

Multiple Gene Expression Classifiers from Different Array Platforms Predict Poor Prognosis of Colorectal Cancer

Yu-Hsin Lin,^{1,2} Jan Friederichs,⁶ Michael A. Black,² Jörg Mages,⁷ Robert Rosenberg,⁶ Parry J. Guilford,^{1,2,5} Vicky Phillips,³ Mark Thompson-Fawcett,³ Nikola Kasabov,⁸ Tumi Toro,⁵ Arend E. Merrie,⁹ Andre van Rij,³ Han-Seung Yoon,⁴ John L. McCall,⁹ Jörg Rüdiger Siewert,⁶ Bernhard Holzmann,⁶ and Anthony E. Reeve^{1,2}

Abstract **Purpose:** This study aimed to develop gene classifiers to predict colorectal cancer recurrence. We investigated whether gene classifiers derived from two tumor series using different array platforms could be independently validated by application to the alternate series of patients. **Experimental Design:** Colorectal tumors from New Zealand ($n = 149$) and Germany ($n = 55$) patients had a minimum follow-up of 5 years. RNA was profiled using oligonucleotide printed microarrays (New Zealand samples) and Affymetrix arrays (German samples). Classifiers based on clinical data, gene expression data, and a combination of the two were produced and used to predict recurrence. The use of gene expression information was found to improve the predictive ability in both data sets. The New Zealand and German gene classifiers were cross-validated on the German and New Zealand data sets, respectively, to validate their predictive power. Survival analyses were done to evaluate the ability of the classifiers to predict patient survival. **Results:** The prediction rates for the New Zealand and German gene-based classifiers were 77% and 84%, respectively. Despite significant differences in study design and technologies used, both classifiers retained prognostic power when applied to the alternate series of patients. Survival analyses showed that both classifiers gave a better stratification of patients than the traditional clinical staging. One classifier contained genes associated with cancer progression, whereas the other had a large immune response gene cluster concordant with the role of a host immune response in modulating colorectal cancer outcome. **Conclusions:** The successful reciprocal validation of gene-based classifiers on different patient cohorts and technology platforms supports the power of microarray technology for individualized outcome prediction of colorectal cancer patients. Furthermore, many of the genes identified have known biological functions congruent with the predicted outcomes.

Authors' Affiliations: ¹Cancer Genetics Laboratory and Departments of ²Biochemistry, ³Medical and Surgical Sciences, and ⁴Pathology, University of Otago; ⁵Pacific Edge Biotechnology Ltd., Centre for Innovation, Dunedin, New Zealand; ⁶Department of Surgery and ⁷Institut für Medizinische Mikrobiologie, Immunologie und Hygiene, Klinikum rechts der Isar, Technische Universität München, Munich, Germany; ⁸Knowledge Engineering and Discovery Research Institute, Auckland University of Technology; and ⁹Department of Surgery, University of Auckland, Auckland, New Zealand

Received 12/14/05; revised 9/28/06; accepted 11/6/06.

Grant support: Health Research Council of New Zealand, the Lottery Grants Board of New Zealand (Y-H. Lin and A.E. Reeve), and the Kommission für Klinische Forschung des Klinikums rechts der Isar (J. Friederichs, R. Rosenberg, J.R. Siewert, B. Holzmann, and J. Mages).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Note: Supplementary data for this article are available at Clinical Cancer Research Online (<http://clincancerres.aacrjournals.org/>).

J.L. McCall, J.R. Siewert, and B. Holzmann contributed equally to this work.

Requests for reprints: Anthony E. Reeve, Cancer Genetics Laboratory and Department of Biochemistry, University of Otago, 710 Cumberland Street, Dunedin, New Zealand. Phone: 64-3-479-7699; Fax: 64-3-479-7738; E-mail: a.reeve@otago.ac.nz.

©2007 American Association for Cancer Research.
doi:10.1158/1078-0432.CCR-05-2734

Colorectal cancer is one of the most common cancers in the developed world and its incidence is increasing in the developing world. Although the progression of colorectal cancer from benign polyp to adenoma to carcinoma is well studied (1), the molecular events influencing the transition and establishment of metastasis are less well understood. The prognosis and treatment of colorectal cancer currently depend on the clinicopathologic stage of disease at the time of diagnosis and primary surgical treatment. Unfortunately, disease stage alone does not allow accurate prediction of outcome for individual patients. If patient outcomes could be predicted more accurately, treatments could be tailored to avoid undertreating patients destined to relapse or overtreating patients who would be cured by surgery alone.

Many attempts have been made to identify markers that predict clinical outcome in colorectal cancer. Until recently, most studies focused on single proteins or gene mutations with limited success in terms of prognostic information (2). Microarray technology enables the identification of multigene classifiers that correlate with cancer outcome and has already been applied to a variety of cancers, including colorectal cancer

(3–5), but methodologic problems and a lack of independent validation has cast doubt over some of the findings (6, 7). Furthermore, doubts about the ability of gene classifiers to predict outcome have arisen because of the poor concordance among predictive genes identified by different groups using different array platforms and methodologies (8).

To improve the current clinicopathologic staging and find robust gene classifiers that can predict disease relapse for colorectal cancer, two research groups independently generated array expression data sets from separate series of primary tumors with clinical follow-up of 5 or more years. Distinct gene classifiers were identified in each data set and reciprocally validated. Furthermore, these genes provided insight into the biological mechanisms underpinning colorectal cancer disease progression.

Materials and Methods

Tumor samples

Two cohorts of patients were included in this study. The New Zealand patients were part of a prospective cohort study that included all disease stages, whereas the German patients were selected from a tumor bank. Clinical information is shown in Table 1, whereas

Table 1. Clinical characteristics of New Zealand and German colorectal tumors

	Relapse free	Relapse
New Zealand data		
No. patients	102	47
Age (SD), y	68.5 (15.1)	69.8 (8.7)
Gender (%)		
Male	48 (47)	22 (47)
Female	54 (53)	25 (53)
Tumor localization (%)		
Right colon	40 (40.2)	18 (38.3)
Left colon	12 (11.8)	4 (8.5)
Sigmoid	32 (30.4)	17 (36.2)
Rectum	18 (17.7)	8 (17.0)
Tumor stage		
I	16	0
II	61	13
III	25	22
IV	0	12*
Median follow-up period/ median recurrence-free period (range), mo	72 (60-80)	15 (0-59)
German data		
No. patients	29	26
Age (SD), y	64.5 (12.5)	62.2 (10.5)
Gender (%)		
Male	17 (55)	16 (61.5)
Female	12 (45)	10 (38.5)
Tumor localization (%)		
Right colon	8 (27.5)	4 (15)
Left colon	7 (24)	5 (20)
Sigmoid	6 (20.5)	7 (27)
Rectum	8 (28)	10 (38)
Tumor stage		
I	5	6
II	24	20
Median follow-up period/ median recurrence-free period (range), mo	82 (64-99)	23.5 (2-59)

*Persisting disease.

Fig. 1 summarizes the experimental design. Primary colorectal tumor samples from 149 New Zealand patients were obtained from patients undergoing surgery at Dunedin Hospital and Auckland Hospital between 1995 to 2000. Tumor samples were snap frozen in liquid nitrogen. All surgical specimens were reviewed by a single pathologist (H.S.Y.) and estimated to contain an average of 85% tumor cells. Among the 149 colorectal cancer patients, 12 had metastatic disease at presentation, 35 developed recurrent disease, and 102 were disease-free after a minimum of 5-year follow-up. Primary colorectal tumor samples from German patients were obtained from patients undergoing surgery at the Surgical Department of the Technical University of Munich between 1995 to 2001. A group of 55 patients with stage I & II colorectal carcinoma was selected from banked tumors, which had been obtained fresh from surgery, snap frozen in liquid nitrogen. Twenty-nine patients were recurrence-free and 26 patients had experienced disease recurrence after a minimum of 5-year follow-up. Tumor content ranged between 70% and 100% with an average of 87%.

RNA extraction and target labeling

New Zealand tumors. Tumors were homogenized and RNA was extracted using TriReagent (Progenz, Auckland, New Zealand) and further purified using RNeasy mini column (Qiagen, Doncaster, Victoria, Australia). Ten micrograms of RNA were labeled with Cy5 dUTP using the indirect amino-allyl cDNA labeling protocol. A reference RNA from 12 different cell lines was labeled with Cy3 dUTP. The fluorescently labeled cDNA was purified using a QiaQuick PCR purification kit (Qiagen, Clifton Hill, Victoria, Australia).

German tumors. Tumors were homogenized and RNA was isolated using RNeasy Mini kit (Qiagen, Hilden, Germany). cRNA preparation was done as described previously (9), purified on RNeasy Columns (Qiagen, Hilden, Germany), and eluted in 55 μ L water. Fifteen micrograms of cRNA were fragmented for 35 min at 95°C, and double-strand cDNA was synthesized with a oligo(dT)-T7 primer (Eurogentec, Koeln, Germany) and transcribed using the Promega RiboMax T7-kit (Promega, Madison, WI) and Biotin-NTP labeling mix (Loxo, Dossenheim, Germany).

Microarray experiments

New Zealand tumors. Hybridization of the labeled target cDNA was done using MWG Human 30K Array oligonucleotides printed slides. Slides were blocked with 1% bovine serum albumin and the hybridization was done in prehybridization buffer at 42°C for at least 12 h followed by a high stringent wash. Slides were scanned with a GenePix Microarray scanner and data were analyzed using GenePix Pro 4.1 (Axon, Foster City, CA).

German tumors. cRNA was mixed with B2 control oligonucleotide, eukaryotic hybridization controls (Affymetrix, Santa Clara, CA), hering sperm (Promega), buffer, and bovine serum albumin to a final volume of 300 μ L and hybridized to one microarray chip (Affymetrix) for 16 h at 45°C. Washing steps and incubation with streptavidin (Roche, Mannheim, Germany), biotinylated goat anti-streptavidin antibody (Serva, Heidelberg, Germany), goat-IgG (Sigma, Taufkirchen, Germany), and streptavidin-phycoerythrin (Molecular Probes, Leiden, the Netherlands) were done in an Affymetrix Fluidics Station. The arrays were then scanned with an Affymetrix scanner and the digitized image data were processed using the Affymetrix Microarray Suite 5.0 software.

Data preprocessing

New Zealand data. Data preprocessing and normalization were done in the R computing environment (10). A log₂ transformation was applied to the foreground intensities from each channel of each array. Data from each spot were used on a per array basis to do print-tip loess normalization via the limma package (11) from the Bioconductor suite of analysis tools (12). Scale normalization (13) was then used to standardize the distribution of log intensity ratios across arrays. Postnormalization cluster analysis revealed the presence of a gene-specific print-run effect present in the data. ANOVA normalization was

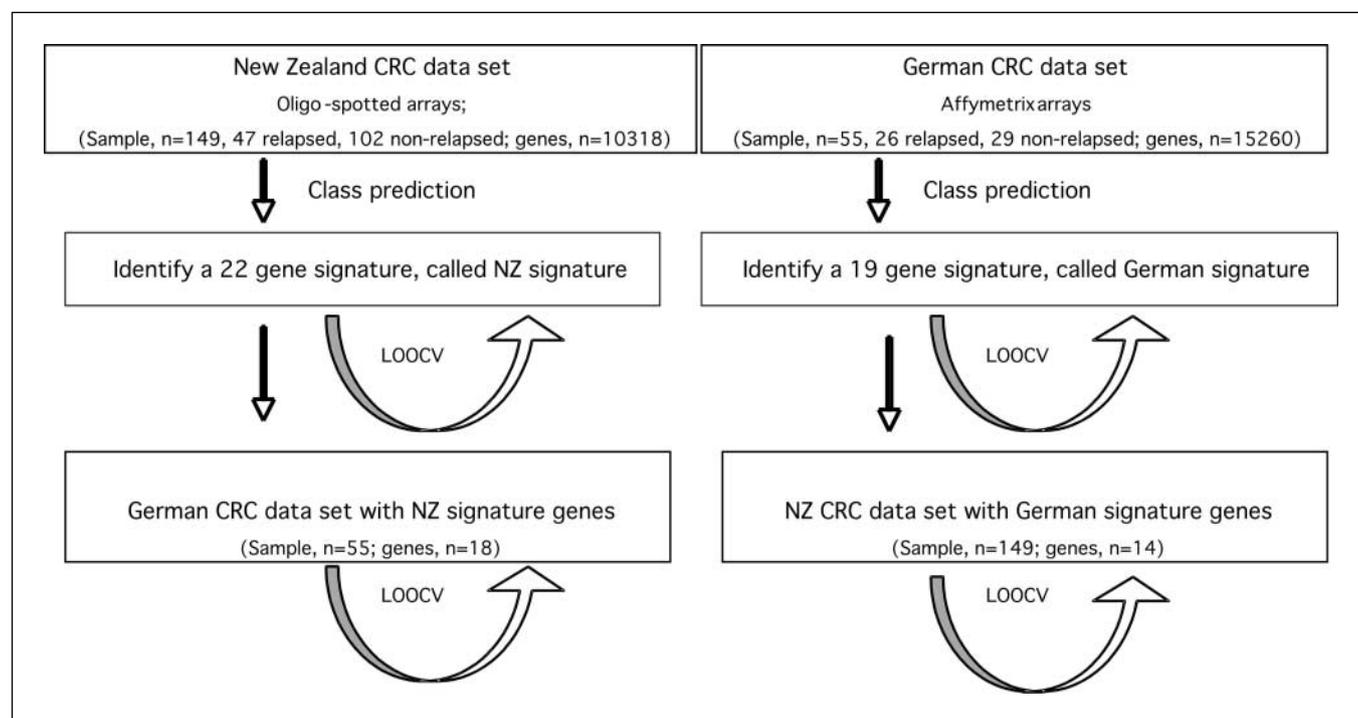


Fig. 1. Flow chart of methods for producing the gene classifier from 149 New Zealand and 55 German colorectal cancer (CRC) samples. New Zealand (NZ) RNA samples were hybridized to oligonucleotide spotted arrays, with a 22-gene classifier produced via LOOCV and then independently validated by LOOCV using the 55-sample German data set. German RNA samples were hybridized to Affymetrix arrays, with a 19-gene classifier produced via LOOCV and then independently validated by LOOCV using the NZ data set.

used to estimate and remove print run effects from the data for each gene. Replicate array data were available for 46 of the 149 samples. Cluster analysis of the entire data set indicated that the duplicate arrays clustered well with each other, suggesting internal consistency of the array platform. Genes with low intensity, large differences between replicates (mean log₂ difference between duplicates higher than 0.5), as well as unknown proteins were removed from the data set. After the initial normalization procedure, a subset of 10,318 probe sets was chosen for further analysis.

German data. All Affymetrix U133A GeneChips passed quality control to eliminate scans with abnormal characteristics. Background correction and normalization were done in the R computing environment (10). Background corrected and normalized expression measures from probe level data (CEL files) were obtained using the robust multiarray average function (14) implemented in the Bioconductor package affy.

After data preprocessing, there are 9,358 and 9,922 distinct genes in New Zealand and German data set, respectively, and 5,343 genes were in common between both data sets.

Clinical information

The following clinical variables were available for both the New Zealand and the German patients: sex, age, tumor localization, vascular invasion, lymphatic invasion, lymphocytic infiltration, and tumor-node-metastasis staging information. Fourteen New Zealand patients were missing data for at least one of vascular invasion (1), lymphatic invasion (13), or lymphocytic infiltration (9). Imputation was used to assign the "most likely" value to missing data points for a given patient, based on the value of other variables for that patient relative to the other patients in the data set. Although this approach does not incorporate information about the uncertainty introduced by missing data, this was not an important consideration, as the clinical information was simply used to establish a baseline predictive ability. In the German data set, vascular invasion was not used in the modeling process as it was only observed in one patient. There were no missing data for the German patients.

Construction of clinical variable classifiers

To establish performance baselines for each of the data sets, logistic regression classifiers based solely on clinical variables were constructed for the prediction of recurrence in both the German and the New Zealand data. This was accomplished using a leave-one-out cross-validation (LOOCV) approach, with variables included in the model in each iteration based on forward selection using the Akaike Information Criterion (AIC). Performance was assessed via area under the receiver operating characteristic (ROC) curve, overall classification rate, sensitivity, specificity, and log odds of recurrence, with 95% confidence intervals (95% CI) adjusted to account for multiple testing. Data analysis was done in the R computing environment (10).

Construction of clinical variable plus gene classifiers

To determine the usefulness of including information about gene expression in the predictive process, the classifiers based on clinical variables were augmented with gene expression data from the two data sets. A LOOCV approach was again taken to construct the predictors. In each iteration, genes were ranked by their significance analysis of microarrays test statistic (15), with 1.05- and 1.1-fold change filters applied to the New Zealand and German data, respectively, and the top 5, 10, or 20 genes then put forward for inclusion in the model. The number of genes put forward at each step was determined by an additional layer of 5-fold cross-validation to avoid selection bias. With the clinical variable model as a baseline, genes were then added via forward selection using the AIC. That is, genes were added to the clinical variable model in a stepwise fashion, only if they added sufficient value (as measured by the AIC), and the model was then used to predict the class of the left out observation. This procedure was repeated for each patient in the data set, with performance again assessed based on area under the ROC curve, overall classification rate, sensitivity, specificity, and log odds of recurrence. Data analysis was done in the R computing environment (10).

Multigene classifiers and cross-validation

Classifiers based solely on gene expression information were constructed using the BRB Array Tools package.¹⁰ Gene selection was done using a random variance model *t* test after an additional log 2 transformation to reduce data skewness. Although use of the random variance model procedure resulted in changes to the gene rankings and thus allowed the selection of different genes when compared with the results of the previous section, it was felt that the increased sophistication of this approach over the significance analysis of microarrays test statistic warranted its use in the BRB Array Tools package. In the German data, 318 genes were found to be differentially expressed when using a significance threshold of 0.001. As most of the differentially expressed genes exhibited relatively small changes in expression, a condition requiring the mean double logged fold differences between the two classes to be higher than 1.1 was added to the gene selection process for the German data. Gene-based classifiers were produced using LOOCV in each of the New Zealand and the German data set. To reduce the problem of overfitting, both gene selection and classifier construction were done during each LOOCV iteration. After LOOCV, the prediction rate was estimated by the fraction of samples correctly predicted, with sensitivity, specificity, and log odds of recurrence also calculated, although these performance estimates should be treated as optimistic, due to the possible introduction of bias through the gene selection and model construction process. ROC curves were unable to be produced for the BRB Array Tools classifiers because of the binary data output produced. To find a gene set that could make the best prediction for unknown samples, different *t* test thresholds using a random variance model were investigated in conjunction with six classification methods: compound covariate classifier, diagonal linear discriminant analysis, 3-nearest neighbors (3-NN), 1-nearest neighbors, nearest centroid, and support vector machines (SVM). A linear Kernel function was used with default variable settings when applying the SVM classifier in BRB Array Tools. The cost and the weight of misclassification were both set to one.

To establish the validity of the New Zealand and German gene classifiers, reciprocal validation was done, with the New Zealand gene classifier validated using the German data set and vice versa. To test the New Zealand genes, probes relating to the 22 genes from the New Zealand classifier were identified in the German data, and LOOCV was used to assess the performance of a classifier for the German samples, based only on these probes. Similarly, probes relating to the 19 genes in the German classifier were identified in the New Zealand data, and LOOCV was used to assess the performance of a classifier for the New Zealand samples. In both cases, a significance threshold of 0.999 was used to ensure that all genes were used in each LOOCV iteration. Differences between the platforms (in particular, log ratio data versus log intensity data) meant that direct application of a prediction rule across data sets was not feasible. The consequence of this is that only the gene sets, and not the prediction rules used, are generalizable to new samples. Although this removes the possibility of gene selection bias from the performance figures, it means that some model selection bias may exist, although this should be minimal, given that LOOCV has again been used in the reciprocal validation.

The significance of the LOOCV prediction results was calculated by permuting the class labels of the samples and finding the proportion of times that the permuted data resulted in a higher LOOCV prediction rate than that obtained for the unpermuted data. All permutation analysis involved 2,000 permutations, with small *P* values indicating that prediction results were unlikely to be due to chance. Additional permutations were done to assess the significance of the collection of genes present in each predictor. This was done by permuting the gene names and repeated selecting sets of random genes and assessing their predictive ability. In total, 2,000 permutations were done for each data set.

Survival analysis

Kaplan-Meier survival analysis for censored data were done using the survival package within the R computing environment (10). Survival was defined to be "disease-free survival" postsurgery. For each analysis, survival curves were constructed, and the log-rank test (16) was used to assess the presence of significant differences between the curves for the two groups in question. Censoring was taken into account for both the New Zealand and the German data sets. For the disease-free survival data, right censoring before 5 years could only occur for nonrecurrent patients as a result of either death, or the last clinical follow-up occurring at <5 years.

Identification of genes coexpressed with chemokine ligands

Genes in the German data, which had a Pearson correlation coefficient >0.75 with at least one of the four chemokines appearing in the predictor in the nonrelapse group, were selected for ontology analysis. Ontology was done using DAVID.¹¹

Results

To determine whether robust gene signatures can be identified that predict disease relapse for colorectal cancer, two research groups from New Zealand and Germany independently generated array expression data sets from separate series of primary tumors with clinical follow-up of 5 or more years. After normalization, each data set was used to generate classifiers based solely on clinical data, classifiers based on both clinical and gene expression data and classifiers based solely on gene expression data, with the goal of predicting recurrence. The gene-based classifiers were subsequently validated on the alternate series of patients.

Clinical data classifier construction. Logistic regression was used to construct a classifier in the New Zealand data set using only clinical variables, as described in Materials and Methods, for the prediction of recurrence. The area under the ROC curve for this classifier was 65%. The maximum LOOCV prediction rate achieved was 77%, with a sensitivity of 34%, a specificity of 97%, and an odds ratio (OR) for disease recurrence of 16.64 (95% CI, 4.7-95.0). Use of the AIC criterion to construct a classifier based on all 149 samples resulted in the selection of T, N₁, and M in the final model, thus reinforcing the relevance of the tumor-node-metastasis staging system.

Construction of a logistic regression classifier based on clinical variables achieved an area under the ROC curve of 65% after LOOCV in the German data. The maximum prediction rate achieved by this classifier was 65%, indicating relatively poor ability to predict using only clinical information. Sensitivity and specificity were 84% and 45%, respectively, and the OR for disease recurrence was 6.02 (95% CI, 1.4-38.3). Construction of a classifier based on all 55 samples yielded a model for which only lymphatic invasion and age were selected by AIC.

Clinical and gene data classifier construction. Adding gene expression information to the New Zealand clinical data classifier resulted in an increase in the area under the ROC curve to 73% (compared with the classifier based solely on clinical variables). This translated to a slight decrease in the

¹⁰ <http://linus.nci.nih.gov/BRB-ArrayTools.html>.

¹¹ <http://apps1.niaid.nih.gov/david/>.

maximum LOOCV prediction rate achieved to 73%, with an increase in sensitivity to 49% and a corresponding drop in specificity to 86%. The OR for disease recurrence was 5.04 (95% CI, 2.2-12.2). Use of the AIC criterion to construct a combined classifier based on both clinical and gene data led to the selection of three genes (*ZBTB6*, *CASP1*, and *CXCL14*) in addition to the clinical variables T, N₁, and M.

In the German data, the addition of gene expression information to the clinical variable classifier led to a marked increase in the area under the ROC curve to 80% and a maximal LOOCV prediction rate of 80%. Sensitivity and specificity were 74% and 86%, respectively, with an OR for disease recurrence of 15.83 (95% CI, 3.7-86.5). Construction of a combined clinical and gene classifier in the full set of 55 German patients led to the selection of three genes according to the AIC criterion: *CA2*, *CXCL9*, and *PBK* (in addition to the clinical variables lymphatic invasion and age).

Based on the results of the logistic regression analysis, it is clear that the addition of gene expression information has the ability to improve the accuracy of classifiers based solely on clinical data. Whereas the improvement for the New Zealand patients was marginal (in that sensitivity was increased at the cost of slightly lowering the overall prediction rate), for the German data, the improvement was marked. The difference between the two results is most likely due to the greater range of tumor stages, and hence the greater ability to stratify based on clinical variables, in the New Zealand data. The performance figures achieved using this approach to predictor construction can be used as a baseline for the more complex classifiers constructed solely from gene expression data using the BRB Array Tools software package.

Gene classifier construction. The New Zealand data set was generated using oligonucleotide printed microarrays. Six different classifiers were constructed, with a SVM using a gene selection threshold of 0.0008 yielding the highest LOOCV prediction rate and producing a 22-gene classifier (77% prediction rate, 53% sensitivity, and 88% specificity; *P* = 0.002; Tables 2 and 3). The New Zealand classifier had an OR

for disease recurrence in the New Zealand patients of 8.4 (95% CI, 3.5-21.4).

The German data set was generated using Affymetrix arrays resulting in a 19-gene (22 probes) and 3-NN classifier (selection threshold of 0.002, log 2 fold change >1.1, 84% classification rate, 85% sensitivity, and 83% specificity; *P* < 0.0001; Tables 2 and 3). The German classifier had an OR for recurrence in the German patients of 24.1 (95% CI, 5.3-144.7). Using Kaplan-Meier analysis, disease-free survival in New Zealand and German patients was significantly different between those predicted to recur or not recur (New Zealand classifier, *P* < 0.0001; Fig. 2A and German classifier, *P* < 0.0001; Fig. 2B).

External validation of the gene classifiers. To validate the New Zealand classifier, the 22 genes were used to construct a SVM classifier in the German data set by LOOCV. A prediction rate of 71% was achieved, which was highly significant (*P* = 0.002; Table 2). The OR for recurrence in German patients, using the New Zealand classifier, was 5.9 (95% CI, 1.6-24.5). We surmise that the reduction in prediction rate, from 77% in New Zealand patients to 71% in German patients, was due to four genes from the New Zealand classifier not being present in the German data because removal of these four genes from the New Zealand data set reduced the LOOCV prediction rate to 71% (Table 2). Disease-free survival for German patients predicted to relapse, according to the New Zealand classifier, was significantly lower than disease-free survival for patients predicted not to relapse (*P* = 0.0049; Fig. 2C).

We next validated the German classifier by using the 19 genes to construct a 3-NN classifier in the New Zealand data set by LOOCV. The prediction rate of 67% was again significant (*P* = 0.046; Table 2), confirming the validity of the German classifier. The OR for recurrence in New Zealand patients, using the German classifier, was 2.6 (95% CI, 1.2-6.0). We consider that the reduction of the prediction rate was due to five genes from German classifier not being present in the New Zealand data. The fact that removal of these five genes from the German data set resulted in a reduction of the LOOCV

Table 2. Assessment of gene classifiers

Data set	Prediction rate	Sensitivity	Specificity	P*	OR
22-gene New Zealand classifier tested on German data					
NZ data (training; SVM)	0.77 (0.66-0.86)	0.53 (0.33-0.73)	0.88 (0.77-0.95)	0.002	8.4 (3.5-21.4)
NZ data minus four genes not found in German data were removed from NZ data set (training; SVM)	0.72	0.38	0.87	0.011	
German data (test; SVM)	0.71 (0.51-0.86)	0.62 (0.32-0.86)	0.79 (0.52-0.95)	0.002	5.9 (1.6-24.5)
19-gene German classifier tested on NZ data					
German data (training; 3-NN)	0.84 (0.65-0.95)	0.85	0.83	<0.0001	24.1 (5.3-144.7)
German data minus five genes not found in NZ data were removed from German data set (training; 3-NN)	0.67	0.65	0.66	0.046	
NZ data (test; 3-NN)	0.67 (0.55-0.78)	0.42 (0.22-0.64)	0.78 (0.65-0.89)	0.045	2.6 (1.2-6.0)

NOTE: Value in parentheses are 95% confidence interval. Abbreviation: NZ, New Zealand. **P* values were calculated from 2,000 permutation of class labels.

Table 3. Genes in CRC classifiers

<i>P</i>	Gene symbol	Description	Accession no.
New Zealand 22-gene classifier			
2.30E-05	<i>FLJ11116</i>	Similar to RAB11-binding protein	XM_010330*
3.30E-05	<i>RBMS1</i>	RNA binding motif, single-stranded interacting protein 1, isoform D	NM_016836
4.60E-05	<i>SACM1L</i>	Ras-GTPase activating protein SH3 domain-binding protein 2	NM_014016
6.80E-05	<i>SOAT1</i>	Sterol <i>O</i> -acyltransferase acyl-coenzyme a: cholesterol acyltransferase 1	NM_003101
7.90E-05	<i>PBK</i>	PDZ-binding kinase	NM_018492
0.00014	<i>G3BP2</i>	Ras-GTPase activating protein SH3 domain-binding protein 2	NM_012297
0.000163	<i>ZBTB20</i>	Zinc finger and BTB domain containing 20	NM_015642
0.000214	<i>ZNF410</i>	Zinc finger protein 410	NM_021188*
0.00022	<i>LOC94794</i>	Similar to hspc042 protein	XM_037699*
0.000293	<i>PSMC1</i>	Proteasome (prosome, macropain) 26s subunit, ATPase, 1	NM_002802
0.000321	<i>COX10</i>	COX10 homologue, cytochrome c oxidase assembly protein, heme A: farnesyltransferase (yeast)	NM_001303
0.000334	<i>GTF3C5</i>	General transcription factor IIIC, polypeptide 5 (63 kDa)	NM_012087
0.000367	<i>HMMR</i>	Hyaluronan-mediated motility receptor (rhamm)	NM_012485
0.000405	<i>UBE2L3</i>	Ubiquitin-conjugating enzyme e2l 3	NM_003347
0.000417	<i>GNAS</i>	Gnas complex locus	NM_000516
0.000467	<i>PPP2R2A</i>	Protein phosphatase 2 (formerly 2a), regulatory subunit b (pr 52), α isoform	NM_002717
0.000493	<i>RNASE2</i>	RNase, RNase a family, 2 (liver, eosinophil-derived neurotoxin)	NM_002934
0.000532	<i>SCOC</i>	Short coiled-coil protein	NM_032547*
0.000578	<i>PSMD9</i>	Proteasome (prosome, macropain) 26s subunit, non-ATPase, 9	NM_002813
0.000593	<i>EIF3S7</i>	Eukaryotic translation initiation factor 3, subunit 7 (ζ , 66/67 kDa)	NM_003753
0.000649	<i>ATP2B4</i>	ATPase, Ca ²⁺ - transporting, plasma membrane 4	U42026
0.000737	<i>ABCC9</i>	ATP-binding cassette, subfamily c, member 9, isoform sur2a- δ -14	NM_020298
German 19-gene classifier			
3.00E-06	<i>CXCL10</i>	Chemokine (C-X-C motif) ligand 10	NM_001565
4.00E-06	<i>FAS</i>	FAS (TNF receptor superfamily, member 6)	NM_000043
8.00E-06	<i>CXCL9</i>	Chemokine (C-X-C motif) ligand 9	NM_002416 [†]
1.20E-05	<i>TLK1</i>	Tousled-like kinase 1	NM_012290
1.30E-05	<i>CXCL11</i>	Chemokine (C-X-C motif) ligand 11	AF030514
2.10E-05	<i>PBK</i>	PDZ-binding kinase	NM_018492
4.20E-05	<i>PSAT1</i>	Phosphoserine aminotransferase 1	NM_021154
7.60E-05	<i>MAD2L1</i>	MAD2 mitotic arrest deficient-like 1 (yeast)	NM_002358
9.80E-05	<i>CA2</i>	Carbonic anhydrase II	M36532
0.000128	<i>GZMB</i>	Granzyme B (granzyme 2, CTL-associated serine esterase 1)	J03189 [†]
0.000177	<i>SLC44A</i>	Solute carrier family 4, sodium bicarbonate cotransporter, member 4	NM_003759
0.000215	<i>DLG7</i>	Discs, large homologue 7 (<i>Drosophila</i>)	NM_014750 [†]
0.000376	<i>TNFRSF11A</i>	Tumor necrosis factor receptor superfamily, member 11a, activator of NF- κ B	NM_003839

*Genes not found in German data.

[†] Genes not found in New Zealand data.

prediction rate from 84% to 67% (Table 2) strongly supports this idea. Disease-free survival for New Zealand patients predicted to relapse, according to the German classifier, was significantly lower than disease-free survival for patients predicted not to relapse ($P = 0.0299$; Fig. 2D).

Permutation test results. To further assess the success of the external validation procedure, 2,000 random sets of 22 genes were selected from the German data and used in LOOCV to construct SVM classifiers. It was found that 90% of such random gene sets produced LOO classification rates <71% (the result achieved in the German data when using the New Zealand signature). Although this is not a particularly impressive figure, it is most likely due to the fact that there were a large number of differentially expressed genes in the German data and thus many possible classifiers. Regardless, the 22 genes found in the New Zealand data are certainly not doing poorly in the German data relative to the bulk of the randomly chosen gene sets. The same procedure was applied to the New Zealand data (2,000 random sets of 19 genes used in a 3-NN classifier), with 94% of such gene sets producing a LOO

classification rate <66% (the result achieved in the New Zealand data when using the German signature), reinforcing the applicability of the 19-gene German signature to the New Zealand data.

Comparison of gene classifiers with current staging system. Significant differences in disease-free survival between patients predicted to relapse or not relapse were also observed within the same clinicopathologic stage (Fig. 3). When patient predictions were stratified according to disease stage, the New Zealand classifier was able to identify patients who were more likely to recur in both stage II subgroup ($P = 0.0013$; Fig. 3A) and stage III subgroup ($P = 0.0295$; Fig. 3A). This was mirrored to a lesser extent when the German classifier was applied to the New Zealand data set, where the difference was only observed for stage III patients ($P = 0.0491$; Fig. 3B). Again, the decreased predictive accuracy of the German classifier was likely due to the absence of five genes from the New Zealand data that decreased the LOOCV prediction rate. The fact that the German predictor achieves reasonable success on the full New Zealand data set but is unable to classify within the stage I or II patients

and achieves only a borderline significance in stage III patients most likely reflects that (a) the relatively high specificity of the German predictor (78%) results in good performance on early stages patients and thus improved performance on the New Zealand data set as a whole and (b) the small sample sizes associated with stages II and III in the New Zealand data make significance difficult to achieve. Although this problem is overcome when the data are combined across all stages, it is still statistically possible that the German predictor would not work in one or more of the stages of interest, even if more samples were available.

Genes in classifiers are related to colorectal cancer disease progression. Several genes in the New Zealand classifier, inclu-

ding *G3BP2* (17), *RBMS1* (18), *HMMR* (19), *UBE2L3* (20), *GNAS* (21), *RNASE2* (22), and *ABCC9* (23), have all been reported to be involved in cancer progression, whereas *RBMS1* (24), *EIF3S7* (25), and *GTF3C5* (26) are related to transcription or translation. *PBK* is a protein kinase, which is involved in the process of mitosis (27) and was the only gene common to the New Zealand and German classifiers. Eleven of 19 genes in the German classifier are involved in the immune response, including four chemokine ligands [*CXCL9*, *CXCL10*, *CXCL11*, and *CXCL13*; (28); *PBK* (29); *INDO* (30); *GBP1* (31); *GZMB* (32); and *KITLG* (33)] and two receptors of the tumor necrosis factor family (*TNFRSF11A* and *FAS*; ref. 34). Eighty-six genes were found to be moderately correlated (Pearson correlation

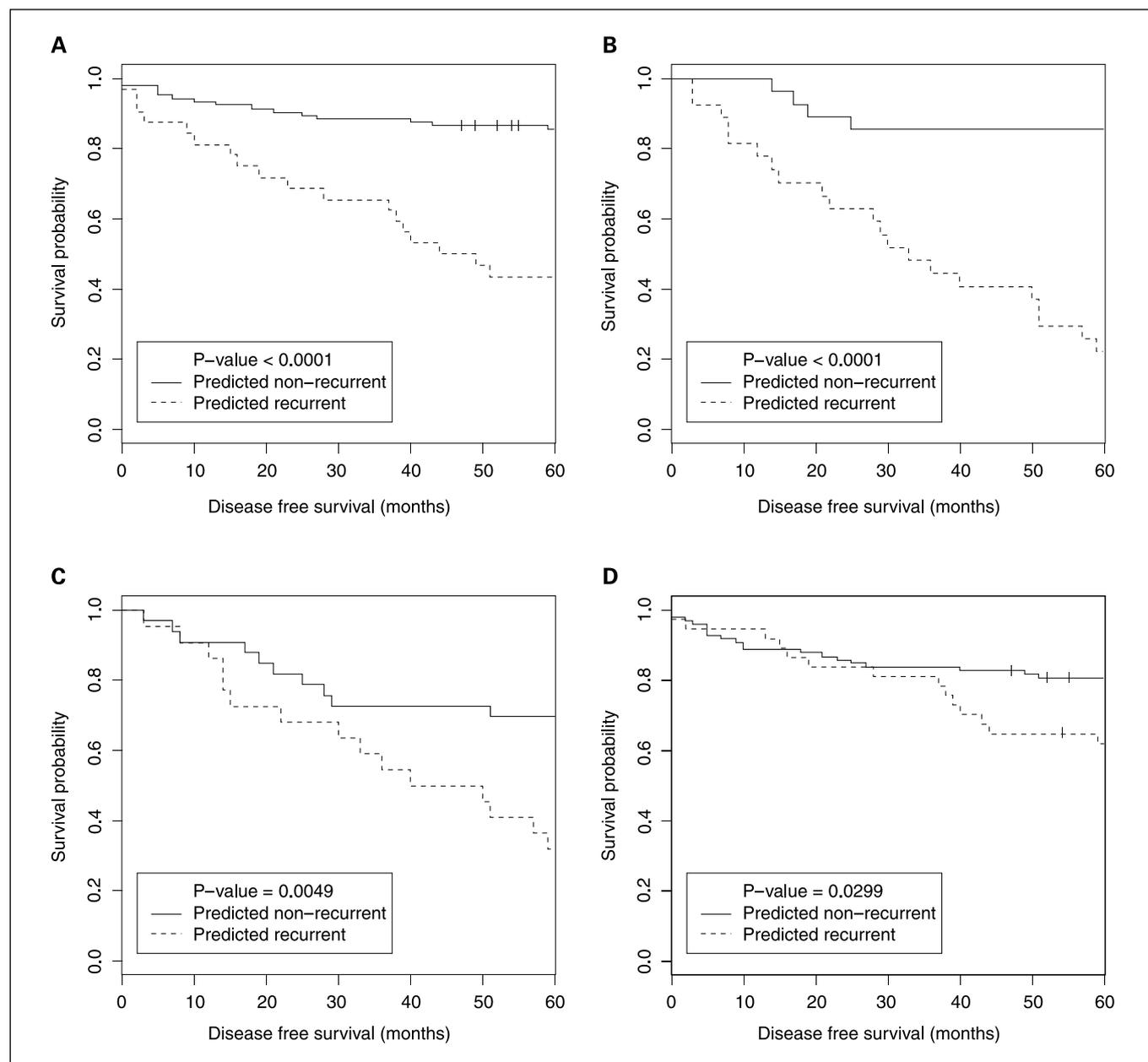


Fig. 2. Kaplan-Meier analysis of disease-free survival time with patients predicted as high versus low risk of tumor recurrence. *A*, using New Zealand 22-gene classifier on 149 tumors from New Zealand patients. *B*, using German 19-gene classifier on 55 tumors from German patients. *C*, New Zealand gene classifier validated on 55 tumors from German patients. *D*, German gene classifier validated on 149 tumors from New Zealand patients. *P* values were calculated using the log-rank test.

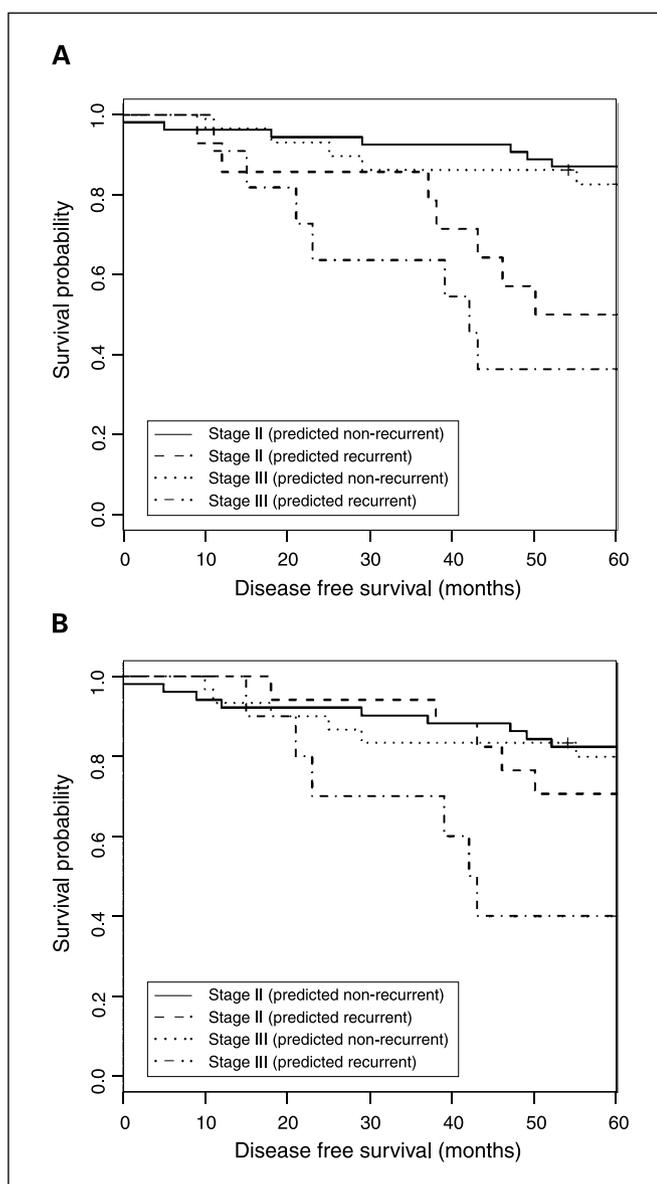


Fig. 3. Kaplan-Meier analysis of disease-free survival time with patients predicted as high versus low risk of tumor recurrence. *A*, using the 22-gene New Zealand classifier on New Zealand patients with stage II and stage III disease. *B*, using the 19-gene German classifier on New Zealand patients with stage II and stage III disease.

coefficient >0.75), with at least one of the four chemokine ligands in the German data (see Supplementary Table S1). Ontology analysis found that 39 of these 65 genes were in the category of immune response ($P < 10^{-26}$; see Supplementary Table S2). This result suggests a key role for the host immune response in determining colorectal cancer metastasis.

Discussion

We have shown that two different multigene classifiers are able to do as well (New Zealand data) or better (German data) than classifiers based solely on clinical information and have the potential to improve the current clinicopathologic-based classification system of colorectal cancer. From a clinical

perspective, the ability to predict outcome within the same clinicopathologic class is the key to individualizing treatment options (35). We have also shown that many genes in the classifiers are biologically related to cancer progression.

There is only one gene (*PBK*) in common in both classifiers and this implies that *PBK* may be important for cancer progression. The adenomatous polyposis coli tumor suppressor protein binds the PDZ2 domain of hDlg, the human homologous of *Drosophila* discs-large tumor suppressor protein. It is very interesting to find that *PBK* also binds to the same domain of hDlg (36). Based on the importance of adenomatous polyposis coli mutation in colorectal progression, further investigation into the role of *PBK* in colorectal cancer and its relationship with *adenomatous polyposis coli* gene is certainly warranted.

It is especially interesting to find that many genes in one of the classifiers are related to the immune response. The immune response has an important role in the progression of different cancers, and T-lymphocyte infiltration in colorectal cancer patients is an indicator of good prognosis (37–39). All of the 11 immune response genes were up-regulated in nonrecurrent patients, which may provide a biological explanation for the good outcome for these patients. To further confirm this result, we chose the four chemokine genes for further analysis. Chemokine ligands not only reflect the activity of the immune system and mediate leukocyte recruitment but also are involved in chemotaxis, cell adhesion and motility, and angiogenesis (40). To investigate the role of immune response, we identified 86 genes coexpressed with the chemokine ligands. Almost half of these genes had a Gene Ontology classification within the “immune response” category, suggesting that the primary function of these genes in the metastatic process is the modulation of the immune response. Furthermore, $CD4^+$ and $CD8^+$ T-cells antigens (*CD8A*, *CD3*, *PRF1*, *TRA@*, and *TRB@*) or function-related antigens (e.g., major histocompatibility molecules, *IFN- γ* -induced proteins, and *IL2RB*) were found in the above immune response coexpressed gene list. The activation of tumor-specific $CD4^+$ and $CD8^+$ T cells has been shown to result in tumor rejection in a mouse colorectal cancer model (38, 41). Collectively, these findings suggest the lymphocytes form part of a tumor-specific host response involved in minimizing the spread of cells from the primary tumor.

Previous investigations to identify colorectal cancer multigene classifiers associated with disease recurrence (3–5) have suffered from small sample sizes (3, 5), may be biased as a result of overfitting the training data (4, 5), lacked external validation (3, 5), or used inappropriate methodology in the external validation (4). To clarify the final point, the external validation from Eschrich et al. (4) did not seem to use the same criterion for class membership as was used for derivation of the gene classifier. Furthermore, the use of hierarchical clustering to determine classes based on observed differences in the classifier genes, rather than survival beyond a fixed time point, makes it difficult to assess the generalizability of this result.

We do not claim that the gene classifiers identified here necessarily represent the best set of genes for predicting disease recurrence. Given the high dimensionality of the data sets, it is very likely that other collections of genes could be found, which are also significantly related to patient survival (42). Low concordance between microarray gene expression

investigations has been a concern, suggesting nonreproducibility arising from different platforms, probe differences, and sample collection (8). In light of these points, it is particularly encouraging to find that the gene classifiers derived here using two different array platforms from two geographically separate series of colorectal cancer patients were both significantly related to tumor recurrence in a two-way reciprocal validation analysis. The essential point is that it does not matter whether different gene classifiers are identical between different studies but whether the classifiers can make an accurate and consistent prediction of clinical behavior (8).

In terms of clinical usefulness, the question of how to construct the 'best' prognostic test is a difficult one. To move forward with the results presented here, rigorous validation of these signatures is required. To accomplish this goal, two independent patient cohorts are required, whose clinical characteristics (particularly disease stage) closely match those of the German and New Zealand patients. Validation of the predictors (which comprise both the genes themselves and the underlying predictive algorithm) will require the use of technology identical to that used here, with validation samples classified blind on a sample by sample basis (as would be the case in routine clinical practice). Although we have shown some degree of generalizability here, it would be overly optimistic to assume that similar levels of performance could be readily achieved in a clinical setting, as issues still exist

involving both reproducibility and sample composition. If an array-based prognostic tool is to enter routine clinical use, it seems likely that it will be based on a single platform and will be application specific, containing only probes that are optimized for detecting the genes contained in the predictor. Issues of technical variability (e.g., variation between sites and operators) will need to be minimized, with centralization into a single national testing facility being one possibility. Given the time lag between tool development and clinical uptake, it is also likely that major technological advances will lead to further performance improvements before such techniques become routine. To obtain accurate estimates of generalizability, there is a need for far larger samples sizes, the makeup of which reflects the population of interest. Whereas the New Zealand data set goes some way toward accomplishing this (as the data were from a prognostic study), the German data only represents a selected subgroup of the early stages disease population. Future validation studies should focus solely on those patients who will benefit most from the application of this technology, namely stage II and III patients from whom accurate prognosis is currently the most difficult.

Acknowledgments

We thank Dr. Richard Simon and Amy Peng Lam for developing BRB Array Tools that were used for doing data analyses.

References

- Arnold CN, Goel A, Blum HE, CR Boland. Molecular pathogenesis of colorectal cancer. *Cancer* 2005;104:2035–47.
- Anwar S, Frayling IM, Scott NA, Carlson GL. Systematic review of genetic influences on the prognosis of colorectal cancer. *Br J Surg* 2004;91:1275–91.
- 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.
- Eschrich S, Yang I, Bloom G, et al. Molecular staging for survival prediction of colorectal cancer patients. *J Clin Oncol* 2005;23:3526–35.
- Barrier A, Lemoine A, Boelle PY, et al. Colon cancer prognosis prediction by gene expression profiling. *Oncogene* 2005;24:6155–64.
- Simon R. Roadmap for developing and validating therapeutically relevant genomic classifiers. *J Clin Oncol* 2005;23:7332–41.
- Michiels S, Koscielny S, Hill C. Prediction of cancer outcome with microarrays: a multiple random validation strategy. *Lancet* 2005;365:488–92.
- Marshall E. Getting the noise out of gene arrays. *Science* 2004;306:630–1.
- Birkenkamp-Demtroder K, Christensen LL, Olesen SH, et al. Gene expression in colorectal cancer. *Cancer Res* 2002;62:4352–63.
- Ihaka R, Gentleman R. R: a language for data analysis and graphics. *J Comput Graph Stat* 1996;5:299–314.
- Smyth GK. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. *Stat Appl Genet Mol Biol* 2004;3:Article 3.
- Gentleman RC, Carey VJ, Bates DM, et al. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol* 2004;5:R80.
- Smyth GK, Speed TP. Normalization of cDNA microarray data. *Methods* 2003;31:265–73.
- 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.
- Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci U S A* 2001;98:5116–21.
- Harrington DP, Fleming TR. A class of rank test procedures for censored survival data. *Biometrika* 1982;69:553–66.
- Barnes CJ, Li F, Mandal M, Yang Z, Sahin AA, Kumar R. Heregulin induces expression, ATPase activity, and nuclear localization of G3BP, a Ras signaling component, in human breast tumors. *Cancer Res* 2002;62:1251–5.
- Niki T, Izumi S, Saegusa Y, et al. MASP promotes ras/myc cooperative cell transforming activity by binding to c-Myc. *Genes Cells* 2000;5:127–41.
- Rein DT, Roehrig K, Schondorf T, et al. Expression of the hyaluronan receptor RHAMM in endometrial carcinomas suggests a role in tumour progression and metastasis. *J Cancer Res Clin Oncol* 2003;129:161–4.
- Fernandez P, Carretero J, Medina PP, et al. Distinctive gene expression of human lung adenocarcinomas carrying LKB1 mutations. *Oncogene* 2004;23:5084–91.
- Frey UH, Eisenhardt A, Lummen G, et al. The T393C polymorphism of the G α s gene (GNAS1) is a novel prognostic marker in bladder cancer. *Cancer Epidemiol Biomarkers Prev* 2005;14:871–7.
- Niini T, Vetteranta K, Hollmen J, et al. Expression of myeloid-specific genes in childhood acute lymphoblastic leukemia—a cDNA array study. *Leukemia* 2002;16:2213–21.
- Yasui K, Mihara S, Zhao C, et al. Alteration in copy numbers of genes as a mechanism for acquired drug resistance. *Cancer Res* 2004;64:1403–10.
- Nomura J, Matsumoto K, Iguchi-Ariga SM, Ariga H. Positive regulation of Fas gene expression by MASP and abrogation of Fas-mediated apoptosis induction in MASP-deficient mice. *Exp Cell Res* 2005;305:324–32.
- Mayeur GL, Fraser CS, Peiretti F, Block KL, Hershey JW. Characterization of eIF3k: a newly discovered subunit of mammalian translation initiation factor eIF3. *Eur J Biochem* 2003;270:4133–9.
- Hsieh YJ, Wang Z, Kovelman R, Roeder RG. Cloning and characterization of two evolutionarily conserved subunits (TFIIIC102 and TFIIIC63) of human TFIIIC and their involvement in functional interactions with TFIIIB and RNA polymerase III. *Mol Cell Biol* 1999;19:4944–52.
- Matsumoto S, Abe Y, Fujibuchi T, et al. Characterization of a MAPKK-like protein kinase TOPK. *Biochem Biophys Res Commun* 2004;325:997–1004.
- Dong VM, McDermott DH, Abdi R. Chemokines and diseases. *Eur J Dermatol* 2003;13:224–30.
- Abe Y, Matsumoto S, Kito K, Ueda N. Cloning and expression of a novel MAPKK-like protein kinase, lymphokine-activated killer T-cell-originated protein kinase, specifically expressed in the testis and activated lymphoid cells. *J Biol Chem* 2000;275:21525–31.
- Logan GJ, Smyth CM, Earl JW, et al. HeLa cells cocultured with peripheral blood lymphocytes acquire an immuno-inhibitory phenotype through up-regulation of indoleamine 2,3-dioxygenase activity. *Immunology* 2002;105:478–87.
- Lubeseder-Martellato C, Guenzi E, Jorg A, et al. Guanylate-binding protein-1 expression is selectively induced by inflammatory cytokines and is an activation marker of endothelial cells during inflammatory diseases. *Am J Pathol* 2002;161:1749–59.
- Phillips SM, Banerjee A, Feakins R, Li SR, Bustin SA, Dorudi S. Tumour-infiltrating lymphocytes in colorectal cancer with microsatellite instability are activated and cytotoxic. *Br J Surg* 2004;91:469–75.
- Oliveira SH, Taub DD, Nagel J, et al. Stem cell factor induces eosinophil activation and degranulation: mediator release and gene array analysis. *Blood* 2002;100:4291–7.
- Xanthoulea S, Pasparakis M, Kousteni S, et al. Tumor necrosis factor (TNF) receptor shedding controls thresholds of innate immune activation that balance

- opposing TNF functions in infectious and inflammatory diseases. *J Exp Med* 2004;200:367–76.
35. Brennan DJ, O'Brien SL, Fagan A, et al. Application of DNA microarray technology in determining breast cancer prognosis and therapeutic response. *Expert Opin Biol Ther* 2005;5:1069–83.
36. Gaudet S, Branton D, Lue RA. Characterization of PDZ-binding kinase, a mitotic kinase. *Proc Natl Acad Sci U S A* 2000;97:5167–72.
37. Canna K, McArdle PA, McMillan DC, et al. The relationship between tumour T-lymphocyte infiltration, the systemic inflammatory response, and survival in patients undergoing curative resection for colorectal cancer. *Br J Cancer* 2005;92:651–4.
38. Pages F, Berger A, Camus M, et al. Effector memory T cells, early metastasis, and survival in colorectal cancer. *N Engl J Med* 2005;353:2654–66.
39. Sandel MH, Dadabayev AR, Menon AG, et al. Prognostic value of tumor-infiltrating dendritic cells in colorectal cancer: role of maturation status and intratumoral localization. *Clin Cancer Res* 2005;11:2576–82.
40. Rossi D, Zlotnik A. The biology of chemokines and their receptors. *Annu Rev Immunol* 2000;18:217–42.
41. Miyazaki M, Nakatsura T, Yokomine K, et al. DNA vaccination of HSP105 leads to tumor rejection of colorectal cancer and melanoma in mice through activation of both CD4 T cells and CD8 T cells. *Cancer Sci* 2005;96:695–705.
42. Ein-Dor L, Kela I, Getz G, Givol D, Domany E. Outcome signature genes in breast cancer: is there a unique set? *Bioinformatics* 2005;21:171–8.

Clinical Cancer Research

Multiple Gene Expression Classifiers from Different Array Platforms Predict Poor Prognosis of Colorectal Cancer

Yu-Hsin Lin, Jan Friederichs, Michael A. Black, et al.

Clin Cancer Res 2007;13:498-507.

Updated version	Access the most recent version of this article at: http://clincancerres.aacrjournals.org/content/13/2/498
Supplementary Material	Access the most recent supplemental material at: http://clincancerres.aacrjournals.org/content/suppl/2007/01/16/13.2.498.DC1

Cited articles	This article cites 42 articles, 15 of which you can access for free at: http://clincancerres.aacrjournals.org/content/13/2/498.full#ref-list-1
Citing articles	This article has been cited by 17 HighWire-hosted articles. Access the articles at: http://clincancerres.aacrjournals.org/content/13/2/498.full#related-urls

E-mail alerts	Sign up to receive free email-alerts related to this article or journal.
Reprints and Subscriptions	To order reprints of this article or to subscribe to the journal, contact the AACR Publications Department at pubs@aacr.org .
Permissions	To request permission to re-use all or part of this article, use this link http://clincancerres.aacrjournals.org/content/13/2/498 . Click on "Request Permissions" which will take you to the Copyright Clearance Center's (CCC) Rightslink site.