

RESEARCH

Open Access

A molecular signature for the prediction of recurrence in colorectal cancer

Lisha Wang^{1,2,3†}, Xiaohan Shen^{1,2,3†}, Zhimin Wang⁵, Xiuying Xiao⁶, Ping Wei^{1,2,3}, Qifeng Wang^{1,2,3}, Fei Ren^{1,2,3}, Yiqin Wang^{1,2,3}, Zebing Liu^{1,2,3}, Weiqi Sheng^{1,2,3}, Wei Huang⁵, Xiaoyan Zhou^{1,2,3} and Xiang Du^{1,2,3,4*}

Abstract

Background: Several clinical and pathological factors have an impact on the prognosis of colorectal cancer (CRC), but they are not yet adequate for risk assessment. We aimed to identify a molecular signature that can reliably identify CRC patients at high risk for recurrence.

Results: Two hundred eighty-one CRC samples (stage II/III) were included in this study. A two-step gene expression profiling study was conducted. First, gene expression measurements from 81 fresh frozen CRC samples were obtained using Affymetrix Human Genome U133 Plus 2.0 Arrays. Second, a focused gene expression assay, including prognostic genes and genes of interest from literature reviews, was performed using 200 fresh frozen samples and a Taqman low-density array (TLDA) analysis. An optimal 31-gene expression classifier for the prediction of recurrence among patients with stage II/III CRC was developed using logistic regression analysis. This gene expression signature classified 58.5% of patients as low-risk and 41.5% as high-risk ($P < 0.001$). The signature was the strongest independent prognostic factor in the multivariate analysis. The five-year relapse-free survival (RFS) rates for the low-risk patients and the high-risk patients were 88.5% and 41.3% ($P < 0.001$), respectively.

Conclusion: We identified a 31-gene expression signature that is closely associated with the clinical outcome of stage II/III CRC patients.

Keywords: Molecular signature, Gene expression, Colorectal cancer, Recurrence

Introduction

Colorectal cancer (CRC) is the third most common type of cancer, with a worldwide annual incidence of over 1.2 million cases and a mortality rate of approximately 50% [1,2]. The growing awareness that CRC is a heterogeneous disease with respect to clinical behavior and prognosis translates into an urgent need for robust molecular subclassifiers in addition to the current parameters. To date, some clinical and pathologic features, such as intestinal perforation/obstruction, adjacent organ involvement (T4), high tumor grade, lymphatic/vascular invasion and inadequate sampling of lymph nodes, can identify a minority of CRC patients who are at a high risk of

recurrence [3,4]. However, these prognostic factors are all relatively weak.

To address this issue, the prognostic potential of molecular markers, including chromosome and microsatellite instability (MSI) and the mutation status of the *KRAS* or *BRAF* genes, has been systematically investigated in CRC [5-8]. Thus far, only *KRAS* mutation analysis has been used in clinical practice as a predictive marker of the effect of *EGFR* antibodies in metastatic disease [5,9]. Analyses of other known critical CRC molecular markers are not currently recommended for screening or for prognostic prediction because they require further validation.

With the recent advent of microarray technology, risk assessment for CRC has been improved by the use of gene expression profiling. DNA microarray technology can measure thousands of mRNA transcripts at once and may be able to describe the complex biology of a tumor more accurately than single markers [10,11]. In the current study, we used gene expression analysis data

* Correspondence: dx2008cn@163.com

†Equal contributors

¹Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai 200032, China

²Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, China

Full list of author information is available at the end of the article

from recurrent and non-recurrent patients with CRC to identify differentially expressed probes. To further validate gene expression, we selected 48 genes that could be assayed using a TaqMan low-density array (TLDA), a real time quantitative PCR (RT-qPCR) based technology, using fresh frozen CRC tissues.

Patients and methods

Patients and tumor samples

Samples were prospectively collected between 2007 and 2009 at Fudan University Shanghai Cancer Center. The inclusion criteria were as follows: primary sporadic colorectal adenocarcinoma (excluding familial adenomatous polyposis (FAP) and hereditary nonpolyposis colorectal cancer (HNPCC)), 18 to 75 years of age, no preoperative chemotherapy and radiotherapy, and similar postoperative chemotherapy regimens. The patients were staged according to the American Joint Committee on Cancer/International Union against Cancer (AJCC/UICC) TNM staging system- seventh edition (2010). Histologic grading (differentiation) was based on the WHO classification of tumors of the digestive system-fourth edition (2010). This study was approved by the Ethical

Committee of our Cancer Center, and written informed consent was obtained from each patient.

Microarray gene expression profiling

Tumor tissues were taken from 81 patients with CRC, rapidly frozen in RNAlater, and stored at -80°C until processing. All samples were visually inspected by two pathologists, who confirmed the presence of tumor cells ($\geq 70\%$) in all samples. RNA was isolated using Trizol (Life Technologies, Carlsbad, California, USA) and purified using RNeasy MinElute Cleanup Kit (QIAGEN, Hilden, Germany) as recommended by the manufacturers. Quantity and quality measurements were carried out using a NanoVue™ Plus Spectrophotometer (GE, London, UK) and an Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, California, USA). Gene expression profiles were determined using Affymetrix HG-U133 Plus 2.0 GeneChips according to the recommendations of the manufacturer.

TaqMan low density array (TLDA)

Two hundred fresh frozen CRC samples were used for TLDA analysis. Pre-designed TaqMan probe and primer

Table 1 Gene expression assays used for configuring the Taqman low-density array card

Probe sets/Ref. sequences	Gene symbol	Assay IDs	Probe sets/Ref. sequences	Gene symbol	Assay IDs
1554997_a_at	<i>PTGS2</i>	Hs01573477_g1	232278_s_at	<i>DEPDC1</i>	AJCSVBJ
1558135_at	<i>TAF11</i>	Hs01051508_g1	232315_at	<i>LOC400713</i>	AJ89JTA
1562921_at	<i>GENE3</i>	AJD1THR	232684_at	<i>LOC253264</i>	AJGJPT7
200632_s_at	<i>NDRG1</i>	Hs00608390_m1	234768_at	<i>GENE28</i>	Hs01078763_m1
203001_s_at	<i>STMN2</i>	Hs00975902_m1	235229_at	<i>GENE29</i>	AJMSGPB
203889_at	<i>SCG5</i>	Hs02559426_s1	238531_x_at	<i>GENE30</i>	AJSIPAM
204886_at	<i>PLK4</i>	Hs00975273_m1	238629_x_at	<i>GENE31</i>	AJN1EVJ
204932_at	<i>GENE8</i>	AJ6RNGU	241607_at	<i>LOC730102</i>	AJBJW5B
204933_s_at	<i>TNFRSF11B</i>	Hs00900360_m1	59437_at	<i>C9orf116</i>	Hs00376168_m1
205828_at	<i>MMP3</i>	Hs00968308_m1	NM_001025366	<i>VEGFA</i>	Hs00900055_m1
205890_s_at	<i>UBD</i>	Hs00197374_m1	NM_000610	<i>CD44</i>	Hs01075861_m1
207808_s_at	<i>PROS1</i>	Hs00165590_m1	NM_004530	<i>MMP2</i>	Hs01548727_m1
211653_x_at	<i>AKR1C2</i>	AJ70LM2	NM_003254	<i>TIMP1</i>	Hs00171558_m1
212315_s_at	<i>NUP210</i>	Hs01090284_m1	NM_004994	<i>MMP9</i>	Hs00234579_m1
215039_at	<i>LOC339524</i>	AJKAKCV	NM_002046	<i>GAPDH</i>	Hs02758991_g1
219054_at	<i>C5orf23</i>	AJHSNOF	NM_004360	<i>CDH1</i>	Hs01023894_m1
219148_at	<i>PBK</i>	Hs00218544_m1	NM_005429	<i>VEGFC</i>	Hs00153458_m1
220295_x_at	<i>GENE18</i>	AJFARNZ	NM_003345	<i>UBC</i>	Hs01871556_s1
221703_at	<i>BRIP1</i>	Hs00908148_m1	NM_001098210	<i>CTNBN1</i>	Hs00355049_m1
226211_at	<i>MEG3</i>	AJLJII3	NM_015675	<i>GADD45B</i>	Hs04188837_g1
226661_at	<i>CDCA2</i>	Hs00401022_m1	NM_002466	<i>MYBL2</i>	Hs00942543_m1
228877_at	<i>RGL3</i>	Hs01004328_g1	NM_000876	<i>IGF2R</i>	Hs00974474_m1
229331_at	<i>SPATA18</i>	Hs01111872_g1	NM_000269	<i>NME1</i>	Hs02621161_s1
230135_at	<i>GENE24</i>	AID1TYD			

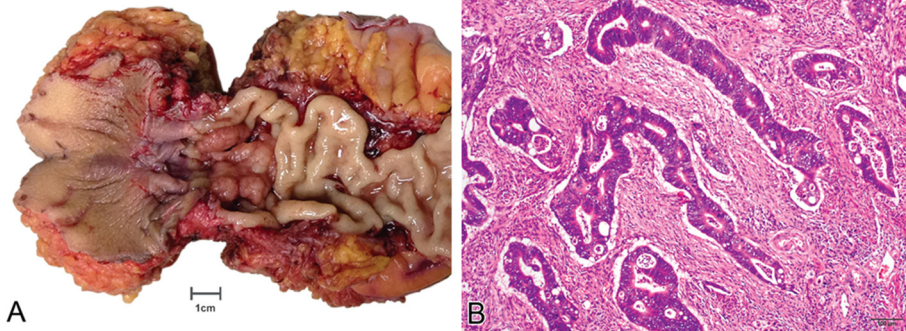


Figure 1 Gross appearance and microscopic findings of colorectal cancer. (A), Gross morphology of an opened Miles' abdominoperineal resection specimen containing an exophytic mass. The bowel lumen can become progressively narrowed and constricted. (B), Representative images of a moderately differentiated colorectal adenocarcinoma (H&E stains, $\times 100$). The tumor cells are arranged in glandular patterns, with a prominent desmoplastic stromal response.

sets for target genes were chosen from an online catalog (Applied Biosystems). Once selected, the primer sets were factory loaded into the 384 wells of TLDA cards. Each TLDA card in this study was configured into 8 identical 48-gene sets (2 samples in duplicate). In our study, 48 genes were chosen based on gene expression profiling analysis and literature reviews [3,12-22] (Table 1). Each set contained *GAPDH*, *UBC*, and *18S*, all of which were used as reference genes for normalization [23,24]. These genes were demonstrated to be expressed in colorectal tissues with little variability. The geometric mean was used for the calculation of the expression of target genes.

Total RNA was reverse-transcribed using the High-Capacity cDNA Reverse Transcription Kit using reaction volumes of 10 μL (Applied Biosystems, Foster City, CA). A total of 100 μL reaction mixture containing 9 μL of cDNA template (corresponding to 100 ng of mRNA) and 50 μL of 2 \times TaqMan Universal PCR Master Mix (Applied Biosystems, Foster City, CA, USA) was added to each line of TLDA after vortex and brief centrifugation. One microliter of the reaction mixture, which contained 1 ng of mRNA, was transferred to each reaction cell. The TLDA plates were sealed with a TLDA sealer before centrifugation in a Thermo Scientific Sorvall ST40 centrifuge. PCR amplification was carried out in

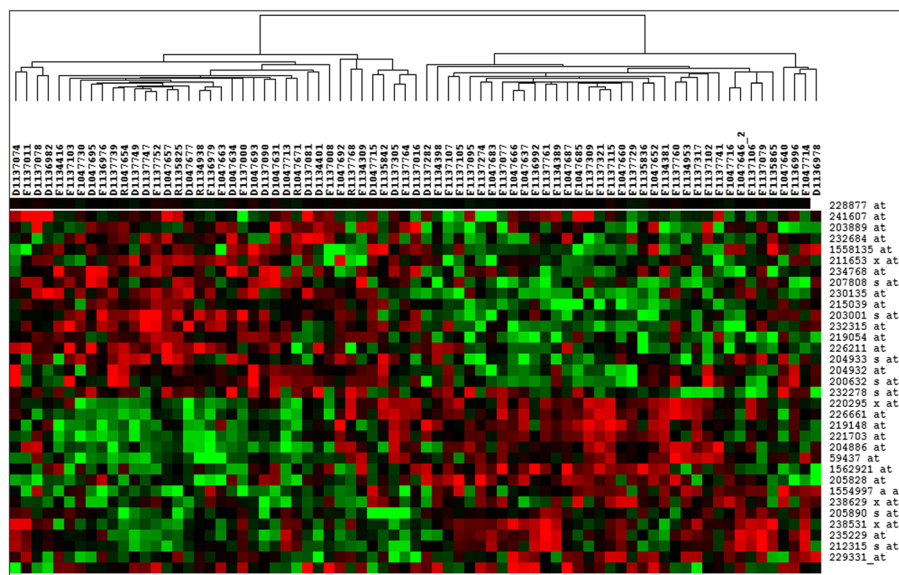


Figure 2 Hierarchical cluster analysis of colorectal cancer (26 recurrent patients vs. 49 non-recurrent patients). The data are presented here in matrix format, in which the rows represent individual genes (31 probes sets) and the columns represent each tissue from CRC patients ("D" = death; "R" = relapse; "F" = survival). The red and green colors reflect high and low expression levels, respectively.

the micro-fluidic card sample block of an ABI Prism® 7900HT sequence detection system (Applied Biosystems, Foster City, CA, USA). The amplification protocol was as follows: 10 min at 94.5°C (activation), 40 cycles of denaturation at 95°C for 15 s, and annealing and extension at 59.7°C for 1 min.

Statistics

The gene expression data were filtered and normalized using Expression Console™ 1.2.0.20 software to remove systematic technical variation before further analysis. Raw data had been deposited in the Gene Expression Omnibus (GEO) database and were accessible through the accession number GSE64857. To reduce variation between individual microarrays, the intensity values for the samples in each microarray were rescaled by means of a quantile normalization method. Each intensity value was log-transformed to a base-2 scale. The intensity value was coded as 1 for expression levels ranked at or

below the 25th percentile of total gene expression, 2 for levels above the 25th percentile and at or below the 50th percentile, 3 for levels above the 50th percentile and at or below the 75th percentile, and 4 for levels above the 75th percentile [25].

The threshold cycle (Ct) generated from TLDA analysis was automatically determined using the SDS2.2 software package. The relative quantities (RQ) were determined using the equation: $RQ = 2^{-\Delta\Delta Ct}$. To reduce variation between TLDA analyses, the Ct values for samples on each TLDA card were also rescaled by means of a quantile normalization method. Each intensity value was log-transformed to a base-2 scale. The expression intensities were ranked as 1, 2, 3, and 4. A binary logistic regression model was performed to discriminate between the recurrent group and the non-recurrent group using a stepwise backward conditional procedure, and probability was set at 0.05 for entry and 0.10 for removal. A patient was classified into the high risk group if the value was greater than 0.5 or into the low risk group if the value was less than 0.5.

Further statistical analyses were performed using SPSS 16.0 software, including independent samples t tests, χ^2 tests, univariate and multivariate Cox proportional hazards analyses (estimation of HR and the corresponding 95% confidence intervals), and Kaplan-Meier survival analysis (differences were assessed by log-rank statistics). For survival analyses, the end-point was relapse-free survival (RFS), which was defined as the probability that patients remain free of recurrence (local/regional or metastatic) as the first event. A two-sided *P*-value of ≤ 0.05 was considered to be significant.

Table 2 mRNA expression in the recurrent group and the non-recurrent group as assayed using the Taqman low-density array assay

Gene symbol	Fold change*	<i>P</i> value	Gene symbol	Fold change*	<i>P</i> value
<i>PTGS2</i>	0.485	0.011	<i>GENE24</i>	0.835	>0.05
<i>TAF11</i>	1.048	>0.05	<i>DEPDC1</i>	0.323	0.0001
<i>GENE3</i>	1.034	>0.05	<i>LOC400713</i>	2.280	0.0419
<i>NDRG1</i>	1.114	>0.05	<i>LOC253264</i>	1.436	>0.05
<i>STMN2</i>	0.994	>0.05	<i>GENE28</i>	1.049	>0.05
<i>SCG5</i>	0.938	>0.05	<i>GENE29</i>	0.346	0.006
<i>PLK4</i>	1.166	>0.05	<i>GENE30</i>	0.836	>0.05
<i>GENE8</i>	N/A	N/A	<i>GENE31</i>	N/A	>0.05
<i>TNFRSF11B</i>	1.049	>0.05	<i>LOC730102</i>	1.159	>0.05
<i>MMP3</i>	1.102	>0.05	<i>C9orf116</i>	1.003	>0.05
<i>UBD</i>	1.005	>0.05	<i>VEGFA</i>	1.228	>0.05
<i>PROS1</i>	0.774	>0.05	<i>CD44</i>	1.031	>0.05
<i>AKR1C2</i>	1.383	>0.05	<i>MMP2</i>	0.654	0.0216
<i>NUP210</i>	0.655	0.0224	<i>TIMP1</i>	2.153	0.0199
<i>LOC339524</i>	1.019	>0.05	<i>MMP9</i>	0.522	0.0018
<i>CSorf23</i>	0.691	>0.05	<i>CDH1</i>	0.712	0.0071
<i>PBK</i>	0.539	0.0028	<i>VEGFC</i>	1.013	>0.05
<i>GENE18</i>	0.617	0.0317	<i>CTNNB1</i>	1.007	>0.05
<i>BRIP1</i>	1.052	>0.05	<i>GADD45B</i>	1.881	0.041
<i>MEG3</i>	2.236	0.0436	<i>MYBL2</i>	1.398	0.041
<i>CDCA2</i>	0.556	0.0011	<i>IGF2R</i>	0.703	0.0278
<i>RGL3</i>	1.903	0.0435	<i>NME1</i>	1.174	>0.05
<i>SPATA18</i>	0.754	>0.05			

*: fold change, the ratio of gene expression in the recurrent group and the non-recurrent group. Bold text denotes *P*<0.05.

Results

Clinicopathologic characteristics

A total of 281 fresh frozen CRC specimens were included in the study (Figure 1), and 81 of these tumors were used for microarray gene expression analysis. Of these 81 patients, 47 were male and 34 were female. The patients' ages ranged from 29 to 75 years (median 56.3 years). The mean follow-up time was 48 months (range 38–68 months). Twenty-six patients had relapsed or died at the time of last follow-up. Six patients were lost to follow-up. The metastatic sites included the liver (*n* = 9), lungs (*n* = 4), abdominal sites (*n* = 2), and other organs.

Two hundred tumors were used for TLDA analysis. These tumors were obtained from 117 male patients and 83 female patients. The ages of the patients ranged from 32 to 75 years (median 57.4 years). The mean follow-up time was 48 months (range 36–68 months). Fifty-five patients relapsed or died at the time of last follow-up. The metastatic sites included the liver (*n* = 17), lungs (*n* = 9), abdominal sites (*n* = 6), bones (*n* = 3), and other organs.

Gene expression analysis

Genome-wide expression data were used to identify genes that correlate with patient prognosis. Thirty-three probe sets were selected according to the following criteria: (1) fold change ≥ 1.5 or ≤ 0.67 ; (2) raw gene expression signal levels above 100 in 60% of the samples; (3) $P \leq 0.01$ (Figure 2). The 33-gene panel was calculated as follows: $y = C + (B_1 \times 1 \dots B_{33} \times 33)$, $P = \text{EXP}(y) / (1 + \text{EXP}(y))$, where C represents the constant ($C = -160.677788$) and B represents the weight coefficient for each gene. The expression levels of these 33 genes correlated with relapse. The probe sets are summarized in Table 1. The sensitivity and specificity of this panel are 92.31% and 71.43%, respectively. The Hosmer-Lemeshow (HL) test showed good fit of the model ($P = 1$). As a continuous variable, the genetic signature classified patients into two groups using 0.5 as a cutoff; group 1 with $P < 0.5$ represented the non-recurrent patients, and group 2 with $P > 0.5$ represented the recurrent patients.

TaqMan low density array analysis

The results of microarray analysis prompted us to implement a TLDA assay with the differentially expressed probes. In total, 48 genes were chosen for the TLDA

assay, including 33 candidate genes that are differentially expressed based on the microarray analysis, 12 relative genes selected from a literature review, and 3 reference genes. Forty-six genes (95.8%, 46/48) were successfully amplified by TLDA. Seventeen genes (*PTGS2*, *NUP210*, *PBK*, *GENE18*, *MEG3*, *CDCA2*, *RGL3*, *DEPDC1*, *LOC400713*, *GENE29*, *MMP2*, *TIMP1*, *MMP9*, *CDH1*, *GADD45B*, *MYBL2*, and *IGF2R*) were differentially expressed in the recurrent group and the non-recurrent group based on the TLDA results (t test, $P < 0.05$). Eleven genes were significantly down-regulated in the recurrent group, while 6 genes were significantly up-regulated. Among these 17 genes (17-gene panel), 10 genes (10-gene panel) were identified from our microarray analysis, and 7 genes (7-gene panel) that were identified by other studies were also confirmed. The fold changes of these genes are summarized in Table 2. The mRNA levels of the target genes were calculated as $2^{-(\Delta\Delta Ct)}$. Figure 3 shows the expression of *CDCA2*, *RGL3*, *CDH1*, and *MYBL2* in the recurrent group and the non-recurrent group.

The 31-gene signature and survival

As not all of the genes amplified in the TLDA assay, a logistic regression analysis was performed in two steps

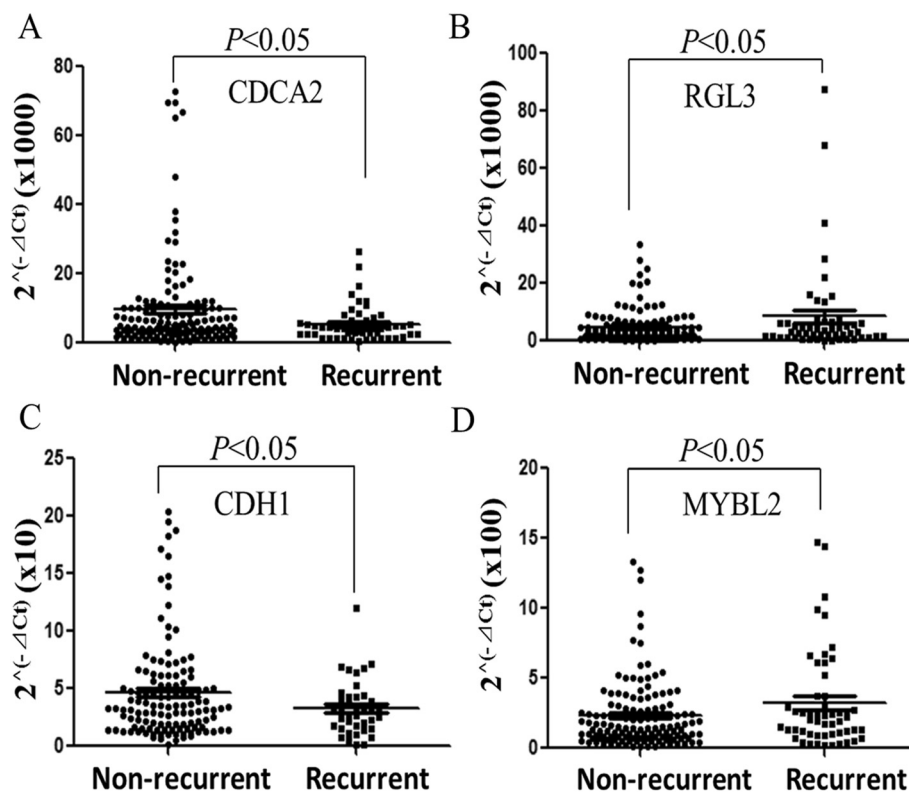


Figure 3 mRNA levels of representative genes in the recurrent group and the non-recurrent group. (A), The level of *CDCA2* mRNA expression was significantly increased in the non-recurrent group compared to the recurrent group. (B), The level of *RGL3* mRNA expression was significantly increased in the recurrent group. (C), The level of *CDH1* was significantly increased in the non-recurrent group. (D), The level of *MYBL2* was significantly increased in the recurrent group.

to avoid the influence of missing values: 79 samples with full TLDA PCR data were first used to generate the formula; then all 200 samples, including 121 samples with partial missing data, were imported and entered into the formula for analysis. Among the 200 patients, 55 were recurrent, and 145 were non-recurrent. A 31-gene panel was developed using logistic regression analysis, each with a new weight coefficient. To compare the prediction performance of the 31-gene, 17-gene, 10-gene, and 7-gene panels, receiver operator characteristic (ROC) curves were drawn, and the area under the curve (AUC) was calculated for each ROC curve based on the standard leave-one-out procedure. The AUC demonstrated that the predicting performance of the 31-gene panel was the best one (0.766 vs 0.576 vs 0.499 vs 0.500, Figure 4). The use of this algorithm resulted in the separation of two groups, the recurrent group and the non-recurrent group. One hundred seventeen out of 200 patients (58.5%) were predicted to be non-recurrent patients (low risk); among these, 106 were verified as non-recurrent. Eighty-three out of 200 patients (41.5%) were predicted to be recurrent patients (high risk); among these, 44 were observed to be recurrent ($P < 0.001$). The positive predictive value,

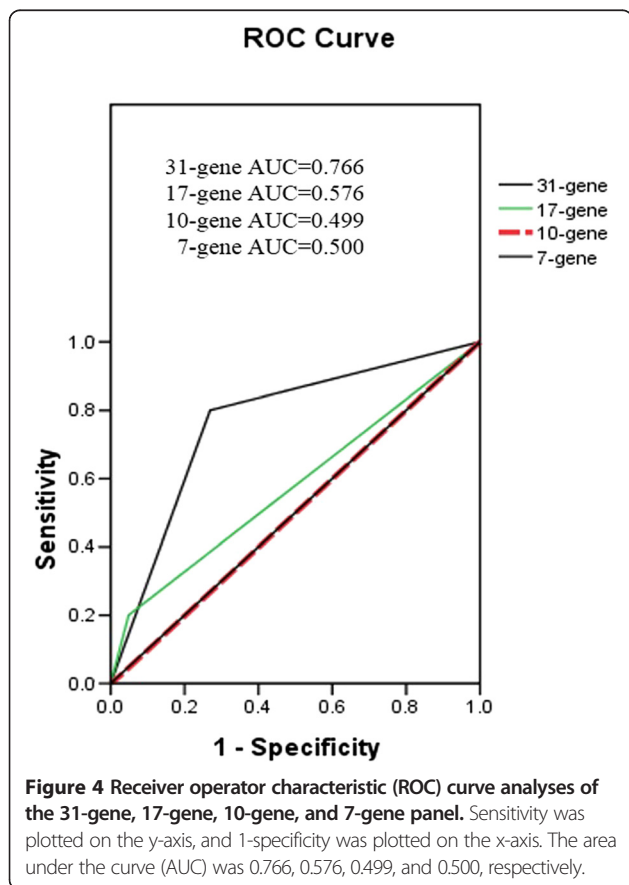
sensitivity and specificity of this set were 90.6%, 73.1% and 80.0%, respectively.

Kaplan-Meier survival curves and the log-rank test were used to compare the survival differences between the two groups. The predicted low-risk group had a longer RFS than the predicted high-risk group (5-year RFS 88.5% vs. 41.3%, $P < 0.001$) (Figure 5). When the signature was separately applied to patients with stage II and stage III disease, it correctly classified the low- and high-risk patients in both groups (stage II, 5-year RFS 98.4% vs. 64.1%, $P < 0.001$; stage III, 5-year RFS 74.6% vs. 24.4%, $P < 0.001$). The univariate and multivariate Cox regression analyses of the clinical parameters and the 31-gene prognostic signature in CRC samples are provided in Table 3. In the univariate Cox regression analysis, TNM stage, T stage, N stage, and the prognostic signature were the strongest predictors of RFS. In the multivariate analysis, only the prognostic signature was a strong independent prognostic factor.

Discussion

CRC is a heterogeneous disease. Even in patients with similar pathological and clinical features, the outcome varies; some are cured, but the cancer recurs in others. Although official guidelines give suggestions for risk stratification, there are no clear recommendations for the administration of adjuvant chemotherapy [26,27]. The staging systems for CRC may have reached their limit of usefulness for predicting outcomes, but molecular methods add value. In the present study, we developed a 31-gene expression signature for the prediction of relapse among patients with stage II/III CRC based on the combined use of gene expression profiling and TLDA analysis. The identification of 31 genes that are closely associated with outcomes in patients with CRC has clinical implications. We propose that patients who have a higher risk of relapse could benefit from adjuvant therapy and that those who have lower risk could be spared what may be unnecessary treatment.

In recent years, microarray analyses for the identification of differences in gene expression patterns have increased our understanding of the molecular genetic events in CRC. Several studies have identified gene expression signatures with a prognostic impact for patients with stage II and III CRC [11,28-31]. In early studies, small sample sizes and the lack of validation of independent sample sets limited the statistical power of the conclusions drawn. However, recent publications have addressed these limitations, and promising gene expression signatures have been suggested; two promising prognostic tests based on the expression levels of different gene panels have been reported. ColoPrint measures the expression levels of 18 genes, whereas Oncotype DX



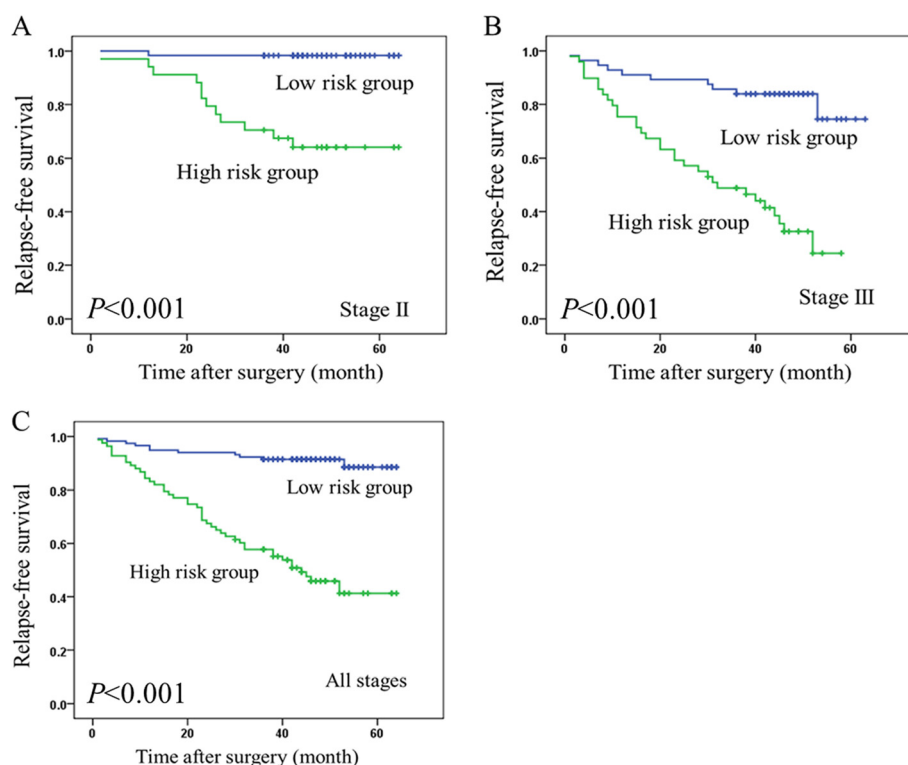


Figure 5 Kaplan-Meier survival curves illustrating relapse-free survival among the recurrent patients and non-recurrent patients.

(A), Stage II, $n = 95$; the 5-year relapse-free survival rate (RFS) for the low-risk patients was 98.4%, and the 5-year RFS of the high-risk patients was 64.1%. (B), Stage III, $n = 105$; the 5-year RFS for low-risk patients was 74.6%, and the 5-year RFS for the high-risk patients was 24.4%. (C), All stages, $n = 200$; the 5-year RFS for the low-risk patients was 88.5%, and the 5-year RFS for the high-risk patients was 41.3%.

includes 12 genes, consisting of 7 recurrence risk genes and 5 reference genes [3,32]. However, there are few data points from Chinese patients. In our previous study, an 18-gene signature (*CD36*, *DHRS13*, *DUSP2*, *FAM198B*, *FKBP5*, *GLT25D2*, *GZMB*, *IL1B*, *ITGAM*, *ITPRIPL2*, *MYBL1*, *NEAT1*, *NUDT16*, *P2RY10*, *PDE4D*, *PDZK1IP1*, *SH2D2A*, and *VSIG10*) was identified that could accurately differentiate between peripheral blood samples of CRC patients and controls. These results open an avenue for the diagnosis and early detection of CRC [33]. Additional data are needed to develop multi-gene signatures to predict prognoses of Chinese patients. Based on the microarray analysis, 33 candidate genes were selected, and algorithms were developed to identify groups of patients with high and low risk of recurrence.

In this study, we have shown that TLDA, a robust methodology based on RT-qPCR, is suitable for the validation of expression profiling of target genes. Hundreds of reactions can be performed simultaneously, enabling a large number of samples to be rapidly assessed [34,35]. This technology would be especially useful for identifying gene signatures with prognostic significance. Gallagher WM et al. [36] combined the use of DNA microarray and a tailored TLDA card and identified multiple molecular

determinants of melanoma progression. In this study, we have developed a TLDA assay that includes 33 genes that are differentially expressed in the recurrent group and the non-recurrent group and 12 relative genes based on a review of the literature. Forty-six genes (95.8%, 46/48) were successfully amplified by TLDA. Seventeen genes (*PTGS2*, *NUP210*, *PBK*, *GENE18*, *MEG3*, *CDCA2*, *RGL3*, *DEPDC1*, *LOC400713*, *GENE29*, *MMP2*, *TIMP1*, *MMP9*, *CDH1*, *GADD45B*, *MYBL2*, and *IGF2R*) were differentially expressed in the recurrent group and the non-recurrent group ($P < 0.05$). Eleven genes were significantly down-regulated in the recurrent group, while 6 genes were significantly up-regulated in the recurrent group. A 31-gene molecular signature was ultimately developed using logistic regression analysis (*PTGS2*, *TAF11*, *GENE3*, *SCG5*, *PLK4*, *TNFRSF11B*, *MMP3*, *UBD*, *AKR1C2*, *NUP210*, *LOC339524*, *C5orf23*, *PBK*, *BRIP1*, *MEG3*, *CDCA2*, *RGL3*, *SPATA18*, *GENE24*, *LOC400713*, *LOC253264*, *CSPP1*, *GENE30*, *C9orf11*, *CD44*, and *MMP2*). The predicted low-risk group had a longer RFS than the predicted high-risk group (5-year RFS 88.5% vs. 41.3%, $P < 0.001$). When the signature was separately applied to patients with stage II and stage III disease, it correctly classified the low- and high-risk patients in both groups. A range of molecular

Table 3 Univariate and multivariate Cox regression analysis of clinical factors and the 31-gene prognostic signature in colorectal samples

Parameters	No. of patients	Univariate		Multivariate	
		P value	HR (95% CI)	P value	HR (95% CI)
Gender		0.303	1.321 (0.778-2.243)	0.940	1.022 (0.571-1.830)
Male	117				
Female	83				
Age		0.889	0.963 (0.656-1.641)	0.161	0.662 (0.372-1.178)
<60	113				
>= 60	87				
Tumor size		0.323	1.323 (0.759-2.308)	0.450	1.260 (0.692-2.296)
<5	141				
>= 5	59				
Location		0.376	0.776 (0.442-1.362)	0.895	0.961 (0.533-1.734)
Colon	77				
Rectal	123				
Gross appearance		0.820	1.062 (0.631-1.788)	0.599	0.863 (0.499-1.493)
Exophytic	39				
Ulcerative	5				
Infiltrative	156				
Differentiation		0.440	1.269 (0.693-2.323)	0.546	1.227 (0.631-2.388)
High	6				
Moderate	153				
Low	41				
Stage		<0.001	3.568 (1.912-6.660)	0.701	1.243 (0.410-3.764)
II	95				
III	105				
T stage		0.025	1.727 (1.069-2.790)	0.074	1.524 (0.960-2.420)
T2	15				
T3	80				
T4	105				
N stage		<0.001	2.569 (1.801-3.665)	0.024	2.101 (1.104-3.997)
N0	96				
N1	75				
N2	29				
Prognostic signature		<0.001	7.445 (3.835-14.454)	<0.001	6.752 (3.347-13.622)
High risk	83				
Low risk	117				

Bold text denotes $P < 0.05$.

markers were observed to correlate with CRC prognosis. The signature was the strongest independent prognostic factor in the univariate and multivariate analyses.

Our strategy has led to the discovery of recurrence risk genes that can be confidently associated with clinical outcome and has confirmed the previously reported associations of several genes with CRC. *PTGS2*, or

cyclooxygenase 2 (*COX2*), is the rate-limiting enzyme in the production of proinflammatory prostaglandins, which contribute to a pro-angiogenic microenvironment in tumors [37]. Several studies have demonstrated that *PTGS2* is overexpressed in various tumors, including CRC [37-39]. Treatment with selective *PTGS2* inhibitors results in reduced colorectal neoplasia risk [40].

However, the data demonstrating a relationship between *PTGS2* expression and patient survival remain inconclusive. Fux R et al. [41] demonstrated that *PTGS2* expression in tumor epithelial cells is unrelated to overall survival and to DFS. Tougeron D et al. [42] reported that aspirin intake is associated with a significant improvement in the survival of CRC patients whose tumors carried mutant but not wild-type copies of the phosphoinositide 3-kinase (*PI3KCA*) gene. In a study by Aziz A et al. [43], overexpression of *PTGS2* is associated with better recurrence-free and disease-specific survival in a large cohort of patients with carcinoma invading bladder muscle treated by cystectomy. Similarly, in our study, *PTGS2* expression was significantly up-regulated in the non-recurrent group compared with the recurrent group. We postulated that *PTGS2* overexpression is involved in the initiation stage of carcinogenesis, but after tumors reach a level of aggressiveness (such as in the stage II/III cancer cases of the present study), the maintenance of *PTGS2* expression may be associated with tumors having a better prognosis. The potential of *PTGS2* as a prognostic marker in CRC should be considered.

Selected genes of interest in the tumor signature, including *VEGFA*, *VEGFC*, *CD44*, *CDH1*, *MMP2*, *MMP9*, *TIMP1*, *CTNBN1*, *GADD45B*, *MYBL2*, *IGF2R*, and *NME1*, are known to be involved in angiogenesis, metastasis, and prognosis. Among these genes, *MMP2*, *TIMP1*, *MMP9*, *CDH1*, *GADD45B*, *MYBL2*, and *IGF2R* were differentially expressed in the recurrent group and the non-recurrent group. To date, the relationships between *MMPs/TIMP* expression and patient survival are controversial. Wong JC [17] reported that *MMP2* was consistently underexpressed in liver metastases compared to primary CRCs. Shorter time to distant metastasis and overall survival occurred in stage III rectal cancer lacking *MMP2*. Furthermore, the authors confirmed that *MMP2* inhibitors promoted cell invasion in CRC cell lines in vitro. However, Hilska M et al. [44] reported that high expression of *MMP2* in the malignant epithelium and the surrounding stroma was associated with reduced survival in colon cancer patients. A multivariate analysis showed that *TIMP3* was the only marker with an independent prognostic value. In our study, the expression of *MMP2* was down-regulated in the recurrent group compared to the non-recurrent group, which was consistent with the study by Wong JC et al. [17]. The expression of *TIMP* was up-regulated in the recurrent group. Similarly, Fong KM et al. [19] reported higher levels of *TIMP1* RNA in adenocarcinomas, which are relatively aggressive non-small cell lung cancers, and observed a striking association between these high levels and an adverse outcome. Holten-Andersen et al. [45] showed that high preoperative plasma levels of *TIMP1* are associated with short survival in CRC patients.

Therefore, instead of inhibiting metastasis, *TIMP* may play a role in determining patient survival, possibly by promoting growth or another function.

Conclusions

In summary, the molecular signature we identified in this study is closely associated with the clinical outcome of stage II/III CRC patients. This classifier is based on the mRNA expression levels of a set of 31 genes and can be used in a standardized assay for the validation of individual patients in a prospective study.

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

XZ and XD designed the study. XZ, XX, and WS conceived and performed the microarray experiments; LW and XS performed the TLDA experiments; LW, ZW, XS, XX, QW, YW, and WH analyzed the data and performed data interpretation; FR and ZL performed the literature search; LW and PW generated the figures and wrote the first manuscript draft. All authors read and approved the final manuscript.

Acknowledgement

This study was supported by the grant from the Science and Technology Commission of Shanghai Municipality (No.10DJ1400501, 10DJ1400502), National Clinical Key Discipline (2011–2015), Priority of Shanghai key discipline of medicine (2013–2015), Shanghai R&D public service platform construction projects (12DZ2295100), and the National Natural Science Foundation of China (81071791, 81401963). We also acknowledge American Journal Experts for significant revisions of this manuscript.

Author details

¹Department of Pathology, Fudan University Shanghai Cancer Center, Shanghai 200032, China. ²Department of Oncology, Shanghai Medical College, Fudan University, Shanghai 200032, China. ³Institute of Pathology, Fudan University, Shanghai 200032, China. ⁴Institutes of Biomedical Sciences, Fudan University, Shanghai 200032, China. ⁵Department of Genetics, Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center and Shanghai Industrial Technology Institute, Shanghai 201203, China. ⁶Department of Oncology, Renji Hospital, School of Medicine, Shanghai Jiaotong University, Shanghai 200127, China.

Received: 16 October 2014 Accepted: 14 January 2015

Published online: 03 February 2015

References

- Cunningham D, Atkin W, Lenz HJ, Lynch HT, Minsky B, Nordlinger B, et al. Colorectal cancer. *Lancet*. 2010;375:1030–47.
- Brenner H, Kloor M, Pox CP. Colorectal cancer. *Lancet*. 2014;383:1490–502.
- O'Connell MJ, Lavery I, Yothers G, Paik S, Clark-Langone KM, Lopatin M, et al. Relationship between tumor gene expression and recurrence in four independent studies of patients with stage II/III colon cancer treated with surgery alone or surgery plus adjuvant fluorouracil plus leucovorin. *J Clin Oncol*. 2010;28:3937–44.
- Gill S, Loprinzi CL, Sargent DJ, Thomé SD, Alberts SR, Haller DG, et al. Pooled analysis of fluorouracil-based adjuvant therapy for stage II and III colon cancer: who benefits and by how much? *J Clin Oncol*. 2004;22:1797–806.
- Roth AD, Tejpar S, Delorenzi M, Yan P, Fiocca R, Klingbiel D, et al. Prognostic role of KRAS and BRAF in stage II and III resected colon cancer: results of the translational study on the PETACC-3, EORTC 40993, SAKK 60–00 trial. *J Clin Oncol*. 2010;28:466–74.
- Watanabe T, Kobunai T, Yamamoto Y, Matsuda K, Ishihara S, Nozawa K, et al. Chromosomal instability (CIN) phenotype, CIN high or CIN low, predicts survival for colorectal cancer. *J Clin Oncol*. 2012;30:2256–64.
- Bertagnoli MM, Niedzwiecki D, Compton CC, Hahn HP, Hall M, Damas B, et al. Microsatellite instability predicts improved response to adjuvant therapy with irinotecan, fluorouracil, and leucovorin in stage III colon

- cancer: Cancer and Leukemia Group B Protocol 89803. *J Clin Oncol*. 2009;27:1814–21.
8. Tejpar S, Saridakis Z, Delorenzi M, Bosman F, Roth AD. Microsatellite instability, prognosis and drug sensitivity of stage II and III colorectal cancer: more complexity to the puzzle. *J Natl Cancer Inst*. 2011;103:841–4.
 9. Pritchard CC, Grady WM. Colorectal cancer molecular biology moves into clinical practice. *Gut*. 2011;60:116–29.
 10. Botling J, Edlund K, Lohr M, Hellwig B, Holmberg L, Lambe M, et al. Biomarker discovery in non-small cell lung cancer: integrating gene expression profiling, meta-analysis, and tissue microarray validation. *Clin Cancer Res*. 2013;19:194–204.
 11. Budinska E, Popovici V, Tejpar S, D'Ario G, Lapique N, Sikora KO, et al. Gene expression patterns unveil a new level of molecular heterogeneity in colorectal cancer. *J Pathol*. 2013;231:63–76.
 12. Terme M, Pernot S, Marcheteau E, Sandoval F, Benhamouda N, Colussi O, et al. VEGFA-VEGFR pathway blockade inhibits tumor-induced regulatory T-cell proliferation in colorectal cancer. *Cancer Res*. 2013;73:539–49.
 13. Stanton MJ, Dutta S, Zhang H, Polavaram NS, Leontovich AA, Hönscheid P, et al. Autophagy control by the VEGF-C/NRP-2 axis in cancer and its implication for treatment resistance. *Cancer Res*. 2013;73:160–71.
 14. Du L, Rao G, Wang H, Li B, Tian W, Cui J, et al. CD44-positive cancer stem cells expressing cellular prion protein contribute to metastatic capacity in colorectal cancer. *Cancer Res*. 2013;73:2682–94.
 15. Hur KI, Toiyama Y, Takahashi M, Balaguer F, Nagasaka T, Koike J, et al. MicroRNA-200c modulates epithelial-to-mesenchymal transition (EMT) in human colorectal cancer metastasis. *Gut*. 2013;62:1315–26.
 16. Wong JC, Chan SK, Schaeffer DF, Sagaert X, Lim HJ, Kennecke H, et al. Absence of MMP2 expression correlates with poor clinical outcomes in rectal cancer, and is distinct from MMP1-related outcomes in colon cancer. *Clin Cancer Res*. 2011;17:4167–76.
 17. Damery S, Nichols L, Holder R, Ward ST, Warmington S, Wilson S, et al. Assessing the value of matrix metalloproteinase 9 (MMP9) in improving the appropriateness of referrals for colorectal cancer. *Br J Cancer*. 2013;108:1149–56.
 18. Fong KM, Kida Y, Zimmerman PV, Smith PJ. TIMP1 and adverse prognosis in non-small cell lung cancer. *Clin Cancer Res*. 1996;2:1369–72.
 19. Curran S, Dundas SR, Buxton J, Leeman MF, Ramsay R, Murray GI. Matrix metalloproteinase/tissue inhibitors of matrix metalloproteinase phenotype identifies poor prognosis colorectal cancers. *Clin Cancer Res*. 2004;10:8229–34.
 20. de la Roche M, Ibrahim AE, Mieszczynek J, Bienz M. LEF1 and B9L shield beta-catenin from inactivation by Axin, desensitizing colorectal cancer cells to tankyrase inhibitors. *Cancer Res*. 2014;74:1495–505.
 21. Williams C, Hoppe HJ, Rezgui D, Strickland M, Forbes BE, Grutzner F, et al. An exon splice enhancer primes IGF2:IGF2R binding site structure and function evolution. *Science*. 2012;338:1209–13.
 22. Garcia I, Mayol G, Ríos J, Domenech G, Cheung NK, Oberthuer A, et al. A three-gene expression signature model for risk stratification of patients with neuroblastoma. *Clin Cancer Res*. 2012;18:2012–23.
 23. Andersen CL, Jensen JL, Orntoft TF. Normalization of real-time quantitative reverse transcription-PCR data: a model-based variance estimation approach to identify genes suited for normalization, applied to bladder and colon cancer data sets. *Cancer Res*. 2004;64:5245–50.
 24. Vandesompele J, De Preter K, Pattyn F, Poppe B, Van Roy N, De Paepe A, et al. Accurate normalization of real-time quantitative RT-PCR data by geometric averaging of multiple internal control genes. *Genome Biol*. 2002;3:RESEARCH0034.
 25. Chen HY, Yu SL, Chen CH, Chang GC, Chen CY, Yuan A, et al. A five-gene signature and clinical outcome in non-small-cell lung cancer. *N Engl J Med*. 2007;356:11–20.
 26. Maughan TS, Adams RA, Smith CG, Meade AM, Seymour MT, Wilson RH, et al. Addition of cetuximab to oxaliplatin-based first-line combination chemotherapy for treatment of advanced colorectal cancer: results of the randomised phase 3 MRC COIN trial. *Lancet*. 2011;377:2103–14.
 27. Seymour MT, Brown SR, Middleton G, Maughan T, Richman S, Gwyther S, et al. Panitumumab and irinotecan versus irinotecan alone for patients with KRAS wild-type, fluorouracil-resistant advanced colorectal cancer (PICCOLO): a prospectively stratified randomised trial. *Lancet Oncol*. 2013;14:749–59.
 28. Zhang JX, Song W, Chen ZH, Wei JH, Liao YJ, Lei J, et al. Prognostic and predictive value of a microRNA signature in stage II colon cancer: a microRNA expression analysis. *Lancet Oncol*. 2013;14:1295–306.
 29. Chang W, Gao X, Han Y, Du Y, Liu Q, Wang L, et al. Gene expression profiling-derived immunohistochemistry signature with high prognostic value in colorectal carcinoma. *Gut*. 2014;63:1457–67.
 30. Agesen TH, Sveen A, Merok MA, Lind GE, Nesbakken A, Skotheim RI, et al. ColoGuideEx: a robust gene classifier specific for stage II colorectal cancer prognosis. *Gut*. 2012;61:1560–7.
 31. Kennedy RD, Bylesjo M, Kerr P, Davison T, Black JM, Kay EW, et al. Development and independent validation of a prognostic assay for stage II colon cancer using formalin-fixed paraffin-embedded tissue. *J Clin Oncol*. 2011;29:4620–6.
 32. Tan IB, Tan P. Genetics: an 18-gene signature (ColoPrint(R)) for colon cancer prognosis. *Nat Rev Clin Oncol*. 2011;8:131–3.
 33. Xu Y, Xu Q, Yang L, Ye X, Liu F, Wu F, et al. Identification and validation of a blood-based 18-gene expression signature in colorectal cancer. *Clin Cancer Res*. 2013;19:3039–49.
 34. Zearo S, Kim E, Zhu Y, Zhao JT, Sidhu SB, Robinson BG, et al. MicroRNA-484 is more highly expressed in serum of early breast cancer patients compared to healthy volunteers. *BMC Cancer*. 2014;14:200.
 35. Mairinger FD, Ting S, Werner R, Walter RF, Hager T, Vollbrecht C, et al. Different micro-RNA expression profiles distinguish subtypes of neuroendocrine tumors of the lung: results of a profiling study. *Mod Pathol*. 2014;27:1632–40.
 36. Gallagher WM, Bergin OE, Rafferty M, Kelly ZD, Nolan IM, Fox EJ, et al. Multiple markers for melanoma progression regulated by DNA methylation: insights from transcriptomic studies. *Carcinogenesis*. 2005;26:1856–67.
 37. Markkula A, Simonsson M, Rosendahl AH, Gaber K, Ingvar C, Rose C, et al. Impact of COX2 genotype, ER status and body constitution on risk of early events in different treatment groups of breast cancer patients. *Int J Cancer*. 2014;135:1898–910.
 38. Haffner MC, Laimer J, Chaux A, Schäfer G, Obrist P, Brunner A, et al. High expression of prostate-specific membrane antigen in the tumor-associated neo-vasculature is associated with worse prognosis in squamous cell carcinoma of the oral cavity. *Mod Pathol*. 2012;25:1079–85.
 39. Cathcart MC, Gray SG, Baird AM, Boyle E, Gately K, Kay E, et al. Prostacyclin synthase expression and epigenetic regulation in nonsmall cell lung cancer. *Cancer*. 2011;117:5121–32.
 40. Vinogradova Y, Coupland C, Hippisley-Cox J. Exposure to cyclooxygenase-2 inhibitors and risk of cancer: nested case-control studies. *Br J Cancer*. 2011;105:452–9.
 41. Fux R, Schwab M, Thon KP, Gleiter CH, Fritz P. Cyclooxygenase-2 expression in human colorectal cancer is unrelated to overall patient survival. *Clin Cancer Res*. 2005;11:4754–60.
 42. Tougeron D, Sha D, Manthravadi S, Sinicropo FA. Aspirin and colorectal cancer: back to the future. *Clin Cancer Res*. 2014;20:1087–94.
 43. Aziz A, Lessard A, Moore K, Hovington H, Latulippe E, Larue H, et al. Improved cancer specific-survival in patients with carcinoma invading bladder muscle expressing cyclo-oxygenase-2. *BJU Int*. 2011;108:531–7.
 44. Hilska M, Roberts PJ, Collan YU, Laine VJ, Kössi J, Hirsimäki P, et al. Prognostic significance of matrix metalloproteinases-1, -2, -7 and -13 and tissue inhibitors of metalloproteinases-1, -2, -3 and -4 in colorectal cancer. *Int J Cancer*. 2007;121:714–23.
 45. Holten-Andersen MN, Stephens RW, Nielsen HJ, Murphy G, Christensen IJ, Stetler-Stevenson W, et al. High preoperative plasma tissue inhibitor of metalloproteinase-1 levels are associated with short survival of patients with colorectal cancer. *Clin Cancer Res*. 2000;6:4292–9.

Submit your next manuscript to BioMed Central and take full advantage of:

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at
www.biomedcentral.com/submit

