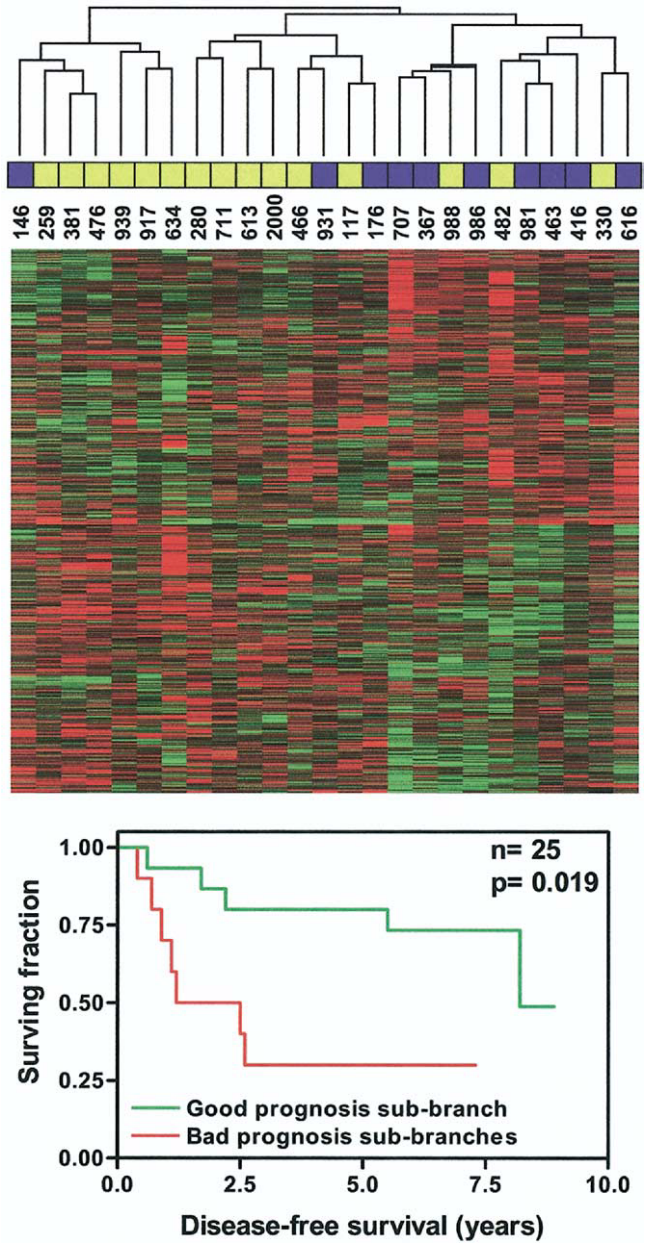

Figure 1. Hierarchical clustering. (*A*) Results of an unsupervised hierarchical clustering of all 25 Dukes' C colorectal tumors with the 5102 genes expressed at detectable levels in all samples. Each column represents a sample, and each row, a gene. Gene expression above and below the median is depicted in red and green, respectively. The blue and yellow squares in the top bar indicate tumors from patients with bad prognosis and good prognosis, respectively. (*B*) Disease-free survival of patients in the sub-branch dominated by bad prognosis patients (*right*) and the other 2 main sub-branches (Kaplan–Meier plot). The log-rank P value is shown.

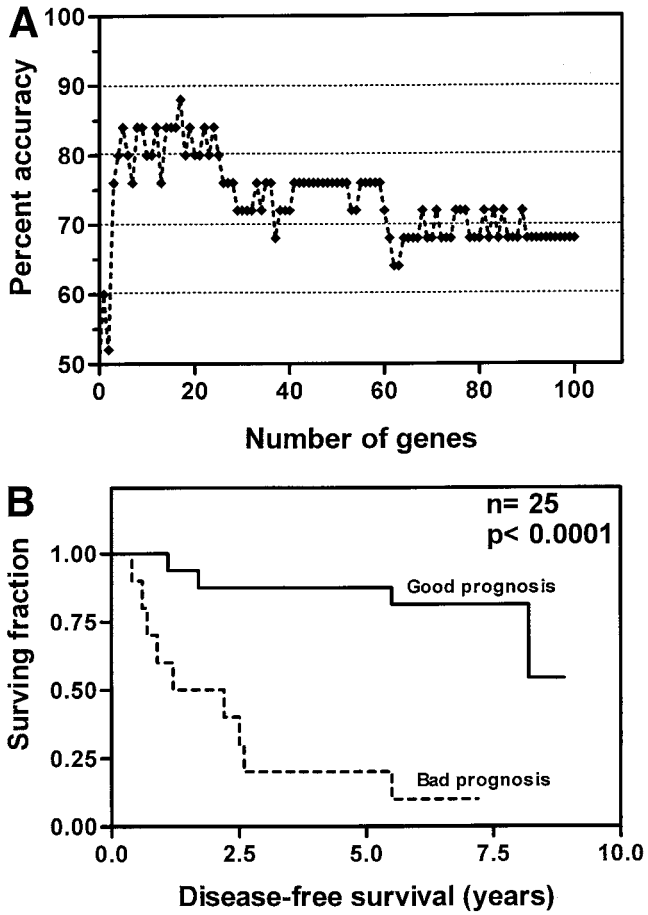

Figure 2. Supervised K-nearest neighbors (KNN) classifier. (*A*) Accuracy of a KNN classifier with 5 neighbors as a function of the number of genes used by the classifier. (*B*) Disease-free survival of patients predicted to have good and bad long-term prognosis by a classifier based on the 5 nearest neighbors and 17 genes (Kaplan–Meier plot). The log-rank *P* value is shown.

differentially expressed in tumors from patients with good and bad prognosis. The expression of these N genes in the sample that was initially left out of the analysis is used to identify the 5 samples with the closest expression profile. The predominant class label (good or bad prognosis) in these 5 nearest neighbors is then given to the sample initially left out and is then compared with the actual class label to assess the accuracy of the classifier. This process is iteratively repeated 25 times, each time leaving out a different sample, and then the observed and predicted prognosis of all 25 patients can be compared.

We varied the number of genes used to build the classifier from 1 to 100 and found that when 17 genes were used, the prognosis of 22 of the 25 samples (88%) was correctly predicted (Figure 2*A* and Table 1). Because every iteration of the cross-validation procedure uses a slightly different group of 17 genes, a total of 72 genes were used at least once in this procedure (Supplementary Materials, Table 3; available at http://research.med.helsinki.fi/cancerbio/gastro). The sensitivity and specificity of this classifier to identify patients with poor prognosis were 80% and 93.3%, respectively, and this was significantly different from what would be expected by chance ($P < .0003$). As an additional control, the class labels were randomly permuted 200 times. As expected, and in contrast with what was observed with the true sample labels, the average accuracy of the classifier in these random groupings was not significantly different from what would be expected by chance (mean accuracy, 53.7%; SD, 15.0%) and was always below the 88% observed with the real class labels. Importantly, this classifier was able to identify 2 groups of patients with significantly different ($P < .0001$) disease-free survival after surgery (Figure 2*B*).

TP53 and *K-RAS* Mutational Status and Allelic Imbalance in 18q Were Not Predictive of Recurrence

We next wanted to compare the accuracy of this prognostic marker with that of other genetic makers previously investigated, such as losses of genetic material in chromosome 18q and the mutational status of *TP53* and *K-RAS*[3–12] in an extended sample set that included the samples entered into the microarray experiments. As shown in Figure 3*A*, no differences in survival were observed in patients whose tumors had a mutant or wild-type p53 gene. In agreement with previous reports,[4,6–8] patients with tumors that had a mutant *K-RAS* gene or with an allelic imbalance in chromosome 18q showed a trend toward a worse prognosis compared with patients who had wild-type *K-RAS* or tumors with no 18q allelic imbalance, respectively, but the differences in survival were not significant (Figure 3*B* and *C*).

Role of RHOA as a Prognostic Marker in Dukes' C Colorectal Cancer

Our expression microarray experiments identified 58 genes with higher and 160 genes with lower expression levels in tumors from patients with poor prognosis compared with tumors from patients with good prognosis (Supplementary Materials, Table 1; available at http://research.med.helsinki.fi/cancerbio/gastro). The *RAS* homologue *RHOA* was one of the genes with the most significant difference in expression between tumors from patients with good and bad prognosis (Supplementary Materials, Table 1; available at http://research.med.helsinki.fi/cancerbio/gastro). Because RAS signaling is one of the processes that is most frequently deregulated in colorectal tumorigenesis, we decided to investigate the potential of this *RAS* homologue as a prognostic marker in Dukes' C colorectal cancer by using an immunohis-

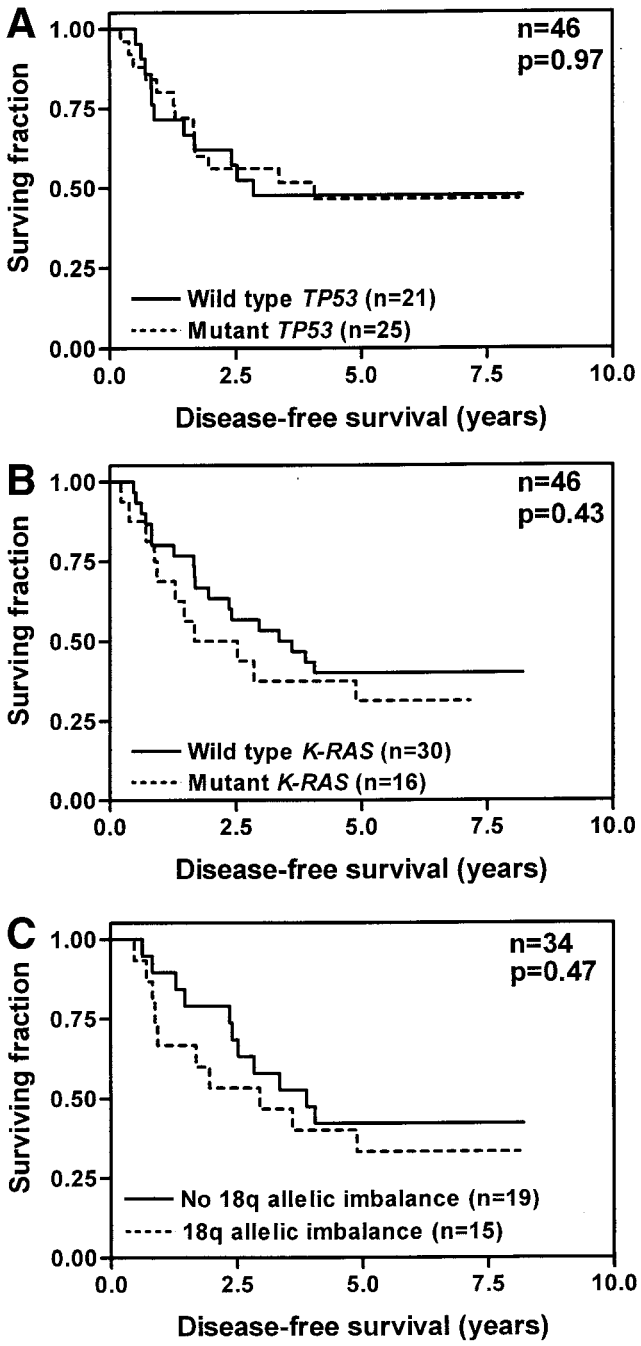

Figure 3. Genetic markers of prognosis. Disease-free survival of Dukes' C colorectal cancer patients by (*A*) *TP53* and (*B*) *K-RAS* mutational status and (*C*) allelic imbalance in chromosome 18q (Kaplan–Meier plots). Log-rank *P* values are shown.

tochemical approach and a completely independent set of samples. For this purpose, we constructed a tissue microarray that contained triplicate formalin-fixed, paraffin-embedded tumor samples from 137 unselected colorectal cancer patients and 16 lymph node metastases. Sections of this tissue microarray were immunostained with a RHOA monoclonal antibody, and RHOA levels were assessed with a semiquantitative scale (Figure 4*A*).

To investigate possible survival differences in patients with tumors with high and low RHOA levels as a function of the intensity cutoff selected to define the groups, we systematically analyzed the survival curves for every possible grouping resulting from increasing the number of patients allocated to the low-RHOA-expression group from 1 to 137. The mean disease-free survival in the group with high RHOA was higher than in the low-RHOA group for 131 of the 137 groupings (Supplementary Materials, Table 4; available at http://research.med.helsinki.fi/cancerbio/gastro). This difference reached statistical significance (log-rank test; $P < .05$) for 28 of these groupings. The dependence of this result on the cutoff selected to allocate samples to the high- and low-RHOA groups will need confirmation in an independent-sample set. Analysis of Kaplan–Meier survival curves generated with a staining intensity threshold of 1.4 showed that patients with tumors expressing low levels of RHOA had significantly worse overall ($P = .03$) and disease-free ($P = .01$) survival than patients with tumors with high RHOA levels (Figure 4*B*). No significant associations were observed between RHOA protein levels and other clinicopathologic parameters such as patient sex, age, histological grade, and tumor location (Supplementary Materials, Table 5; available at http://research.med.helsinki.fi/cancerbio/gastro).

To investigate the possible role of genomic *RHOA* mutations in the levels of protein expression, 5 of the tumor samples showing the highest RHOA protein levels and 5 of the tumors with the lowest expression were screened for mutations in the complete RHOA coding sequences. Consistent with previous studies,[36,37] no mutations were found in these 10 selected tumors.

Discussion

Although histopathologic staging of colorectal tumors provides a useful tool to identify groups of patients who vary in their prognosis, considerable variability exists in the long-term survival of patients within each class. Approximately 50% of the Dukes' C patients will have disease recurrence and die as a result of their disease after potentially curative surgery, whereas the other half are surgically cured, and it is currently not possible to distinguish between these 2 subgroups of patients.[1,2] The advent of routine adjuvant treatment for these patients over the last decade has significantly improved the survival of 10%–20% of these patients. However, it has also complicated the study of prognostic factors that predict recurrence after surgery, because chemotherapy confuses the interpretation of results obtained with recent tumor collections. Future progress in iden-

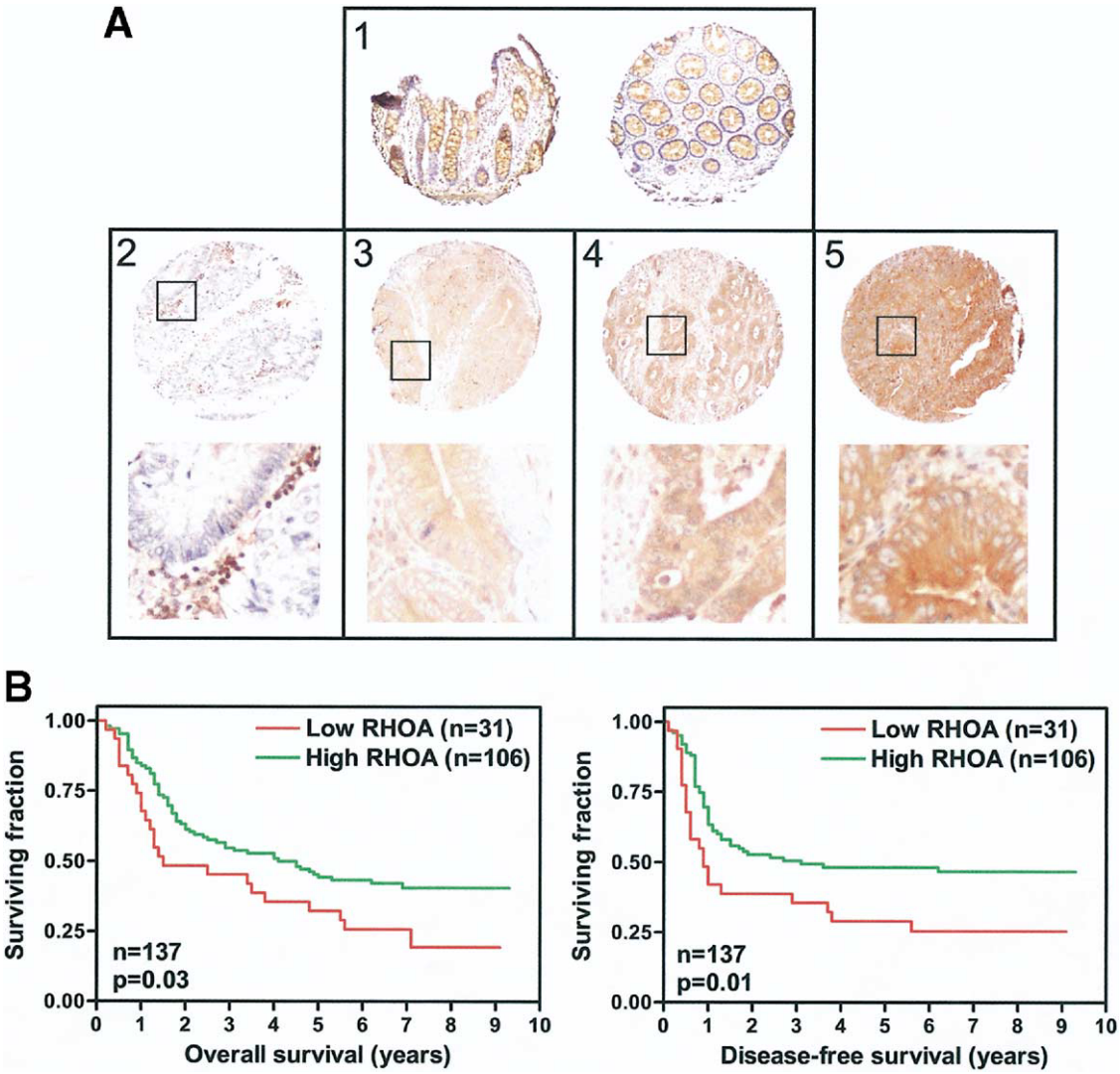

Figure 4. RHOA immunohistochemical staining. (*A*) Representative examples of normal colon samples (1) and tumors showing low (2), intermediate (3 and 4), or high (5) RHOA immunostaining (original magnification, 200×). Higher-power magnification is shown in the lower part. (*B*) Overall and disease-free survival of 137 Dukes' C colorectal cancer patients with high and low RHOA tumor levels (Kaplan–Meier plots). Log-rank *P* values are shown.

tifying new markers capable of predicting the probability of disease recurrence after surgery heavily relies on archived tumor samples from patients who did not receive adjuvant treatment. Microarray analysis is a powerful new technique that allows the molecular characterization of tumors, but paraffin-embedded archival materials are currently not suitable for this type of analysis. In this study, we used a unique collection of fresh-frozen tumor samples collected from 1994 to 1998 (minimum follow-up, 6 years; mean, 8.9 years) and used high-density oligonucleotide microarray analysis to assess the level of expression of 22,283 sequences to identify profiles of expression that characterize Dukes' C colorectal tumors from patients at high risk of recurrence. Starting from the 1042 tumors available in our tumor collection, we selected tumors from Dukes' C patients (218 patients) who had surgery as the only form of treatment (91 patients). Optimal-quality RNA could be extracted from 25 of the 47 tumor samples that had at least 70% tumor cells.

A total of 218 distinct genes showed a significant difference in expression in tumors from patients with good and bad prognosis. Gene ontology annotations were used to classify genes into functionally related categories, and a functional group enrichment analysis was used to identify biologically relevant categories enriched in the number of genes differentially expressed. This approach found a significant difference between tumors from patients with good and bad prognosis in the level of expression of 20 genes involved in protein transport. Of

these 20 genes, 17 showed higher expression levels in tumors from patients with good prognosis, and the remaining 3 genes had higher levels in tumors from patients who had disease recurrence within 5 years of surgery (Supplementary Materials, Table 2; available at http://research.med.helsinki.fi/cancerbio/gastro). Most of these genes (*ARCN1*, *ARF1*, *ARF4*, *CLTA*, *CLTC*, *GDI2*, *RH0A*, *SDCBP*, *SEC24D*, *VCP*, and *VPS16*) are involved in vesicle trafficking, a process that regulates multiple signaling mechanisms and has previously been linked to tumorigenesis.[38,39] The remaining genes in this category are involved in nuclear transport (*KPNA2*, *KPNB1*, *RAN*, and *RANBP2*), mitochondrial transport (*HSPD1* and *TIMM23*), or endoplasmic reticulum trafficking (*SSR1*). All these are processes previously shown to be deregulated in colorectal cancer.[40–42] Protein folding is a related category in which a coordinate underexpression of at least 10 genes was observed in tumors from patients with poor prognosis. Genes such as *HSPD1*, *PPID*, and *CALR* are key to ensuring that proteins are correctly folded to be functionally active and play an important role in tumorigenesis.[43–45]

Another category with a significant enrichment in the number of genes differentially expressed between patients with good and bad prognosis was tRNA aminoacylation for protein translation. At least 4 of the enzymes that charge tRNAs with their cognate amino acids showed lower expression levels in tumors from patients with poor prognosis (AARS, EPRS, NARS, and TARS). This family of proteins is essential in protein synthesis but has recently been realized to have important additional functions in key processes such as tRNA processing, RNA splicing and trafficking, ribosomal RNA synthesis, apoptosis, angiogenesis, and inflammation.[46,47]

Remarkably, unsupervised hierarchical clustering of all tumor samples identified an expression signature associated with tumors from patients with bad prognosis. Seven of the 10 patients with poor prognosis clustered together within 1 of the main 3 sub-branches of the tree. The survival of the patients in this sub-branch was significantly ($P = .019$) worse than that of the patients in the other 2 sub-branches. This indicates that primary Dukes' C colorectal tumors from patients at high risk of recurrence have a distinct gene-expression signature at the time of surgery that determines the course of disease. To better discriminate patients with good and bad prognosis, we used a supervised K-nearest neighbors approach. A classifier using the 5 nearest neighbors and 17 genes, correctly classified 88% (22 of 25) of the patients as having a high or low probability of recurrence. A leave-one-out cross-validation approach was able to distinguish 2 subgroups of patients who had a highly significant difference in their long-term disease-free survival ($P < .0001$). Moreover, microarray-based expression profiling outperformed other genetic markers previously investigated, such as *TP53* and *K-RAS* status or allelic imbalance in chromosome 18q,[3–12] which in this study showed limited prognostic power in an extended sample set. Collectively, these results formally show for the first time the potential of gene-expression profiling to predict the probability of recurrence of Dukes' C colorectal cancer after surgery.

The RAS family of small guanosine triphosphate–binding proteins regulates cell growth and the actin cytoskeleton, and RAS signaling deregulation is a common early event in colorectal cancer progression.[48,49] One of the genes with the most significantly reduced expression in tumors from patients with bad prognosis compared with tumors from good-prognosis patients was the *RAS* homologue gene *RH0A*. This gene has been shown to regulate a signal transduction pathway that links plasma membrane receptors to the assembly of focal adhesions and actin stress fibers.[50] Recently it has been reported that high levels of RHOA can inhibit cell motility.[51,52] This led us to investigate the potential of RHOA as a prognostic marker in Dukes' C colorectal cancer. Using immunohistochemistry and a tissue microarray, we assessed the level of expression of RHOA in an independent set of 137 formalin-fixed paraffin-embedded tumor samples from Dukes' C patients. Patients with low RHOA tumor levels had significantly worse overall ($P = .03$) and disease-free ($P = .01$) survival compared with patients whose tumors had high RHOA protein levels. In agreement with this, lymph node metastases showed significantly ($P = .017$) lower levels of RHOA protein than the primary tumors; this further suggests that reduced RHOA expression could favor tumor spread. Moreover, the shorter survival of patients with low RHOA tumor protein levels could be observed in patients who had surgery as the only form of treatment and in those who, in addition, received 5-FU–based adjuvant chemotherapy. These results indicate that RHOA tumor expression levels can be a useful marker for predicting the probability of recurrence of Dukes' C patients treated with or without 5-FU–based chemotherapy. Therefore, RHOA could be used to identify a subset of patients with a higher probability of recurrence and for whom a more aggressive treatment may be justified. Although adjuvant treatment with 5-FU has become standard treatment of Dukes' C colorectal cancer, Irinotecan and oxaliplatin have now clearly been shown to be effective alternative agents.[53–57] Reduced RHOA levels could therefore identify a group of patients who could

benefit from combined treatment with 5-FU and CPT-11 and/or oxaliplatin.

In conclusion, in this study we show that primary Dukes' C colorectal tumors that are likely to form distant metastases display a distinctive expression signature; high-density oligonucleotide microarray analysis accurately distinguished patients with good and bad prognosis after surgery. Moreover, we show that microarray-based genome-wide screening for genes with different levels of expression at the messenger RNA level can be used to identify single prognostic markers for these patients. RHOA was found to be one such marker that could be used to identify a subset of patients with poor prognosis who could benefit from more aggressive treatment.

References

1. Moertel CG, Fleming TR, Macdonald JS, Haller DG, Laurie JA, Goodman PJ, Ungerleider JS, Emerson WA, Tormey DC, Glick JH, et al. Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. N Engl J Med 1990;322:352–358.
2. Moertel CG, Fleming TR, Macdonald JS, Haller DG, Laurie JA, Tangen CM, Ungerleider JS, Emerson WA, Tormey DC, Glick JH, et al. Fluorouracil plus levamisole as effective adjuvant therapy after resection of stage III colon carcinoma: a final report. Ann Intern Med 1995;122:321–326.
3. Bell SM, Scott N, Cross D, Sagar P, Lewis FA, Blair GE, Taylor GR, Dixon MF, Quirke P. Prognostic value of p53 overexpression and c-Ki-ras gene mutations in colorectal cancer. Gastroenterology 1993;104:57–64.
4. Benhattar J, Losi L, Chaubert P, Givel JC, Costa J. Prognostic significance of K-ras mutations in colorectal carcinoma. Gastroenterology 1993;104:1044–1048.
5. Goh HS, Yao J, Smith DR. p53 point mutation and survival in colorectal cancer patients. Cancer Res 1995;55:5217–5221.
6. Ogunbiyi OA, Goodfellow PJ, Herfarth K, Gagliardi G, Swanson PE, Birnbaum EH, Read TE, Fleshman JW, Kodner IJ, Moley JF. Confirmation that chromosome 18q allelic loss in colon cancer is a prognostic indicator. J Clin Oncol 1998;16:427–433.
7. Lanza G, Matteuzzi M, Gafa R, Orvieto E, Maestri I, Santini A, del Senno L. Chromosome 18q allelic loss and prognosis in stage II and III colon cancer. Int J Cancer 1998;79:390–395.
8. Watanabe T, Wu TT, Catalano PJ, Ueki T, Satriano R, Haller DG, Benson AB III, Hamilton SR. Molecular predictors of survival after adjuvant chemotherapy for colon cancer. N Engl J Med 2001;344: 1196–1206.
9. Laurent-Puig P, Olschwang S, Delattre O, Remvikos Y, Asselain B, Melot T, Validire P, Muleris M, Girodet J, Salmon RJ, et al. Survival and acquired genetic alterations in colorectal cancer. Gastroenterology 1992;102:1136–1141.
10. Jen J, Kim H, Piantadosi S, Liu ZF, Levitt RC, Sistonen P, Kinzler KW, Vogelstein B, Hamilton SR. Allelic loss of chromosome 18q and prognosis in colorectal cancer. N Engl J Med 1994;331:213–221.
11. Diep CB, Thorstensen L, Meling GI, Skovlund E, Rognum TO, Lothe RA. Genetic tumor markers with prognostic impact in Dukes' stages B and C colorectal cancer patients. J Clin Oncol 2003;21:820–829.
12. Cohn KH, Ornstein DL, Wang F, LaPaix FD, Phipps K, Edelsberg C, Zuna R, Mott LA, Dunn JL. The significance of allelic deletions and aneuploidy in colorectal carcinoma. Results of a 5-year follow-up study. Cancer 1997;79:233–244.
13. Schena M, Shalon D, Davis RW, Brown PO. Quantitative monitoring of gene expression patterns with a complementary DNA microarray. Science 1995;270:467–470.
14. Augenlicht LH, Kobrin D. Cloning and screening of sequences expressed in a mouse colon tumor. Cancer Res 1982;42:1088–1093.
15. Wodicka L, Dong H, Mittmann M, Ho MH, Lockhart DJ. Genome-wide expression monitoring in *Saccharomyces cerevisiae*. Nat Biotechnol 1997;15:1359–1367.
16. Mahadevappa M, Warrington JA. A high-density probe array sample preparation method using 10- to 100-fold fewer cells. Nat Biotechnol 1999;17:1134–1136.
17. Golub TR, Slonim DK, Tamayo P, Huard C, Gaasenbeek M, Mesirov JP, Coller H, Loh ML, Downing JR, Caligiuri MA, Bloomfield CD, Lander ES. Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. Science 1999;286:531–537.
18. Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, Armitage JO, Warnke R, Staudt LM, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature 2000;403:503–511.
19. Kihara C, Tsunoda T, Tanaka T, Yamana H, Furukawa Y, Ono K, Kitahara O, Zembutsu H, Yanagawa R, Hirata K, Takagi T, Nakamura Y. Prediction of sensitivity of esophageal tumors to adjuvant chemotherapy by cDNA microarray analysis of gene-expression profiles. Cancer Res 2001;61:6474–6479.
20. Wang Y, Jatkoe T, Zhang Y, Mutch MG, Talantov D, Jiang J, McLeod HL, Atkins D. Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. J Clin Oncol 2004;22:1564–1571.
21. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci U S A 2001;98:5116–5121.
22. Zeeberg BR, Feng W, Wang G, Wang MD, Fojo AT, Sunshine M, Narasimhan S, Kane DW, Reinhold WC, Lababidi S, Bussey KJ, Riss J, Barrett JC, Weinstein JN. GoMiner: a resource for biological interpretation of genomic and proteomic data. Genome Biol 2003;4:R28.
23. Eisen MB, Spellman PT, Brown PO, Botstein D. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A 1998;95:14863–14868.
24. Lloyd SP. Least square quantization in PCM. Trans Information Theory 1982;IT-28:129–137.
25. Arango D, Wilson AJ, Shi Q, Corner GA, Arañes MJ, Nicholas C, Lesser M, Mariadason JM, Augenlicht LH. Molecular mechanisms of action and prediction of response to oxaliplatin in colorectal cancer cells. Br J Cancer 2004;91:1931–1946.
26. Mariadason JM, Arango D, Shi Q, Wilson AJ, Corner GA, Nicholas C, Aranes MJ, Lesser M, Schwartz EL, Augenlicht LH. Gene expression profiling-based prediction of response of colon carcinoma cells to 5-fluorouracil and camptothecin. Cancer Res 2003; 63:8791–8812.
27. Arango D, Mariadason JM, Willson AJ, Yang W, Corner GA, Arañes MJ, Nicholas C, Augenlicht LH. c-Myc overexpression sensitizes colon cancer cells to camptothecin-induced apoptosis. Br J Cancer 2003;89:1757–1765.
28. Alhopuro P, Alazzouzi H, Sammalkorpi H, Salovaara R, Hemminki A, Järvinen H, Mecklin JP, Aaltonen LA, Arango D. SMAD4 levels and response to 5-fluorouracil in colorectal cancer. Clin Cancer Res (in press).
29. Servomaa K, Kiuru A, Kosma VM, Hirvikoski P, Rytomaa T. p53 and K-ras gene mutations in carcinoma of the rectum among Finnish women. Mol Pathol 2000;53:24–30.

30. Laiho P, Launonen V, Lahermo P, Esteller M, Guo M, Herman JG, Mecklin JP, Jarvinen H, Sistonen P, Kim KM, Shibata D, Houlston RS, Aaltonen LA. Low-level microsatellite instability in most colorectal carcinomas. Cancer Res 2002;62:1166–1170.
31. Hsu IC, Metcalf RA, Sun T, Welsh JA, Wang NJ, Harris CC. Mutational hotspot in the p53 gene in human hepatocellular carcinomas. Nature 1991;350:427–428.
32. Alazzouzi H, Alhopuro P, Salovaara R, Sammalkorpi H, Järvinen H, Mecklin JP, Schwartz SJ, Aaltonen LA, Arango D. SMAD4 as a prognostic marker in colorectal cancer. Clin Cancer Res 2005; 11:2606–2611.
33. Canzian F, Salovaara R, Hemminki A, Kristo P, Chadwick R, Aaltonen L, de la Chapelle A. Semiautomated assessment of loss of heterozygosity and replication error in tumors. Cancer Res 1996;56:3331–3337.
34. Horiuchi A, Imai T, Wang C, Ohira S, Feng Y, Nikaido T, Konishi I. Up-regulation of small GTPases, RhoA and RhoC, is associated with tumor progression in ovarian carcinoma. Lab Invest 2003; 83:861–870.
35. Mariadason JM, Arango D, Corner GA, Aranes MJ, Hotchkiss KA, Yang W, Augenlicht LH. A gene expression profile that defines colon cell maturation in vitro. Cancer Res 2002;62:4791–4804.
36. Rihet S, Vielh P, Camonis J, Goud B, Chevillard S, de Gunzburg J. Mutation status of genes encoding RhoA, Rac1, and Cdc42 GTPases in a panel of invasive human colorectal and breast tumors. J Cancer Res Clin Oncol 2001;127:733–738.
37. Moscow JA, He R, Gnarra JR, Knutsen T, Weng Y, Zhao WP, Whang-Peng J, Linehan WM, Cowan KH. Examination of human tumors for rhoA mutations. Oncogene 1994;9:189–194.
38. Cheng KW, Lahad JP, Kuo WL, Lapuk A, Yamada K, Auersperg N, Liu J, Smith-McCune K, Lu KH, Fishman D, Gray JW, Mills GB. The RAB25 small GTPase determines aggressiveness of ovarian and breast cancers. Nat Med 2004;10:1251–1256.
39. Fukata M, Kaibuchi K. Rho-family GTPases in cadherin-mediated cell-cell adhesion. Nat Rev Mol Cell Biol 2001;2:887–897.
40. Kau TR, Way JC, Silver PA. Nuclear transport and cancer: from mechanism to intervention. Nat Rev Cancer 2004;4:106–117.
41. Garrido C, Gurbuxani S, Ravagnan L, Kroemer G. Heat shock proteins: endogenous modulators of apoptotic cell death. Biochem Biophys Res Commun 2001;286:433–442.
42. Rao RV, Ellerby HM, Bredesen DE. Coupling endoplasmic reticulum stress to the cell death program. Cell Death Differ 2004;11: 372–380.
43. Brunagel G, Shah U, Schoen RE, Getzenberg RH. Identification of calreticulin as a nuclear matrix protein associated with human colon cancer. J Cell Biochem 2003;89:238–243.
44. Schubert A, Grimm S. Cyclophilin D, a component of the permeability transition-pore, is an apoptosis repressor. Cancer Res 2004;64:85–93.
45. Xanthoudakis S, Roy S, Rasper D, Hennessey T, Aubin Y, Cassady R, Tawa P, Ruel R, Rosen A, Nicholson DW. Hsp60 accelerates the maturation of pro-caspase-3 by upstream activator proteases during apoptosis. EMBO J 1999;18:2049–2056.
46. Han JM, Kim JY, Kim S. Molecular network and functional implications of macromolecular tRNA synthetase complex. Biochem Biophys Res Commun 2003;303:985–993.
47. Francklyn C, Perona JJ, Puetz J, Hou YM. Aminoacyl-tRNA synthetases: versatile players in the changing theater of translation. RNA 2002;8:1363–1372.
48. Bos JL. ras oncogenes in human cancer: a review. Cancer Res 1989;49:4682–4689.
49. Oliveira C, Westra JL, Arango D, Ollikainen M, Domingo E, Ferreira A, Velho S, Niessen R, Lagerstedt K, Alhopuro P, Laiho P, Veiga I, Teixeira MR, Ligtenberg M, Kleibeuker JH, Sijmons RH, Plukker JT, Imai K, Lage P, Hamelin R, Albuquerque C, Schwartz S Jr, Lindblom A, Peltomaki P, Yamamoto H, Aaltonen LA, Seruca R, Hofstra RM. Distinct patterns of KRAS mutations in colorectal carcinomas according to germline mismatch repair defects and hMLH1 methylation status. Hum Mol Genet 2004;13: 2303–2311.
50. Jaffe AB, Hall A. Rho GTPases in transformation and metastasis. Adv Cancer Res 2002;84:57–80.
51. Vial E, Sahai E, Marshall CJ. ERK-MAPK signaling coordinately regulates activity of Rac1 and RhoA for tumor cell motility. Cancer Cell 2003;4:67–79.
52. Cox EA, Sastry SK, Huttenlocher A. Integrin-mediated adhesion regulates cell polarity and membrane protrusion through the Rho family of GTPases. Mol Biol Cell 2001;12:265–277.
53. de Gramont A, Figer A, Seymour M, Homerin M, Hmissi A, Cassidy J, Boni C, Cortes-Funes H, Cervantes A, Freyer G, Papamichael D, Le Bail N, Louvet C, Hendler D, de Braud F, Wilson C, Morvan F, Bonetti A. Leucovorin and fluorouracil with or without oxaliplatin as first-line treatment in advanced colorectal cancer. J Clin Oncol 2000;18:2938–2947.
54. Giacchetti S, Perpoint B, Zidani R, Le Bail N, Faggiuolo R, Focan C, Chollet P, Llory JF, Letourneau Y, Coudert B, Bertheaut-Cvitkovic F, Larregain-Fournier D, Le Rol A, Walter S, Adam R, Misset JL, Levi F. Phase III multicenter randomized trial of oxaliplatin added to chronomodulated fluorouracil-leucovorin as first-line treatment of metastatic colorectal cancer. J Clin Oncol 2000;18: 136–147.
55. Cunningham D, Pyrhonen S, James RD, Punt CJA, Hickish TF, Heikkila R, Johannesen TB, Starkhammar H, Topham CA, Awad L. Randomised trial of irinotecan plus supportive care versus supportive care alone after fluorouracil failure for patients with metastatic colorectal cancer. Lancet 1998;352:1413–1418.
56. Douillard JY, Cunningham D, Roth AD, Navarro M, James RD, Karasek P, Jandik P, Iveson T, Carmichael J, Alakl M, Gruia G, Awad L, Rougier P. Irinotecan combined with fluorouracil compared with fluorouracil alone as first-line treatment for metastatic colorectal cancer: a multicentre randomised trial. Lancet 2000; 355:1041–1047.
57. Arango D, Augenlicht LH. New approaches to colorectal cancer treatment. Recent research developments in cancer. Volume 3. Trivandrum, India: Transworld Research Network, 2001: 385–395.

Received November 10, 2004. Accepted June 9, 2005.

Address requests for reprints to: Diego Arango, PhD, Molecular Biology and Biochemistry Research Center (CIBBIM), Valle Hebron University Hospital, Passeig Vall d'Hebron 119-129, Barcelona 08035, Spain. e-mail: darango@vhebron.net; fax: (34) 93-489-4040.

Supported by grants from the Sigrid Juselius Foundation (to D.A.; 2003 and 2004) and grants from the Academy of Finland (to L.A.A.; grant 44870, Finnish Center of Excellence Program 2000 to 2005). H.A. is supported by a fellowship from the Agéncia de Gestió d'Ajuts Universitaris i de Recerca de la Generalitat de Catalunya.

The authors thank Sini Martinen and Iina Vuoristo for excellent technical assistance and Tuula Lehtinen and Kirsi Pylvänäinen for their assistance collecting the tumor samples.