JOURNAL OF CLINICAL ONCOLOGY

O R I G I N A L   R E P O R T

Development and Independent Validation of a Prognostic Assay for Stage II Colon Cancer Using Formalin-Fixed Paraffin-Embedded Tissue

Richard D. Kennedy, Max Bylesjo, Peter Kerr, Timothy Davison, Julie M. Black, Elaine W. Kay, Robert J. Holt, Vitali Proutski, Miika Ahdesmaki, Vadim Farztdinov, Nicolas Goffard, Peter Hey, Fionnuala McDyer, Karl Mulligan, Julie Mussen, Eamonn O'Brien, Gavin Oliver, Steven M. Walker, Jude M. Mulligan, Claire Wilson, Andreas Winter, Diarmuid O'Donoghue, Hugh Mulcahy, Jacintha O'Sullivan, Kieran Sheahan, John Hyland, Rajiv Dhir, Oliver F. Bathe, Ola Winqvist, Upender Manne, Chandrakumar Shanmugam, Sridhar Ramaswamy, Eduardo J. Leon, William I. Smith Jr, Ultan McDermott, Richard H. Wilson, Daniel Longley, John Marshall, Robert Cummins, Daniel J. Sargent, Patrick G. Johnston, and D. Paul Harkin

See accompanying editorial doi: 10.1200/JCO.2011.37.8646 and article doi: 10.1200/JCO.2010.32.8732

Corresponding author: Paul Harkin, PhD, Centre for Cancer Research & Cell Biology, Queen's University Belfast, 97 Lisburn Rd, Belfast, BT9 7BL, Northern Ireland, United Kingdom; e-mail: paul.harkin@almacgroup.com.

A B S T R A C T

Purpose
Current prognostic factors are poor at identifying patients at risk of disease recurrence after surgery for stage II colon cancer. Here we describe a DNA microarray–based prognostic assay using clinically relevant formalin-fixed paraffin-embedded (FFPE) samples.

Patients and Methods
A gene signature was developed from a balanced set of 73 patients with recurrent disease (high risk) and 142 patients with no recurrence (low risk) within 5 years of surgery.

Results
The 634–probe set signature identified high-risk patients with a hazard ratio (HR) of 2.62 ($P < .001$) during cross validation of the training set. In an independent validation set of 144 samples, the signature identified high-risk patients with an HR of 2.53 ($P < .001$) for recurrence and an HR of 2.21 ($P = .0084$) for cancer-related death. Additionally, the signature was shown to perform independently from known prognostic factors ($P < .001$).

Conclusion
This gene signature represents a novel prognostic biomarker for patients with stage II colon cancer that can be applied to FFPE tumor samples.

J Clin Oncol 29. © 2011 by American Society of Clinical Oncology

INTRODUCTION

Approximately one quarter of patients with colon cancer present with stage II disease, indicating invasion through the bowel wall but not metastatic spread to local lymph nodes.[1] The 5-year survival for these patients after surgery is approximately 75% to 80%. Nevertheless, approximately 20% of patients will develop recurrent disease within their lifetime.[2] On the basis of several large clinical studies, current guidelines advise against the standard use of adjuvant chemotherapy for stage II colon cancer. Some studies demonstrate a small benefit or trend toward better outcome after chemotherapy in this group, indicating that a proportion of higher risk patients may benefit. The Quick and Simple and Reliable (QUASAR) study, for example, reported a 3.5% ab-

solute improvement in overall survival for adjuvant chemotherapy in stage II disease.[3] Patients with tumors that display increased-risk features such as T4 or high-grade disease are often offered chemotherapy,[1,4] but these prognostic factors are relatively weak.[4] A number of prognostic molecular markers are also under investigation, including loss of heterozygosity at 18q and presence of microsatellite instability.[5] These, however, still require validation and are not part of standard clinical practice. DNA microarray technology can measure several thousand mRNA transcripts at once and may be able to capture the complex biology that underlies colon cancer recurrence better than single gene markers.[6] Several studies have been published, applying DNA microarray technology to the identification of prognostic gene signatures in stage II colon

cancer.[7-11] Although these studies illustrate the ability of microarray profiling to identify potential prognostic assays, no assay has been rigorously tested or validated for clinical use because of the lack of suitable fresh frozen tumor material. Most available samples are formalin-fixed and paraffin-embedded (FFPE) because this has been the standard storage method for several decades. However, to date, DNA microarray technology has not been able to reliably analyze expression profiles from this material for the purpose of assay development and delivery. In this study, we demonstrate the development and validation of the first prognostic microarray-based gene signature for patients with stage II colon cancer using clinically relevant FFPE tumor samples.

PATIENTS AND METHODS

Sample Selection

Samples were collected retrospectively with the following eligibility criteria: stage II colon adenocarcinoma only, with no evidence of residual disease; patient age 45 years or older at time of primary surgery; six or more regional lymph nodes assessed; a minimum of 50% tumor cells present in the tissue section; no family history of colon cancer; no preoperative or postoperative cancer therapy within 1 year of surgery (although therapy given after recurrence was acceptable); and minimum patient follow-up of 5 years for low-risk patients. Low-risk patients were defined as those with no cancer recurrence within 5 years of primary surgery. High-risk patients were defined as those with metastatic cancer recurrence within 5 years of primary surgery. Patients with local disease recurrence were excluded because this recurrence may have been a result of local residual disease after surgery rather than metastatic tumor. Samples were collected from 12 centers (Data Supplement). All samples underwent independent histopathologic review by an expert pathologist (E.W.K.). The data set was compared with the Surveillance, Epidemiology, and End Results database to ensure it represented a general population with stage II colon cancer (Data Supplement).

Gene Expression Profiling From FFPE Tissue

Total RNA was extracted from FFPE tumor samples using the Roche High Pure RNA Paraffin Kit (Roche, Basel, Switzerland). Total RNA was amplified using the NuGEN WT-Ovation FFPE System (NuGEN, San Carlos, CA). The amplified product was hybridized to the Almac Colorectal Cancer DSA (Almac, Craigavon, United Kingdom) on the Affymetrix 7G scanner (Affymetrix, Santa Clara, CA). A sample profile scheduling strategy was used that involved the stratification of samples into batches that were randomized against targeted clinical and sample property factors in addition to operators, reagent, and material lots. Quality control criteria were applied, and biologic and technical factors were balanced between low- and high-risk samples. Additional information regarding sample processing, balancing, and quality control is available in the Data Supplement.

Classifier Model Identification

Model development started with 5,014 probe sets identified as stable and/or having comparable longitudinal stability under FFPE fixation (Data Supplement) to avoid the issue of differential degradation of probe sets. Signature generation was subsequently performed using the partial least squares classification method[12] with selection of important features based on recursive feature elimination (RFE)[13] during 10 repeats of five-fold cross validation. All aspects of the model development were appropriately nested within the cross validation, including an initial filtering to remove 50% of the probe sets with the lowest variance and intensity, reference-based robust multichip averaging (RefRMA) normalization and summarization,[14] and RFE discarding the least important 10% of probe sets at each iteration. The total number of features to include in the final model was determined by the feature length with the highest average area under the receiver operating characteristics curve (AUC) under cross validation. The threshold for dichotomization of the predictions from each model was selected based on the maximum of the sum of sensitivity and specificity (minimum of the Youden J statistic[15]) from cross-validated training data. In the case of multiple thresholds with largely identical performance, the hazard ratio (HR) from Cox proportional hazards regression[16] was used as a tiebreaker to favor higher HR values.

The precision of the predictions was evaluated by predicting technical replicates of a colorectal cancer cell line (HCT116) embedded in FFPE, which was profiled concurrently with the clinical samples. The repeated technical measurements of this sample were not included in model development but were predicted by all 50 cross-validation training subsets as an independent test set with a view to select models with high repeatability and reproducibility (Data Supplement). Additionally, a permutation test was performed[17] where the true class labels were reshuffled randomly 100 times followed by complete model development. This was done to assess what classification performance one can expect by chance from a data set with these characteristics and to reveal any bias in the signature generation procedure (Data Supplement).

The independence of the final model in the context of known clinical factors was evaluated using univariate and multivariate Cox proportional hazards regression. The input used was the predicted dichotomized class labels together with tumor stage, patient tumor grade, tumor location, patient age, patient sex, mucinous/nonmucinous subtype, and number of lymph nodes retrieved. Microsatellite instability was not included as a factor because this information was not available for the majority of the samples.

Gene Ontology annotation[18] and enrichment of Gene Ontology biologic processes and molecular functions were performed using an internally developed tool based on the genes in the final signature. The hypergeometric distribution with false discovery rate multiple testing correction[19] was used to determine functional classes of genes significantly enriched.[20] The pathway analysis was generated through the use of Ingenuity Pathway Analysis (Ingenuity Systems, Redwood City, CA).

Signature Validation

The procedure of applying the signature to new samples is described in the Data Supplement, and the predicted signature output for the training and validation sets is also available in the Data Supplement.

RESULTS

Development of a Stage II Colon Cancer Prognostic Signature From FFPE Tissues

We used disease-free survival at 5 years as the primary end point for this study. After balancing for clinical factors and applying quality control criteria to the initial data set (Data Supplement), a training set of 215 patients (142 low-risk and 73 high-risk patients) was identified (Data Supplement). Fifty percent variance-intensity filtering, RefRMA normalization, RFE feature selection, and partial least squares classification were performed under 10 repeats of five-fold cross validation for estimation of the classification performance. Cross validation indicated a 634-transcript signature to be optimal for prognostic classification (Data Supplement). A receiver operating characteristic curve with an AUC of 0.68 ($P < .001$) was generated, indicating a significant association between signature score and prognosis (Fig 1A). The observed AUC was significantly higher than random in the permutation analysis (Data Supplement) and displayed a low variance in the evaluation of the precision from technical replicates (Data Supplement). A threshold of 0.465 for dichotomization of the signature prediction scores was established from the Youden J statistics (Data Supplement), yielding an HR of 2.62 ($P < .001$; Fig 1B). Table 1 contains a summary of the classification performance over the signature generation during cross validation.
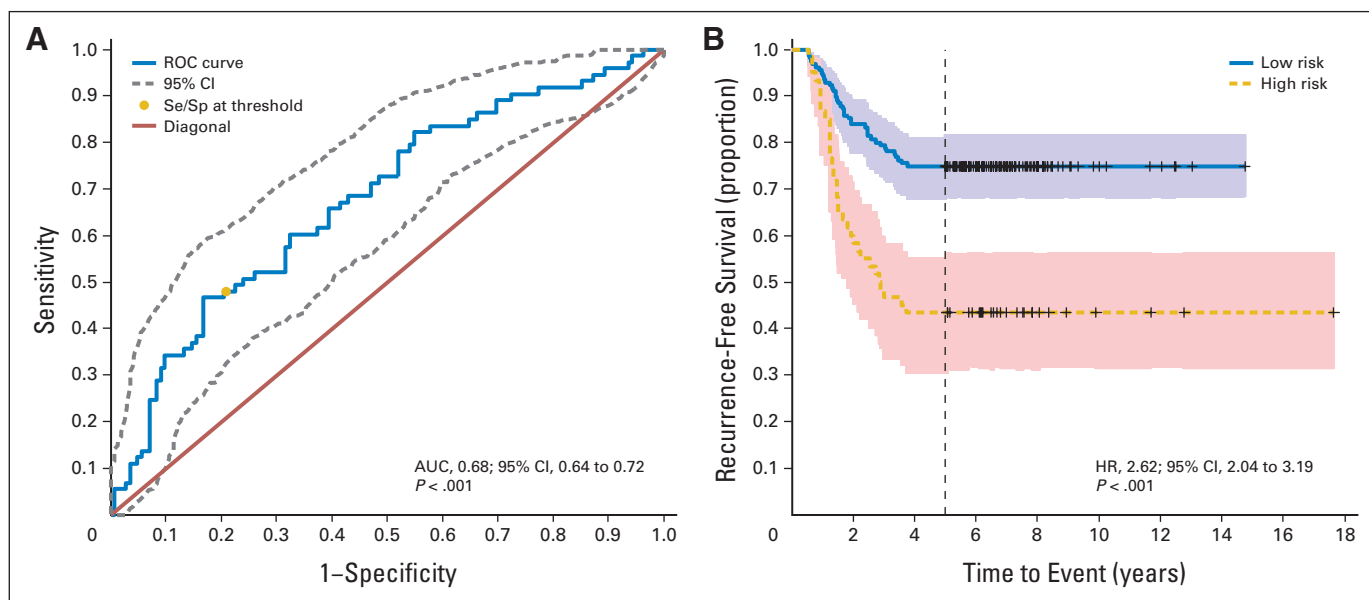
Fig 1. (A) Receiver operating characteristic (ROC) curve of 10 cross-validation (CV) repeats from the 215-sample training set. The displayed area under the curve (AUC) curve represents the average over the 10 CV repeats with 95% CIs estimated using bootstrapping. (B) Kaplan-Meier plot of the 10 cross-validation repeats during model generation from the 215-sample training set. Crosses mark censored event. The dashed vertical line denotes the low-/high-risk cutoff point of 5 years. The displayed hazard ratio (HR) represents the average over the 10 CV repeats with 95% CIs estimated using bootstrapping. Se, sensitivity; Sp, specificity.

Independent Validation of the Stage II Colon Cancer Prognostic Signature

The prognostic signature was applied to an independent validation set of 144 patients enriched for recurrence (85 low-risk and 59 high-risk patients) using the threshold score identified in the training set. The sample analysis was run separately and at a later time to the training set. The signature predicted disease recurrence with an HR of 2.53 ($P < .001$) in the high-risk group (Fig 2 and Table 1). The signature also predicted cancer-related death with an HR of 2.21 ($P = .0084$) in the high-risk group (Fig 3).

Assessment of Signature Independence From Known Prognostic Factors

For a prognostic assay to be useful, it must perform independently from known prognostic factors used in the clinic. Therefore, we assessed the independence of the assay in both a univariate and multivariate analysis (Table 2). The prediction of prognosis was significant in both the univariate ($P < .001$) and multivariate ($P < .001$) analysis, demonstrating that the signature provided prognostic information in addition to conventional risk factors. Furthermore, we assessed the

independence of the signature with the addition of lymphovascular invasion in the samples where this had been recorded (100 of 144 samples in the validation set). The signature performed independently in the univariate ($P < .001$) and multivariate analysis ($P < .001$; Data Supplement).

Functional Analysis of the Genes in the Prognostic Signature

Next we asked if the assay detected biologic processes known to be relevant to colon cancer recurrence. The 634 probe sets were analyzed using Ingenuity Pathway Analysis, and a list of statistically significant pathways were identified (Table 3), the most significant of which was *IGF-1* signaling.

DISCUSSION

In this study, we developed a DNA microarray–based assay that identifies patients at higher risk of recurrence after surgery for stage II colon cancer. Specifically, the signature identified a high-risk cohort

---

Table 1. Classification Performance of the Training and Independent Validation Sets

Data Set	AUC	Se	Sp	NPV	PPV	HR
Training	0.682	0.478	0.791	0.858	0.365	2.618
95% CI	0.643 to 0.720	0.407 to 0.549	0.737 to 0.845	0.845 to 0.872	0.317 to 0.413	2.041 to 3.195
Validation	0.684	0.559	0.718	0.867	0.331	2.526
95% CI	0.594 to 0.761	0.423 to 0.673	0.617 to 0.811	0.828 to 0.900	0.250 to 0.434	1.536 to 4.154

NOTE. The 95% CIs are ± 2 standard deviations from cross validation (training set) or bootstrapping with 1,000 repeats (validation set); 80% and 20% priors have been used when calculating the NPVs and PPVs, respectively. The threshold $t = 0.465$ was used for dichotomization of the signature score.
Abbreviations: AUC, area under the receiver operating characteristics curve; HR, hazard ratio; NPV, negative predictive value (negative is low risk); PPV, positive predictive value (positive is high risk); Se, sensitivity; Sp, specificity.
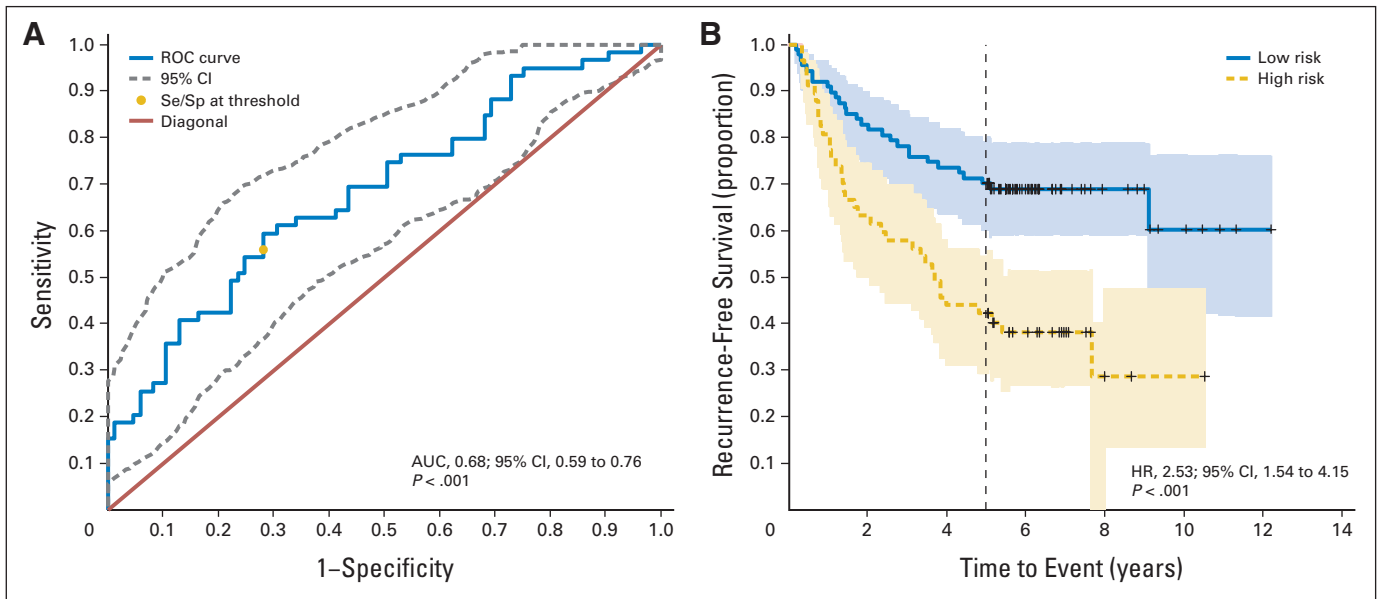

Fig 2. (A) Receiver operating characteristic (ROC) curve of the 144-sample independent validation set. The 95% CIs of the area under the curve (AUC) have been estimated using bootstrapping. (B) Kaplan-Meier plot of time to recurrence in the independent validation set. Crosses mark censored event. The dashed vertical line denotes the low-/high-risk cutoff point of 5 years. The 95% CIs have been estimated using bootstrapping. HR, hazard ratio; Se, sensitivity; Sp, specificity.

with an HR of recurrence of 2.53 and an HR of cancer-related death of 2.21 in an independent validation set. Validation of a prognostic assay using a completely separate set is necessary to avoid overestimations of the performance of the signature from the training set.[21] The HR of 2.53 for recurrence compares favorably with histologic factors currently used to make decisions in the clinic, which typically have an HR of approximately 1.5 or less.[22] Moreover, the signature does not require individual interpretation and may offer a more standardized approach than conventional histopathologic factors. Importantly, the
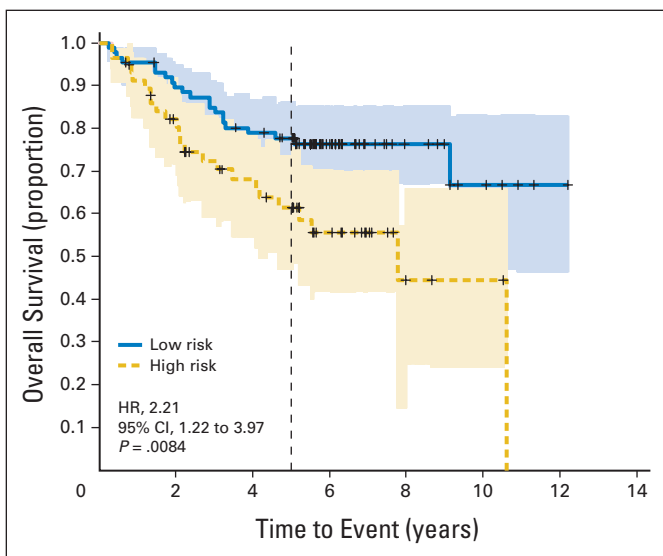

Fig 3. Kaplan-Meier plot of time to cancer-related death in the independent validation set. Forty-five patients were recorded with cancer-related death in the validation set. Crosses mark censored event. The dashed vertical line denotes the low-/high-risk cutoff point of 5 years. The 95% CIs have been estimated using bootstrapping. HR, hazard ratio.

assay is performed on FFPE tissue and, therefore, is easily applied in current medical practice.

Although several DNA microarray–based prognostic tests in several cancer types have been published, only one has been introduced into clinical practice,[12] and to date, none is used in colon cancer. This may be a result of two major factors. First, many of the signatures have been developed from fresh or frozen tissue. Second, inappropriate study methodology has resulted in a failure to validate the test in an independent data set.

Regarding the use of frozen tissue samples, although this tissue type provides excellent microarray data, a test generated from this tissue is unlikely to perform adequately in FFPE tissue. This can create difficulty in collecting enough samples to develop and independently validate a prognostic test. In addition, implementation of fresh tissue–based assays requires a change in clinical practice, because samples need to be collected at the time of surgery. However, several fresh tissue–based prognostic studies for stage II colon cancer have been published,[7-10] and one has undergone validation in an independent frozen tissue set of 144 patients with stage II colon cancer with an HR of 3.34 (95% CI, 1.24 to 9.00).[23]

FFPE is the standard for tumor archiving, and numerous tumor banks already exist for assay development. Importantly, no change in sample collection and processing is required for the development and clinical implementation of FFPE-based assays.

O'Connell et al[24] recently published an 18-gene quantitative polymerase chain reaction test composed of seven genes prognostic in stage II colon cancer, six genes predictive of benefit from adjuvant chemotherapy, and five shared reference genes. Kerr et al[25] presented retrospective validation of this test in the QUASAR study at the 2009 Annual Meeting of the American Society of Clinical Oncology, reporting that it predicted high and low risk of recurrence for stage II colon cancer with an HR of 1.48 at 3 years but did not predict benefit from adjuvant therapy.

Table 2. Comparison of 634-Transcript Signature to Standard Pathologic Parameters in the Independent Validation Set

Parameter	Univariate			Multivariate		
	HR	95% CI	P	HR	95% CI	P
Tumor stage (T4 v T3)	1.230	0.667 to 2.269	.5067	1.617	0.840 to 3.110	.1501
Patient age	1.039	1.010 to 1.069	.0086	1.046	1.014 to 1.078	.0041
Tumor grade						
2	0.815	0.456 to 1.456	.4895	1.274	0.480 to 3.383	.6265
3	1.326	0.654 to 2.689	.4340	2.161	0.636 to 7.339	.2169
Tumor location (proximal v distal)	1.766	1.075 to 2.901	.0248	2.158	1.224 to 3.804	.0078
Sex	1.165	0.713 to 1.901	.5426	0.971	0.549 to 1.720	.9204
Mucinous subtype	0.825	0.418 to 1.627	.5787	0.896	0.433 to 1.856	.7682
No. of nodes retrieved	1.007	0.983 to 1.032	.5678	1.014	0.988 to 1.041	.2824
Prognostic signature	2.526	1.536 to 4.154	< .001	2.551	1.471 to 4.423	< .001

NOTE. Both the univariate and multivariate analyses have been performed using Cox proportional hazards regression with P values coming from a log-likelihood test. For tumor grade, grade 1 has been used as the reference point for calculating the HR. Patient age and number of nodes retrieved are analyzed as continuous factors. The interpretation of the HR of patient age is the increased risk for a change in 1 year of age, and correspondingly, the interpretation of the HR of number of nodes retrieved is the increased risk for an increase of one retrieved node.
Abbreviation: HR, hazard ratio.

Our assay was developed to work with FFPE tissue but using a DNA microarray platform, thereby vastly increasing the number of detectable mRNA transcripts and biologic processes relative to quantitative polymerase chain reaction technology. As a result of using FFPE material with a microarray platform, several methodologic issues needed to be considered. Formalin fixation results in the degradation of mRNA transcripts through the cross linking of RNA to protein. Most of this degradation occurs immediately, but some transcripts continue to degrade with time. The DNA microarray platform used for the study has probe sets designed to the 3′ end of mRNA transcripts to enhance the ability to detect degraded transcripts.[26] In addition, we analyzed a separate set of colon cancer samples over time to ensure we did not incorporate probe sets that detected unstable or differentially stable mRNA transcripts as part of the signature.

The predictive value of the signature is above and beyond known prognostic clinical covariates. This performance can largely be attributed to the initial balancing of prognosis against biologic and technical factors that was performed as part of establishing a suitable training set. Biologic factors considered include known prognostic factors such as pT stage and grade, as well as other nonprognostic factors that may have affected gene expression including tumor location, patient age, and sex. Technical factors such as FFPE block age and the contributing center were also balanced between high- and low-risk samples in the training set. In addition, randomization of operators and reagent kits was performed to avoid confounding between technical factors and known clinical factors. This minimized the risk that the assay was dependent on the operator or relied on the use of samples from specific centers or the use of specific batches of reagents. Because the assay was developed to be independent from known prognostic factors, we believe that it may be possible to develop a multiparametric test that incorporates several factors to give an even more accurate prognostic indicator.

Functional analysis of the gene signature revealed that *IGF-1* signaling, *TGF-β* signaling, and *HMGB1* signaling were among the

Table 3. Most Significant Molecular Pathways Measured by the Assay

Ingenuity Canonical Pathways	P	Genes
IGF-1 signaling	.0011	PTK2, JUN, CTGF, IGF1, FOXO3, PRKAR2A, RASA1, IGFBP2
PXR/RXR activation	.0055	PPARA, SCD, FOXO3, PRKAR2A, INSR, CYP3A5
TGF-β signaling	.0081	JUN, BMP2, SMURF2, VDR, SERPINE1, SMURF1
Estrogen receptor signaling	.0089	PRKDC, POLR2A, MED1, TAF15, NCOR1, GTF2F1, MED12, SMARCA4
DNA methylation and transcriptional repression signaling	.0100	CHD3, MTA2, RBBP4
Aryl hydrocarbon receptor signaling	.0120	CTSD, JUN, NFIX, CCND2, MED1, GSTO2, CYP1B1, SMARCA4
Glucocorticoid receptor signaling	.0138	ICAM1, POLR2A, JUN, HMGB1 (includes EG:3146), MED1, TAF15, NCOR1, NR3C2, GTF2F1, STAT1, SERPINE1, SMARCA4
HMGB1 signaling	.0209	ICAM1, JUN, RHOQ, HMGB1 (includes EG:3146), RHOJ, SERPINE1
Molecular mechanisms of cancer	.0219	PRKDC, BMP2, PRKAR2A, RHOJ, HIF1A, RALBP1, PTK2, JUN, CCND2, RHOQ, RABIF, FANCD2, ARHGEF2, RASA1
Chemokine signaling	.0224	PTK2, CALM3, JUN, CAMK1D, LIMK2
Purine metabolism	.0257	PKM2, TJP2, HSPD1, RALBP1, DLG3, SMARCA4, INO80, POLR2A, PICK1, POLR3A, BAT1, ATP5G1, NT5E
Sonic hedgehog signaling	.0275	HKR1, ARRB2, PRKAR2A
Semaphorin signaling in neurons	.0302	PTK2, RHOQ, RHOJ, LIMK2
RhoA signaling	.0355	PTK2, IGF1, ARHGAP4, ARHGAP12, LIMK2, ARHGAP8

most significant pathways identified. All of these have been previously reported to confer a poor prognosis in colon cancer through promoting tumor growth, invasion, and metastasis and preventing apoptosis.[27-29]

In conclusion, to our knowledge, we are the first investigators to develop and validate a robust prognostic DNA microarray signature for stage II colon cancer from FFPE stored tumor tissue. We now plan a further retrospective validation of the test in a large cohort of stage II colon cancer samples collected as part of a clinical trial. We believe that this study demonstrates the value in analyzing other large retrospective FFPE-based tumor banks to develop similar tests for other cancer types.

AUTHORS' DISCLOSURES OF POTENTIAL CONFLICTS OF INTEREST

Although all authors completed the disclosure declaration, the following author(s) indicated a financial or other interest that is relevant to the subject matter under consideration in this article. Certain relationships marked with a "U" are those for which no compensation was received; those relationships marked with a "C" were compensated. For a detailed description of the disclosure categories, or for more information about ASCO's conflict of interest policy, please refer to the Author Disclosure Declaration and the Disclosures of Potential Conflicts of Interest section in Information for Contributors.
Employment or Leadership Position: Richard D. Kennedy, Almac Diagnostics (C); Timothy Davison, Almac Diagnostics (C); Peter Kerr, Almac Diagnostics (C); Julie M. Black, Almac Diagnostics (C); Max Bylesjo, Almac Diagnostics (C); Robert J. Holt, Almac Diagnostics (C); Vitali Proutski, Almac Diagnostics (C); Miika Ahdesmaki, Almac Diagnostics (C); Vadim Farztdinov, Almac Diagnostics (C); Nicolas Goffard, Almac Diagnostics (C); Peter Hey, Almac Diagnostics (C); Fionnuala McDyer, Almac Diagnostics (C); Jude M. Mulligan, Almac Diagnostics (C); Karl Mulligan, Almac Diagnostics (C); Julie Mussen, Almac Diagnostics (C); Eamonn O'Brien, Almac Diagnostics (C); Gavin Oliver, Almac Diagnostics (C); Steven M. Walker, Almac Diagnostics (C); Claire Wilson, Almac Diagnostics (C); Andreas Winter, Almac Diagnostics (C); Patrick G. Johnston, Almac Diagnostics (C); D. Paul Harkin, Almac Diagnostics (C) **Consultant or Advisory Role:** Elaine W. Kay, Almac Diagnostics (C); Daniel J. Sargent, Almac Diagnostics (C) **Stock Ownership:** Karl Mulligan, Almac Diagnostics; Patrick G. Johnston, Almac Diagnostics; D. Paul Harkin, Almac Diagnostics **Honoraria:** Patrick G. Johnston, sanofi-aventis, Chugai Pharmaceuticals, Pfizer **Research Funding:** Oliver F. Bathe, Almac Diagnostics; Sridhar Ramaswamy, Almac Diagnostics **Expert Testimony:** None **Other Remuneration:** None

AUTHOR CONTRIBUTIONS

Conception and design: Richard D. Kennedy, Timothy Davison, Peter Kerr, Julie M. Black, Max Bylesjo, Robert J. Holt, Vitali Proutski, Jude M. Mulligan, Karl Mulligan, Julie Mussen, Eamonn O'Brien, Steven M. Walker, Patrick G. Johnston, D. Paul Harkin
Provision of study materials or patients: Diarmuid O'Donoghue, Hugh Mulcahy, Jacintha O'Sullivan, Kieran Sheahan, John Hyland, Rajiv Dhir, Oliver F. Bathe, Ola Winqvist, Upender Manne, Chandrakumar Shanmugam, Sridhar Ramaswamy, Eduardo J. Leon, William I. Smith Jr, Ultan McDermott, Richard H. Wilson, Daniel Longley, John Marshall, Robert Cummins
Collection and assembly of data: Julie M. Black, Elaine W. Kay, Jude M. Mulligan, Diarmuid O'Donoghue, Hugh Mulcahy, Jacintha O'Sullivan, Kieran Sheahan, John Hyland, Rajiv Dhir, Oliver F. Bathe, Ola Winqvist, Upender Manne, Chandrakumar Shanmugam, Sridhar Ramaswamy, Eduardo J. Leon, William I. Smith Jr, Ultan McDermott, Richard H. Wilson, Daniel Longley, John Marshall, Robert Cummins
Data analysis and interpretation: Richard D. Kennedy, Timothy Davison, Peter Kerr, Julie M. Black, Max Bylesjo, Robert J. Holt, Vitali Proutski, Miika Ahdesmaki, Vadim Farztdinov, Nicolas Goffard, Peter Hey, Fionnuala McDyer, Julie Mussen, Eamonn O'Brien, Gavin Oliver, Steven M. Walker, Claire Wilson, Andreas Winter, Daniel J. Sargent, Patrick G. Johnston, D. Paul Harkin
Manuscript writing: All authors
Final approval of manuscript: All authors

REFERENCES

1. Benson AB 3rd, Schrag D, Somerfield MR, et al: American Society of Clinical Oncology recommendations on adjuvant chemotherapy for stage II colon cancer. J Clin Oncol 22:3408-3419, 2004

2. Edge S, Byrd D, Compton C, et al: AJCC Cancer Staging Manual (ed 7). New York, NY, Springer, 2010

3. Gray R, Barnwell J, McConkey C, et al: Adjuvant chemotherapy versus observation in patients with colorectal cancer: A randomised study. Lancet 370:2020-2029, 2007

4. Van Cutsem E, Oliveira J: Primary colon cancer: ESMO clinical recommendations for diagnosis, adjuvant treatment and follow-up. Ann Oncol 20:49-50, 2009 (suppl 4)

5. Gangadhar T, Schilsky RL: Molecular markers to individualize adjuvant therapy for colon cancer. Nat Rev Clin Oncol 7:318-325, 2010

6. Harkin DP: Uncovering functionally relevant signaling pathways using microarray-based expression profiling. Oncologist 5:501-507, 2000

7. Bertucci F, Salas S, Eysteries S, et al: Gene expression profiling of colon cancer by DNA microarrays and correlation with histoclinical parameters. Oncogene 23:1377-1391, 2004

8. 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 22:1564-1571, 2004

9. Eschrich S, Yang I, Bloom G, et al: Molecular staging for survival prediction of colorectal cancer patients. J Clin Oncol 23:3526-3535, 2005

10. Arango D, Laiho P, Kokko A, et al: Gene-expression profiling predicts recurrence in Dukes' C colorectal cancer. Gastroenterology 129:874-884, 2005

11. Salazar R, Bender RA, Bruin S, et al: Development and validation of a robust high-throughput gene expression test (Coloprint) for risk stratification of colon cancer patients. 2010 Gastrointestinal Cancers Symposium, Orlando, FL, January 22-24, 2010 (abstr 295)

12. DeJong S: SIMPLS: An alternative approach to partial least squares regression. Chemometrics Intell Lab Syst 18:251-263, 1993

13. Guyon I, Weston J, Barnhill S, et al: Gene selection for cancer classification using support vector machines. Mach Learn 46:389-422, 2002

14. Harbron C, Chang KM, South MC: RefPlus: An R package extending the RMA algorithm. Bioinformatics 23:2493-2494, 2007

15. Youden WJ: Index for rating diagnostic tests. Cancer 3:32-35, 1950

16. Cao D, Hou M, Guan YS, et al: Expression of HIF-1alpha and VEGF in colorectal cancer: Association with clinical outcomes and prognostic implications. BMC Cancer 9:432, 2009

17. Mukherjee S, Tamayo P, Rogers S, et al: Estimating dataset size requirements for classifying DNA microarray data. J Comput Biol 10:119-142, 2003

18. Ashburner M, Ball CA, Blake JA, et al: Gene ontology: Tool for the unification of biology—The Gene Ontology Consortium. Nat Genet 25:25-29, 2000

19. Benjamini Y, Drai D, Elmer G, et al: Controlling the false discovery rate in behavior genetics research. Behav Brain Res 125:279-284, 2001

20. Huang da W, Sherman BT, Lempicki RA: Bioinformatics enrichment tools: Paths toward the comprehensive functional analysis of large gene lists. Nucleic Acids Res 37:1-13, 2009

21. Dupuy A, Simon RM: Critical review of published microarray studies for cancer outcome and guidelines on statistical analysis and reporting. J Natl Cancer Inst 99:147-157, 2007

22. Gill S, Loprinzi CL, Sargent DJ, et al: Pooled analysis of fluorouracil-based adjuvant therapy for stage II and III colon cancer: Who benefits and by how much? J Clin Oncol 22:1797-1806, 2004

23. Salazar R, Roepman P, Capella G, et al: Gene expression signature to improve prognosis prediction of stage II and III colorectal cancer. J Clin Oncol 29:17-24, 2011

24. O'Connell MJ, Lavery I, Yothers G, et al: Relationship between tumor gene expression and recurrence in four independent studies of patients

with stage II/III colon cancer treated with surgery alone or surgery plus adjuvant fluorouracil plus leucovorin. J Clin Oncol 28:3937-3944, 2010

25. 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 27:169s, 2009 (suppl 15s; abstr 4000)

26. Farragher SM, Tanney A, Kennedy RD, et al: RNA expression analysis from formalin fixed paraffin embedded tissues. Histochem Cell Biol 130:435-445, 2008

27. Nosho K, Yamamoto H, Taniguchi H, et al: Interplay of insulin-like growth factor-II, insulin-like growth factor-I, insulin-like growth factor-I receptor, COX-2, and matrix metalloproteinase-7, play key roles in the early stage of colorectal carcinogenesis. Clin Cancer Res 10:7950-7957, 2004

28. Tsushima H, Kawata S, Tamura S, et al: High levels of transforming growth factor beta 1 in patients with colorectal cancer: Association with disease progression. Gastroenterology 110:375-382, 1996

29. Völp K, Brezniceanu ML, Bosser S, et al: Increased expression of high mobility group box 1 (HMGB1) is associated with an elevated level of the antiapoptotic c-IAP2 protein in human colon carcinomas. Gut 55:234-242, 2006

Affiliations

D. Paul Harkin, Richard D. Kennedy, Patrick G. Johnston, Richard Wilson, Daniel Longley, and Ultan McDermott, Centre for Cancer Research and Cell Biology, Queen's University of Belfast, Belfast, Northern Ireland; Elaine W. Kay, Royal College of Surgeons in Ireland, Beaumont Hospital; Diarmuid O'Donoghue, Hugh Mulcahy, Jacintha O'Sullivan, Kieran Sheahan, and John Hyland, Centre for Colorectal Disease, Education and Research Centre, St Vincent's Hospital, Dublin, Ireland; D. Paul Harkin, Richard D. Kennedy, Max Bylesjo, Timothy Davison, Peter Kerr, Julie M. Black, Vadim Farztdinov, Fionnuala McDyer, Nicolas Goffard, Miika Ahdesmaki, Gavin Oliver, Peter Hey, Eamonn O'Brien, Julie Mussen, Vitali Proutski, Claire Wilson, Steven Walker, Andreas Winter, Karl Mulligan, Robert J. Holt, and Patrick G. Johnston, Almac Diagnostics, Craigavon, United Kingdom; Upender Manne and Chandrakumar Shanmugam, University of Alabama at Birmingham, Birmingham, AL; William I. Smith Jr, Suburban Hospital, Bethesda, MD; John Marshall, Lombardi Cancer Center, Georgetown University, Washington, DC; Sridhar Ramaswamy and Eduardo J. Leon, Massachusetts General Hospital Cancer Center, Richard B. Simches Research Center, Boston, MA; Rajiv Dhir, University of Pittsburgh Medical Center, Shadyside-Presbyterian Hospital, Pittsburgh, PA; Daniel J. Sargent, Mayo Clinic, Rochester, MN; Oliver F. Bathe, Tom Baker Cancer Center, Calgary, Alberta, Canada; and Ola Winqvist, Karolinska Institutet, Karolinska Hospital, Stockholm, Sweden.

■ ■ ■