A gene signature of 8 genes could identify the risk of recurrence and progression in Dukes' B colon cancer patients

EVA BANDRÉS1*, RAQUEL MALUMBRES1*, ELENA CUBEDO1, BEATRIZ HONORATO1, RUTH ZARATE1, ALBERTO LABARGA2, UNAI GABISU2, JESÚS JAVIER SOLA3 and JESÚS GARCÍA-FONCILLAS1

1Laboratory of Pharmacogenomics, Cancer Research Program (Center for Applied Medical Research), University of Navarra, Av. Pio XII 55, 31008 Pamplona; 2Imagina Biotek, Edificio CEIN, Polígono Mocholí, 31110 Noain (Navarra); 3Department of Pathology, University of Navarra, 31008 Pamplona, Spain

Received November 21, 2006; Accepted January 17, 2007

Abstract. The benefit of postoperative adjuvant chemotherapy in patients with Dukes' B colorectal cancer is still uncertain and its routine use is not recommended. The five-year relapse rate is ~25-40% and the identification of patients at high risk of recurrence would represent an important strategy for the use of adjuvant chemotherapy. We retrospectively analyzed gene expression profiles in frozen tumor specimens from patients with Dukes' B colorectal cancer by using high density oligonucleotide microarrays. Our results show a subset of 48 genes differentially expressed with an associated probability <0.001 in the t-test. Another statistical procedure based on the Fisher criterion resulted in 11 genes able to separate both groups. We selected the 8 genes present in both subsets. The differential expression of five genes (CHD2, RPS5, ZNF148, BRI3 and MGC23401) in colon cancer progression was confirmed by real-time PCR in an independent set of patients of Dukes' B and C stages.

Introduction

Colorectal cancer (CRC) is one of the major causes of cancer death worldwide (1). To date, radical surgery followed by adjuvant chemotherapy in patients with node-positive disease (Dukes' C stage), is the mainstay of therapy (2). In the node-negative Dukes' B stage disease no adjuvant chemotherapy is used after tumor resection, although 25-40% of patients will develop recurrent disease; therefore, the role of adjuvant chemotherapy in this setting is still unclear (3). Although the pathological TNM (tumor-node-metastasis) staging system represents the main tool for identifying prognostic differences (4), this system is not sufficient for predicting recurrence in Dukes' B stage (5). There is a need to identify predictive factors, in addition to TNM staging, to guide the identification of Dukes' B patients that are likely to relapse. This would be of great help in improving treatment strategies in the node-negative disease.

In this regard, several molecules have been proposed as predictive markers for colorectal cancer relapse; however, data for most markers remain inconclusive (6). Thus, there are technological limitations for accurately predicting recurrence by traditional methods. High inter-patient heterogeneity of colorectal carcinoma can also limit the predictive ability of tests.

The development of microarray technologies, which allow parallel analyses of many genes, has led to a new era in medical science (7). Several studies have demonstrated that gene expression signatures could predict clinical outcome (8). In colorectal cancer, a study using Affimetrix technology showed a 23-gene-set that represents a prognostic signature inversely associated with a higher risk of tumor recurrence in Dukes' B stage (9). In the present study, we report a gene expression analysis of 16 patients with Dukes' B colon cancer by using spotted microarrays containing 19000 oligonucleotide sequences. Data analysis of microarrays is still a matter of discussion among the scientific community. One of the most worrying issues is the high risk of obtaining false positives. In order to minimize this problem, we used two different statistical procedures in the search of significant differences of gene expression associated with relapse. Initially, we used a permutation t-test for two means and secondly a Fisher test with a variation of the ‘leave one out’ iteration procedure, based on the study presented by Iizuka et al (10). Only the genes selected by both procedures were considered as differentially expressed between relapsed and non-relapsed patients. We obtained a gene-set of 8 genes differentially expressed between both groups.

Validation by real-time PCR was performed for those genes whose probes hybridized with exon regions. We show that down-regulation of BRI3, CHD2, MGC23401, ZNF148 and RPS5 genes is associated with colorectal cancer progression.

Correspondence to: Dr Eva Bandrés, Laboratory of Pharmacogenomics, Center for Applied Medical Research (CIMA), University of Navarra, Av. Pio XII 55, 31008 Pamplona, Spain
E-mail: ebandres@unav.es

*Contributed equally

Key words: colonic neoplasms, recurrence, disease progression, oligonucleotide array, sequence analysis, tumor markers
Patients and methods

Samples. This study includes 16 samples from Dukes' B colorectal tumors which had undergone surgery at the Department of Surgery of the University Clinic of Navarra, between 1997 and 1999. None of the patients received adjuvant chemotherapy or radiotherapy before surgery or after tumor resection. Six patients developed recurrent disease and the other 10 remained free of cancer 5 years after surgery. For confirmation of the association with recurrence and progression of some genes obtained by gene expression profiling, we analyzed the expression of five genes in an independent prospective set of 27 samples with Dukes' B or C stage by real-time PCR. All tumor samples were collected at resection time and immediately frozen. Staging was performed according to American Joint Committee on Cancer criteria (Greene, FL). The ethics committee of the University Clinic of Navarra approved this study and all patients gave their informed consent.

DNA microarray analysis for gene expression profile. Extraction of total RNA was performed with Trizol® total RNA isolation reagent (Gibco BRL, Life Technologies, Gaitherburg, MD, USA) according to the manufacturer's instructions. All the RNA samples used in this study were cleaned up with RNeasy mini kit (Qiagen, Valencia, CA) and were exhaustively treated with RNase-free DNase I (Qiagen) to remove residual DNA. The concentration was quantified using RiboGreen quantification kit (Molecular Probes, Leiden, The Netherlands) and quality control of RNA was performed by electrophoresis and ethidium bromide staining on a 2% agarose gel.

The corresponding cDNA probes were prepared using the Micromax system (NEN, Perkin Elmer, Boston, MA) according to the manufacturer's protocol. The labeled cDNA was pre-heated to 50°C and hybridized to Human 19K Oligo array slides (60 mers) (Center for Applied Genomics, University of Medicine of New Jersey).

After hybridization at 48°C for 16 h in a slide cassette (Telechem, Sunnyvale, CA), slides were washed sequentially in a series of solutions of increasing stringency. Immediately after washing, the presence of fluorescent-labeled cDNAs on the microarray was detected using a fluorescent anti-fluorescein antibody conjugate and TSA detection (Micromax) according to the manufacturer's protocol with appropriate modifications.

Microarray data normalization and analysis. The GMS 418 scanner (Genetic Microsystems, Woburn, MA), a confocal scanning instrument containing 2 laser sources and high-resolution photo multiplier tubes (10 micron resolution) was used for scanning the hybridized microarrays. After image acquisition, the scanned images were imported into ImaGene 4.1 software (BioDiscovery, Marina del Re, CA) to quantify the signal intensities.

The intensity value associated to each spot is the result of subtracting a Gaussian function of the noise from the foreground values (11). After this background subtraction, base 2 logarithms of all data were calculated and genes with more than two missing values were excluded from the analysis. The remaining missing values were replaced by using the KNN imputation method (12). Then, the quantile normalization method (13) was applied to normalize the data.

Table 1. Clinical and pathological characteristics of the patients and their tumors.*

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Disease free</th>
<th>Relapsed</th>
<th>P-values</th>
</tr>
</thead>
<tbody>
<tr>
<td>Age</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean</td>
<td>64.4</td>
<td>62.7</td>
<td>0.77 (ns)</td>
</tr>
<tr>
<td>Sex</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>4</td>
<td>1</td>
<td>0.34 (ns)</td>
</tr>
<tr>
<td>Male</td>
<td>6</td>
<td>5</td>
<td></td>
</tr>
<tr>
<td>T stage</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>3</td>
<td>1 (ns)</td>
</tr>
<tr>
<td>3</td>
<td>7</td>
<td>3</td>
<td></td>
</tr>
<tr>
<td>Tumor size</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>&lt;5</td>
<td>7</td>
<td>6</td>
<td>0.37 (ns)</td>
</tr>
<tr>
<td>≥5</td>
<td>3</td>
<td>0</td>
<td></td>
</tr>
<tr>
<td>Localization</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>LC</td>
<td>1</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>RC</td>
<td>3</td>
<td>1</td>
<td>0.81 (ns)</td>
</tr>
<tr>
<td>SC</td>
<td>6</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>

*LC, left colon; RC, right colon; SC, colon-sigma; (ns), not significant.

The data discussed in this publication have been deposited in NCBI's Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo/) and are accessible through GEO Series accession number GSE2630.

We decided to use two different statistical procedures in the search of significant differences in gene expression between the two groups of patients. Only the genes selected by both statistical procedures were considered as differentially expressed between relapsed and non-relapsed patients. The first test was a permutation t-test for comparison of two means (15) and the second one a variation of the Fisher test based on the work presented by Iizuka et al (10), in which they searched for the optimal number of genes that could differentiate between two groups of samples. Briefly, we used the same algorithms for the calculation of the Fisher criterion, then we tried different numbers of candidate genes and selected 30, a number that yielded very good classification results when evaluated by means of Fast ICA (11) and Hierarchical Clustering (17). The procedure consisted of selecting the 30 genes with the highest Fisher criterion value in 6 rounds of iteration, each round leaving one sample of the relapsed group and two of the non-relapsed group out of the calculations (a variant of the ‘leave one out’ method). We selected the genes present in at least 3 of the iterations.

In silico hybridizations. In order to confirm the identity of the genes able to hybridize with the probes of the genetic signature, we performed two in silico hybridizations with the sequence of the human genome, one with the BLAT algorithm (18) of the University of California Santa Cruz Genome Browser and
and detection of normalize emissions. Primers and probes used for amplification
Real-time PCR was carried out using Taqman Universal PCR
ABI PRISM 7700 sequence detector (PE Applied Biosystems).
incubated at 42˚C for 30 min, followed by incubation at 72˚C
Biosystems, Foster City, CA). The reaction mixture was
used for confirmation of the results by real-time PCR. Total
27 patients (15 with Dukes’ B and 12 with Dukes’ C stage) was
Real-time PCR. RNA extracted from an independent set of
or intergenic regions.
the only possible hybridization of the probes was with introns
In case of several possible hybridizations, we selected the
ones involving exons of well known genes, but in some cases
the only possible hybridization of the probes was with introns or intergenic regions.

Real-time PCR. RNA extracted from an independent set of
27 patients (15 with Dukes’ B and 12 with Dukes’ C stage) was
used for confirmation of the results by real-time PCR. Total
RNA (2 µg) from each sample were used to generate cDNA
using the Taqman reverse transcription reagent kit (PE Applied
Biosystems, Foster City, CA). The reaction mixture was
incubated at 42˚C for 30 min, followed by incubation at 72˚C
for 10 min.
Each cDNA sample was analyzed in triplicate using the
ABI PRISM 7700 sequence detector (PE Applied Biosystems).
Real-time PCR was carried out using Taqman Universal PCR
master mix (Applied Biosystems), containing ROX to
normalize emissions. Primers and probes used for amplification
and detection of CHD2, BRI3, MGC23401, ZNF148 and 18S
ribosomal RNA were purchased from Applied Biosystems as
‘Assay on demand’ (Assay ID: Hs_00172280_m1, Assay ID:
Hs00854645_g1, Assay ID: Hs00299246_m1, Assay ID: Hs00222661_m1 and Assay ID: Hs_99999907_m1). Primers
and Taqman probe for amplification and detection of RPS5
were designed using the Primer Express 1.0 software (Applied
Biosystems) (Forward: CTCATGACTGTGCGCATCGT,
Reverse: CACTTGGATGATGGCGTTCA and Probe: TG
CCTTCGAGATCATACACCTGCTCACA). For thermal
cycling, the following conditions were applied: 10 min at
95˚C and 40 cycles of 15 sec at 95˚C and 1 min at 59˚C.
Expression levels were normalized to the Ct value of the
ribosomal RNA. Fold induction was calculated using the
formula \(2^{-\Delta\Delta C_{t}}\), where \(\Delta C_{t}\) = target gene Ct-‘housekeeping’
gene Ct.

Results

Patient and tumor characteristics. Clinical and pathological
features of the patients and their tumors are summarized in
Table I. Information on age, gender, TNM stage, grade, tumor
size and tumor location were collected. The patient and tumor
characteristics did not differ significantly between relapsed and
non-relapsed patients.

Gene expression profile. Microarray technology was used to
analyze a series of 16 stage B colorectal tumors with different
outcome in terms of relapse. In order to identify gene markers
that can discriminate between the relapsed and the non-
relapsed patients, we performed a permutation t-test for
comparison of two means. Genes were selected regarding
their associated probability, considering 0.001 as the threshold
P-value for significant differences. According to this criterion,
48 genes are differentially expressed between relapsed and
non-relapsed patients. Further, we applied a variation of the
Fisher test. We tried different numbers of candidate genes and
selected 30 as an appropriate input for the selection procedure,
obtaining 11 genes that appeared in three or more iteration
rounds and yielded very good separation results when
evaluated by means of Fast ICA (Fig. 1) and unsupervised
Hierarchical Clustering (Fig. 2). Finally, we considered as a
good gene signature the genes selected by both statistical
procedures, a total of 8, 1 was up-regulated and 7 down-
regulated. This prognostic signature associates with a higher
risk of recurrence (Table II).

In silico hybridizations. In order to confirm the identity of the
genes able to hybridize with the eight probes, we performed
in silico hybridizations of the probes with the sequence of the
human genome, using the BLAT algorithm and the SSAHA
algorithm. By combination of the results of both algorithms
we assigned a unique hybridization site for each probe. Five
probes hybridized with exons of the genes chomodomain-
helicase DNA binding protein 2 (CHD2), ribosomal protein
S5 (RPS5), zinc finger DNA binding protein of 148 kDa
(ZNF148), brain protein I3 (BRI3) and the hypothetical
gene MGC23401. All these genes appear down-regulated in relapse
in our gene signature. Two other probes hybridize with introns of
two well known genes, H+ transporting lysosomal ATPase
V0 subunit A isoform 1 gene (ATP6V0A1) and ELK4, the
cofactor of the serum response factor. ATP6V0A1 is the only
gene in our gene signature that is up-regulated in relapse.
Finally, one probe hybridized only with the intergenic region
between the genes glucocorticoid modulatory element B1
(GMEB1) and high glucose regulated protein (HGRP8 or
YTHDF2).

Validation of microarray data by real-time PCR. In order to
confirm the role of these genes in CRC progression, we
measured the mRNA expression levels of 5 out of 8 genes
present in the putative prognostic signature (BRI3, ZNF148,
RPS5, MGC23401 and CHD2). RNA from an independent
set of patients was tested by real-time PCR. The results of PCR
analysis are in agreement with the microarray data (Fig. 3)
and confirm that this gene-set signature is associated with
colorectal cancer progression.
Discussion

The gold standard for predicting clinical outcome of most cancers has been clinical and pathological staging of tumors after surgery. However, patients with Dukes' B colorectal cancer at the same tumor stages may show different outcome, indicating that the conventional staging procedures may be unable to precisely predict cancer prognosis. Therefore, the search for new prognostic factors able to identify high-risk patients and modulate cancer treatment options is still actively
ongoing. Although many markers have been extensively described, data for individual molecules failed to elaborate the complex patterns of carcinogenesis and cancer progression. In this regard, the development of microarray technologies, that allow undertaking of parallel analyses of many genes, could help to identify molecular factors involved in cancer progression. In this study, we report an 8-gene signature derived from microarray gene expression data with the aim to identify molecular markers associated with relapse in Dukes' B colon cancer patients. Our study suggests that colon cancer prognosis can be derived from gene expression profile of the primary tumor. Probably, small changes of different genes in primary tumors are responsible for tumor progression rather than large changes in only one gene.

Probe sequences in silico hybridization with the human genome allowed us to confirm the identity of the genes whose expression was being detected by the probes of our gene signature: CHD2, RPS5, BR13, ZNF148, MGC23401 (hypothetical protein), ELK4 and GMEB1 or YTFDH2 would be down-regulated in relapse and ATPV0A1 up-regulated. Although most of the probes (5 out of 8) hybridize with exons of the corresponding genes, two of the probes hybridize with introns of two well known genes, ATP6V0A1 and ELK4. Our probe possibly detects a new isoform of these genes in which the introns would become part of exons by alternative splicing. Besides, one probe hybridized only with the intergenic region between the genes glucocorticoid modulatory element B1 (GMEB1) and high glucose regulated protein (HGRP8 or YTHDF2). One of these two genes may present alternative isoforms that include this probe in an exon, or another gene still not described could be between these two genes.

CHD2 is one of the most interesting genes in our gene signature as it has also been related to Dukes' B colon cancer relapse by an independent group (9) using microarray technology. The CHD gene family is a group of highly conserved proteins sharing sequence motifs and functional domains associated with the regulation of chromatin and gene transcription. It has been recently described that CHD1 specifically interacts with the methylated lysine 4 mark on histone 3 in yeast recruiting histone acetylase activity, what is associated with a remodeling of chromatin that facilitates transcription (20). On the other hand, CHD1 co-immunoprecipitates with the transcriptional corepressor NcoR, with histone deacetylases and also with RNA splicing proteins (21). Thus, CHD1 is a key factor in the regulation of transcription by post-translational modification of histones and also influences alternative splicing. Some authors suggest that the whole chromodomain family may have similar recognition properties; therefore, CHD2 can be expected to participate in the same or similar processes and its down-regulation may affect epigenetics and the whole transcription regulation in the cell.

ZNF148 is a zinc finger transcription factor that is universally expressed (22). It has been shown that ZNF148 binds to GC-rich DNA elements in a variety of promoters involved in growth regulation (23-25). Moreover, for the rat pituitary adenoma cell line GH4, Bai and Merchant (26) showed that elevated expression of ZNF148 inhibits cell proliferation and promotes growth arrest through stabilization of the p53 protein. Furthermore, tumor cells with a mutation in the p53 gene are resistant to ZNF148-mediated stabilization and are associated with hepatocellular carcinoma recurrence (27). In our study, ZNF148 is down-regulated in patients with...
involved in tumor necrosis factor-
proteins with broad tissue expression. This protein is 
TNF is down-regulated in relapsed patients, suggesting that the 
and of S5 ribosomal protein in colorectal cancer progression. 
proteins as checkpoints in the carcinogenesis process in general 
research will be required to determine the role of ribosomal 
with relapse in comparison with non-relapsed patients. 
Moreover, their expression is also down-regulated in Dukes' C 
patients in comparison with Dukes' B patients. Additional 
research will be required to determine the role of ribosomal 
proteins as checkpoints in the carcinogenesis process in general 
and of S5 ribosomal protein in colorectal cancer progression. 
The BRI3 gene belongs to a family of integral membrane 
proteins with broad tissue expression. This protein is 
involved in tumor necrosis factor-α-induced cell death in 
murine fibrosarcoma cells L929 (30). Its overexpression 
induces apoptosis in these cells and the presence of BRI3 is 
required in the TNF-α-induced cell death. In our study, BRI3 is 
down-regulated in relapsed patients, suggesting that the 
TNF-α-induced apoptosis pathway could be altered in colorectal 
cancer progression. 

Independently of the confirmation of these hypotheses, 
our 8 probes for gene expression are able to discriminate 
between our relapsed and non-relapsed patients. Furthermore, 
in the present study we showed that the down-regulation of 
5 out of 8 genes is associated with colon cancer progression. 
Thus, a study on the functions of the genes present in this 
gene signature could also improve our understanding of the 
molecular mechanisms involved in colorectal cancer progression.

Acknowledgements

Eva Bandrés was financially supported by the Fondo de 
Investigación Sanitaria F.I.S. (ISICIB), Grant FIS: 01/0816. We 
thank the Center for Applied Genomics and the New Jersey 
Commission on Science and Technology for providing the 
Oligo 19K human array. We also thank Maria Dolores Otero 
for critical reading and discussion of the study and Inés López 
for excellent technical assistance.

References

1. Saha D, Roman C and Beauchamp RD: New strategies for 
colorectal cancer prevention and treatment. World J Surg 26: 
762-766, 2002.
3. MacDonald JS: Adjuvant therapy of colon cancer. CA Cancer J 
4. Graziano F and Cascini S: Prognostic molecular markers for 
planning adjuvant chemotherapy trials in Dukes' B colorectal 
cancer patients: how much evidence is enough? Ann Oncol 14: 
5. Buyse M and Piedbois P: Should Dukes' B patients receive 
adjuvant therapy? A statistical perspective. Semin Oncol (Suppl 1): 
6. McLeod HL and Murray GI: Tumour markers of prognosis in 
7. Schena M, Shalon D, Davis RW and Brown PO: Quantitative 
mapping of gene expression patterns with a complementary 
of colon cancer by DNA microarrays and correlation with histo-
and activating p53 function has been demonstrated (28). 
Variable expression of this gene in colorectal cancers 
compared to adjacent normal tissues has also been observed 
(29), although so far no correlation between the level of 
expression and the severity of the disease has been reported. 
In this study, we found a lower PRS5 expression in patients 
with relapse in comparison with non-relapsed patients. 
Moreover, their expression is also down-regulated in Dukes' C 
patients in comparison with Dukes' B patients. Additional 
research will be required to determine the role of ribosomal 
proteins as checkpoints in the carcinogenesis process in general 
and of S5 ribosomal protein in colorectal cancer progression. 
The BRI3 gene belongs to a family of integral membrane 
proteins with broad tissue expression. This protein is 
involved in tumor necrosis factor-α-induced cell death in 
murine fibrosarcoma cells L929 (30). Its overexpression 
induces apoptosis in these cells and the presence of BRI3 is 
required in the TNF-α-induced cell death. In our study, BRI3 is 
down-regulated in relapsed patients, suggesting that the 
TNF-α-induced apoptosis pathway could be altered in colorectal 
cancer progression. 

Independently of the confirmation of these hypotheses, 
our 8 probes for gene expression are able to discriminate 
between our relapsed and non-relapsed patients. Furthermore, 
in the present study we showed that the down-regulation of 
5 out of 8 genes is associated with colon cancer progression. 
Thus, a study on the functions of the genes present in this 
gene signature could also improve our understanding of the 
molecular mechanisms involved in colorectal cancer progression.

Acknowledgements

Eva Bandrés was financially supported by the Fondo de 
Investigación Sanitaria F.I.S. (ISICIB), Grant FIS: 01/0816. We 
thank the Center for Applied Genomics and the New Jersey 
Commission on Science and Technology for providing the 
Oligo 19K human array. We also thank Maria Dolores Otero 
for critical reading and discussion of the study and Inés López 
for excellent technical assistance.

References

1. Saha D, Roman C and Beauchamp RD: New strategies for 
colorectal cancer prevention and treatment. World J Surg 26: 
762-766, 2002.
3. MacDonald JS: Adjuvant therapy of colon cancer. CA Cancer J 