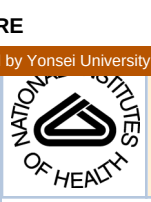

Published in final edited form as:
Gut. 2012 September ; 61(9): 1291–1298. doi:10.1136/gutjnl-2011-300812.

Prognostic gene expression signature associated with two molecularly distinct subtypes of colorectal cancer

Sang Cheul Oh[1,2], **Yun-Yong Park**[1], **Eun Sung Park**[1,3], **Jae Yun Lim**[1,4], **Soo Mi Kim**[5], **Sang-Bae Kim**[1], **Jongseung Kim**[1], **Sang Cheol Kim**[6], **In-Sun Chu**[6], **J Joshua Smith**[7,8], **R Daniel Beauchamp**[7,8], **Timothy J Yeatman**[9], **Scott Kopetz**[10], and **Ju-Seog Lee**[1]

[1]Department of Systems Biology, Division of Cancer Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

[2]Division of Hemato-Oncology, Department of Internal Medicine, Korea University Medical Center, Korea University College of Medicine, Seoul, Korea

[3]Institute for Medical Convergence, Yonsei University College of Medicine, Seoul, Korea

[4]Department of Medical Oncology, Yonsei University College of Medicine, Seoul, Korea

[5]Department of Physiology, Chonbuk National University Medical School and Hospital, Jeonju, Korea

[6]Korean Bioinformation Center, Korea Research Institute of Bioscience and Biotechnology, Daejeon, Korea

[7]Department of Surgery, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

[8]Department of Cell and Developmental Biology, Vanderbilt University School of Medicine, Nashville, Tennessee, USA

[9]Department of Surgery, Moffitt Cancer Center and Research Institute, Tampa, Florida, USA

[10]Department of Gastrointestinal Medical Oncology, Division of Cancer Medicine, The University of Texas MD Anderson Cancer Center, Houston, Texas, USA

Abstract

Aims—Despite continual efforts to develop prognostic and predictive models of colorectal cancer by using clinicopathological and genetic parameters, a clinical test that can discriminate between patients with good or poor outcome after treatment has not been established. Thus, the authors aim to uncover subtypes of colorectal cancer that have distinct biological characteristics associated with prognosis and identify potential biomarkers that best reflect the biological and clinical characteristics of subtypes.

Correspondence to Dr Ju-Seog Lee, Department of Systems Biology, Division of Cancer Medicine, Unit 950, The University of Texas MD Anderson Cancer Center, 1515 Holcombe Blvd, Houston, TX 77030, USA; jlee@mdanderson.org.

Competing interests None.

Contributors SCO, YYP and JSL participated in the design and performance of the study. JJS, RDB, TJY and SK provided clinical data. SCO, YYP, JSL, JYL, SMK, SBK, JK, ESP and SK participated in analysis and interpretation of the data. SCK, ISC and JSL performed statistical analysis. The manuscript was drafted by SCO, YYP and JSL and reviewed by all authors. All authors read and approved the final manuscript.

SC Oh and Y-Y Park contributed equally to this study.

Provenance and peer review Not commissioned; externally peer reviewed.

Methods—Unsupervised hierarchical clustering analysis was applied to gene expression data from 177 patients with colorectal cancer to determine a prognostic gene expression signature. Validation of the signature was sought in two independent patient groups. The association between the signature and prognosis of patients was assessed by Kaplan–Meier plots, log-rank tests and the Cox model.

Results—The authors identified a gene signature that was associated with overall survival and disease-free survival in 177 patients and validated in two independent cohorts of 213 patients. In multivariate analysis, the signature was an independent risk factor (HR 3.08; 95% CI 1.33 to 7.14; p=0.008 for overall survival). Subset analysis of patients with AJCC (American Joint Committee on Cancer) stage III cancer revealed that the signature can also identify the patients who have better outcome with adjuvant chemotherapy (CTX). Adjuvant chemotherapy significantly affected disease-free survival in patients in subtype B (3-year rate, 71.2% (CTX) vs 41.9% (no CTX); p=0.004). However, such benefit of adjuvant chemotherapy was not significant for patients in subtype A.

Conclusion—The gene signature is an independent predictor of response to chemotherapy and clinical outcome in patients with colorectal cancer.

INTRODUCTION

Colorectal cancer is one of the most common cancers in the USA and the rest of the world, accounting for an estimated 146 900 new cases and 49 920 deaths in 2009 in the USA alone.[12] Although surgical resection is highly effective for patients with early-stage colon cancers, a high proportion of patients have relapse after complete surgical resection, with 40% to 50% of patients with stage III disease experiencing such relapse within 5 years.[34] Development of various chemotherapy regimens including 5-fluorouracil, oxaliplatin and irinotecan as the preferred treatment has considerably improved tumour response rate and median overall survival (OS).[56] The use of recently developed targeted drugs such as cetuximab and bevacizumab in combination with chemotherapy has further improved the survival of patients with advanced colon cancer.[78] However, there is considerable clinicopathological heterogeneity among the tumours. In addition, tumours with similar histopathological appearance can follow significantly different clinical courses. Approximately 50% of patients with advanced colon cancer have a radiographic response after systemic chemotherapy.[5] Thus, for better patient care and management, it is important to understand any molecular heterogeneity significantly associated with this differential response to chemotherapy and to develop models to predict those patients who would benefit the most or least. Clinicopathological staging systems (both American Joint Committee on Cancer (AJCC) and Dukes staging) have been the gold standard for prognostication.[9] However, they offer little information about response to treatment in individual patients or about potential therapeutic targets.

Recent advances in technology allow us unbiased genome-wide screening of potential markers or gene expression signatures that might well reflect prognosis. This approach has shown potential success in the study of breast cancer.[10] In the current study, by applying a genome-wide survey of gene expression data, we attempted to distinguish subtypes of colon cancer that have distinct biological characteristics associated with prognosis and to identify potential biomarker genes or a gene expression signature that best reflects the biological and clinical characteristics of each subtype. We further tried to establish a prediction model to help guide treatment strategies for patients after surgery, which can select the patients who need further treatment due to the aggressive biological characteristics of their disease. Here, we report on a limited number of genes whose expression patterns can predict the survival of patients as well as their response to chemotherapy.

METHODS

Patients and gene expression data

All clinical and gene expression data are available from the National Center for Biotechnology Information Gene Expression Omnibus database (http://www.ncbi.nlm.nih.gov/geo). Gene expression data from the Moffit Cancer Center (Moffit cohort, GSE17536, n=177) were used as the exploration data set.[11] Gene expression data from the Vanderbilt Medical Center (GSE17537, n=55) and Max Planck Institute (GSE12945, n=62) were pooled and used as the first validation data set (Vanderbilt and Max Planck (VMP) cohort, n=117).[1112] Gene expression data from the Royal Melbourne Hospital that is part of GSE14333 (n=96) were used as the second validation data set.[13] Gene expression data of these patients were redeposited as an independent data set to Gene Expression Omnibus (GSE29971). To test the prognostic significance of gene expression signatures, we used only gene expression data with available patient survival data. Although three prognostic variables (OS, disease-specific survival and disease-free survival (DFS)) were available for the Moffit cohort, only OS and DFS data were available for the VMP and Melbourne cohorts, respectively.

Adjuvant chemotherapy data were available only for the Moffit, Vanderbilt Medical Center and Melbourne cohorts. Of the 328 patients in the Moffit, Vanderbilt and Melbourne cohorts, 147 (2 in AJCC stage I, 28 in stage II, 81 in stage III and 36 in stage IV) had received standard adjuvant chemotherapy (either single-treatment 5-fluorouracil/capecitabine or a combination of 5-fluorouracil and oxaliplatin). The remaining patients did not receive chemotherapy (n 168) or treatment data were not available (n=13). DFS was defined in a previous study as the time from surgery to the first confirmed relapse and censored when a patient died or was alive without recurrence at last contact.[13]

Statistical analysis of microarray data

BRB-ArrayTools (http://linus.nci.nih.gov/BRB-ArrayTools.html) were used primarily for statistical analysis of gene expression data,[14] and all other statistical analyses were performed in the R language environment (http://www.r-project.org). All gene expression data were generated by using the Affymetrix U133 version 2.0 platform except for the Max Planck Institute cohort. Data from the Max Planck Institute were generated by using the Affymetrix U133A platform. Raw data were downloaded from public databases and normalised using a robust multi-array averaging method.[15] We identified genes that were differentially expressed among the two classes using a random-variance t test. Differences of gene expression between two classes were considered statistically significant if their p value was less than 0.001. A stringent significance threshold was used to limit the number of false-positive findings. We also performed a global test of whether the expression profiles differed between the classes by permuting the labels of which arrays corresponded to which classes. For each permutation, the p values were recomputed and the number of genes significant at the 0.001 level was noted. The proportion of the permutations that gave at least as many significant genes as with the actual data was the significance level of the global test. Cluster analysis was performed with Cluster and TreeView.[16]

To predict the class of the independent patient cohort, we adopted a previously developed model using different algorithms.[17–19] In order to integrate each data set for constructing prediction models, we aligned gene features in each data set according to Affymetrix probe ID. Although 114 probes were identified as candidate probe set for constructing prediction models in t test analysis, we constructed master prediction models by using gene expression data from 80 probes for both validation sets because only 80 probes are shared in both U133A and U133 v2.0 due to the different probe size of microarray platforms. In addition,

we used the full probe (114 probes) model for the second validation set (Melbourne) to assess the difference between two prediction models. Gene expression data from different cohorts were independently centralised by subtracting mean expression value across samples before pooling them together for building prediction models. Then, gene expression data in the training set (Moffit cohort) were combined to form a classifier according to a given algorithm (compound covariate predictor (CCP),[20] linear discriminant analysis (LDA)[21] or Bayesian compound covariate predictor (BCCP)).[22] Full mathematical description of three prediction models is available in the supplementary methods and the assembled data set used for constructing prediction models is also available as supplementary data. The robustness of the classifier was estimated by the misclassification rate determined during the leave-one-out cross-validation (LOOCV) in the training set. For each prediction, training of classifier was done independently and misclassification rate was calculated during each training. When applied to the independent validation sets (VMP and Melbourne cohorts), prognostic significance was estimated by Kaplan–Meier plots and log-rank tests between two predicted subgroups of patients. After LOOCV, the sensitivity and specificity of prediction models were estimated by the fraction of samples correctly predicted. Multivariate Cox proportional hazard regression analysis was used to evaluate independent prognostic factors associated with survival, and gene signature, tumour stage and pathological characteristics were used as covariates. Cox proportional hazard regression model was also used to analyse the interaction between subgroups and adjuvant chemotherapy treatment. A p value of less than 0.05 was considered to indicate statistical significance, and all tests were two tailed. All subset analyses are based on predicted outcome with the 80-probe model.

Gene set and gene network analysis

Ingenuity Pathways Analysis (IPA; Ingenuity Systems, http://www.ingenuity.com) was used for gene set enrichment analysis and gene network analysis. Gene set enrichment analysis was carried out to identify the most significant gene sets associated with disease process, molecular and cellular functions and normal physiological and development condition in 114 prognostic genes as described in the instruction from Ingenuity Systems. The significance of over-represented gene sets was estimated by the right-tailed Fisher's exact test. Gene network analysis was carried out by using a global molecular network developed from information contained in the Ingenuity Knowledge Base. Eight hundred and eighty-two gene features were mapped to the Ingenuity Knowledge Base. Identified gene networks were ranked according to scores provided by IPA. The score is the likelihood of a set of genes being found in the networks due to random chance. For example, a score of 3 indicates that there is a 1/1000 chance that the focus genes are in a network due to random chance.

RESULTS

Significant association of prognosis with two subgroups found by hierarchical clustering

To uncover potential subgroups of colorectal cancer, we applied hierarchical clustering analysis to gene expression data (table 1, Moffit cohort, n=177) and found two major subgroups of colorectal cancer (supplementary figure S1). When the association of the two subgroups with clinical variables was examined, significant association with the two subgroups was apparent only for patient survival (supplementary table S1, 53% vs 29%, p=0.0007, by χ^2 test). As expected, Kaplan–Meier plots and the log-rank test showed significant differences in all prognostic variables including OS and DFS ($p=5.6\times10^{-4}$ and p=0.01, respectively, by log-rank test; figure 1).

We next sought to identify a limited number of genes whose expression is tightly associated with the two subgroups. By applying a stringent threshold cut-off (p<0.001 and twofold difference), we identified 114 gene features (supplementary figure S2 and supplementary

table S2). Interestingly, expressions of *SPP1* and *POSTN*, previously reported as metastasis genes and associated with poor prognosis in colon cancer,[2324] were much higher in the poor prognosis subgroup B. To uncover biological characteristics enriched in 114 genes, we carried out gene set enrichment analysis with IPA. As expected, it revealed enrichment of genes whose function is highly associated with cancer, cell growth and proliferation and tumour morphology (supplementary figure S3), indicating that selected genes might well reflect the biological characteristics of the two subgroups of colon cancer.

Validation of the prognostic gene expression signature in independent cohorts

Having in hand a distinct gene expression signature (114 genes) that well reflects the prognosis of patients with colorectal cancer, we next sought to validate the association of the signature with prognosis in independent patient cohorts. For this validation, gene expression data from the Vanderbilt Medical Center (n=55) and Max Planck Institute (n=62) were pooled (VMP cohort, n=117), and previously established data training and prediction methods were applied and gene expression data from the Moffit cohort were used for training of classifiers. Because 80 probes are common to all training and validation sets, gene expression data from 80 probes were used for constructing the prediction model. When patients with colon cancer in the VMP cohort were stratified according to a cluster-specific gene expression signature, Kaplan–Meier plots showed significant differences (p=0.032 by log-rank test) in OS of patients between the two subtypes that were predicted by CCP (figure 2). Specificity and sensitivity for correctly predicting subtype B in the test set (Moffit cohort) during LOOCV were 0.826 and 0.868, respectively. When CCP was replaced by different prediction algorithms such as LDA and BCCP, the prognostic difference between the two subgroups remained significant (p=0.04 and p=0.005 by log-rank test for LDA and BCCP, respectively).

We next performed subset analysis only with patients with stage II or stage III cancer. In subset analysis, the gene expression signature successfully identified patients with poor survival among those with stage II and stage III cancer (supplementary figure S4), supporting the notion that the gene expression signature is independent of the current staging systems. Univariate and multivariate Cox proportional hazards regression analyses were undertaken in the VMP cohort to evaluate the prognostic value of the gene expression signature in combination with other clinical variables including patient age at diagnosis, AJCC stage, sex and grade. In the univariate analysis, the gene signature and AJCC stage were significantly associated with OS (p=0.036 and $p=1.16\times10^{-7}$, respectively). In the multivariate analysis, AJCC stage and gene signature still maintained the significance ($p=1.59310^{5}$ and p=0.008), and the grade showed marginal significance (p=0.04) (table 2).

To further test the robustness of the gene expression signature, we applied the 80-probe prediction model to a second validation cohort (Melbourne cohort, n=96). Similar to the results for the two previous cohorts, there is a substantial trend of association with prognosis (supplementary figure S5B). When the full probe model (114 probes) was applied, the association of signature with prognosis became more significant (supplementary figure S5B). In the univariate analysis with predicted outcome from the 114-probe model, the gene signature and AJCC stage were significantly associated with DFS (p=0.048 and p=0.002, respectively), and significance remains similar in the multivariate analysis (p=0.049 and 0.002, respectively) (supplementary table S3).

Significant association of the signature with DFS of patients after adjuvant chemotherapy

Adjuvant chemotherapy data were available for 328 of the 390 patients from the three cohorts. We therefore sought to determine the association of the new prognostic gene expression signature (80-probe model) with response to chemotherapy. As expected, the

difference in DFS between the two subtypes remained significant in the combined subset of patients (3-year rate, 89.1% (A) vs 74.6% (B); $p=7.8\times10^{-4}$ by log-rank test, figure 3A) after excluding patients with stage IV cancer. The HR of subtype B for relapse was 2.7 (95% CI 1.47 to 4.79; p=0.001). To examine the association of the signature with response to adjuvant chemotherapy, we performed subset analysis with patients in AJCC stage III (n=109), a stage for which the benefit of adjuvant chemotherapy has been well established.[25–27] Patients with stage III disease were subdivided into two subtypes (A or B), and the difference in DFS was independently assessed. Adjuvant chemotherapy significantly affected DFS in patients in subtype B (3-year rate, 71.2% (CTX) vs 41.9% (no CTX); p=0.004 by log-rank test, figure 3C). However, such benefit of adjuvant chemotherapy was not significant for patients in subtype A (3-year rate, 85.4% (CTX) vs 75.2% (no CTX); p=0.5 by log-rank test, figure 3D). When Cox regression model was applied, the interaction of subtypes with adjuvant chemotherapy reached a significance level of 0.36 (supplementary figure S6). However, consistent with the Kaplan–Meier plot and log-rank test, the estimated HR for adjuvant chemotherapy in subgroup B was 0.31 (95% CI 0.14 to 0.73; p=0.007), while HR for relapse for adjuvant chemotherapy in subtype A was 0.67 (95% CI 0.19 to 2.34; p=0.5).

Biological insight of prognostic and predictive signature

Although the 114-gene expression signature was robust enough to discriminate between patients in all three different cohorts, the number of genes was too small to develop a gene network analysis, because we applied an extremely stringent cut-off (p<0.001 and twofold) to avoid any potential false-positive problem during signature-based prediction. Thus, to explore the biological characteristics of subgroups with poorer prognosis (patients in subtype B), we attempted to identify genes whose expression patterns were conserved in all three cohorts. To maximise the compatibility of the three data sets, we included only gene expression data generated by using the Affymetrix U133 v2.0. Gene lists X, Y and Z (figure 4A) represent the 882 genes that were differentially expressed between subgroups A and B in all three cohorts.

We next performed pathway analysis on the 882 genes in the extended gene list (figure 4B) using the IPA tool (supplementary table S4). This analysis revealed a series of putative networks. Functional connectivity of the top network revealed a strong over-representation of TGFβ pathways (supplementary figure S7), suggesting that its activation might be a key genetic determinant associated with poorer survival of patients with colon cancer in subtype B. In addition, genes involved in the activation of the NF-κB pathway, a key survival pathway in many cancers, are overexpressed in subtype B colon cancer (supplementary figure S8), suggesting that activation of NF-κB might be, in part, accountable for poorer survival of patients in subtype B. Interestingly, two genes in the NF-κB network are FYN and LYN–part of SRC tyrosine kinase family–whose activity are frequently increased and well associated with metastasis potential and poorer outcome in colon cancer.[28,29]

DISCUSSION

By analysing gene expression data from colorectal cancer tissues, we identified a limited number of genes that are significantly associated with prognosis. Robustness of the signature was validated in two independent cohorts. Subset analysis with only patients with stage III cancer indicated that the signature might be associated with the benefit of adjuvant chemotherapy. Current staging systems and biomarkers are very limited in providing therapeutic guidance to patients with colorectal cancer. Our gene expression signature may open up new opportunities to develop molecular stratification of patients and provide treatment guidance.

We applied two independent but complimentary methods to construct the signature, test its robustness and validate its association with clinical outcomes. In the first approach, we used unsupervised clustering to identify subgroups that differed with respect to gene expression patterns. Subsequent analysis with clinical data revealed that the two subgroups differed significantly in OS and DFS, indicating that biological characteristics reflected in gene expression patterns may well represent clinical heterogeneity that has not been properly addressed in the current staging systems. In a second approach, we applied supervised prediction models to validate the association of the signature with clinical outcomes in two independent patient cohorts. The robustness of the 114-gene signature that we studied was supported by the high sensitivity (>0.8) and spec-ificity (>0.8) during training of prediction models within the Moffit cohort and a significant association of predicted outcome with patient prognosis in both test cohorts (figure 2 and supplementary figure S5).

Subset analysis of patients with available chemotherapy data suggested that the 114-gene signature might be able to predict which patients would benefit from adjuvant chemotherapy. In patients with stage III disease, 5-fluorouracil-based chemotherapy was significantly associated with improved outcome for patients in subtype B (HR 0.31; 95% CI 0.14 to 0.73; p=0.007), whereas its benefit was not statistically significant for patients in subtype A. Thus, our newly identified gene signature showed a mixed prognostic and predictive association in the current study. This mixed association is in good agreement with previous findings in breast cancer showing that a strong prognostic recurrence score based on the Oncotype DX assay is highly predictive for response to chemotherapy as well.[1030]

Although an OS benefit for 5-fluorouracil-based adjuvant chemotherapy has been established for patients with AJCC stage III cancer,[252631] the use of adjuvant chemotherapy remains controversial for patients with AJCC stage II disease. The QUick and Simple And Reliable (QUASAR) trial reported a modest but statistically significant improvement in DFS in patients with stage II cancer treated with 5-fluorouracil-based adjuvant chemotherapy compared with observation.[32] However, no additional trials have been performed appropriately to validate it in patients with stage II cancer, owing in part to the large sample size that would be required. It has been suggested that at least 9680 patients per group would be required to detect a 2% survival difference between the treatment and control arms, with 90% power and a 0.05 significance level.[33] The findings of QUASAR and other pooled analyses suggest that there might be a subset of patients with stage II cancer at high risk of recurrence who might benefit from 5-fluorouracil-based adjuvant chemotherapy. Our analysis was limited to patients with stage III cancer because the number of patients with stage II disease who received chemotherapy was too small in our current study cohort. The use of new predictive gene signatures may help reduce the number of patients needed in prospective clinical trials to estimate the benefit of chemotherapy in patients with stage II cancer by identifying patients at higher risk in advance of treatment.

Molecular functions of genes upregulated in subtype B were in good agreement with clinical characteristics of that subtype. Interestingly, network analysis revealed that many of these genes were under the control of the TGFβ pathway, which is frequently associated with metastasis in many types of cancers by promoting epithelial and mesenchymal transition.[3435] Activation of *SRC* family kinases in poorer prognostic subtype is in good agreement with previous studies showing that SRC or its related kinase activity increases in colorectal tumours relative to adjacent mucosa, with the highest activity observed in metastases, and correlates inversely with patient survival.[2836] Therefore, these genes overexpressed in patients with subtype B well reflect the aggressiveness of colorectal cancer cells. SRC-targeted agents, which are now in advanced clinical development for patients with solid tumours, might be good candidates for targeted treatment for patients in subtype B.[37]

Several previous studies have tried to identify prognostic gene expression signatures. Twenty-three-gene and 30-gene prognostic signatures were independently developed to predict recurrence in patients with AJCC stage II disease.[3839] However, these signatures have not been validated in independent patient groups by other investigators and cannot predict the response to adjuvant chemotherapy. A recent study developed a risk score of recurrence based on evolutionarily conserved 34 genes.[11] Although its independence over the use of stage has not been firmly established, a 128-gene signature was identified as a marker for the genomic stage of colorectal cancer and was well associated with prognosis.[13] A new multi-gene expression assay for colon cancer, known as Oncotype DX, has been introduced with the aim of improving treatment decisions, especially for patients with stage II disease.[40] Although this seven-gene prognostic marker was validated in the QUASAR cohort, the chemotherapy-benefit gene signature was not validated in the same cohort. Interestingly, of the seven genes in the prognostic Oncotype DX, three (*FAP, INHBA* and *BGN*) are present in our extended gene list (figure 4), suggesting that there might be partial overlap between our prognostic subtypes (A and B) and the high- and low-risk groups identified by the Oncotype DX recurrence score.

Our new signature may overcome the current limitation of biomarkers of colorectal cancer. Among several interesting biomarkers linked with clinical outcomes of patients with colorectal cancer,[41] microsatellite instability (MSI) is the only prognostic marker validated in multiple studies and independently of stage.[4243] Although the predictive values of MSI to adjuvant chemotherapy and of *KRAS* mutations to the use of epidermal growth factor receptor inhibitors have been established, these markers are only useful as negative markers for treatments.[4445] Thus, these markers fail to predict which patients will benefit from treatments.

While it is interesting to see the association of the signature with the potential benefit of adjuvant chemotherapy in patients with stage III colorectal cancer, the predictive nature of the signature is not firmly established yet since interaction of subgroup with adjuvant chemotherapy (or heterogeneity of two subgroups over adjuvant chemotherapy) did not reach significance level (supplementary figure S6). However, due to the small number of patients used in the analysis, it would be too premature to draw a strong conclusion for the predictive nature of the signature. Although significant, our multivariate analysis (table 2) has also some limitations since other known predictors of prognosis in colorectal cancer such as the refined TN substages, MSI status and number of examined nodes are not included. Thus, the significance and robustness of the signature as prognostic markers and predictive markers for adjuvant chemotherapy remain to be determined in future studies.

In conclusion, we identified two new prognostic subtypes of colorectal cancer that showed a significant difference in the survival of patients. The gene signature could predict the response to adjuvant chemotherapy. This study clearly demonstrated that our gene signature reflects the molecular characteristics of patients with colorectal cancer and provides us an opportunity for the rational design of future clinical trials for testing the benefit of adjuvant chemotherapy for patients with stage II and ultimately stage III colorectal cancer. This result, if confirmed in prospective studies, might improve patient care by providing more practical guidance for different treatments.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We are grateful to Dr Richard Simon for his help in preparing the mathematical description of prediction models.

Funding This study was supported by the G. S. Hogan Gastrointestinal Research Fund from The University of Texas MD Anderson Cancer Center, and partially by CA069457, CA095103, CA068485, and GM088822.

REFERENCES

1. Parkin DM, Bray F, Ferlay J, et al. Global cancer statistics, 2002. CA Cancer J Clin. 2005; 55:74–108. [PubMed: 15761078]
2. Jemal A, Siegel R, Ward E, et al. Cancer statistics, 2009. CA Cancer J Clin. 2009; 59:225–49. [PubMed: 19474385]
3. Carlsson U, Lasson A, Ekelund G. Recurrence rates after curative surgery for rectal carcinoma, with special reference to their accuracy. Dis Colon Rectum. 1987; 30:431–4. [PubMed: 3595361]
4. Midgley R, Kerr D. Colorectal cancer. Lancet. 1999; 353:391–9. [PubMed: 9950460]
5. Midgley RS, Yanagisawa Y, Kerr DJ. Evolution of nonsurgical therapy for colorectal cancer. Nat Clin Pract Gastroenterol Hepatol. 2009; 6:108–20. [PubMed: 19153564]
6. Kopetz S, Chang GJ, Overman MJ, et al. Improved survival in metastatic colorectal cancer is associated with adoption of hepatic resection and improved chemotherapy. J Clin Oncol. 2009; 27:3677–83. [PubMed: 19470929]
7. Cunningham D, Humblet Y, Siena S, et al. Cetuximab monotherapy and cetuximab plus irinotecan in irinotecan-refractory metastatic colorectal cancer. N Engl J Med. 2004; 351:337–45. [PubMed: 15269313]
8. Kabbinavar F, Hurwitz HI, Fehrenbacher L, et al. Phase II, randomized trial comparing bevacizumab plus fluorouracil (FU)/leucovorin (LV) with FU/LV alone in patients with metastatic colorectal cancer. J Clin Oncol. 2003; 21:60–5. [PubMed: 12506171]
9. Mamounas E, Wieand S, Wolmark N, et al. Comparative efficacy of adjuvant chemotherapy in patients with Dukes' B versus Dukes' C colon cancer: results from four National Surgical Adjuvant Breast and Bowel Project adjuvant studies (C-01, C-02, C-03, and C-04). J Clin Oncol. 1999; 17:1349–55. [PubMed: 10334518]
10. Paik S, Shak S, Tang G, et al. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med. 2004; 351:2817–26. [PubMed: 15591335]
11. Smith JJ, Deane NG, Wu F, et al. Experimentally derived metastasis gene expression profile predicts recurrence and death in patients with colon cancer. Gastroenterology. 2010; 138:958–68. [PubMed: 19914252]
12. Staub E, Groene J, Heinze M, et al. An expression module of WIPF1-coexpressed genes identifies patients with favorable prognosis in three tumor types. J Mol Med. 2009; 87:633–44. [PubMed: 19399471]
13. Jorissen RN, Gibbs P, Christie M, et al. Metastasis-associated gene expression changes predict poor outcomes in patients with Dukes stage B and C colorectal cancer. Clin Cancer Res. 2009; 15:7642–51. [PubMed: 19996206]
14. Simon R, Lam A, Li MC, et al. Analysis of gene expression data using BRB-ArrayTools. Cancer Inform. 2007; 3:11–17. [PubMed: 19455231]
15. Irizarry RA, Hobbs B, Collin F, et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostatistics. 2003; 4:249–64. [PubMed: 12925520]
16. Eisen MB, Spellman PT, Brown PO, et al. Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci U S A. 1998; 95:14863–8. [PubMed: 9843981]
17. Lee JS, Heo J, Libbrecht L, et al. A novel prognostic subtype of human hepatocellular carcinoma derived from hepatic progenitor cells. Nat Med. 2006; 12:410–16. [PubMed: 16532004]
18. Lee JS, Chu IS, Mikaelyan A, et al. Application of comparative functional genomics to identify best-fit mouse models to study human cancer. Nat Genet. 2004; 36:1306–11. [PubMed: 15565109]
19. Lee JS, Chu IS, Heo J, et al. Classification and prediction of survival in hepatocellular carcinoma by gene expression profiling. Hepatology. 2004; 40:667–76. [PubMed: 15349906]

20. Radmacher MD, McShane LM, Simon R. A paradigm for class prediction using gene expression profiles. J Comput Biol. 2002; 9:505–11. [PubMed: 12162889]
21. Dudoit S, Fridlyand F, Speed TP. Comparison of discrimination methods for classification of tumors using DNA microarrays. J Am Stat Assoc. 2002; 97:77–87.
22. Wright G, Tan B, Rosenwald A, et al. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. Proc Natl Acad Sci U S A. 2003; 100:9991–6. [PubMed: 12900505]
23. Bao S, Ouyang G, Bai X, et al. Periostin potently promotes metastatic growth of colon cancer by augmenting cell survival via the Akt/PKB pathway. Cancer Cell. 2004; 5:329–39. [PubMed: 15093540]
24. Yeatman TJ, Chambers AF. Osteopontin and colon cancer progression. Clin Exp Metastasis. 2003; 20:85–90. [PubMed: 12650611]
25. Laurie JA, Moertel CG, Fleming TR, et al. Surgical adjuvant therapy of large-bowel carcinoma: an evaluation of levamisole and the combination of levamisole and fluorouracil. The North Central Cancer Treatment Group and the Mayo Clinic. J Clin Oncol. 1989; 7:1447–56. [PubMed: 2778478]
26. Moertel CG, Fleming TR, Macdonald JS, et al. Levamisole and fluorouracil for adjuvant therapy of resected colon carcinoma. N Engl J Med. 1990; 322:352–8. [PubMed: 2300087]
27. Moertel CG, Fleming TR, Macdonald JS, et al. Fluorouracil plus levamisole as effective adjuvant therapy after resection of stage III colon carcinoma: a final report. Ann Intern Med. 1995; 122:321–6. [PubMed: 7847642]
28. Yeatman TJ. A renaissance for SRC. Nat Rev Cancer. 2004; 4:470–80. [PubMed: 15170449]
29. Kopetz S, Shah AN, Gallick GE. Src continues aging: current and future clinical directions. Clin Cancer Res. 2007; 13:7232–6. [PubMed: 18094400]
30. Sparano JA, Paik S. Development of the 21-gene assay and its application in clinical practice and clinical trials. J Clin Oncol. 2008; 26:721–8. [PubMed: 18258979]
31. Wolmark N, Rockette H, Fisher B, et al. The benefit of leucovorin-modulated fluorouracil as postoperative adjuvant therapy for primary colon cancer: results from National Surgical Adjuvant Breast and Bowel Project protocol C-03. J Clin Oncol. 1993; 11:1879–87. [PubMed: 8410113]
32. Gray R, Barnwell J, McConkey C, et al. Quasar Collaborative Group. Adjuvant chemotherapy versus observation in patients with colorectal cancer: a randomised study. Lancet. 2007; 370:2020–9. [PubMed: 18083404]
33. Benson AB III, Schrag D, Somerfield MR, et al. American Society of Clinical Oncology recommendations on adjuvant chemotherapy for stage II colon cancer. J Clin Oncol. 2004; 22:3408–19. [PubMed: 15199089]
34. Wendt MK, Allington TM, Schiemann WP. Mechanisms of the epithelialmesenchymal transition by TGF-beta. Future Oncol. 2009; 5:1145–68. [PubMed: 19852727]
35. Robson H, Anderson E, James RD, et al. Transforming growth factor beta 1 expression in human colorectal tumours: an independent prognostic marker in a subgroup of poor prognosis patients. Br J Cancer. 1996; 74:753–8. [PubMed: 8795578]
36. Lieu C, Kopetz S. The SRC family of protein tyrosine kinases: a new and promising target for colorectal cancer therapy. Clin Colorectal Cancer. 2010; 9:89–94. [PubMed: 20378502]
37. Saad F, Lipton A. SRC kinase inhibition: targeting bone metastases and tumor growth in prostate and breast cancer. Cancer Treat Rev. 2010; 36:177–84. [PubMed: 20015594]
38. Wang Y, Jatkoe T, Zhang Y, et al. Gene expression profiles and molecular markers to predict recurrence of Dukes' B colon cancer. J Clin Oncol. 2004; 22:1564–71. [PubMed: 15051756]
39. Barrier A, Boelle PY, Roser F, et al. Stage II colon cancer prognosis prediction by tumor gene expression profiling. J Clin Oncol. 2006; 24:4685–91. [PubMed: 16966692]
40. Kerr D, Gray R, Quirke P, et al. A quantitative multigene RT-PCR assay for prediction of recurrence in stage II colon cancer: selection of the genes in four large studies and results of the independent, prospectively designed QUASAR validation study. J Clin Oncol. 2009; 27(15 Suppl):4000.
41. Walther A, Johnstone E, Swanton C, et al. Genetic prognostic and predictive markers in colorectal cancer. Nat Rev Cancer. 2009; 9:489–99. [PubMed: 19536109]

42. Popat S, Hubner R, Houlston RS. Systematic review of microsatellite instability and colorectal cancer prognosis. J Clin Oncol. 2005; 23:609–18. [PubMed: 15659508]

43. Thibodeau SN, Bren G, Schaid D. Microsatellite instability in cancer of the proximal colon. Science. 1993; 260:816–19. [PubMed: 8484122]

44. Karapetis CS, Khambata-Ford S, Jonker DJ, et al. K-ras mutations and benefit from cetuximab in advanced colorectal cancer. N Engl J Med. 2008; 359:1757–65. [PubMed: 18946061]

45. Ribic CM, Sargent DJ, Moore MJ, et al. Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer. N Engl J Med. 2003; 349:247–57. [PubMed: 12867608]

Significance of this study

What is already known on this subject?

▶ Colorectal cancer is a clinically heterogeneous disease, and its heterogeneity has not been characterised at the molecular level

▶ The benefit of adjuvant chemotherapy for patients with stage II colorectal cancer is not well established

▶ There are no clinically useful biomarkers that can reliably predict the prognosis and response to adjuvant chemotherapy

What are the new findings?

▶ There seem to be distinct subtypes of colorectal cancer that are well associated with prognosis of patients

▶ Expression signature of 114 genes that best reflect the difference between two subtypes can predict the likelihood of the benefit of adjuvant chemotherapy especially in patients with AJCC (American Joint Committee on Cancer) stage III disease

▶ Prognostic and potentially predictive gene expression signature is present at the time of diagnosis

How might it impact on clinical practice in the foreseeable future?

▶ The use of gene expression profiling can improve the molecular classification of patients with colorectal cancer by adding to the existing classifications

▶ The gene expression signature might be useful markers for identifying patients with stage II disease who would have the benefit of adjuvant chemotherapy in a prospective study

▶ It will open up new opportunities for personalised treatment of patients with colorectal cancer based on stratification by prognostic and predictive gene expression signature

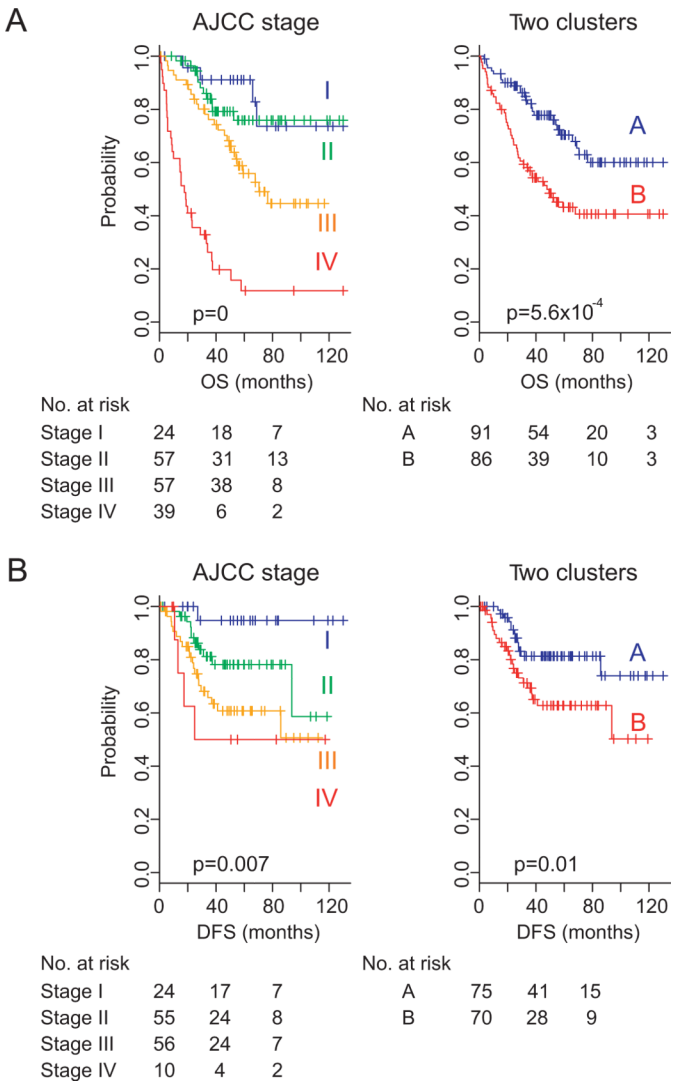

Figure 1.

Kaplan–Meier plots of the prognosis of patients with colon cancer in the Moffit cohort. Patients were stratified according to American Joint Committee on Cancer (AJCC) stage or gene expression patterns (two clusters). Disease free survival data from 32 patients are not available.

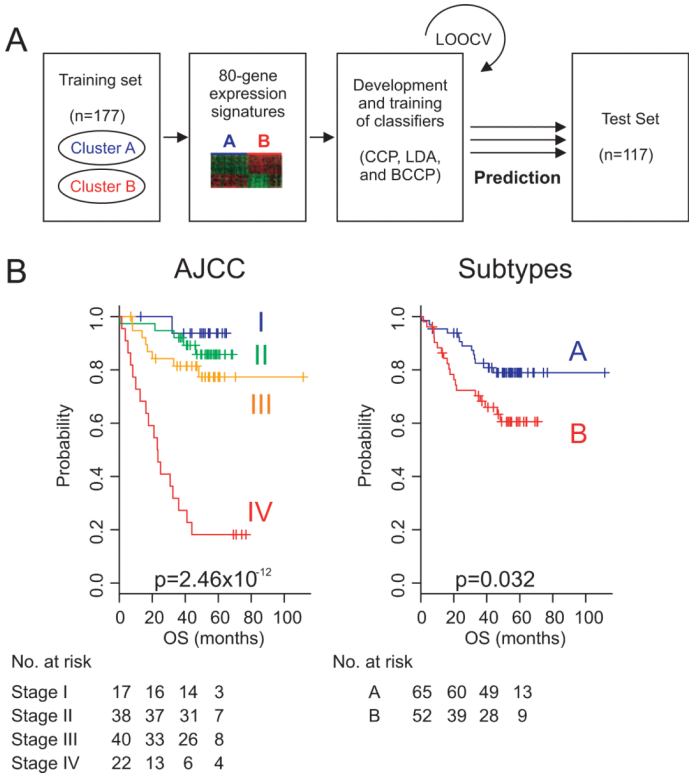

Figure 2.

Construction of prediction model in the test cohort according to gene expression signatures from the Moffit cohort. (A) Schematic overview of the strategy used for the construction of prediction models and evaluation of predicted outcomes based on gene expression signatures. (B) Kaplan–Meier plots of OS. Patients were stratified according to AJCC stage or two subgroups predicted by CCP. p Values were obtained from the log-rank test. The '+' symbols in the panels indicate censored data. AJCC, American Joint Committee on Cancer; CCP, compound covariate predictor; OS, overall survival.

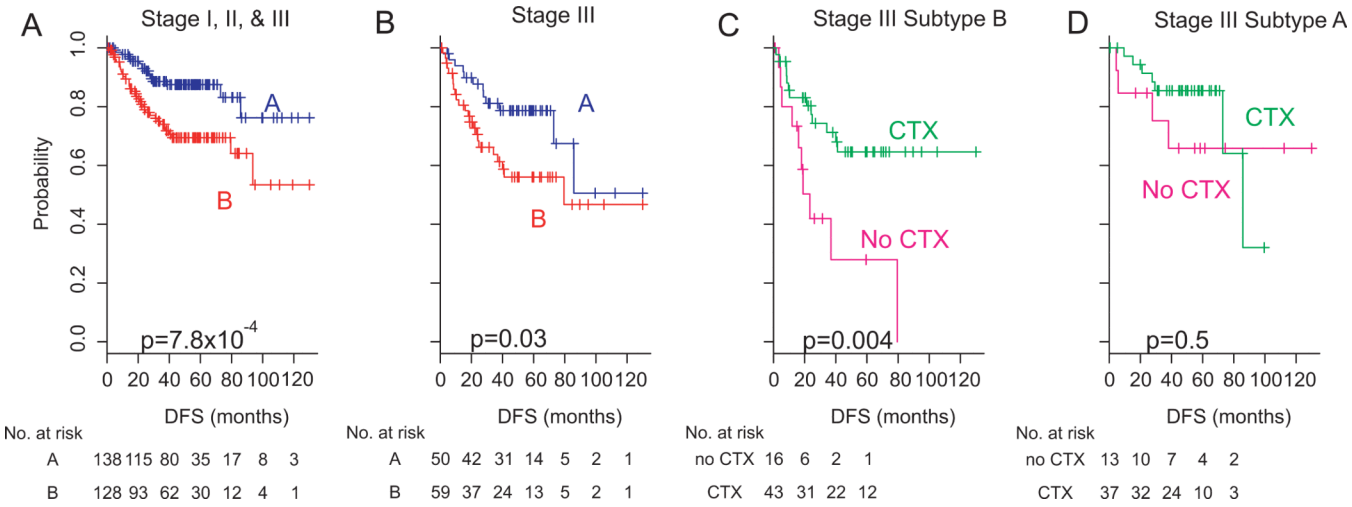

Figure 3.

Significant association of two subtypes with adjuvant chemotherapy. (A) Kaplan–Meier plots of disease-free survival (DFS) of patients with colorectal cancer in the combined cohort. Patients were plotted according to the prognostic expression signature of 80 genes (two subtypes). Patients in stage I, those in stage II and those in stage III with available adjuvant chemotherapy data were included for analysis (n=266). (B) Kaplan–Meier plots of DFS of patients with colorectal cancer in the combined cohort (patients in stage III, n=109). (C) Kaplan–Meier plots of patients in subtype B with stage III disease (n=59). Patients were plotted according to presence and absence of adjuvant chemotherapy (CTX). (D) Kaplan–Meier plots of patients in subtype A with stage III disease (n=50). Patients were plotted according to presence and absence of adjuvant chemotherapy (CTX).

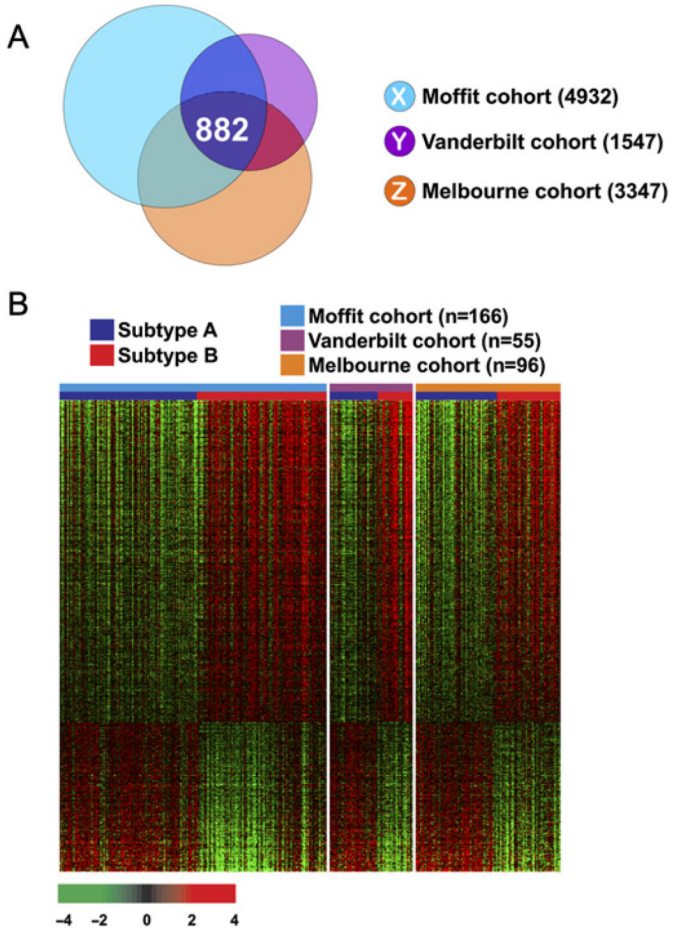

Figure 4.

Subtype-specific gene expression patterns conserved in all three cohorts of patients with colorectal cancer. (A) Venn diagram of genes with expression that differed significantly between patients with colorectal cancer in subtypes A and B in the three different cohorts. Univariate test (two-sample t test) with multivariate permutation test (10 000 random permutations) was applied. In each comparison, we applied a cut-off p value of less than 0.001 to retain genes with expression that differed significantly between the two groups of tissues examined. (B) Expression patterns of selected genes shared in the three colon cancer cohorts. The expressions of only 882 genes were commonly upregulated or downregulated in all three cohorts. Colored bars at the top of the heat map represent samples as indicated.

Table 1

Clinical and pathological features of patients with colorectal cancer

Variable	Moffit cohort	VMP cohort	Melbourne cohort
Patients (n)	177	117	96
Male, n (%)	96 (54)	60 (51)	52 (54)
Female, n (%)	81 (46)	57 (49)	44 (46)
Age (years)			
Median	66	63	68
Range	26–92	23–94	30–92
Location, n (%)			
Colon	177 (100)	84 (72)	72 (75)
Rectum	0(0)	33 (28)	24 (25)
AJCC stage, n (%)			
I	24 (14)	17 (15)	19 (20)
II	57 (32)	38 (21)	42 (70)
III	57 (32)	40 (34)	35 (36)
IV	39 (22)	22 (19)	0 (0)
Grade, n (%)			
1	16 (9)	1 (1)	
2	134 (76)	63 (54)	
3	27 (15)	34 (29)	
NA		19 (16)	96 (100)
Adjuvant chemotherapy, n (%)			
Yes	82 (46)	39 (33)	26 (27)
No	82 (46)	16 (14)	70 (73)
NA	13 (8)	62 (53)	0
Number of deaths	72	32	20 *
Median follow-up (months)	42.2	48.2	40.2

AJCC, American Joint Committee on Cancer; NA, not applicable; VMP, Vanderbilt and Max Planck.

* Recurrence.

Table 2

Univariate and multivariate Cox proportional hazard regression analyses of OS in the VMP cohort (n=117)

	Univariate		Multivariate	
	HR (95% CI)	**p Value**	**HR (95% CI)**	**p Value**
Gender (M or F)	0.68 (0.34 to 1.38)	0.29	1.07 (0.43 to 2.61)	0.88
Age (>70)	1.44 (0.97 to 2.9)	0.3	1.68 (0.72 to 3.9)	0.22
AJCC stage (I, II, III, IV)	3.6 (2.24 to 5.79)	1.16×10^{-7}	2.9 (1.82 to 4.93)	1.59×10^{-5}
Grade (1, 2, 3)	1.72 (0.8 to 3.69)	0.16	2.31 (1.03 to 5.2)	0.04
Gene signature (A or B)	2.1 (1.0 to 4.3)	0.036	3.08 (1.33 to 7.14)	0.008

AJCC, American Joint Committee on Cancer; F, female; M, male; OS, overall survival; VMP, Vanderbilt and Max Planck.